







HyRAD-X Exome Capture Museomics Unravels Giant Ground Beetle Evolution

Emmanuel F.A. Toussaint ^{1,*†}, Jérémy Gauthier ^{1,†}, Julia Bilat¹, Conrad P.D.T. Gillett ², Harlan M. Gough ³, Håkan Lundkvist⁴, Mickael Blanc¹, Carlos P. Muñoz-Ramírez ^{5,6}, and Nadir Alvarez ^{1,7}

¹Natural History Museum of Geneva, Switzerland

²University of Hawai'i Insect Museum, University of Hawai'i at Mānoa, Honolulu, Hawaii, USA

³Florida Natural History Museum, University of Florida, Gainesville, Florida, USA

⁴Naturcentrum AB, Stenungsund, Sweden

⁵Instituto de Entomología, Universidad Metropolitana de Ciencias de la Educación, Santiago, Chile

⁶Centro de Investigación en Biodiversidad y Ambientes Sustentables (CIBAS), Universidad Católica de la Santísima Concepción, Chile

⁷Department of Genetics and Evolution, University of Geneva, Switzerland

*Corresponding author: E-mail: emmanuel.toussaint@ville-ge.ch.

Accepted: 10 May 2021

† These authors are joint first authors.

Abstract

Advances in phylogenomics contribute toward resolving long-standing evolutionary questions. Notwithstanding, genetic diversity contained within more than a billion biological specimens deposited in natural history museums remains recalcitrant to analysis owing to challenges posed by its intrinsically degraded nature. Yet that tantalizing resource could be critical in overcoming taxon sampling constraints hindering our ability to address major evolutionary questions. We addressed this impediment by developing phyloHyRAD, a new bioinformatic pipeline enabling locus recovery at a broad evolutionary scale from HyRAD-X exome capture of museum specimens of low DNA integrity using a benchtop RAD-derived exome-complexity-reduction probe set developed from high DNA integrity specimens. Our new pipeline can also successfully align raw RNAseq transcriptomic and ultraconserved element reads with the RAD-derived probe catalog. Using this method, we generated a robust timetree for Carabinae beetles, the lack of which had precluded study of macroevolutionary trends pertaining to their biogeography and wing-morphology evolution. We successfully recovered up to 2,945 loci with a mean of 1,788 loci across the exome of specimens of varying age. Coverage was not significantly linked to specimen age, demonstrating the wide exploitability of museum specimens. We also recovered fragmentary mitogenomes compatible with Sanger-sequenced mtDNA. Our phylogenomic timetree revealed a Lower Cretaceous origin for crown group Carabinae, with the extinct *Aplothorax* Waterhouse, 1841 nested within the genus *Calosoma* Weber, 1801 demonstrating the junior synonymy of *Aplothorax* **syn. nov.**, resulting in the new combination *Calosoma burchellii* (Waterhouse, 1841) **comb. nov.** This study compellingly illustrates that HyRAD-X and phyloHyRAD efficiently provide genomic-level data sets informative at deep evolutionary scales.

Key words: ancient DNA, Carabidae phylogenomics, Coleoptera evolution, hybridization capture, phyloHyRAD, RAD-sequencing.

Introduction

The field of phylogenetics is being reinvigorated by methodological advances that continue to benefit from developments in high-throughput sequencing, increasingly allowing for the acquisition of large genomic fractions from nonmodel

organisms (Allio et al. 2020; Blair and Ané 2020). Of particular interest is the potential of such methods to take advantage of over a billion biological specimens housed in museum collections across the globe because these vast repositories of biodiversity have hitherto remained largely neglected in this

Significance

Museum specimens housed in natural history collections represent a tremendous wealth of genetic information that is not yet fully unlocked because their DNA can be highly degraded and difficult to sequence. The recently introduced HyRAD-X method alleviates this impediment by allowing the capture of thousands of loci from the genome of both adequately preserved and old Museum specimens. In this study, we developed a new phyloHyRAD bioinformatics pipeline and used it to produce the first robust, dated phylogenomic backbone for Carabinae giant ground beetles. Our new method makes it possible to perform DNA read alignments at a deep evolutionary scale and to retrieve phylogenomic information from even very old specimens (>50 years), providing a useful tool for future systematic and evolutionary work.

respect. Methods such as whole-genome sequencing or transcriptomics usually rely on fresh, bespokenly sampled source material containing well-preserved DNA to produce large-scale genomic data (but see for instance Staats et al. 2013; Yao et al. 2020). Yet, it is often practically impossible to obtain such specimens because many species are rare, extinct, or known from only very few specimens in museums (Deng et al. 2019; Wells et al. 2019; Jin et al. 2020). Hence, developing methods such as genomic capture that may unlock the potential of existing museum specimens to generate genetic data is necessary. Approaches such as exon capture (Lemmon et al. 2012; Mayer et al. 2016; Knyshov et al. 2019) or ultraconserved elements (UCEs, Faircloth et al. 2012) have proven very efficient in generating genomic data from museum specimens (e.g., Blaimer et al. 2016; McCormack et al. 2016; Ruane and Austin 2017; St Laurent et al. 2018; Toussaint et al. 2018).

The major drawback when applying genomic techniques to museum specimens is the generally poor preservation inherent in their constituent DNA templates (Colella et al. 2020; Jin et al. 2020), hindering or precluding application of standard high-throughput sequencing methods. Because DNA extracted from these specimens is often degraded and highly fragmented, approaches relying upon genomic complexity reduction (i.e., reducing the portion of the genome that is studied) are only applicable provided that certain initial conditions, such as mild levels of DNA fragmentation (with DNA molecules >5 kb remaining in the extract), are met (Tin et al. 2014; Burrell et al. 2015). Because such methods further fragment DNA molecules having already been fragmented through postmortem endonuclease activity, a minimum amount of starting DNA comprising a significant proportion of long fragments is crucial to allow for the greatest possible number of restriction sites to be conserved across all samples. However, for many old or otherwise deteriorated samples, these conditions will not be met, and consequently, methods such as RADseq (Miller et al. 2007) will often be inapplicable. Alternatively, employing hybridization capture of target DNA using liquid- or solid-phase probes (Carpenter et al. 2013) allows access to molecules of low integrity typical of those present in historical specimens, that is, those that are too short

and/or degraded to be retrieved from a sample using conventional methods. Most such methods rely upon commercially synthesized probes, applying either exome capture (Bi et al. 2013) or capture of a subset of loci (Jones and Good 2016). A broad catalog of possible genomic targets can permit exploration of a wide array of evolutionary scales. For instance, UCEs are well-suited for inferring relationships from suprageneric (e.g., Wood et al. 2018) to interspecific levels (e.g., Manthey et al. 2016), whereas capture methods based on RADseq loci can reconstruct phylogeographic scenarios or population processes at the intraspecific level (e.g., Ali et al. 2016; Hoffberg et al. 2016). More seldomly, RADseq approaches have enabled reconstruction of evolutionary patterns at deeper scales (e.g., Leaché et al. 2015).

Among the latter techniques is Hybridization Capture Using RAD Probes (HyRAD, Suchan et al. 2016), which in contrast to all other hybridization capture methods based on RADseq loci that require prior knowledge of genome sequences, does not require commercially synthesized probes, instead relying only on probes that can be synthesized within researchers' laboratories. For this method, the probes are built through a double-digest RADseq (Peterson et al. 2012) protocol using either a high-integrity DNA template from fresh specimens of the target species (i.e., HyRAD *sensu stricto*) or a messenger RNA template (i.e., HyRAD-X, Schmid et al. 2017). So far, both methods have proven reliable in uncovering evolutionary processes at the intraspecific level using both historical (Schmid et al. 2018; Gauthier et al. 2020) and ancient samples (Schmid et al. 2017). However, their application at broader evolutionary scales in a phylogenetic framework has not yet been evaluated. Because HyRAD-X uses a messenger RNA template to design the double-digestion RAD-sequencing (ddRAD)-based probes onto which DNA of the specimens of interest will hybridize, it should theoretically perform similarly to an exome capture approach in producing data suitable for phylogenetic reconstruction. A major advantage being that prior acquisition of commercially synthesized probes representative of the target lineage is unnecessary. Instead, production of a set of probes from a limited number of fresh specimens spanning the expected phylogenetic diversity of the target clade should sufficiently encompass the extant diversity of that clade to enable

capture of all fragments homologous to the reduced representation in the messenger RNA template. Bioinformatic treatment of HyRAD data has been made more accessible thanks to the release of popHyRAD, a suite of scripts dedicated to treating HyRAD data at a shallow evolutionary scale (Gauthier et al. 2020). However, to analyze HyRAD-X data at deeper evolutionary scales, several modifications should be implemented, such as alignment of genomic data to an exomic template, the merging of reads aligned to different probes into a single alignment encompassing all taxa, and the output of final alignments as well as single nucleotide polymorphism (SNP) data.

The cosmopolitan ground beetle subfamily Carabinae (Insecta, Coleoptera, Carabidae) is a useful model with which to apply and test the versatility of HyRAD-X. Macroevolutionary trends in this group, for example pertaining to the evolution of wing morphologies (i.e., apterous, brachypterous, macropterous, see Imura et al. 2018; Toussaint and Gillett 2018), are currently obscured by the lack of a robust phylogenetic framework and divergence times. This group comprises ca. 1500 species distributed in four tribes: the Holarctic Cydrini (four genera, ca. 230 species), the South American Ceroglossini (one genus, ca. 10 species), the Australasian Pamborini (two genera, ca. 20 species), and the cosmopolitan Carabini (two genera, ca. 1100 species) (Jiroux 2006; Takami and Sota 2006; Cavazzuti 2010; Bruschi 2013; Deuve 2019). Furthermore, the monotypic genus *Aplothorax* Waterhouse, 1841 endemic to the Atlantic island of St Helena, considered extinct since the 1970s (Gray et al. 2019), is classified within the subfamily but without a robust tribal placement (Jeannel 1940; Prüser and Mossakowski 1998; Bruschi 2013; but see Sota et al. 2020). In addition to ranking among the most popular of beetles owing to their often elegant and attractive appearance (Fabre 1901; Chatenet 1986), the Carabinae include some of the largest and bulkiest ground beetles, including species in the subgenus *Carabus* (*Procerus*) Dejean, 1826 that attain 6 cm in length (Cavazzuti 1989), and can be termed ‘giant ground beetles’ accordingly. This subfamily was the focus of early studies during the dawn of molecular phylogenetics (Su et al. 1996a, 1996b); all clades having since received attention from phylogenetic studies incorporating Sanger sequencing (Kim, Tominaga, et al. 2000; Kim, Zhou, et al. 2000; Takami 2000; Tominaga et al. 2000; Okamoto et al. 2001; Sota and Vogler 2001, 2003; Su et al. 2003, 2005; Osawa et al. 2004; Sota and Ishikawa 2004; Su, Imura, Okamoto, Kim, et al. 2004; Su, Imura, Okamoto, and Osawa 2004; Sota et al. 2005; Andújar et al. 2012a; Andújar et al. 2012b; Deuve et al. 2012; Deuve and Faille 2013; Muñoz-Ramírez 2015; Muñoz-Ramírez et al. 2016; Imura et al. 2018; Toussaint and Gillett 2018; Sota et al. 2020). An early attempt to employ RAD-sequencing within the genus *Carabus* L., 1758 aimed to construct a molecular phylogeny of the charismatic subgenus *Carabus* (*Chysocarabus*) Thomson, 1875, indicating that a potential strength of such an approach

is the inference of relationships at the intrageneric level (Cruaud et al. 2014). This approach was successfully pursued for phylogenetic inference within Japanese *Carabus* (e.g., Takahashi et al. 2014; Fujisawa et al. 2019). However, very few studies have focused on intertribal relationships (but see Osawa et al. 2004; Toussaint and Gillett 2018; Sota et al. 2020). Based upon preliminary analyses of Sanger sequencing or mitogenome data, the snail-eating beetles (Cydrini) have been shown to be sister to the rest of the subfamily, followed by Ceroglossini and Pamborini, which together form an apparent Gondwanan stock, that is in turn sister to Carabini (comprising the diverse genera *Carabus* and *Calosoma* Weber 1801) (Imura et al. 2018; Toussaint and Gillett 2018; Sota et al. 2020). Jeannel (1940) hypothesized that *Aplothorax* is a close relative of the Afrotropical subgenus *Calosoma* (*Ctenosta*) Motschulsky, 1865 and this placement was recently supported by analysis of mitogenomes (Sota et al. 2020). The genus *Calosoma* is unique in this subfamily of predominantly brachypterous flightless species, for containing a significant proportion of fully winged and flight-capable species (Bruschi 2013). Charles Darwin himself observed the impressive flight capabilities of *Calosoma* when he saw them flying onto the H.M.S. Beagle whilst many miles off the South American coast (Darwin 1845). Major obstacles to obtaining a robust and comprehensive phylogeny of Carabinae include the substantial species richness of the group, the widespread geographic distribution of constituent clades across diverse latitudes and elevations, combined with the extreme localization of many species.

Although reliable fossils of Carabinae exist (e.g., Nel 1987, 1988; Deuve 1998; Farrés and Altimiras 2012; Yahiro et al. 2018; Kirejtshuk et al. 2019), the age of the subfamily has been contentious, since the early work of Jeannel (1940) to recent times (Andújar et al. 2012a; Andújar et al. 2014; Toussaint and Gillett 2018; Opgenoorth et al. 2020; Sota et al. 2020). This uncertainty stems from the disputed sister group to Carabinae, an absence of reliable fossils outside of the tribe Carabini, and the frequent past use of biogeographic constraints in clock calibrations, which have likely biased results toward younger divergence times. Current estimates for the crown age of Carabinae range between as recent as ca. 35 Ma (Andújar et al. 2012a; Andújar et al. 2014) to as old as ca. 170 Ma (Toussaint and Gillett 2018). Larger-scale phylogenomic studies of beetles have also provided insights into the evolution of Carabinae. For instance, the most comprehensive attempt at unraveling the evolutionary history of beetles to date (McKenna et al. 2019) recovered an age of ca. 35 Ma for the crown of *Calosoma* and *Carabus*, and an age of ca. 120 Ma for the stem (i.e., the split with *Elaphrus* [Carabidae, Elaphrinae] in that study). However, the placement of Carabinae within ground beetles remains controversial in light of recent phylogenomic work (Gustafson et al. 2020) and our understanding of Carabinae evolutionary history is therefore hampered by the lack of a robust phylogeny and requires an

in-depth phylogenomic reappraisal. Considering that giant ground beetles were some of the first nonmodel insects to have been studied using molecular phylogenetic approaches, now seems a timely opportunity to develop new methods employing a modern toolkit, to further unravel the evolutionary history of these charismatic beetles.

We apply HyRAD-X to a representative set of giant ground beetles and infer divergence times at an unprecedented evolutionary scale compared with previous museomics studies based on a set of RAD-like loci. We analyze RAD-like loci obtained from 48 specimens selected to represent all recognized genera of Carabinae and recently inferred clades within the genus *Calosoma* (Toussaint and Gillett 2018). We also introduce a novel, dedicated bioinformatic pipeline, phyloHyRAD, which complements the recently published popHyRAD suite (Gauthier et al. 2020), allowing for construction of a reference catalog, the alignment of reads, and the identification of individual sequences in a phylogenetic framework. Our aims are 1) to generate a new exome capture toolkit to infer the evolutionary history of Carabinae at different taxonomic scales, from intersubfamilial to intrageneric relationships; 2) to introduce a new bioinformatic pipeline allowing the optimization of data recovery from HyRAD-X raw reads in a phylogenetic context (fig. 1); 3) to infer a robust phylogenomic backbone for Carabinae using the resulting HyRAD-X output combined with external data produced through RNAseq and UCE approaches; and 4) to estimate divergence times among giant ground beetles.

Results and Discussion

Genomic Data Recovery

After DNA extraction of 48 specimens, we applied the HyRAD-X protocol, and two independent Illumina sequencing runs to obtain a mean of 15,737,206 reads for each sample (SD 19,657,890). After cleaning of these raw reads (see below), 96.4% (SD 5.2%) of reads were retained (supplementary table S1, Supplementary Material online). These results highlight considerable heterogeneity in the amount of genetic information recovered across samples. From the initial sampling, 13 specimens produced a very low number of reads and recovered loci, with a correspondingly low genomic coverage (usually <5 kb in total) (fig. 2) preventing their inclusion in the final analyses, although in preliminary testing they were recovered in their correct tribes (e.g., *Cychrus morawitzi* Gehin, 1863, *Calosoma atrovirens* Chaudoir, 1869, *Calosoma burtoni* Alluaud, 1913, and *Calosoma chinense* Kirby 1818). Overall, 35 out of 48 sequenced taxa were included in the final analyses, covering all currently recognized genera except for the Nearctic genus *Scaphinotus* Dejean, 1826, whose phylogenetic placement within Cychrini remains doubtful (Su et al. 2004; Imura et al. 2018; Toussaint and Gillett 2018; Sota et al. 2020).

phyloHyRAD pipeline

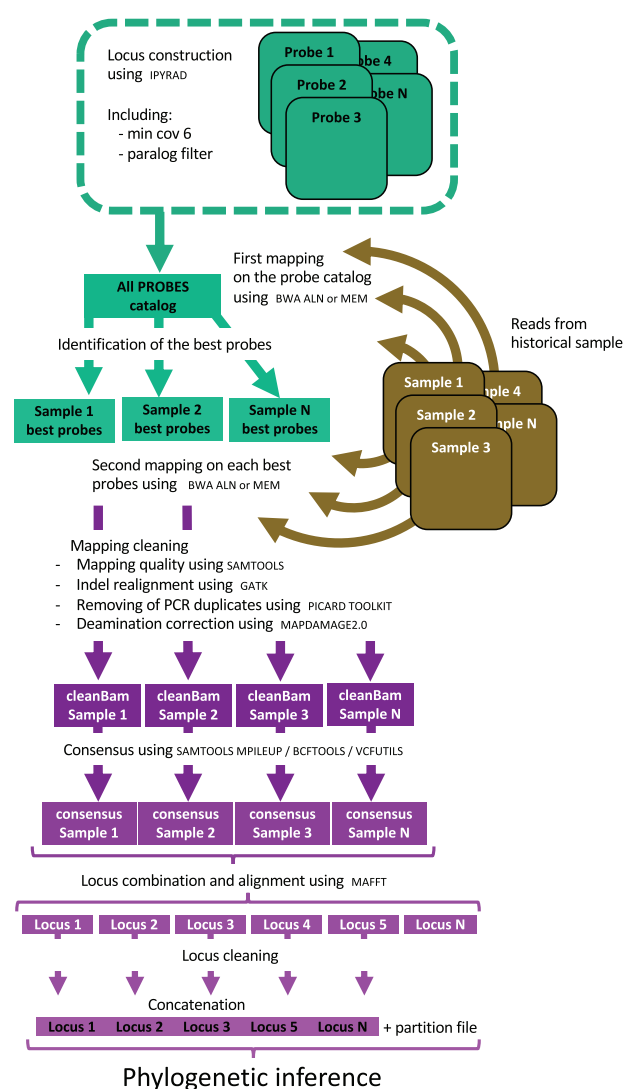


Fig. 1.—Schematic representation of the phyloHyRAD pipeline. In green is shown the construction of the probe loci used as a catalog for the mapping of the sample reads (shown in brown). The purple section indicates the mapping and cleaning steps that lead to construction of the combined and aligned loci for each sample, which are then concatenated and used in the phylogenetic analyses.

For the mapping step, although BWA-ALN is the recommended tool when analyzing ancient DNA (Schubert et al. 2012), we evaluated the influence of the mapping tools by also using a second, more flexible approach, namely BWA-MEM (Li 2013). Indeed, at deep evolutionary scales, where the genetic distance between the probes and the captured samples can sometimes be significant, and which is a crucial parameter for both the capture process and the mapping onto reference probes, a higher flexibility in mapping seems to be essential to retrieve sufficient genetic information. We observed a higher percentage of mapped reads using BWA-

HyRAD-X loci (BWA-MEM)

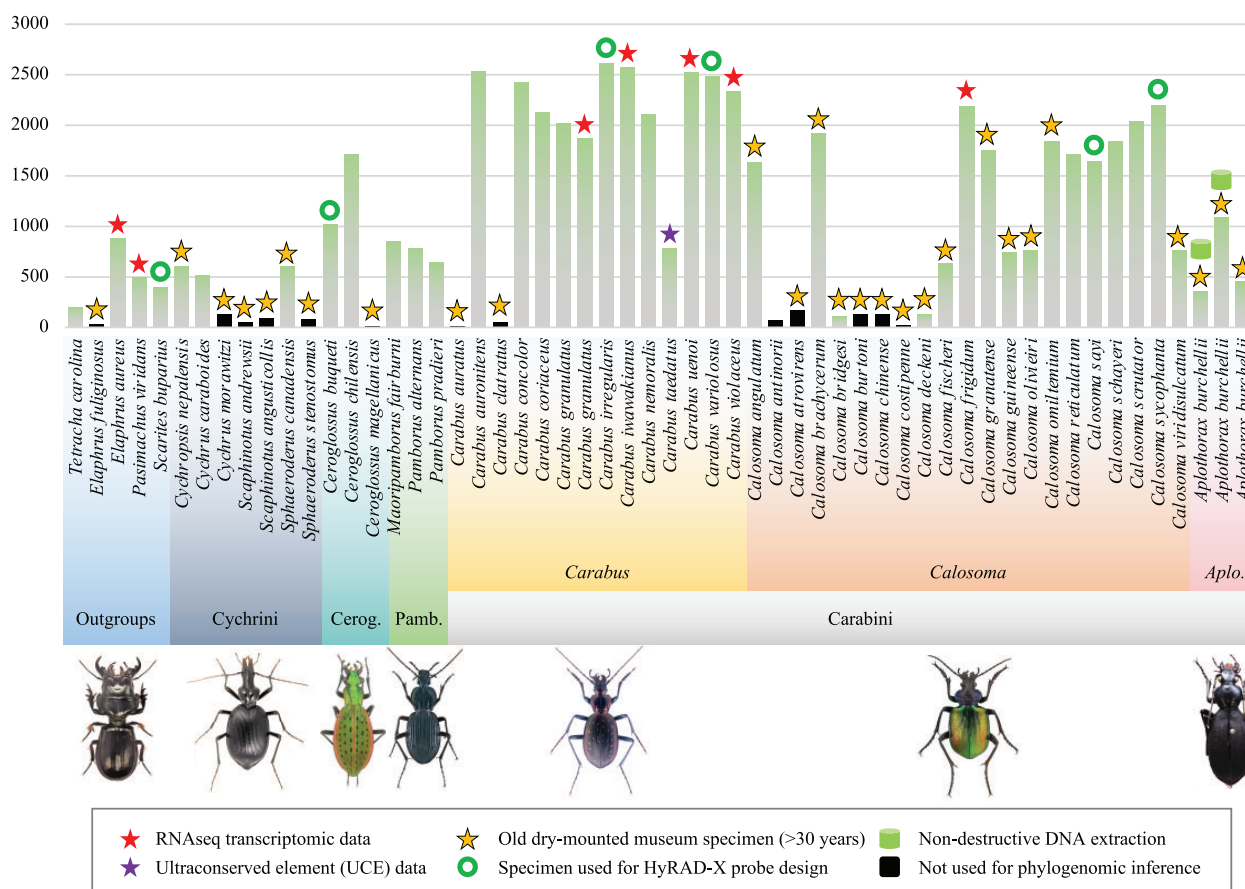


Fig. 2.—Graphical representation of locus recovery for each taxon. The histogram represents the number of final assembled HyRAD-X loci obtained with BWA-MEM for each taxon, including those that were not retained for final phylogenomic inferences. Taxa are ordered taxonomically with each major clade being illustrated by a habitus photograph of a representative species.

Table 1

Alignment Statistics for Each Data Set, Including the Alignment Length, the Percentage of Missing Data, the Numbers, and Percentages of Variable Sites and of Parsimony Informative Sites, and the GC Content

Data Set	Mapping Tool	No. of Taxa	Loci	Alignment Length	Missing Data	Variable Sites	Parsimony Informative Sites	GC Content
A	BWA-MEM	43	2,945	474,696	67.9%	102,929 (21.7 %)	57,387 (12.1 %)	0.456
B	BWA-MEM	40	2,945	474,696	65.7%	102,802 (21.7 %)	57,285 (12.1 %)	0.455
C	BWA-ALN	40	1,227	222,732	80.7%	24,772 (11.1 %)	9,922 (4.5 %)	0.465
D	BWA-MEM	43	100	27,749	56.5%	6,946 (25.0%)	4,177 (15.1%)	0.448

MEM (34.9% [SD 12.9%] before cleaning and 4.6% [SD 2.4%] after cleaning) than when using BWA-ALN (2.6% [SD 2.8%]) of mapped reads before cleaning and 0.2% [SD 0.1%] after cleaning) (details in [supplementary table S1, Supplementary Material](#) online). Differences in percentages of mapped reads affect the mean coverage ([supplementary fig. S1, Supplementary Material](#) online) and although lower coverage still allows recovery of some loci, these are characterized by increased missing data. The mean number of

retrieved loci obtained by each mapping method is similar, that is, 1,398 loci with BWA-MEM (SD 816) ([supplementary fig. S1, Supplementary Material](#) online) and 1,178 loci with BWA-ALN (SD 808) ([supplementary table S1, Supplementary Material](#) online), although the proportion of missing data is higher with BWA-ALN ([table 1](#)). This is particularly reflected in the number of Parsimony-informative sites (i.e., those with a minor-frequency allele shared by at least two specimens) present in the final alignments, where five times more

informative sites were present in the BWA-MEM derived alignment (57,285 sites) than in that obtained with BWA-ALN (9,922 sites) (table 1). This variation explains differences in resolution and branch support in the phylogeny across analyses (fig. 3).

Data from the eight taxa obtained from external repositories were included in the final alignments. Mapping of these data revealed a low percentage of mapped reads: a mean of 6.8% (SD 3.2%) before cleaning and 1.3% (SD 0.9%) after cleaning with BWA-MEM, and a mean of 0.4% (SD 0.4%) before cleaning and 0.1% (SD 0.1%) after cleaning with BWA-ALN (supplementary table S1, Supplementary Material online). This was expected because these data resulted from approaches such as RNA-Seq or UCEs that do not specifically target the loci captured by our HyRAD-X approach. Despite these relatively low mapping levels, locus recovery from the raw reads for these taxa remained excellent, with a mean recovery of 1,788 loci (SD 868) with BWA-MEM and 530 loci (SD 515) with BWA-ALN.

The effect of missing data on phylogeny reconstruction has been intensively studied for nearly three decades (Novacek, 1992; Wiens, 1998; Dell'Ampio et al. 2014; Sann et al. 2018), including studies restricted to the application of RADseq-type approaches (Rubin et al. 2012; Hosner et al. 2016; Huang and Knowles 2016; Crotti et al. 2019). Most studies agree that data sets containing a certain level of missing data, even of the order of 50% or 60%, can still provide useful phylogenetic information. On the other hand, overly permissive strategies, that include loci consisting of more than 80% or 90% missing data, have revealed suboptimal results (Crotti et al. 2019). In addition, the inclusion of sequences that are too short may also prove unpredictable, due to gene trees that are literally too short to allow optimization of the species tree (Hosner et al. 2016). However, Streicher et al. (2016) have shown that in concatenation approaches, optimality can be achieved despite a relatively high level of missing data, provided that the taxon sampling is high.

Our most well-supported result was achieved using the most flexible aligner (BWA-MEM), which yielded both the highest number of retrieved loci and the lowest level of missing data.

Phylogenetic Inference of Giant Ground Beetles

Results of the maximum likelihood phylogenomic inferences are summarized in figure 3 (see supplementary figs. S2–S6, Supplementary Material online for more details). The phylogenetic resolution and associated branch support across the topologies differed greatly according to the type of loci (BWA-ALN vs. BWA-MEM) and the taxon sampling. The reduced number of loci obtained using BWA-ALN (Data set C) resulted in lower branch support and inconsistent phylogenetic relationships compared with other analyses (fig. 3). Similarly, including taxa represented by low coverage resulted in decreased branch

support (fig. 3, Data set A). Removing the outgroup *Tetracha carolina* (Linnaeus, 1763), represented by low loci coverage (206 loci with BWA-MEM, fig. 3), resulted in strong branch support across the backbone of the tree (fig. 3, Data set B). However, RogueNaRok analyses did not reveal a high number of rogue taxa nor did they identify *T. carolina* as such. Only *Calosoma deckeni* (Gerstaecker, 1867) in Data set A and *Carabus nemoralis* Müller, 1764 in Data set E were identified as potential rogue taxa (see supplementary table S3, Supplementary Material online for detailed results of the RogueNaRok analyses). Finally, the phylogenetic hypothesis derived from the analysis of Data set D (fig. 3) is largely compatible with the one derived from the most comprehensive analysis of Data set A despite being based on a substantially lower number of loci. Below we summarize and discuss the results of the analysis based on Data set B, which we consider the most robust inference.

We recover Carabinae as monophyletic and divided into three main clades, in line with previous phylogenetic reconstructions (Imura et al. 2018; Toussaint and Gillett 2018; Sota et al. 2020). The first clade corresponds to the tribe Cychnini, wherein *Sphaeroderus* Dejean, 1831 is recovered as sister to *Cychrus* Fabricius, 1794 and *Cychropsis* Boileau, 1901, with strong branch support. The missing genus *Scaphinotus* could not be included in our final analyses and its placement within Cychnini remains uncertain (Su et al. 2004; Toussaint and Gillett 2018). The second clade comprises the tribes Ceroglossini and Pamborini that are recovered as sister clades with strong branch support. Within Pamborini, we recover *Maoripamborus* Brookes, 1944 as sister to *Pamborus* Latreille, 1817, with strong branch support. The third clade includes the tribe Carabini, with the subtribes Carabina and Calosomatina recovered as reciprocally monophyletic, with strong branch support.

Within *Carabus*, we recover three main clades. The first one comprises the subgenera *Carabus* (*Hygrocarabus*) Thomson, 1875 and *Carabus* (*Platycarabus*) Morawitz, 1886, as proposed by Deuve et al. (2012), who also recovered *Carabus* (*Chaetocarabus*) Thomson, 1875 (not included in this study) within it (i.e., the Arcifera group *sensu* Deuve 2019). We then recover a clade comprising *Carabus* (*Orinocarabus*) Kraatz, 1878 and *Carabus* (*Tanaocarabus*) Reitter, 1896 as sister to the remainder of *Carabus*. Although our taxon sampling is limited in this highly diverse genus, which contains almost 100 subgenera (Deuve 2019), the strongly supported backbone we infer will serve as a basis for future phylogenetic work.

Our analyses indicate with strong branch support, and in concordance with Sota et al. (2020), that *Aplothorax* is a lineage nested within the cosmopolitan genus *Calosoma*. The former genus therefore becomes a junior subjective synonym of the latter, *Aplothorax* Waterhouse, 1841 **syn. nov.** of *Calosoma* Weber, 1801, thereby creating the new combination: *Calosoma burchellii* (Waterhouse, 1841) **comb. nov.** This placement confirms Jeannel's (1940) early hypothesis based on morphological features. Further studies are needed to understand the placement of this unique lineage with

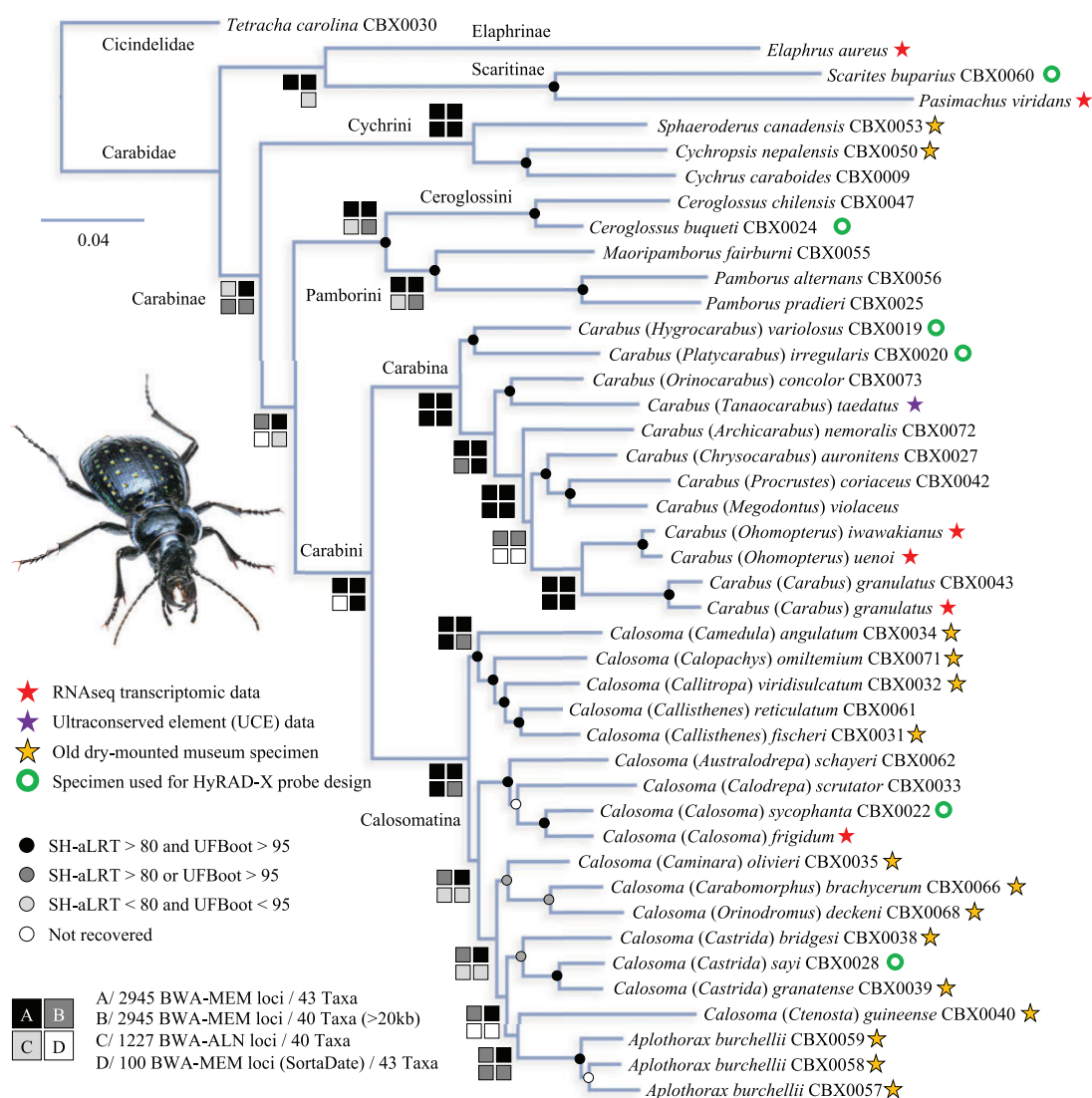


Fig. 3.—Best Maximum likelihood tree inferring Carabinae relationships. Best scoring maximum likelihood tree based on Data set A (2,945 loci, 43 taxa, BWA-MEM mapper). Branch support from this analysis is shown for all branches. Branch support retrieved in different analyses is shown for major branches according to the inserted caption. A live photograph of the Mongolian desert endemic *Calosoma (Callisthenes) fischeri* Fischer, 1842 is presented (Photograph credit: Lily Kumpke).

respect to the winged species of the subgenus *Calosoma* (*Ctenosta*). Within *Calosoma*, our phylogenomic inference recovers the same clades revealed in Toussaint and Gillett (2018) despite a reduced taxon sampling, but with strong branch support across the backbone and the main clades. However, our results contradict, to some extent, the results of Sota et al. (2020) who studied mitogenomes across Carabinae and recovered a different hypothesis within *Calosoma*, albeit with lower branch support. Our results provide a robust backbone for this genus, intimating that taxonomic reappraisal is necessary. The phylogenomic tree presented in this study confirms multiple gains and losses of wings across *Calosoma*, which likely resulted from its complex and dynamic biogeographical history (fig. 4).

Evolutionary History of Carabinae Beetles

Our BEAST analyses all converged successfully, with effective sample size values well above 200 for all estimated parameters. Overall, divergence times broadly overlapped between analyses using different clock and tree priors (table 2). Comparison of marginal likelihoods reveals that the analysis based upon seven Bayesian uncorrelated lognormal relaxed clocks and a Yule tree model was best supported, although support was not significantly better than with the same number of clocks and a birth–death model (table 2). We present the results of this analysis in figure 4 and discuss them hereafter (see supplementary fig. S7, Supplementary Material online for more details). We recover a split between tiger beetles

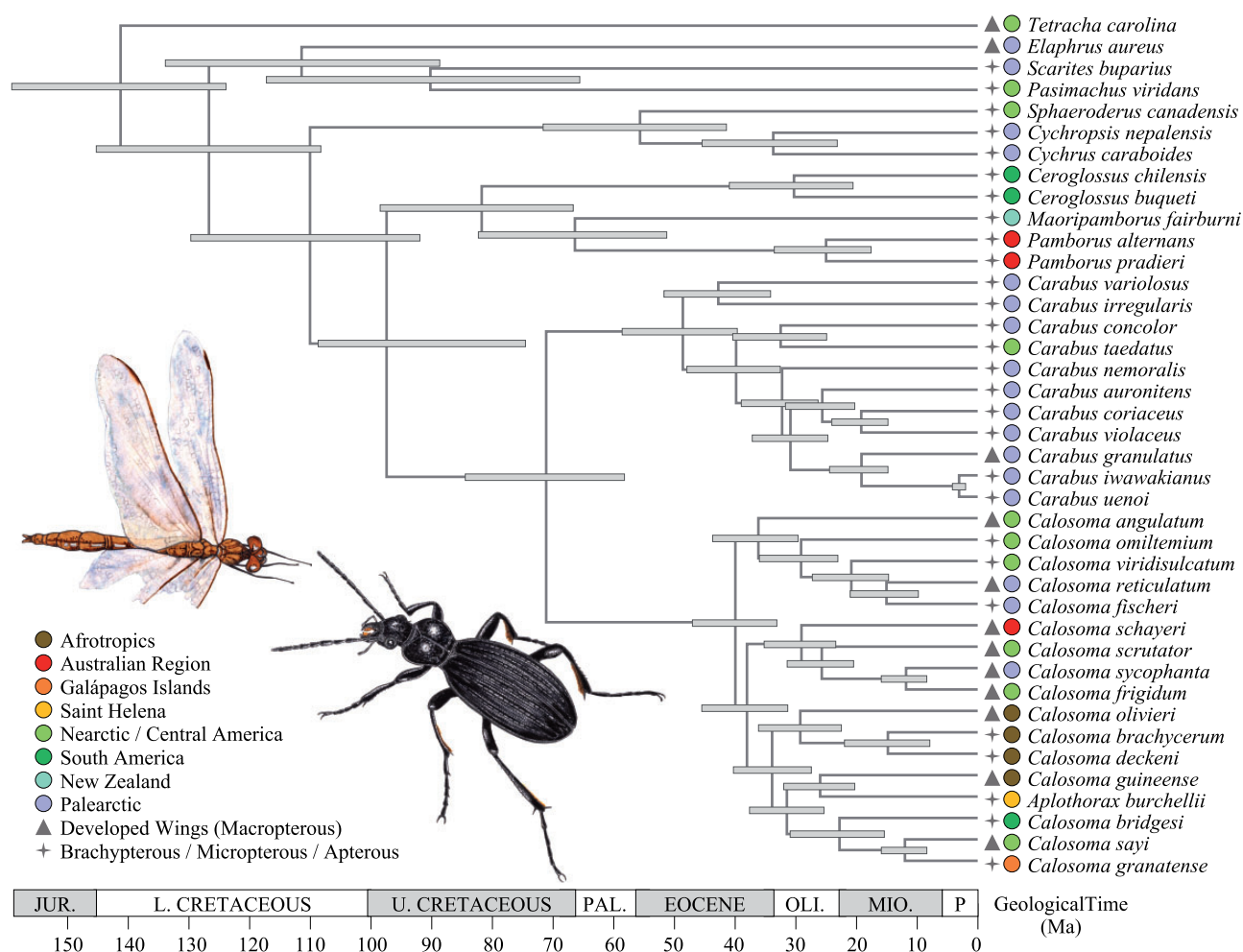


Fig. 4.—Bayesian divergence time estimates for Carabinae. Maximum clade credibility tree obtained from a BEAST analysis using seven Bayesian lognormal relaxed clocks and a Yule pure birth tree model. Node estimates are postburn in median ages, with 95% credibility intervals represented by a grey horizontal bar for each node. An illustration of *Aplothorax burchellii* scavenging on an endemic St. Helena Darter *Sympetrum dilatatum* (Calvert, 1892) is presented.

Table 2

BEAST Median Divergence Times and 95% Credibility Intervals with MLE Comparison

Analysis	Clocks	Tree Model	SS MLE	PS MLE	Root Median [95% CI]	Carabidae Median [95% CI]	Carabini Median [95% CI]
A1	1	Yule	−110,761.73	−110,764.85	147.0 [129.0–167.3]	126.1 [105.8–145.7]	71.0 [53.8–88.5]
A2	1	Birth–death	−110,759.26	−110,765.45	147.0 [127.9–166.1]	125.6 [105.8–145.7]	69.9 [53.1–87.3]
A3	7	Yule	−110,706.32	−110,706.93	141.7 [125.4–160.8]	127.3 [109.1–145.9]	71.4 [58.2–84.5]
A4	7	Birth–death	−110,706.19	−110,706.75	142.0 [124.4–159.8]	127.4 [108.7–145.9]	71.4 [58.6–85.0]

NOTE.—SS, stepping-stone sampling; MLE, marginal likelihood estimate; PS, path-sampling; 95% CI, 95% credibility interval from posterior BEAST estimates.

(Cicindelidae) and ground beetles in the Lower Cretaceous ca. 140 Ma. The crown age of Carabinae is estimated to lie in the Lower Cretaceous, ca. 110 Ma (95% HPD = 92–130 Ma). Following that split, we recover a crown age for Cychrini of ca. 55 Ma (95% HPD = 42–72 Ma). We infer a split between Carabini and the Ceroglossini+Pamborini clade ca. 100 Ma (95% HPD = 58–85 Ma), at the Lower to Upper Cretaceous boundary. The split between Ceroglossini and Pamborini is

estimated to have occurred ca. 80 Ma (95% HPD = 67–99 Ma), suggesting a possible vicariant event between the faunas of South America and Australia, which could be linked to the glaciation of Antarctica. The split between *Maoripamborus* and *Pamborus* is dated at ca. 70 Ma (95% HPD = 52–83 Ma), which is consistent with the hypothesis of vicariance between Australia and Zealandia (Schellart et al. 2006; Neall and Trewick 2008). The split between *Calosoma* and *Carabus*

is also dated at ca. 70 Ma (95% HPD = 58–85 Ma), whilst the crown ages of *Carabus* and *Calosoma* are, respectively, dated at ca. 50 Ma (95% HPD = 40–59 Ma) and ca. 40 Ma (95% HPD = 33–48 Ma). Without denser sampling of the few Nearctic *Carabus* species, it is difficult to discuss the alternative biogeographic hypotheses that led to the colonization of this region from the Palearctic. Within *Calosoma*, most splits cannot be explained by vicariant events, and therefore the biogeographical history of the genus must have been very dynamic, with multiple range expansion and dispersal events at continental scales. The colonization of St Helena by the ancestors of *Aplothorax* took place between the Oligocene and Miocene, most likely out of Africa, where the closely-related and fully winged subgenus *Calosoma* (*Ctenosta*) occurs. The oldest rocks on that remote oceanic island of volcanic origin have been dated at ca. 14.5 Ma (Baker 1973), which is in line with our estimates. Considering the remoteness of St Helena from continental Africa, we hypothesize that ancestors of *Aplothorax* colonized that island by active flight (potentially wind-assisted). This possibly occurred in a stepping-stone manner across seamounts of the St Helena/Guinea chain, which may previously have been emerged (Peyve 2011), the beetles subsequently losing their flight capacity through adaptation to the island ecosystems by evolution of progressive brachyptery, as is well documented in other insular insects (e.g., Roff 1990).

Our results predate the divergence times proposed by Andújar et al. (2014), who recovered a crown age for Carabinae ca. 30 Ma, compared with ca. 110 Ma in this study. Their approach based on Bayes Factor Cluster Analysis led them to select comparatively recent biogeographic and fossil calibrations, whilst simultaneously excluding older ones (i.e., the vicariance between Australia and New Zealand), resulting in very recent age estimates. Relying on a chimeric approach, Sota et al. (2020) constrained the crown of Carabini with a secondary calibration derived from that of Andújar et al. (2014) but also included the vicariance between Australia and New Zealand as a minimum age for the split between *Maoripamborus* and *Pamborus*, resulting in a likely unrealistic stem age for Carabini (ca. 120 Myr). Our results present a more balanced pattern, with a crown age for Carabinae broadly in line with that of Sota et al. (2020) despite our reliance on fossil and secondary calibrations. Correspondingly, our divergence time estimates for the crown of *Carabus* significantly predate those of Andújar et al. (2014) and Sota et al. (2020), who proposed a range of ca. 25–35 Ma. Similarly, our results postdate the estimates of Toussaint and Gillett (2018) who relied on a single deep secondary calibration between *Trachypachus* Motschulsky, 1844 and Carabinae, resulting in older divergence times. However, our estimates are consistent with recent reappraisal of ages in this group based upon new biogeographic and fossil calibrations (Opgenoorth et al. 2020). We believe that our Bayesian relaxed clock dating exercise, based on the most

comprehensive review of the fossil record to date and incorporating recent secondary calibrations obtained from large-scale phylogenomic studies, is more likely to represent a good estimate of Carabinae evolution. Future inferences of comprehensively sampled and robust phylogenomic Adephaga timetrees based on multiple fossil calibrations will eventually unveil the sister group to Carabinae with more certainty.

Efficiency and Future of HyRAD-X-Like Approaches

The HyRAD-X approach has proved efficient in generating a comprehensive genomic data set leading to inference of a new phylogenomic hypothesis for Carabinae (fig. 3). The success of HyRAD-X appears to be a product of its flexibility. For instance, during the hybridization step, it is possible for historical DNA target sequences with some mismatch to nevertheless hybridize to the template probes. Moreover, during the phyloHyRAD bioinformatic mapping step, the use of the more relaxed BWA-MEM mapping algorithm instead of the more stringent BWA-ALN allowed for the recovery of much more data than anticipated (Ziemann 2016). Our strategy of distributing the probe-producing samples across the expected diversity in the phylogeny has made it possible to obtain homologous sequences for species that are phylogenetically distant. One exception was the outgroup *T. carolina*, for which comparatively fewer loci were recovered (258 with BWA-MEM and 197 with BWA-ALN), resulting in reduced phylogenetic resolution (fig. 3).

To investigate factors involved in the recovery of sequences and loci from historical samples, we estimated correlations between the number of sequenced reads or mapped loci and sample specificities. Firstly, we evaluated the influence of the genetic distance to a probe by estimating the phylogenetic distance between each sample and the phylogenetically closest sample used as a probe. A significant correlation was observed between this distance and the number of mapped loci obtained using BWA-ALN. A similar tendency was also observed for the number of loci obtained with BWA-MEM from the fresh samples (in black on fig. 5A) but no correlation was observed for historical samples (in yellow on fig. 5A). These results reveal a limited influence of probe diversity on their ability to hybridize to and recover sequences from fresh samples for which we had obtained DNA of high integrity. This result is in line with results from other approaches such as UCEs where locus recovery decreases with phylogenetic distance from taxa used in probe design (e.g., Faircloth et al. 2015; Gustafson et al. 2020). In contrast, for historical samples containing comparatively fragmented DNA, loci recovery success is likely dependent on other factors, such as the age of the sample or the storage conditions (e.g., when preserved in constant climatic conditions, DNA can still be better preserved), which may induce heterogeneity between samples. In addition, the fact that historical samples are characterized by overall smaller fragments than fresh samples might

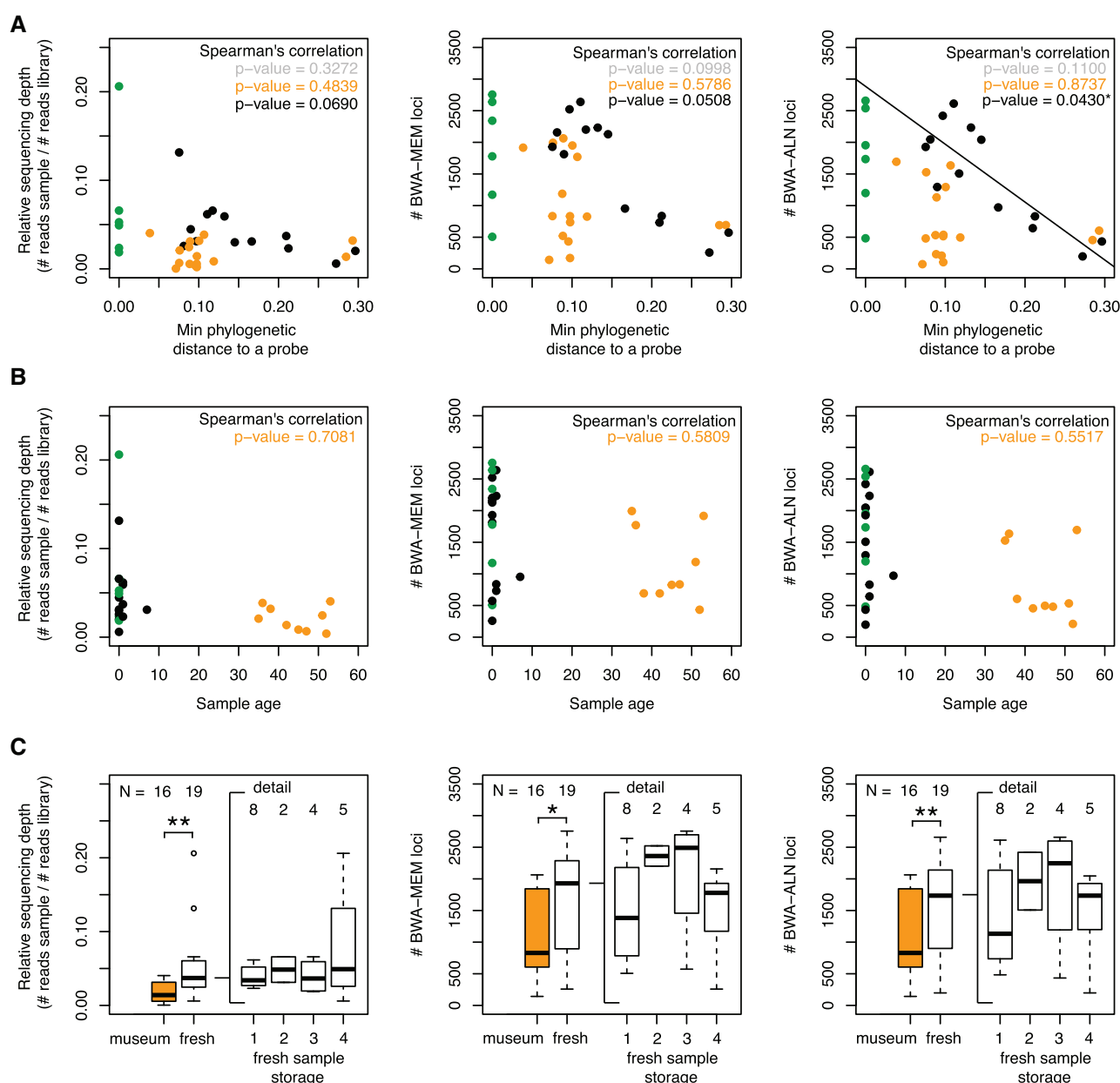


Fig. 5.—Statistical summary of locus recovery. (A) Plot representing the relationship between the minimal phylogenetic distance to a probe and three outcomes: the relative sequencing depth, the number of BWA-MEM mapped loci, and the number of BWA-ALN mapped loci. In each plot, fresh samples are shown in black, fresh samples also used for probes in green, and samples from museums with an age > 30 years in orange. Correlations were tested with Spearman's correlation tests on fresh samples in black, museum samples in orange, and combined fresh and museum samples in gray. (B) Plots representing the relation between the age of the sample and the three previous outcomes. (C) Boxplots comparing the three previous outcomes between the museum samples and the fresh samples (including fresh samples also used for probes). Significance was evaluated using Wilcoxon rank tests. For the fresh samples additional boxplots show variation according to the sample storage history: 1) in absolute ethanol at room temperature, 2) in absolute ethanol at -20°C , 3) dry at -80°C , and 4) in RNAlater at -80°C .

increase the likelihood of hybridization during the capture phase, even to probes that are phylogenetically distant (i.e., reduction of molecular stoichiometry during hybridization). Similarly, for these historical samples, no correlation between the age of the sample and the amount of retrieved genetic data, reads, and loci was observed (in yellow on fig.

5B). This result is surprising because such a correlation is usually recovered when applying capture approaches to museum specimens (e.g., Blaimer et al. 2016; McCormack et al. 2016). Nevertheless, a significant difference exists in the amount of retrieved genetic data, reads, and loci, between fresh and historical samples, with, as

expected, a higher observed recovery success for fresh samples (fig. 5C).

The development of the new phyloHyRAD pipeline (fig. 1) is a major step forward for rapid and accurate locus assembly. This pipeline allows for the construction of a catalog from probe data, the mapping of reads from historical samples, data cleaning, and the construction of consensus sequences for loci that are subsequently combined and aligned to obtain the final alignments. Moreover, this pipeline is particularly flexible because it makes possible the integration of external data from extremely diverse capture methods, including RNA-Seq and UCEs. In this study, it resulted in the acquisition of sufficient loci for successful analysis, with a mean of 1,788 loci (SD 759) obtained with BWA-MEM and 530 loci (SD 421) with BWA-ALN from our samples (table 1). These features ensure that, in combination with the HyRAD-X approach, the pipeline is broadly applicable, allowing for the combination and integration of existing legacy data to further enrich data sets.

Deep sequencing of the libraries resulted in the sequencing of DNA fragments present in large quantities accompanying the captured and amplified fragments in the sample libraries. The HyRAD-X approach allowed for recovery of these high copy number sequences, including mitochondrial and ribosomal DNA, resulting in sequences for 13 mitochondrial genes from our historical samples. When combined with existing mitogenomes available for Carabinae (Sota et al. 2020) and analyzed under a maximum likelihood criterion, these data resulted in the inference of a large mitogenome phylogeny (supplementary figs. S8 and S9, Supplementary Material online), further illustrating the potential for HyRAD-X to enhance data accessibility and analysis possibilities from combining existing next generation or classical barcode sequence data. This tree recovers moderate branch support as well as inconsistent phylogenetic relationships with the analyses based on genome-scale data (fig. 3). This result is not surprising considering the high evolutionary rate of mitochondrial DNA and the potentially corresponding high levels of saturation and homoplasy in older lineages (Rubinoff and Holland 2005; DeSalle et al. 2017). Nevertheless, this phylogeny is likely to represent the most comprehensive one of Carabinae to date with relevant intrageneric phylogenetic placements (supplementary figs. S8 and S9, Supplementary Material online). However, we emphasize that the phylogenomic hypothesis presented in figure 3 is the most robust global subfamily-level estimate to date for the evolutionary history of Carabinae.

Conclusion

The HyRAD-X approach is a powerful and versatile addition to the phylogenomic toolbox, allowing for the generation of large data sets that are compatible with barcoding, mitogenomic, target capture, or RNA-seq data. Introduction of the new phyloHyRAD bioinformatic pipeline ensures that this

approach is even more tractable, efficient, and accessible. Using this combined methodology, we infer a robust dated phylogenomic hypothesis for Carabinae giant ground beetles derived largely from museum collection specimens, thereby demonstrating that genomic-level DNA information preserved in historical specimens can be unlocked and widely exploited in evolutionary studies.

Materials and Methods

Taxon Sampling

We sampled 48 taxa for the purpose of this study, including two outgroups based on recent ground beetle phylogenomic evidence (Gough et al. 2020), namely *Scarites buparius* (Forster, 1771) (Coleoptera, Carabidae, Scaritinae) and the tiger beetle *T. carolina* (Coleoptera, Cicindelidae). The latter was used to root the trees. Fieldwork was conducted in Chile, Corsica, continental France, Sweden, and the United States to gather fresh specimens, which were collected either into 96% ethanol or RNAlater (Thermo Fisher Scientific) for probe design. In total six specimens were used to design the probe set: *Calosoma (Castrida) sayi* Dejean, 1826, *Calosoma (Calosoma) sycophanta* (L., 1758), *Carabus (Platycarabus) irregularis* Fabricius, 1792, *Carabus (Hygrocarabus) variolosus* Fabricius, 1787, *Ceroglossus buqueti* (Laporte, 1834), and *S. buparius*. The rest of the sampled specimens were either dry-pinned or ethanol-preserved museum specimens (supplementary table S1, Supplementary Material online). We selected representatives of all Carabinae genera, as well as most subgenera of *Calosoma*, to test Jeannel's (1940) hypothesis that *Aplothorax* is nested within *Calosoma*, and to establish a robust backbone within that genus. Since one of our aims was to place the St Helena endemic *Aplothorax burchellii* (see Lorenz 2021 for reference on the correct spelling of this species) within the phylogeny of Carabinae, we sampled three historical specimens of this presumably extinct species from the Natural History Museum of Geneva collection. Because the monophyly of most *Calosoma* subgenera is doubtful (Su et al. 2005; Toussaint and Gillett 2018; Sota et al. 2020), we also selected representatives of all the main clades inferred for that genus by Toussaint and Gillett (2018).

RNA Extraction and Probe Preparation

The full HyRAD-X protocol for exome capture as applied to the subfamily Carabinae is available in supplementary file S1, Supplementary Material online. First, RNA was extracted from the six specimens used to design the probe set with a modified version of QIAGEN's RNeasy protocol for Purification of Total RNA from Animal Tissues. Double-stranded cDNA (ds cDNA) was then synthesized, its concentration measured, and quality assessed with Fragment Analyzer (Advanced Analytical Technologies). We used ddRAD (Peterson et al. 2012) following modifications by Mastretta-Yanes et al. (2015). After

testing for digestion profiles using several couples of restriction enzymes, ds cDNA was eventually sheared with MseI and PstI. Two dsDNA adapters were ligated to the digested ds cDNA, each of them harboring an overhang compatible with either MseI or PstI-overhang. One of the adapters (ligated to the PstI-overhang) comprised a T7-promoter sequence necessary for final transcription of probes into RNA. The ligation product was PCR-amplified to obtain libraries compatible with sequencing on Illumina platforms and to add a unique index specific to each specimen used for probe design. A different PCR primer (annealing to the MseI, or P5, end) with a unique 6-bp index was used for further amplification. Libraries were size-selected using PippinPrep (Sage Science) on a 2% agarose cassette (SageScience) in range mode 200–600 bp and then underwent a second round of amplification. The six individual libraries were then sequenced on a lane of an Illumina MiSeq with 300-bp reads. These six libraries were transcribed into RNA and biotinylated in a single reaction using HiScribe T7 High Yield RNA Synthesis Kit (New England Biolabs). We measured the concentrations of each of the six RNA probes in a Qubit assay (Thermo Fisher Scientific) and used these values to prepare an equimolar pool solution for the subsequent hybridization capture step.

DNA Extractions and Shotgun Library Preparation

DNA extractions of one ground-up leg from each specimen were performed using a modified OmniPrep (G-Biosciences) protocol (see [supplementary material, Supplementary Material](#) online). Because specimens of the presumably extinct *A. burchellii* are particularly rare in museum collections, we performed a nondestructive DNA extraction on two specimens. The purified DNA was quantified, and its quality assessed with Fragment Analyzer. Shearing with Fragmentase (New England Biolabs) was performed for fresh samples before library preparation. We used a modified version of the protocol used in [Suchan et al. \(2016\)](#) for the preparation of shotgun libraries, based upon [Tin et al. \(2014\)](#) and [Meyer and Kircher \(2010\)](#), who employed methods specific to library preparation from single-stranded DNA of museum specimens and adapted to the preparation of multiplexed libraries, respectively. Purified DNA was first phosphorylated with T4 Polynucleotide Kinase, then heat-denatured and quickly chilled on an ice and water mix. G-tailing was performed with Terminal Transferase and second strand DNA was synthesized with Klenow Fragment (3'→5'exo-) using a poly-C oligonucleotide. Blunt-end reaction was performed with T4 DNA Polymerase and barcoded adapters were ligated to the phosphorylated end (opposite from the poly-C end). After adapter fill-in with *Bst* DNA Polymerase (Large Fragment), two PCR replicates were run independently using Phusion U Hot Start DNA Polymerase (Thermo Scientific) and indexed PCR primers. The two PCR shotgun libraries replicates were pooled together, purified, and quantified in a Picogreen

assay (Invitrogen). Libraries were pooled in equimolar quantities based upon their respective concentrations.

Hybridization Capture and Sequencing

Hybridization capture for enrichment of shotgun libraries was based on the MYbaits protocol (Arbor Biosciences), but with a two-step capture at different temperatures as suggested by [Li et al. \(2013\)](#). Here, we used an inverse-touchdown approach, with a first capture at 50°C, whose product was then captured a second time at 65°C to improve the stringency of the reaction (Orlando L, personal communication). Two independent library sequencing runs were performed on Illumina NovaSeq SP (Fasteris, Switzerland) and Illumina HiSeq2500 sequencers, using a paired-end protocol on rapid run mode (Lausanne Genomic Technologies Facility, Switzerland).

phyloHyRAD Pipeline

The first step of the phyloHyRAD pipeline (<https://github.com/JeremyLGauthier/PHYRAD/tree/master/phyloHyRAD>) (fig. 1) is the construction of a loci reference catalog. The sequence pairs generated from the ddRAD probe libraries were cleaned using AdapterRemovalv2 ([Schubert et al. 2016](#)) and Cutadapt ([Martin 2011](#)) to remove adaptors, bases with a quality score lower than 20 and reads smaller than 30 bp. Read quality was first checked using FastQC (Babraham Institute) before loci construction was performed using ipyrad ([Eaton and Overcast 2020](#)) with a minimum depth of 6 and a clustering threshold of 0.70 (following testing with values 0.70, 0.80, and 0.90). Ultimately, loci shared by at least two probes were retained in a reference catalog, which was evaluated for contamination using the metagenomic sequence classifier Centrifuge ([Kim et al. 2016](#)). Contrary to UCE analyses using phyluce ([Faircloth 2016](#)), the phyloHyRAD pipeline does not perform individual loci assembly from historical samples but capitalizes on the sequencing of probe sequences to generate a reliable catalog. During this step, ipyrad considers and removes putative paralogous loci if more than 50% of the shared SNP are heterozygous, parameter “max_shared_Hs_locus.” Reads from each historical sample were trimmed and cleaned using Cutadapt ([Martin 2011](#)) to remove barcodes, adaptors, bases with a quality lower than 20, and reads smaller than 30 bp. Terminal poly-Cs were removed using a custom Perl script (DropBpFastq_polyC.pl) before read quality was checked using FastQC.

Cleaned reads from each historical sample were individually mapped onto the loci catalog generated above using both a strict algorithm, BWA-ALN ([Li and Durbin 2009](#)), and a more relaxed one, BWA-MEM ([Li 2013](#)). Indels were realigned using the GATK IndelRealigner ([McKenna et al. 2010](#)) and PCR duplicates were removed using MarkDuplicates from the Picard toolkit (<http://broadinstitute.github.io/picard>). Nucleotide mis-incorporation patterns were investigated using MapDamage2.0 ([Jónsson et al. 2013](#)), and base quality

scores were rescaled according to their probability of representing a postmortem DNA deamination event, to reduce the impact of DNA decay on downstream analyses. Finally, individual consensus sequences were generated for each locus using the following scripts from samtools suite: mpileup, bcftools, and vcfutils.pl (Li et al. 2009). Consensus sequences were cleaned using seqtk (<https://github.com/lh3/seqtk>) to retain bases with a phred quality > 30. Cleaned consensus sequences were combined using a custom script and aligned with MAFFT using the –auto option to automatically select an appropriate strategy for the alignment (Katoh and Standley 2013).

Integration of External Data

Raw reads from eight additional taxa were gathered from the SRA database and other online repositories to test the compatibility of HyRAD-X data with established phylogenomic data. Specifically, we collected raw reads from the following species: *Calosoma* (*Calosoma*) *frigidum* Kirby, 1837 (transcriptome, SRR2083640), *Carabus* (*Carabus*) *granulatus* L., 1758 (transcriptome, SRR596983), *Carabus* (*Ohomopterus*) *iwawakianus* (Nakane, 1953) (transcriptome, DRR089198), *Carabus* (*Tanaocarabus*) *taedatus* (UCEs, SRR10334070), *Carabus* (*Ohomopterus*) *uenoi* Ishikawa, 1960 (transcriptome, DRR089202), *Carabus* (*Megodontus*) *violaceus* L., 1758 (transcriptome, SRR10675209), *Elaphrus* *aureus* Müller, 1821 (Coleoptera, Carabidae, Elaphrinae) (transcriptome, SRR2083660) and *Pasimachus* *viridans* LeConte, 1858 (Coleoptera, Carabidae, Scaritinae) (transcriptome, <https://doi.org/10.5061/dryad.8wv9ghx3h9>). These raw reads were integrated directly into the phyloHyRAD pipeline at the mapping step and treated in the same manner as the other samples, apart from omitting the MapDamage step because these data derived from nondegraded fresh samples.

Alignment Post-treatment

All assembled loci recovered from the relaxed algorithm were further cleaned in Geneious R11 (Biomatters) to remove short or problematic sequences. We relied on four main nucleotide data sets to infer phylogenetic relationships among Carabinae lineages. The first data set was composed of 2,945 loci processed using the relaxed BWA-MEM algorithm (see above) and 43 taxa (Data set A). Several taxa had a very low number of loci and were removed to prevent spurious phylogenetic inferences (i.e., specimens with <5 kb of data). A second data set was derived from the Data set A, with the exception that taxa represented by less than 20 kb of data were pruned, resulting in a final matrix of 40 taxa (Data set B). The third data set was produced using all recovered and clean loci from the strict algorithm, BWA-ALN, for a total of 1,227 loci and 40 taxa (Data set C). A fourth data set was generated with the loci selected for the BEAST analyses (Data set D, see below) to perform the divergence time analyses. For each data set, loci

were then concatenated using AMAS (Borowiec 2016) to produce global alignments and partition files. Evaluation of the capture efficiency and its influence on the final data sets was performed by estimating the proportion of parsimony-informative sites across the entire alignments using AMAS and for each sample using FASconCAT-G (Kück and Longo 2014). Complementary alignment statistics have been estimated on each data set using Alistat (Wong et al. 2020). The phylogenetic distance between historical samples and probes was estimated using distTips from the adephylo package (Jombart et al. 2010). To investigate the impact of sample history on data recovery, relative sequencing depth (i.e., number of reads for each sample divided by the number of reads in the library), number of loci for each method and sample characteristics were analyzed and Spearman's correlations were performed and plotted using R (R Project).

Recovery of Mitochondrial Genes from HyRAD-X Data

Mitochondrial sequences from the cleaned reads of each historical sample were assembled using MitoFinder (Allio et al. 2020), using the following publicly available mitochondrial sequences from GenBank as references: *Calosoma* sp. (GU176340), *C. granulatus* (MN122870), *S. buparius* (MF497822), and *Tetracha* sp. (MG253284). We thereafter built an additional data set (Data set E), combining these new mitogenomic data with the mitogenome data set created by Sota et al. (2020), resulting in a data matrix consisting of a total of 60 taxa, 14 loci (all protein-coding mitochondrial genes in addition to the 16S rRNA locus), corresponding to ca. 12 kbp.

Phylogenetic Inference

All phylogenetic analyses using the different nucleotide alignments were performed in IQ-TREE 2.0.5 (Minh et al. 2020) using the edge-linked partition model (Chernomor et al. 2016). Because simultaneously estimating the best models of nucleotide substitution and partitioning schemes was too computationally demanding, we used PartitionFinder 2.1.1 to first estimate the best partitioning schemes a priori, beginning with one partition per locus for all data sets. To conduct tractable analyses on a local cluster and on the CIPRES Science Gateway cluster (Miller et al. 2010), we used the rcluster algorithm under the Akaike Information Criterion corrected (AICc), with rcluster-max = 1,000 and rcluster-percent = 20. The resulting partitioning schemes were then used in IQ-TREE to select corresponding models of nucleotide substitution using ModelFinder (Kalyaanamoorthy et al. 2017) and the AICc across all available models in IQ-TREE. To avoid local optima, we performed 100 independent tree searches for each data set in IQ-TREE with default options. To estimate branch support, we calculated 1,000 ultrafast bootstraps along with 1,000 SH-aLRT tests in IQ-TREE (Guindon et al. 2010; Hoang et al. 2018). We used the hill-climbing nearest-

neighbor interchange topology search strategy to avoid severe model violations leading to biased ultrafast bootstrap estimations (Hoang et al. 2018). We also conducted a separate analysis of the preferred Data set B (see Results) with 100 regular bootstraps to assess branch support. The best tree for each analysis was selected based on the comparison of maximum likelihood scores. Finally, we performed a RogueNaRok analysis (Aberer et al. 2013) to detect potential rogue taxa for each data set using the best-scoring ML tree and the 1,000 ultrafast bootstrap with default options.

Divergence Time Estimation

Divergence time estimation using RAD-sequencing data is complex not only because of the large number and short lengths of recovered RAD loci but also because of the large percentage of inherent missing data. Downsizing the initial data set is therefore necessary to facilitate divergence time computation in a probabilistic framework. To generate a data set that would be tractable for Bayesian inference of divergence times using relaxed clocks, we first excluded all loci not represented in at least 20 taxa and lacking a minimum length of 200 bp. From the initial set of 2,945 loci (BWA-MEM), this first limiting step resulted in a pool of 558 candidate loci. We then estimated phylogenetic trees for each resulting locus using IQ-TREE and a model of nucleotide substitution selected using ModelFinder. These phylogenetic trees and the best scoring tree from the species tree ML inference conducted with Data set A (see above) were used to conduct a gene shopping approach as developed by Smith et al. (2018) in SortaDate, using some underlying UNIX code and programs implemented in phyx (Brown et al. 2017). RAD loci were filtered using the following three criteria: 1) clock-likeness, 2) tree length, and 3) least topological conflict with the species tree. We selected the 100 best scoring RAD loci based on this filtering and concatenated them to produce the final alignment used in the divergence time estimation.

The Carabinae fossil record is relatively scarce. To calibrate the relaxed clocks implemented in BEAST, we relied on a selection of both fossil and secondary calibration while carefully avoiding circularity (i.e., the retained fossils were not used to obtain the ages that were in turn used as secondary calibrations). We used two fossils of Carabinae to calibrate the clocks using soft exponential prior distributions. Firstly, we used †*Calosoma agassizi* Barthélemy-Lapommeraye 1846 (ca. 27.8–33.9 Ma, Rupelian) from France (Nel 1988), hitherto surprisingly overlooked despite a series of well-preserved specimens. This fossil cannot be placed with certainty within *Calosoma* and is therefore used here as a crown calibration considering the comprehensive sampling of *Calosoma* in this study. An older and recently described fossil possibly belonging to *Calosoma* from the Vic Formation (ca. 33.9–37.8 Ma, Priabonian) in Spain (Farrés and Altimiras 2012), was not

used because it did not provide a significantly different calibration and was less reliable based on morphological grounds. Secondly, we used *Carabus (Ohomopterus)* sp. from the Tatsumi-tôge formation in Tottori Prefecture, Japan (Hiura 1971); the fossil elytra have already been used by Deuve et al. (2012) to constrain the stem of *Carabus (Ohomopterus)* Reitter, 1896, and therefore we used the same placement in our analyses. Other good *Calosoma* and *Carabus* fossils exist (e.g., Deuve 1998; Yahiro et al. 2018; Kirejtshuk et al. 2019) but were not relevant considering the above-mentioned fossils and the taxon sampling in our final phylogeny (see Results). The three other tribes, Ceroglossini, Cydrini and Pamborini are not represented by any reliable usable fossils described to date, with many older descriptions being inaccurate (Nel A, personal communication). To avoid divergence time estimation biases, we included two secondary calibrations derived from the recent dated transcriptomic tree of beetles constructed by McKenna et al. (2019). Specifically, we used the crown ages of Cicindelidae+Carabidae (158 Ma, 95% CI = 138–184 Ma) and of Carabidae (121 Ma, 95% CI = 97–143 Ma) to constrain the two corresponding nodes in our topology with lognormal priors spanning the 95% credibility intervals of the estimates from McKenna et al. (2019). Importantly, we therefore avoided circularity by selecting secondary calibration from a study (McKenna et al. 2019) that did not use the two fossils we used to estimate divergence times.

All analyses were performed in BEAST 1.10.4 (Suchard et al. 2018). The best partitioning scheme and models of substitution were determined with PartitionFinder2 (Lanfear et al. 2017) using the rclusterf algorithm with parameters rcluster-max = 1,000, rcluster-percent = 20 and min-subset-size = 200, and the Bayesian Information Criterion algorithm to select between competing models. Because this algorithm relies on only three models (GTR, GTR+G, and GTR+I + G), we re-estimated a posteriori the best models using all those included in BEAST. The data set was partitioned a priori by locus for a total of 100 initial partitions. We implemented clock partitioning by conducting analyses with 1) a single clock for all partitions and 2) one clock for each partition (seven in total, see Results). We assigned a Bayesian lognormal relaxed clock model to the different clock partitions. We also tested different tree models by using a Yule (pure birth) or a birth–death model. The analyses consisted of 100 million generations with parameter and tree sampling every 5,000 generations. We estimated marginal likelihood estimates (MLE) for each analysis using path-sampling and stepping-stone sampling (Baele et al. 2012), with 1,000 path steps, and chains running for 1 million generations with a log-likelihood sampling every 1,000 cycles. The Maximum Clade Credibility tree of each analysis with median divergence age estimates were generated in TreeAnnotator 1.10.4 (Suchard et al. 2018) after removing the first 25 million generations as burn-in.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

The authors are indebted to David Seldon (University of Auckland, Auckland, New Zealand) and Geoff Monteith (Queensland Museum, Brisbane, Australia) for sending critical specimens used in this study. They thank Louis Toussaint (Bazemont, France) for assistance in fieldwork, and Michael Balke (ZSM-SNSB, Munich, Germany) for sending samples of specimens housed in the ZSM collections. They also warmly thank André Nel (MNHN, Paris, France) for discussions and details on the fossil fauna of Carabinae ground beetles. They thank Lily Kumpe for permission to reproduce one of her photographs. They also thank four anonymous reviewers and the editor Mario dos Reis Barros for insightful suggestions. The authors declare that there is no conflict of interest.

Data Availability

The data underlying this article (final alignments, trees, and input files) are available at: <https://doi.org/10.5281/zenodo.4748993>. Raw reads are available on the SRA repository BioProject ID PRJNA706141. The whole pipeline including custom scripts are available at the Github repository <https://github.com/JeremyLGauthier/PhyRAD/tree/master/phyloHyRAD>.

Literature Cited

- Aberer AJ, Krompass D, Stamatakis A. 2013. Pruning rogue taxa improves phylogenetic accuracy: an efficient algorithm and webservice. *Syst Biol*. 62(1):162–166.
- Ali OA, et al. 2016. RAD capture (Rapture): flexible and efficient sequence-based genotyping. *Genetics* 202:389–400.
- Allio R, et al. 2020. MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. *Mol Ecol Res*. 20:892–905.
- Andújar C, Gomez-Zurita J, Rasplus JY, Serrano J. 2012a. Molecular systematics and evolution of the subgenus *Mesocarabus* Thomson, 1875 (Coleoptera: Carabidae: *Carabus*), based on mitochondrial and nuclear DNA. *Zool J Linn Soc*. 166(4):787–804.
- Andújar C, Serrano J, Gómez-Zurita J. 2012b. Winding up the molecular clock in the genus *Carabus* (Coleoptera: Carabidae): assessment of methodological decisions on rate and node age estimation. *BMC Evol Biol*. 12:40.
- Andújar C, Soria-Carrasco V, Serrano J, Gómez-Zurita J. 2014. Congruence test of molecular clock calibration hypotheses based on Bayes factor comparisons. *Methods Ecol Evol*. 5(3):226–242.
- Baele G, et al. 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol Biol Evol*. 29:2157–2167.
- Baker PE. 1973. Islands of the South Atlantic. In: Nairn AEM, Stehli FG, editors. *The ocean basins and margins*, vol. 1, The South Atlantic. New York (NY), London: Plenum Press. p. 493–553.
- Bi K, et al. 2013. Unlocking the vault: next-generation museum population genomics. *Mol Ecol*. 22:6018–6032.
- Blaimer BB, Lloyd MW, Guillory WX, Brady SG. 2016. Sequence capture and phylogenetic utility of genomic ultraconserved elements obtained from pinned insect specimens. *PLoS One* 11(8):e0161531.
- Blair C, Ané C. 2020. Phylogenetic trees and networks can serve as powerful and complementary approaches for analysis of genomic data. *Syst Biol*. 69:593–601.
- Borowiec ML. 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* 4:e1660.
- Brown JW, Walker JF, Smith SA. 2017. Phyx: phylogenetic tools for unix. *Bioinformatics* 33:1886–1888.
- Bruschi S. 2013. *Calosoma of the world: Coleoptera, Carabidae*. Bologna: Natura edizioni scientifiche.
- Burrell AS, Disotell TR, Bergery CM. 2015. The use of museum specimens with high-throughput DNA sequencers. *J Hum Evol*. 79:35–44.
- Carpenter M, et al. 2013. Pulling out the 1%: Whole-Genome Capture for the Targeted Enrichment of Ancient DNA Sequencing Libraries. *Am J Hum Genet*. 93(5):852–864.
- Cavazzuti P. 1989. *Monografia del Genere Procerus* (Coleoptera, Carabidae, Carabini). Savigliano: Edizione L'Artistica.
- Cavazzuti P. 2010. World catalogue of the tribe Cychrini (Coleoptera, Carabidae, Carabinae). *Ann Mus Civ Stor Nat Giacomo Doria*. 102:203–294.
- Chatenet G. D. 1986. *Guide des coléoptères d'Europe*. Neuchâtel-Paris: Delachaux et Niestlé.
- Chernomor O, Von HA, Minh BQ. 2016. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst Biol*. 65:992–1008.
- Colella JP, Tiago A, MacManes MD. 2020. A linked-read approach to museomics: higher quality de novo genome assemblies from degraded tissues. *Mol Ecol Resour*. 20:856–870.
- Crotti M, Barratt CD, Loader SP, Gower DJ, Streicher JW. 2019. Causes and analytical impacts of missing data in RADseq phylogenetics: insights from an African frog (*Afrixalus*). *Zool Scr*. 48(2):157–167.
- Cruaud A, et al. 2014. Empirical assessment of RAD sequencing for interspecific phylogeny. *Mol Biol Evol*. 31:1272–1274.
- Darwin C. 1845. *Journal of researches into the natural history and geology of the countries visited during the voyage of H.M.S. Beagle round the world, under the Command of Capt. Fitz Roy, R.N.* 2nd ed. London: John Murray.
- Dell'Ampio E, et al. 2014. Decisive data sets in phylogenomics: lessons from studies on the phylogenetic relationships of primarily wingless insects. *Mol Biol Evol*. 31(1):239–249.
- Deng J, Guo Y, Cheng Z, Lu C, Huang X. 2019. The prevalence of single-specimen/locality species in insect taxonomy: an empirical analysis. *Diversity* 11(7):106–114.
- DeSalle R, Schierwater B, Hadrys H. 2017. MtDNA: the small workhorse of evolutionary studies. *Front Biosci*. 22:873–887.
- Deuve T. 1998. Trois fossiles remarquablement conservés du Miocène de France, appartenant aux genres *Carabus* L. et *Ledouxnebria* nov. (Coleoptera, Carabidae et Nebriidae). *Bull Soc Entomol Fr*. 103:229–236.
- Deuve T. 2019. Classification du genre *Carabus* L., 1758. *Liste Blumenthal 2018–2019* (Coleoptera, Carabidae). *Coléoptères* 25:33–102.
- Deuve T, Cruaud A, Genson G, Rasplus JY. 2012. Molecular systematics and evolutionary history of the genus *Carabus* (Col. Carabidae). *Mol Phylogenet Evol*. 65:259–275.
- Deuve T, Faille A. 2013. Structure secondaire de l'ARN 18S et phylogénie basale du genre *Carabus* L., 1758 (Coleoptera: Carabidae). *Ann Soc Entomol Fr (N.S.)*. 49(4):430–445.
- Eaton DA, Overcast I. 2020. ipyrad: interactive assembly and analysis of RADseq datasets. *Bioinformatics* 36(8):2592–2594.
- Fabre J-H. 1901. *Souvenirs entomologiques, études sur l'instinct et les moeurs des insectes*, Septième Série. Paris: Librairie Ch. Delagrave.

- Faircloth BC. 2016. PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics* 32(5):786–788.
- Faircloth BC, Branstetter MG, White ND, Brady SG. 2015. Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among Hymenoptera. *Mol Ecol Resour*. 15(3):489–501.
- Faircloth BC, et al. 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Syst Biol*. 61(5):717–726.
- Farrés F, Altimiras J. 2012. Insectos eocenos de la comarca de Vic (Barcelona). *Nota preliminar. Batalleria (Barcelona) Rev Paleontol*. 17:73–79.
- Fujisawa T, Sasabe M, Nagata N, Takami Y, Sota T. 2019. Genetic basis of species-specific genitalia reveals role in species diversification. *Sci Adv*. 5(6):eaav9939.
- Gauthier J, et al. 2020. Museomics identifies genetic erosion in two butterfly species across the 20th century in Finland. *Mol Ecol*. 20(5):1191–1205.
- Gough HM, Allen JM, Toussaint EFA, Storer CG, Kawahara AY. 2020. Transcriptomics illuminate the phylogenetic backbone of tiger beetles. *Biol J Linn Soc*. 129(3):740–751.
- Gray A, et al. 2019. The status of the invertebrate fauna on the South Atlantic island of St Helena: problems, analysis, and recommendations. *Biodivers Conserv*. 28(2):275–296.
- Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 59:307–321.
- Gustafson GT, Baca SM, Alexander AM, Short AE. 2020. Phylogenomic analysis of the beetle suborder Adephaga with comparison of tailored and generalized ultraconserved element probe performance. *Syst Entomol*. 45(3):552–570.
- Hiura I. 1971. Discovery of *Apotomopterus* from the Japanese Neogene (Coleoptera: Carabidae). *Occas Pap Osaka Mus Nat Hist*. 1:45–50.
- Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol*. 35:518–522.
- Hoffberg SL, et al. 2016. RAD cap: sequence capture of dual-digest RAD seq libraries with identifiable duplicates and reduced missing data. *Mol Ecol Resour*. 16(5):1264–1278.
- Hosner PA, Faircloth BC, Glenn TC, Braun EL, Kimball RT. 2016. Avoiding missing data biases in phylogenomic inference: an empirical study in the landfowl (Aves: Galliformes). *Mol Biol Evol*. 33(4):1110–1125.
- Huang H, Knowles LL. 2016. Unforeseen consequences of excluding missing data from next-generation sequences: simulation study of RAD sequences. *Syst Biol*. 65:357–365.
- Imura Y, et al. 2018. Evolutionary history of Carabid ground beetles with special reference to morphological variations of the hind-wings. *Proc Jpn Acad Ser B Phys Biol Sci*. 94:360–371.
- Jeannel R. 1940. Les Calosomes. *Mém Mus Natl Hist Nat Sér A Zool*. 13:1–240.
- Jin M, Zwick A, Ślipiński A, de Keyser R, Pang H. 2020. Museomics reveals extensive cryptic diversity of Australian prionine longhorn beetles with implications for their classification and conservation. *Syst Entomol*. 45(4):745–770.
- Jiroux E. 2006. Le genre *Ceroglossus*, collection systématique. Vol. 14. Verneuil-sur-Seine: Magellanes.
- Jombart T, Balloux F, Dray S. 2010. Adephylo: new tools for investigating the phylogenetic signal in biological traits. *Bioinformatics* 26(15):1907–1909.
- Jones MR, Good JM. 2016. Targeted capture in evolutionary and ecological genomics. *Mol Ecol*. 25:185–202.
- Jónsson H, Ginolhac A, Schubert M, Johnson PL, Orlando L. 2013. mapDamage2. 0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29(13):1682–1684.
- Kalyaanamoorthy S, Minh BQ, Wong TK, von Haeseler A, Jermin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 14:587.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30(4):772–780.
- Kim CG, Tominaga O, Su ZH, Osawa S. 2000. Differentiation within the genus *Leptocarabus* (excl. *L. kurilensis*) in the Japanese Islands as deduced from mitochondrial ND5 gene sequences (Coleoptera, Carabidae). *Genes Genet Syst*. 75:335–342.
- Kim CG, et al. 2000. Pattern of morphological diversification in the *Leptocarabus* ground beetles (Coleoptera: Carabidae) as deduced from mitochondrial ND5 gene and nuclear 28S rDNA sequences. *Mol Biol Evol*. 17:137–145.
- Kim D, Song L, Breitwieser FP, Salzberg SL. 2016. Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Res*. 26(12):1721–1729.
- Kirejtshuk AG, et al. 2019. The beetle (Coleoptera) fauna of the Insect Limestone (late Eocene), Isle of Wight, southern England. *Earth Environ Sci Trans R Soc Edinb*. 110(3–4):405–492.
- Knyshov A, Gordon ER, Weirauch C. 2019. Cost-efficient high throughput capture of museum arthropod specimen DNA using PCR-generated baits. *Methods Ecol Evol*. 10(6):841–852.
- Kück P, Longo GC. 2014. FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Front Zool*. 11(1):81.
- Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. 2017. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol Biol Evol*. 34:772–773.
- Leaché AD, et al. 2015. Phylogenomics of phrynosomatid lizards: conflicting signals from sequence capture versus restriction site associated DNA sequencing. *Genome Biol Evol*. 7(3):706–719.
- Lemmon AR, Emme SA, Lemmon EM. 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. *Syst Biol*. 61(5):727–744.
- Li C, Hofreiter M, Straube N, Corrigan S, Naylor GJ. 2013. Capturing protein-coding genes across highly divergent species. *Biotechniques* 54:321–326.
- Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv Preprint*. arXiv:1303.3997.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25(14):1754–1760.
- Li H, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
- Lorenz W. 2021. CarabCat: Global database of ground beetles (version Oct 2017). In: Catalogue of Life, Roskov Y, et al., editors. Species 2000 & ITIS catalogue of life, 2021-05-07. Species 2000: Naturalis, Leiden, the Netherlands. ISSN 2405–8858. Digital resource at www.catalogueoflife.org.
- Manthey JD, Campillo LC, Burns KJ, Moyle RG. 2016. Comparison of target-capture and restriction-site associated DNA sequencing for phylogenomics: a test in cardinalid tanagers (Aves, Genus: *Piranga*). *Syst Biol*. 65(4):640–650.
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*. 17(1):10–12.
- Mastretta-Yanes A, et al. 2015. Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Mol Ecol Resour*. 15:28–41.
- Mayer C, et al. 2016. BaitFisher: a software package for multispecies target DNA enrichment probe design. *Mol Biol Evol*. 33(7):1875–1886.
- McCormack JE, Tsai WL, Faircloth BC. 2016. Sequence capture of ultraconserved elements from bird museum specimens. *Mol Ecol Resour*. 16(5):1189–1203.

- McKenna A, et al. 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20(9):1297–1303.
- McKenna DD, et al. 2019. The evolution and genomic basis of beetle diversity. *Proc Natl Acad Sci USA.* 116:24729–24737.
- Meyer M, Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc.* 2010(6):t5448.
- Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: *Proceedings of the Gateway Computing Environments Workshop (GCE)*; 2010 Nov 14. New Orleans (LA): IEEE. p. 1–8.
- Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res.* 17:240–248.
- Minh BQ, et al. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol.* 37:1530–1534.
- Muñoz-Ramírez C. 2015. The phylogenetic position of *Ceroglossus ochsenii* GERMAIN and *Ceroglossus guerini* GERMAIN (Coleoptera: Carabidae), two endemic ground beetles from the Valdivian forest of Chile. *Rev Chil Entomol.* 40:19–26.
- Muñoz-Ramírez CP, Bitton J-P, Doucet SM, Knowles LL. 2016. Mimics here and there, but not everywhere: müllerian mimicry in *Ceroglossus* ground beetles? *Biol Lett.* 12(9):20160429.
- Neall VE, Trewick SA. 2008. The age and origin of the Pacific islands: a geological overview. *Philos Trans R Soc B.* 363:3293–3308.
- Nel A. 1987. Sur la présence du genre *Carabus* L. (1758) dans les calcaires du Stampien de Cereste (Alpes-de-Haute-Provence) (Coleoptera Carabidae). *L'Entomologiste* 43:243–246.
- Nel A. 1988. Les Calosomes fossiles de l'Oligocène du sud-est de la France [Col. Carabidae]. *Bull Soc Entomol Fr.* 93:257–268.
- Novacek MJ. 1992. Fossils, topologies, missing data, and the higher level phylogeny of eutherian mammals. *Syst Biol.* 41(1):58–73.
- Okamoto M, Kashiwai N, Su ZH, Osawa S. 2001. Sympatric convergence of the color pattern in the Chilean *Ceroglossus* ground beetles inferred from sequence comparisons of the mitochondrial ND5 gene. *J Mol Evol.* 53:530–538.
- Opgenoorth L, Hofmann S, Schmidt J. 2020. Rewinding the molecular clock in the genus *Carabus* (Coleoptera: Carabidae): Revisiting Andujar et al. in light of new fossil evidence and the Gondwana split. *bioRxiv* doi:10.1101/2020.02.19.912543.
- Osawa S, Su ZH, Imura Y. 2004. Phylogeny and distribution of the subfamily Carabinae. In: Osawa S, Su ZH, Imura Y, editors. *Molecular phylogeny and evolution of Carabid ground beetles*. Tokyo: Springer. p. 25–32.
- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* 7(5):e37135.
- Peyve AA. 2011. Seamounts in the east of South Atlantic: origin and correlation with Mesozoic-Cenozoic magmatic structures of West Africa. *Geotecton* 45(3):195–209.
- Prüser F, Mossakowski D. 1998. Conflicts in phylogenetic relationships and dispersal history of the subtribe Carabinae (Coleoptera: Carabidae). In: Ball GE, Casale A, Vigna Taglianti A, editors. *Phylogeny and classification of Caraboidea* (Coleoptera: Adephaga). Torino: Museo Regionale di Scienze Naturali. p. 297–328.
- Roff DA. 1990. The evolution of flightlessness in insects. *Ecol Monogr.* 60(4):389–421.
- Ruane S, Austin CC. 2017. Phylogenomics using formalin-fixed and 100+ year-old intractable natural history specimens. *Mol Ecol Resour.* 17(5):1003–1008.
- Rubin BE, Ree RH, Moreau CS. 2012. Inferring phylogenies from RAD sequence data. *PLoS One* 7(4):e33394.
- Rubinfeld D, Holland BS. 2005. Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. *Syst Biol.* 54(6):952–961.
- Sann M, et al. 2018. Phylogenomic analysis of Apoidea sheds new light on the sister group of bees. *BMC Evol Biol.* 18(1):1–15.
- Schellart WP, Lister GS, Toy VG. 2006. A Late Cretaceous and Cenozoic reconstruction of the Southwest Pacific region: tectonics controlled by subduction and slab rollback processes. *Earth-Sci Rev.* 76(3–4):191–233.
- Schmid S, et al. 2017. HyRAD-X, a versatile method combining exome capture and RAD sequencing to extract genomic information from ancient DNA. *Methods Ecol Evol.* 8(10):1374–1388.
- Schmid S, et al. 2018. Spatial and temporal genetic dynamics of the grasshopper *Oedaleus decorus* revealed by museum genomics. *Ecol Evol.* 8(3):1480–1495.
- Schubert M, et al. 2012. Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics.* 13(1):178.
- Schubert M, Lindgreen S, Orlando L. 2016. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res Notes.* 9(1):1–7.
- Smith SA, Brown JW, Walker JF. 2018. So many genes, so little time: a practical approach to divergence-time estimation in the genomic era. *PLoS One* 13(5):e0197433.
- Sota T, Ishikawa R. 2004. Phylogeny and life-history evolution in *Carabus* (subtribe Carabina: Coleoptera, Carabidae) based on sequences of two nuclear genes. *Biol J Linn Soc.* 81(1):135–149.
- Sota T, Takami Y, Monteith GB, Moore BP. 2005. Phylogeny and character evolution of endemic Australian Carabid beetles of the genus *Pamborus* based on mitochondrial and nuclear gene sequences. *Mol Phylogenet Evol.* 36:392–404.
- Sota T, Vogler AP. 2001. Incongruence of mitochondrial and nuclear gene trees in the Carabid beetles *Ohomopterus*. *Syst Biol.* 50:39–59.
- Sota T, Vogler AP. 2003. Reconstructing species phylogeny of the Carabid beetles *Ohomopterus* using multiple nuclear DNA sequences: heterogeneous information content and the performance of simultaneous analyses. *Mol Phylogenet Evol.* 26:132–154.
- Sota T, et al. 2020. The origin of the giant ground beetle *Aplothorax burchelli* on St Helena Island. *Biol J Linn Soc.* 131(1):50–60.
- St Laurent RA, Hamilton CA, Kawahara AY. 2018. Museum specimens provide phylogenomic data to resolve relationships of sack-bearer moths (Lepidoptera, Mimatonoidea, Mimatonoidea). *Syst Entomol.* 43(4):729–761.
- Staats M, et al. 2013. Genomic treasure troves: complete genome sequencing of herbarium and insect museum specimens. *PLoS One* 8(7):e69189.
- Streicher JW, Schulte JA, Wiens JJ. 2016. How should genes and taxa be sampled for phylogenomic analyses with missing data? An empirical study in iguanian lizards. *Syst Biol.* 65(1):128–145.
- Su ZH, Imura Y, Kim CG, Okamoto M, Osawa S. 2003. Phylogenetic relationships in the division Lipastromorpha (Coleoptera, Carabidae) of the world as deduced from mitochondrial ND5 gene sequences. *Genes Genet Syst.* 78(1):37–51.
- Su ZH, Imura Y, Okamoto M, Osawa S. 2004. Pattern of phylogenetic diversification of the Cychrini ground beetles in the world as deduced mainly from sequence comparisons of the mitochondrial genes. *Gene* 326:43–57.
- Su ZH, Imura Y, Osawa S. 2005. Evolutionary history of Calosomina ground beetles (Coleoptera, Carabidae, Carabinae) of the world as deduced from sequence comparisons of the mitochondrial ND5 gene. *Gene* 360:140–150.

- Su ZH, et al. 1996a. Phylogenetic relationships and evolution of the Japanese Carabinae ground beetles based on mitochondrial ND5 gene sequences. *J Mol Evol.* 42:124–129.
- Su ZH, et al. 1996b. Geography-linked phylogeny of the *Damaster* ground beetles inferred from mitochondrial ND5 gene sequences. *J Mol Evol.* 42(2):130–134.
- Su ZH, et al. 2004. Phylogeny and evolution of Digitulati ground beetles (Coleoptera, Carabidae) inferred from mitochondrial ND5 gene sequences. *Mol Phylogenet Evol.* 30(1):152–166.
- Suchan T, et al. 2016. Hybridization capture using RAD probes (hyRAD), a new tool for performing genomic analyses on collection specimens. *PLoS One* 11(3):e0151651.
- Suchard MA, et al. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* 4(1):vey016.
- Takahashi T, Nagata N, Sota T. 2014. Application of RAD-based phylogenetics to complex relationships among variously related taxa in a species flock. *Mol Phylogenet Evol.* 80:137–144.
- Takami Y, Sota T. 2006. Four new species of the Australian Pamborus Latreille (Coleoptera, Carabidae) carabid beetles. *Aust J Entomol.* 45(1):44–54.
- Takami Y. 2000. Phylogeny of the subgenus *Ohomopter* (Coleoptera, Carabidae, genus Carabus): a morphological aspect. *Tokyo Metrop Univ Bull Nat Hist.* 4:12–32.
- Tin MM-Y, Economo EP, Mikhayev AS. 2014. Sequencing degraded DNA from non-destructively sampled museum specimens for RAD-tagging and low-coverage shotgun phylogenetics. *PLoS One* 9(5):e96793.
- Tominaga O, et al. 2000. Formation of the Japanese Carabina fauna inferred from a phylogenetic tree of mitochondrial ND5 gene sequences (Coleoptera, Carabidae). *J Mol Evol.* 50(6):541–549.
- Toussaint EFA, et al. 2018. Anchored phylogenomics illuminates the skipper butterfly tree of life. *BMC Evol Biol.* 18(1):101.
- Toussaint EFA, Gillett CPDT. 2018. Rekindling Jeannel's Gondwanan vision? Phylogenetics and evolution of Carabinae with a focus on *Calosoma* caterpillar hunter beetles. *Biol J Linn Soc.* 123(1):191–207.
- Wells A, Johanson KA, Dostine P. 2019. Why are so many species based on a single specimen? *Zoosymposia* 14(1):32–38.
- Wiens JJ. 1998. Does adding characters with missing data increase or decrease phylogenetic accuracy? *Syst Biol.* 47(4):625–640.
- Wong TK, et al. 2020. A minimum reporting standard for multiple sequence alignments. *NAR Genom Bioinform.* 2(2):lqaa024.
- Wood HM, González VL, Lloyd M, Coddington J, Scharff N. 2018. Next-generation museum genomics: phylogenetic relationships among palpid spider species using sequence capture techniques (Araneae: Palpimanoidea). *Mol Phylogenet Evol.* 127:907–918.
- Yahiro K, Sugiyama K, Hayashi M. 2018. Late Pliocene of fossil *Calosoma* (Coleoptera, Carabidae) from the Koka Formation, Kobiwako Group in Shiga Prefecture, Japan. *Elytra New Series.* 8:227–233.
- Yao L, et al. 2020. Population genetics of wild *Macaca fascicularis* with low-coverage shotgun sequencing of museum specimens. *Am J Phys Anthropol.* 173(1):21–33.
- Ziemann M. 2016. Accuracy, speed and error tolerance of short DNA sequence aligners. *bioRxiv.* doi:10.1101/053686 1–19.

Associate editor: Mario dos Reis