A peer-reviewed version of this preprint was published in PeerJ on 30 April 2019.

Matos-Maraví P, Duarte Ritter C, Barnes CJ, Nielsen M, Olsson U, Wahlberg N, Marquina D, Sääksjärvi I, Antonelli A. 2019. Biodiversity seen through the perspective of insects: 10 simple rules on methodological choices and experimental design for genomic studies. PeerJ 7:e6727 https://doi.org/10.7717/peerj.6727

Author Cover Page

Article submission to PeerJ

Manuscript category: Literature Review

Collection: "*Endless forms: Advances in evolutionary analyses of biodiversity*"

Article title: **Biodiversity seen through the perspective of insects: 10 simple rules on methodological choices, common challenges, and experimental design for genomic studies.**

Authors: Pável Matos-Maraví* [1,2], Camila Duarte Ritter [1,2], Christopher J. Barnes [3], Martin Nielsen [3,4], Urban Olsson [1,2], Niklas Wahlberg [5], Daniel Marquina [6,7], Ilari Sääksjärvi [8], Alexandre Antonelli [1,2,9,10]

Affiliations:

[1] Gothenburg Global Biodiversity Centre, SE-40530, Gothenburg, Sweden.

[2] Department of Biological and Environmental Sciences, University of Gothenburg, SE-40530, Gothenburg, Sweden.

[3] Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark

[4] Section for Evolutionary Genomics, Natural History Museum of Denmark, University of Copenhagen, 1350 Copenhagen, Denmark

[5] Department of Biology, Lund University, 223 62 Lund, Sweden

[6] Department of Bioinformatics and Genetics, Swedish Museum of Natural History, SE-10405, Stockholm, Sweden

[7] Department of Zoology, Stockholm University, SE-10691, Stockholm, Sweden

[8] Biodiversity Unit, University of Turku, 20014, Turku, Finland

[9] Gothenburg Botanical Garden, Gothenburg, Sweden.

[10] Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02139, USA.

*Corresponding author: Pável Matos-Maraví, E-mail: pavelm14@gmail.com

Abstract: The study of biodiversity within the spatiotemporal continuum of evolution, e.g., studying local communities, population dynamics, or phylogenetic diversity, has been important to properly identify and describe the current biodiversity crisis. However, it has become clear that a multi-scale approach – from the leaves of phylogenetic trees to its deepest branches – is necessary to fully comprehend, and predict, biodiversity dynamics. Massive parallel DNA sequencing opens up opportunities for bridging multiple dimensions in biodiversity research, thanks to its efficiency to recover millions of nucleotide polymorphisms, both under neutral or selective pressure. Here we aim to identify the current status, discuss the main challenges, and look into future perspectives on biodiversity genomics research focusing on insects, which arguably constitute the most diverse and ecologically important group of metazoans. We suggest 10 simple rules that every biologist could follow to 1) provide a succinct step-by-step guide and best-practices to anyone interested in biodiversity research through insect genomics, 2) review and show relevant literature on biodiversity and evolutionary research in the field of entomology, and 3) make available a perspective on biodiversity studies using insect genomics. Our compilation is targeted at researchers and students who may not yet be specialists in entomology or genomics, but plan to carry out own research in insect genomics. We foresee that the genomic revolution and its application to the study of non-model insect lineages will represent a major leap to our understanding of insect diversity, and by consequence the largest portion of Earth's biodiversity, and its evolution in time and space.

Keywords: Biodiversity, evolution, NGS, museomics, taxonomic impediment

Introduction

The global decline in biodiversity is unquestionable (Barnosky et al., 2011). The rate of species diversity loss is comparable to those of ancient mass-extinction events (Ceballos et al., 2015), but our understanding on the spatiotemporal continuum of biodiversity is still limited (Fig. 1). Not only the current methodologies to quantify biodiversity at different temporal and spatial scales need to be profoundly revised (Vellend, 2017), but also a multi-disciplinary effort is necessary to comprehend species diversity and its evolution. High-throughput DNA technologies, including massive parallel DNA sequencing, have been used during the past two decades to study biodiversity. Thus, researchers nowadays have access to standard tools which can generate unprecedented amounts of genomic information, being cost- and time-efficient (Lugg et al., 2018), and with the potential to integrate their results with previous efforts.

In this article, we aim to briefly review and provide a guideline on the usage of massive parallel DNA sequencing technologies to bridge the study of biodiversity at different scales, with a focus on the largest biotic radiation on Earth: insects. These six-legged invertebrates represent more than half of all known eukaryotic species (Grimaldi & Engel, 2005; Mora et al., 2011; Stork et al., 2015; Stork, 2018) and they are one of the most important components of eukaryotic biodiversity in terms of abundance and ecology. However, as much as 80% of insect diversity, and therefore much of the Earth's biodiversity, remains to be formally described (Hamilton et al., 2010; Scheffers et al., 2012; Stork, 2018). While there is so much undescribed insect diversity in the nature, a significant number may already be deposited within museum collections in need of formal description (Suarez & Tsutsui, 2004; Veijalainen et al., 2012). Therefore, the study of biodiversity through insect genomics, using entomological mass-sampling techniques in the field and the archived material, is timely and represents a significant opportunity to advance our

74 understanding of life on Earth. This article aims to fill a gap in the literature on a simple
75 guideline to study biodiversity through insect genomics, thus this review is primarily targeted at
76 researchers and students who may not yet be experts in entomology or genomics.

77

78 **Survey Methodology**

79 We reviewed published literature related to biodiversity and evolutionary research using insect
80 genomics, including but not limited to methods for review and original articles on collecting
81 insects, specimen preservation and storage, genomic DNA isolation from archived material, and
82 post-sequencing approaches. We ensure a comprehensive and unbiased procedure by using
83 primarily PubMed and Google Scholar to search for articles, complemented with searches in
84 Scopus and Web of Science. We also used a combination of keywords, such as "insect
85 genomics", "museum DNA", "high-throughput sequencing", and "biodiversity assessment".

86

87 **Ten Simple Steps to Study Biodiversity through Insect Genomics**

88 We structure this article in 10 simple rules (Fig. 2) that every biologist should understand to 1)
89 better interpret the results and conclusions coming from insect biodiversity research, and 2) start
90 planning a multi-dimensional study of biodiversity using insects as target group and high-
91 throughput genomic tools. Overall, we briefly review the current state in biodiversity and
92 evolution research through the study of insect genomics, by revisiting pioneering studies that aim
93 to bridge different spatial and temporal scales. We identify a series of limitations and challenges

currently faced by insect genomics, but we also find hopeful approaches to comprehend the origin and dynamics of extant biodiversity.

Rule 1: Define the questions and scope of the study

Producing genomic data is no longer a major challenge for many labs. Instead, many researchers seem to be producing large amounts of data, without having a clear idea of how to use it afterwards. Although it may seem obvious, we consider important to stress that careful thinking is required to define the research questions and hypothesis of any study, and how to best address them. A few projects might be totally discovery-driven with no prior expectations, but in general it can be very useful to clearly define the hypotheses to be tested, and how. This will then inform on the whole chain of methods and analyses, since there is no 'one size fits all' when it comes to biodiversity and evolutionary studies.

With massive-parallel DNA sequencing technologies, the study of evolutionary relations can be complemented with fast quantification of diversity, abundances, and species interactions such as studies on host-parasite interactions (Toju, 2015), in environmental samples (Shokralla et al., 2012) or even from the ethanol used for preservation of historical specimens (Linard et al., 2016). However, economical limitations exist regarding the number of specimens and the extent of their genomes that could be sequenced (Wachi, Matsubayashi & Maeto, 2018). Therefore, researchers should choose from a series of available sequencing approaches that better suits their research questions. For example, if the focus is on finding potential loci involved in adaptation and speciation, a reduced representation of the genomes might be cost-efficient because several individuals from different populations could be pooled in one sequencing experiment, or if the

aim is to profile many organisms within insect communities, DNA metabarcoding could provide a fast quantification of diversity and relative abundances.

Rule 2: Set up your collaborations strategically

A major challenge in the study of evolution from populations to species is the lack of non-genomic data, including taxonomic, paleontological, and ecological information. Despite the abundance of genomic information that can nowadays be generated, evolutionary biologists are facing the necessity to 1) increase field expeditions in search of the unknown diversity, 2) incorporate fossil data in the Tree of Life, and 3) study the phenotypes and life history data in specimen collections. Naturally, the most efficient direction to integrate such different perspectives is to establish and strengthen a collaborative network. For example, working along with paleontologists might bring a temporal perspective in the study of evolution and biodiversity dynamics (Marshall, 2017). In some insect groups, such as Hymenopterans (sawflies, wasps, bees, ants) and Coleopterans (beetles), the morphologies of ancient, extinct lineages might be better preserved as inclusions in amber, and even ancient ecologies could be preserved (Johnson et al., 2001). Other groups such as Lepidoptera (butterflies and moths) are on the contrary hardly fossilized (Labandeira & Sepkoski, 1993), thus representing a challenge, for instance, in divergence time analyses when working with phylogenomic data (Wheat & Wahlberg, 2013). However, other sources of information, such as larval host-plant ages, may improve divergence time analyses when fossil record is scarce (Chazot et al., 2018).

Collaborating closely with ecologists would strengthen the study of adaptation, differentiation, and the mechanisms of speciation, and a comprehensive knowledge of life history data, insect

ecologies, or common garden experiments are ideal to tease apart adaptive from non-adaptive variation. Moreover, Natural History Museums (NHMs) are the repositories of our natural world and include not only archived specimens but also valuable historical, demographic, life-history, and genetic data that can add another dimension to evolutionary research (Burrell, Disotell & Bergey, 2015; Buerki & Baker, 2016). Research on biodiversity will benefit by working closely along with curators and research assistants at NHMs. For example, population range expansion in historical times, host-parasite interaction changes after human disturbances, or the effect of current climate change on the structure of populations, are topics that could be directly benefited by incorporating the information from NHM collection records (Burrell, Disotell & Bergey, 2015). Moreover, the information that curators might hold on the collection and preservation methods of specimens is valuable when selecting which specimens should undergo high-throughput sequencing (Kanda et al., 2015; Short, Dikow & Moreau, 2018).

Rule 3: Go to the field

We are worried that the rapid increase of genetic data in public databases might discourage students and researchers from generating novel data. Instead, we argue that field work is absolutely essential to the advancement of our field, and should be part of every biologist's education as well as routine in advanced careers. Fieldwork will also benefit museum collections, and vice-versa, museum collections –through genetic and morphological studies based on specimens– will benefit fieldwork. Of course, there might be lines of research that do not demand fieldwork, but even taxonomists, method developers, and researchers in other disciplines may profit from the experience of studying and responsibly collecting specimens in

nature. Extensive field surveys are often required to obtain a representative inventory of insect assemblages at both local and regional scales; but such surveys represent only a minority within entomological field studies. Most studies aiming to understand insect diversity patterns only target a small portion of the species present in a single study location. This is true given the high species richness and varying abundance, habits and seasonality of insects, including parasitoids, predators, scavengers, leaf-chewers, sap-suckers, among others. A careful selection of field sampling methods, along with proper understanding of their function and targeted groups, is thus critical (Noyes, 1989) (see Table 1 for a non-comprehensive overview of mass-sampling methods).

For some cases, such as in biodiversity assessment, it may be enough to conduct simple and rapid field surveys. However, in other cases, such as in exhaustive inventories or when studying diversity dynamics through time and space, greater mass-sampling efforts may be needed, requiring a combination of multiple methods, longer term inventories and wide expertise, together with effective ways to estimate true species richness based on collected samples (Vogel, 2017). For example, in a recent tropical large-scale species inventory, Borkent & Brown (2015) investigated local species richness of cloud forest Diptera (true flies) for more than one year by using two Malaise traps, a wide range of supplementary collecting methods, and a one-week intensive “Diptera-Blitz” conducted by a large network of experts. In another case study, Gómez et al. (2018) sampled the Western Amazonian local parasitoid wasp diversity by using 41 Malaise traps, with a total sampling effort of 230 Malaise-trap months (one Malaise-trap month corresponds to one trap collecting in the field for a period of one month). In this case, despite the tremendous sampling effort, cumulative species curves suggested that a significant portion of the local parasitoid diversity remained unobserved; a fact that may be generalized for many other

tropical insect groups as well. Reviews of entomological collection methods for both mass-sampling and group-specific research are available in the literature and are essential reading before field collections (Agosti et al., 2000; Basset et al., 2003; Lamarre et al., 2012; Larsen, 2016).

Needless to say, be a sensible collector; many insects are rare and threatened, so every collecting effort should be associated with a risk assessment, even informally if not required. There are also many federal and international regulations to follow, such as those stipulated under the Nagoya Protocol under the Convention on Biological Diversity (https://www.cbd.int/abs/about/) and the CITES legislation (https://www.cites.org/). Researchers should all follow all good practices for Access and Benefit Sharing (e.g., https://naturwissenschaften.ch/organisations/biodiversity/abs/goodpractice).

Rule 4: Treat your specimens well to enhance its use

The amount and quality of isolated genomic DNA from insect collections depend on a myriad of factors, including killing reagents, preservation of specimens in the field, and final voucher storage conditions (Kanda et al., 2015; Short, Dikow & Moreau, 2018). For example, Dillon et al., (1996) (see also Reiss, Schwert & Ashworth, 1995; Gilbert et al., 2007b) found that specimens killed with ethanol yielded significantly higher quantities of higher quality DNA compared to other killing/preservation agents such as ethyl acetate vapor, formalin or ethylene glycol. Moreover, rapid and effective drying of the specimens in the field, especially in the tropics, is a very important step for voucher preservation and may be an alternative to cryopreservation (Prendini, Hanner & DeSalle, 2002). Preservation of specimens in ethanol and

at low temperatures would be ideal, but may cause logistic problems during transportation and make the collections highly flammable. Propylene glycol may be a safer alternative and logistically easier to transport than ethanol (Ferro & Park, 2013), and it might even be used to attract certain arthropod species (Höfer et al., 2015). Moreover, initiatives to implement large cryobanks are important (Koebler, 2013), though these technologies are yet restricted to very few and large NHMs (Corthals & Desalle, 2005).

The use of ethylene glycol may provide reasonable amounts of DNA regardless of specimen age, and with lesser risks in the field (Dillon, Austin & Bartowsky, 1996), though the age of specimen seems not to be a critical factor for obtaining minimal amount of DNA for massive-parallel sequencing (Ruane & Austin, 2017) (see Table 2 for an overview of published studies using archived insects). In fact, ancient beetle genomic DNA (ca. 560,000 to 5,960 years old) has been successfully isolated (Heintzman et al., 2014). However, due to the fragmented nature of ancient DNA, PCR-based techniques are overall not successful to recover genetic data. Fortunately, evidence suggests that fragmented DNA due to age or preservation reagents does not dramatically affect the performance of PCR-free, massive-parallel sequencing (Tin, Economo & Mikheyev, 2014; Timmermans et al., 2016). However, the success of current sequencing approaches still depends in some cases on the quality of isolated DNA, such as in RAD-seq, thus minimal damage in the field and during storage is advisable.

Rule 5: Work closely with taxonomists

Genomic data is only one component of biodiversity, but in order to reach the general public biodiversity needs as well to be tangible. The tasks of taxonomists, including the identification,

description, and classification of species in meaningful groupings, are unfortunately sometimes neglected. The high diversity and density of insects, coupled with laborious taxonomic assessment and lack of resources for taxonomists, makes the morphological identification of every specimen sampled by mass-collecting techniques a difficult and high resource-consuming task. The so-called "taxonomic impediment" (di Castri, Vernhes & Younes, 1992) encompasses two general difficulties: 1) not enough resources and training are allocated to taxonomic work and 2) few people are working in taxonomy (thus slowing down the rate of species discovery, identification, and classification), which may be a consequence of the former difficulty (Wheeler, Raven & Wilson, 2004; de Carvalho et al., 2007; Ebach, Valdecasas & Wheeler, 2011; Audisio, 2017). Indeed, we may be in the midst of a revolution in taxonomy, as evidenced by the intense debate on its epistemological and methodological grounds (Dubois, 2011; Ceríaco et al., 2016; Garnett & Christidis, 2017; Raposo et al., 2017; Thorpe, 2017), but it is also clear that in the meantime entomological research must use complementary approaches to reliably estimate diversity through time and among places and environments. Therefore, taxonomists should be part of biodiversity studies using insect genomics, and the DNA sequences generated by such studies should be seen as a necessary supplement to the work of taxonomists.

Rule 6: Isolate DNA in the right way

Most recent studies using massive-parallel DNA sequencing, even those on ancient insects, have used commercial kits for DNA isolation, thus reducing time, complexity, and health risks in laboratory procedures (Staats et al., 2013; Heintzman et al., 2014; Kanda et al., 2015; Blaimer et al., 2016; Pitteloud et al., 2017). However, in-house methods might be more effective than

commercial kits when working with ancient samples having little and highly-degraded DNA (Gilbert et al., 2007c; Meyer et al., 2016). On the other hand, non-destructive protocols for DNA isolation are preferable when working with valuable, archived specimens or with bulk samples such as those coming from insect mass-collecting techniques that later need to be taxonomically curated. However, there is surprisingly little data available comparing the efficiency of destructive vs. non-destructive protocols applied to insects (Gilbert et al., 2007a; Nieman et al., 2015). A number of non-destructive DNA isolation protocols have been published (e.g., (Thomsen et al., 2009; Castalanelli et al., 2010; Tin, Economo & Mikheyev, 2014), but in general they vary depending on the targeted insect group. For example, insects whose external structure are not delicate, including Diptera, Hymenoptera and Coleoptera, tend to be more resistant to submergence of whole specimen in extraction buffers, giving better results (Heintzman et al., 2014; Tin, Economo & Mikheyev, 2014). In other more delicate groups such as Lepidoptera, the use of abdomens is advisable, given that in many cases the abdomens need to be removed from the individual for genitalia preparation (Knölke et al., 2005). In other insect groups that hold sufficient starting material for DNA isolation in particular tissues, such as muscles in the massive legs of Orthoptera (grasshoppers, locusts, crickets), grinding one leg might not be a significant loss to the collection (Tagliavia et al., 2011).

Although curators at NHMs may be reluctant to provide specimens for molecular studies, with valid reasons considering that most species might consist of singletons or very rare collections (Lim, Balke & Meier, 2012), the design of selective sampling, minimizing the damage of collections especially of those very rare specimens, is therefore crucial. As a side note, there has not been any discussion in the literature about the suitability for massive-parallel sequencing using the hundreds of thousands, or perhaps millions, DNA aliquots generated in the past three

decades for Sanger-sequencing work. In principle, old DNA aliquots of low quantities and potentially fragmented DNA may face the same constraints of using archived specimens from NHMs, and might thus be processed with protocols designed for archived samples (e.g., library preparation, sequencing approach) (Tin, Economo & Mikheyev, 2014; Kanda et al., 2015; Suchan et al., 2016; Timmermans et al., 2016).

Rule 7: Revise your DNA sequencing approach

At this point, you should already have decided which sequencing approach will be best suitable to address your research question(s), but now you should carefully evaluate the quality of DNA that you *de facto* were able to obtain, and decide on which sequencing approach to really follow.

Reviews on massive-parallel DNA sequencing approaches can be found in the literature (Mamanova et al., 2010; Metzker, 2010; Mardis, 2017). Below, we categorize and briefly describe available massive-parallel DNA sequencing technologies of potential interest for entomological biodiversity research (see Table 3 for a summary of such methods and key publications). The current leading short-read DNA sequencing technology is Illumina. We have grouped the main approaches used in the study of entomological biodiversity into three categories: 1) targeted-sequencing, 2) non-targeted, reduced-representation of whole genome, and 3) whole-genome skimming. In addition, single-molecule DNA sequencing technologies such as Oxford Nanopore and PacBio can accelerate the amount of DNA data recovery in real time and from highly-degraded starting DNA material (Thompson & Milos, 2011). We thus consider these technologies as promising, and we briefly introduce them here.

Target Sequencing: It is highly-efficient when the aim is to recover only DNA markers with a particular rate of evolution (fast and slow) or under different selective pressures (Lemmon & Lemmon, 2013). Moreover, because it targets only a tiny subset of the whole genome, targeted sequencing is cost-effective as tens or hundreds of specimens can be pooled together in a single sequencing experiment (Mamanova et al., 2010). This is particularly useful when working with environmental samples, such as those coming from mass-sampling techniques (Morinière et al., 2016). For example, metabarcoding, an approach that targets a barcoding region such as the COI mitochondrial gene, can be useful in the study of evolution among local environments and in biodiversity assessment because it might be more reliable, fast and replicable than traditional biodiversity surveys (Ji et al., 2013; Zhou et al., 2013; Vesterinen et al., 2016).

There are two usual ways to target particular loci, 1) through PCR or 2) by using "baits"-based in-vitro capture. PCR has the advantage of being cheap but the development of universal primers is the main limitation because sequence specificity to desired loci decreases through mutation and long divergence times among lineages. Target capture using "baits" can be expensive ("baits" need to be specially synthesized) but has the advantages of 1) simplify laboratory procedures (one can pool several specimens for the capture experiment), 2) target a wider range of lineages despite evolutionary distance among them, and 3) reduce amplification biases due to PCR primer design and relative abundance of DNA molecules in a pool of specimens. However, prior genomic information either published annotated genomes or transcriptomes are needed to design target-enrichment probes. Therefore, probe kits targeting conserved regions primarily for phylogenomic purposes have been published for those insect orders having good reference databases (Faircloth et al., 2015; Faircloth, 2016; Young et al., 2016; Breinholt et al., 2018). Furthermore, recent attempts to integrate "baits"-based capture into metabarcoding have had

disparate degree of success; such as the sequencing of non-target organisms or pseudogenes on the negative side (Shokralla et al., 2016), or the recovery of sequences of very rare species in a pool of samples and the quantification of relative abundance and biomass on the positive side (Dowle et al., 2016).

Random reduced-representation of genome: Restriction-site-associated DNA (RAD) sequencing has proven to be an efficient and cheap approach to obtain millions of single nucleotide polymorphisms (SNPs), both neutral and under selection (Andrews et al., 2016). However, restriction enzyme sites may not be conserved for a long evolutionary time, thus this approach seems to be restricted to population or species complexes. However, a recent protocol targeting RAD-seq markers (hyRAD) may ameliorate the lack of phylogenetic conservation of restriction enzyme sites across divergent lineages (Suchan et al., 2016). The amount and quality of DNA might impose a second limitation to RAD-seq. For example, Tin, Economo & Mikheyev (2014) using ant specimens as old as 100 years were able to recover SNPs but were unsuccessful at genome mapping due to the extremely short DNA fragments and imprecise DNA size selection. Overall, RAD-seq is promising in the study of insect diversity and evolution because it generates a large amount of SNPs, is cheap and can be run on pools of specimens, as long as genomic DNA is of good quality (long DNA fragments are needed for an efficient restriction enzyme activity) and taxa of interest are not evolutionarily distant.

Whole-genome skimming: It is the simplest approach in terms of sequence library preparation. It consists of randomly, shallow sequencing the whole-genome of an individual, including both mitochondria and nuclear content. Furthermore, when working with historical specimens with highly-fragmented DNA, one can skip the step of fragmentation in library preparation (Suchan et al., 2016; Timmermans et al., 2016). Whole-genome skimming has been applied in a number of

339 insect studies, proving that the method is fast and can recover entire mitochondrial genomes
340 from even old museum material (Staats et al., 2013), and low-copy nuclear protein-coding genes
341 (Maddison & Cooper, 2014; Kanda et al., 2015). With the expected decrease in sequencing
342 prices, target sequencing approaches may no longer be a cost-effective choice in the future. For
343 instance, recent studies have identified the benefits of mitochondrial metagenomics (MMG),
344 including longer barcodes with larger amount of SNPs (use of mitogenomes instead of the COI
345 fragment), and PCR-free library preparation (no amplicons would be needed) with the
346 advantages of using highly-fragmented DNA from old specimens and a more reliable
347 quantification of relative abundance (biomass) in mass-sampling collections (Crampton-Platt et
348 al., 2015, 2016; Cicconardi et al., 2017; Gómez-Rodríguez et al., 2017). However, it was noted
349 that having a reference genome is important to improve mapping and discovery of homologous
350 SNPs in the nuclear genome (Tin, Economo & Mikheyev, 2014), which may yet restrict the use
351 of whole-genome skimming and the recovery of nuclear data in insect groups with poor genomic
352 knowledge.

353 Single-molecule sequencing approaches such as PacBio and Oxford Nanopore. The two main
354 advantages are related to a better assembly of genomes with low quantities of poor quality DNA
355 and the portability of some devices (e.g., MinION), which can generate DNA sequences in real-
356 time and in any place in the world, including remote field locations. Moreover, laboratory
357 protocols are simplified and DNA amplification is not at all necessary, which is beneficial for a
358 more accurate quantification of DNA molecules present in the sample pool (Thompson & Milos,
359 2011). However, technological improvement is needed to reduce the high sequencing error rates
360 (>10 %) (Mardis, 2017; Shendure et al., 2017). Nonetheless, single-molecule sequencing
361 promises to drastically reduce sequencing costs, thus the age where having complete genome

362 sequences for any living insect may be even closer (Kelley et al., 2014). Moreover, the long
363 reads that single-molecule sequencing approaches generate might help resolve long repeat
364 elements in the genome, thus providing invaluable scaffold for short reads to improve accuracy
365 in assembly and annotation of insect genomes (Richards & Murali, 2015).

366

367 *Rule 8: Choose the most suitable tools for data analyses*

368 Although genomic sequencing is becoming easier and more affordable, processing the data
369 generated remains a major bottleneck in many projects. Bioinformatic pipelines have been
370 implemented during the past two decades of massive parallel sequencing, thus researchers
371 nowadays count with standard procedures to analyze genomic DNA. However, there exist a
372 number of limitations and challenges that remain to be explored. For example, genomic missing
373 data, as in supermatrices for phylogenomic studies, might hinder statistical power in the
374 inference of species relations, but its effects in systematic bias is yet unclear (Misof et al.,
375 2014a,b). Moreover, taxonomic sampling in phylogenomics is usually lower than in the
376 published Sanger-sequencing work, a fact that might bias systematic inference in insect higher-
377 level phylogenies (Behura, 2015). On the other hand, a number of pipelines have been published
378 for analyzing target-sequencing data from environmental samples (Schloss et al., 2009; Caporaso
379 et al., 2010; Boyer et al., 2016). Such programs provide a delimitation of Operational Taxonomic
380 Units (OTUs), the analogs of species, derived from sequence similarity of typically 97 %.
381 However, assigning thresholds to define analogs of species is problematic because 1) there is a
382 risk to artificially increase or decrease local diversity, 2) inflated OTU richness might be related
383 to sequence chimeras and sequencing errors, and 3) there exist a lack of standardization of

384 threshold values (Huse et al., 2010; Oliver et al., 2015; Alberdi et al., 2018). The shortcomings
385 of using thresholds to define OTUs might even escalate when studying the entomofauna of
386 hyperdiverse regions such as the tropics. In those cases, there are usually no good estimates of
387 genetic variability between species and a large portion of tropical insects are not represented in
388 reference databases. However, recent advances might alleviate in part the variability caused by
389 sequencing errors by incorporating other observed patterns such as OTUs co-occurrence (Frøslev
390 et al., 2017). In any case, the preservation and morphological study of vouchers are critical to
391 validate taxonomic assignments and thresholds.

392 Mitochondrial metagenomics (MMG) could in principle improve OTU assignment and species
393 delimitation because contigs would span different barcode regions (COI, ND2, 16S rDNA) (Liu
394 et al., 2016) and risks of primer-related biases are ameliorated (Taberlet et al., 2012; Tang et al.,
395 2014). Moreover, whilst approaches such as log-binomial normalizations (through DeSeq2 and
396 CSS) have attempted to normalize metabarcoding data (McMurdie & Holmes, 2014), results via
397 PCR-based approaches remain semi-quantitative at best (Pawluczyk et al., 2015). However,
398 metagenomic studies of insects have generally been limited only to their microbiomes (Cox-
399 Foster et al., 2007; Suen et al., 2010; Shi et al., 2013), but it is difficult to assess the convenience
400 of metagenomics in more complex environmental insect samples (but see (Crampton-Platt et al.,
401 2015, 2016; Cicconardi et al., 2017; Gómez-Rodríguez et al., 2017). Nonetheless, studies on soil
402 and fecal MMG have shown that the assembly of mitogenomes from environmental DNA
403 (eDNA) remains challenging and is not yet as cost-effective as compared to target sequencing
404 (Tedersoo et al., 2015; Srivathsan et al., 2016).

405

406 *Rule 9: Make your data and results publicly available*

407 From a practical viewpoint, what is not in a database does not exist (or nearly so). Databases are
408 not only the repositories of genomic information, but also an indispensable tool in the study of
409 biodiversity and evolution. They also allow the reproduction of results and use for other purposes
410 such as in biodiversity assessment. Furthermore, the study of populations, their mechanisms of
411 adaptation and speciation, is an active field given the opportunities that represent the hundreds of
412 insect genome projects published and registered in GenBank (Yeates et al., 2016). Moreover, in
413 the study of species interactions, such as host-parasite and feeding habits, the recovery of cryptic
414 diversity necessitates a reference database because in many cases the identification of taxa
415 through morphological comparison becomes impossible; for example, internal parasites
416 (Schoonvaere et al., 2016), gut microbiota (Hammer et al., 2017), and highly-degraded organic
417 material such as dietary content in gut (Pompanon et al., 2012).

418 Furthermore, initiatives such as BOLD (Ratnasingham & Hebert, 2007) and the usage of the COI
419 barcode can shed more lights in the assignments of OTU thresholds when studying tropical
420 communities (García-Robledo et al., 2013). However, building local databases including several
421 markers might complement metabarcoding studies in the identification and delimitation of
422 species (Deagle et al., 2014). National reference databases have also been implemented, such as
423 the newly initiated DNAmark project in Denmark. This project aims to provide a reference
424 database for 1,000 species with full mitochondrial sequences, along with nuclear sequences
425 derived from shotgun sequencing (http://dnamark.ku.dk/). Other initiatives to catalogue national
426 biodiversity have also been put forward in Germany (Hendrich et al., 2015), Norway (NorBOL,
427 http://www.norbol.org/) and Finland (FinBOL; http://www.finbol.org/), to further expand the
428 BOLD project.

Rule 10: Disseminate your findings

Research articles are the standard way to communicate results to the scientific community. However, misinterpretations of scientific findings can be commonly found in the literature aimed for the general public and decision-makers. Thus, public outreach of our findings needs to be explicitly considered as part of our project design. Moreover, because scientific research is a collaborative enterprise (see Rule 2), it is important to discuss and reach a consensus with collaborators before spreading findings to the general public. This is particularly important given the recent misunderstandings on biodiversity research that have been reported, and the urge to include both factual evidence and ethical arguments in communications to the general public (Antonelli & Perrigo, 2018).

Given that diversity estimates can fluctuate significantly depending on the way results are treated (e.g., as in metabarcoding (Frøslev et al., 2017; Alberdi et al., 2018)) special care should be taken when presenting these findings, and in general we advocate for a conservative approach which does not artificially inflate diversity estimates. Furthermore, the access of scientific knowledge and data by governmental bodies is still restricted, especially in low and lower-middle income countries. Biodiversity research is a cornerstone for Environmental Impact Assessments, but hyperdiverse animal groups such as insects remain underrepresented in biodiversity assessments in species-rich countries (Ritter et al., 2017).

Perspectives and Conclusions

In this article we have identified general challenges, including: 1*) insufficient evaluation of non-destructive methods* applied to insects to get good quantity and quality DNA from fresh, mass-collections and archived specimens, *2) limitations to genomic data analysis*, including assembly of reads, missing genomic information from datasets, methods for estimating diversity and abundance in environmental samples, and inference of population and species evolution, *3) taxonomic, ecological, and life history knowledge* generation is not at the same pace as the genomic revolution.

Insects are ideal study organisms because they show remarkable diversity in species numbers and ecologies, being the dominant eukaryotic group in terrestrial and freshwater environments. The integration of ecology and evolution is achievable with the new genomic techniques, which offer the possibility to generate datasets that can be used in the study of biodiversity at different spatiotemporal scales. For example, the evolutionary framework of local insect communities can now be inferred in a single sequencing effort (Crampton-Platt et al., 2015), while the study of populations and speciation using massive-parallel sequencing can be better understood with a comprehensive knowledge of local variations (Jiggins, 2016). Altogether, we expect that the increase of molecular data together with taxonomic and ecological studies will allow a better biodiversity and evolutionary comprehension which is essential for conservation and to understand biological dynamics.

Acknowledgements

This article is a product of the workshop organized by the Gothenburg Global Biodiversity Centre (GGBC), in Gothenburg, Sweden, in March 2017. We are grateful to Sven Toresson for his help in organizing the logistics of the workshop.

References

Agosti D., Majer JD., Alonso LE., Schultz TR. 2000. *Ants. Standard methods for measuring and monitoring biodiversity*. Washington, DC.: Smithsonian Institution Press.

Alberdi A., Aizpurua O., Gilbert MTP., Bohmann K. 2018. Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods in Ecology and Evolution* 9:134–147. DOI: 10.1111/2041-210X.12849.

Andrews KR., Good JM., Miller MR., Luikart G., Hohenlohe PA. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics* 17:81–92. DOI: 10.1038/nrg.2015.28.

Antonelli A., Perrigo A. 2018. The science and ethics of extinction. *Nature Ecology & Evolution*:In Press. DOI: 10.1038/s41559-018-0500-z.

Audisio P. 2017. Insect taxonomy, biodiversity research and the new taxonomic impediments. *Fragmenta Entomologica* 49:121–124. DOI: 10.4081/fe.2017.252.

Barnosky AD., Matzke N., Tomiya S., Wogan GOU., Swartz B., Quental TB., Marshall C., McGuire JL., Lindsey EL., Maguire KC., Mersey B., Ferrer EA. 2011. Has the Earth's sixth mass extinction already arrived? *Nature* 471:51–57. DOI: 10.1038/nature09678.

Basset Y., Novotny V., Miller SE., Kitching RL. 2003. Methodological advances and limitations in canopy entomology. In: Basset Y, Novotny V, Miller SE, Kitching RL eds. *Arthropods of Tropical Forests: Spatio-temporal dynamics and resource use in the canopy*. Cambridge, UK: Cambridge University Press, 7–16.

Behura SK. 2015. Insect phylogenomics. *Insect Molecular Biology* 24:403–411. DOI: 10.1111/imb.12174.

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:e0161531. DOI: 10.1371/journal.pone.0161531.

Borkent A., Brown B V. 2015. How to inventory tropical flies (Diptera) - One of the megadiverse orders of insects. *Zootaxa* 3949:301–322. DOI: 10.11646/zootaxa.3949.3.1.

Boyer F., Mercier C., Bonin A., Le Bras Y., Taberlet P., Coissac E. 2016. OBITOOLS: a UNIX-inspired software package for DNA metabarcoding. *Molecular Ecology Resources* 16:176–182. DOI: 10.1111/1755-0998.12428.

Breinholt JW., Earl C., Lemmon AR., Lemmon EM., Xiao L., Kawahara AY. 2018. Resolving relationships among the megadiverse butterflies and moths with a novel pipeline for anchored phylogenomics. *Systematic Biology* 67:78–93. DOI: 10.1093/sysbio/syx048.

Buerki S., Baker WJ. 2016. Collections-based research in the genomic era. *Biological Journal of the Linnean Society* 117:5–10. DOI: 10.1111/bij.12721.

Burrell AS., Disotell TR., Bergey CM. 2015. The use of museum specimens with high-throughput DNA sequencers. *Journal of Human Evolution* 79:35–44. DOI:

10.1016/j.jhevol.2014.10.015.

Caporaso JG., Kuczynski J., Stombaugh J., Bittinger K., Bushman FD., Costello EK., Fierer N., Peña AG., Goodrich JK., Gordon JI., Huttley GA., Kelley ST., Knights D., Koenig JE., Ley RE., Lozupone CA., Mcdonald D., Muegge BD., Pirrung M., Reeder J., Sevinsky JR., Turnbaugh PJ., Walters WA., Widmann J., Yatsunenko T., Zaneveld J., Knight R. 2010. QIIME allows analysis of high- throughput community sequencing data. *Nature Methods* 7:335–336. DOI: 10.1038/nmeth0510-335.

de Carvalho MR., Bockmann FA., Amorim DS., Brandao CRF., de Vivo M., de Figueiredo JL., Britski HA., de Pinna MCC., Menezes NA., Marques FPL., Papavero N., Cancello EM., Crisci J V., McEachran JD., Schelly RC., Lundberg JG., Gill AC., Britz R., Wheeler QD., Stiassny MLJ., Parenti LR., Page LM., Wheeler WC., Faivovich J., Vari RP., Grande L., Humphries CJ., DeSalle R., Ebach MC., Nelson GJ. 2007. Taxonomic impediment or impediment to taxonomy? A commentary on systematics and the cybertaxonomic-automation paradigm. *Evolutionary Biology* 34:140–143. DOI: 10.1007/s11692-007-9011-6.

Castalanelli MA., Severtson DL., Brumley CJ., Szito A., Foottit RG., Grimm M., Munyard K., Groth DM. 2010. A rapid non-destructive DNA extraction method for insects and other arthropods. *Journal of Asia-Pacific Entomology* 13:243–248. DOI: 10.1016/j.aspen.2010.04.003.

di Castri F., Vernhes JR., Younes T. 1992. The network approach for understanding global biodiversity. *Biology International. The News Magazine of the International Union of Biological Sciences (IUBS)* 25:3–9.

533 Ceballos G., Ehrlich PR., Barnosky AD., García A., Pringle RM., Palmer TM. 2015. Accelerated
534 modern human–induced species losses: entering the sixth mass extinction. *Sciences*
535 *Advances* 1:e1400253. DOI: 10.1126/sciadv.1400253.

536 Ceríaco LMP., Gutiérrez EE., Dubois A., Abdala CS., Alqarni AS., Adler K., Adriano EA.,
537 Aescht E., Agarwal I., Agatha S., Agosti D., Aguiar AJC., Aguiar JJM., Ahrens D., Aleixo
538 A., Alves MJ., Do Amaral FR., Ananjeva N., Andrade MC., De Andrade MB., Andreone F.,
539 Aquino PPU., Araujo PB., Arnaud H., Arroyave J., Arthofer W., Artois TJ., Astúa D.,
540 Azevedo C., Bagley JC., Baldo D., Barber-James HM., Bärmann E V., Bastos-Silveira C.,
541 Bates MF., Bauer AM., Bauer F., Benine RC., Bennett DJ., Bentlage B., Berning B., Bharti
542 D., Biondo C., Birindelli J., Blick T., Boano G., Bockmann FA., Bogdanowicz W., Böhme
543 W., Borgo E., Borkin L., Bornschein MR., Bour R., Branch WR., Brasileiro CA., Braun
544 JK., Bravo GA., Brendonck L., Brito GRR., Britto MR., Buckup PA., Burckhardt D.,
545 Burkhardt U., Busack SD., Campos LA., Canard A., Cancello EM., Caramaschi U.,
546 Carpenter JM., Carr M., Carrenho R., Cartaxana A., Carvajal MA., Carvalho GS., De
547 Carvalho MR., Chaabane A., Chagas C., Chakrabarty P., Chandra K., Chatzimanolis S.,
548 Chordas SW., Christoff AU., Cianferoni F., Claramunt S., Cogãlniceanu D., Collette BB.,
549 Colli GR., Colston TJ., Conradie W., Constant J., Constantino R., Cook JA., Cordeiro D.,
550 Correia AM., Cotterill FPD., Coyner B., Cozzuol MA., Cracraft J., Crottini A., Cuccodoro
551 G., Curcio FF., D'Udekem D'Acoz C., D'Elía G., D'Haese C., Das I., Datovo A., Datta-
552 Roy A., David P., Day JG., Daza JD., De Bisthoven LJ., De La Riva De La Viña IJ., De
553 Muizon C., De Pinna M., Piacentini VDQ., De Sá RO., De Vivo M., Decher J., Dekoninck
554 W., Delabie JHC., Delfino M., Delmastro GB., Delsinne T., Denys C., Denzer W.,
555 Desutter-Grandcolas L., Deuti K., De Resbecq TD., Di Dario F., Dinets V., DoNascimiento

C., Donoso DA., Doria G., Drewes RC., Drouet E., Duarte M., Durette-Desset MC., Dusoulier F., Dutta SK., Engel MS., Epstein M., Escalona M., Esselstyn JA., Eto K., Faivovich J., Falaschi RL., Falin ZH., Faundez EI., Feijó A., Feitosa RM., Fernandes DS., Fikáček M., Fisher BL., FitzPatrick MJ., Forero D., Franz I., Freitag H., Frétey T., Fritz U., Gallut C., Gao S., Garbino GST., Garcete-Barrett BR., García-Prieto L., García FJ., Garcia PCA., Gardner AL., Gardner SL., Garrouste R., Geiger MF., Giarla TC., Giri V., Glaubrecht M., Glotzhober RC., Godoi FSP., Gofas S., Gonçalves PR., Gong J., Gonzalez VH., González-Orej JA., González-Santillán E., González-Soriano E., Goodman SM., Grandcolas P., Grande L., Greenbaum E., Gregorin R., Grillitsch H., Grismer LL., Grootaert P., Grosjean S., Guarino FM., Guayasamin JM., Guénard B., Guevara L., Guidoti M., Gupta D., Gvoždík V., Haddad CFB., Hallermann J., Hassanin A., Hausmann A., Heaney LR., Heinicke MP., Helgen KM., Henle K., Hirschmann A., Holmes MW., Hołyńska M., Hołyński R., Hormiga G., Huber BA., Hugot JP., Hutterer R., Iskandar D., Iverson JB., Jäger P., Janssen R., Jerep F., Jocqué R., Jungfer KH., Justine J Lou., Kamei RG., Kamiński MJ., Karner M., Kearney T., Khot R., Kieckbusch M., Köhler J., Koepfli KP., Kondorosy E., Krogmann L., Krolow TK., Krüger M., Kucharzewski C., Kullander SO., Kumar S., Kupfer A., Kuramoto M., Kurina O., Kury A., Kvist S., La Marca E., La Terza A., LaVal R., Lacher TE., Lamas CJE., Lambert MR., Landry B., Langeani F., Langone JA., Lattke JE., Lavilla EO., Leenders T., Lees DC., Leite YLR., Lehmann T., Lhano MG., Lim BK., Lin X., Löbl I., De Lucena CAS., De Lucena ZMS., Lucinda P., Lujan NK., Luporini P., Luz DR., Lynch JD., Machado LF., Mahony S., Malabarba LR., Manuel-Santos M., Marinho-Filho J., Marini M., Marques AC., Marques MP., Mateus O., Matsui M., Mazuch T., McCranie J., McKellar RC., McMahan CD., Mecke S., Meißner K.,

579 Mendoza-Becerril MA., Mendoza-Palmero CA., Merker S., Mezzasalma M., Midgley JM.,
580 Miller J., Miller MJ., Mincarone MM., Minet J., Miralles A., Miranda TP., Missoup AD.,
581 Modrý D., Molinari J., Monadjem A., Montreuil O., Moratelli R., Moreira CR., Moreira
582 FFF., Mourer-Chauviré C., Mulieri PR., Munroe TA., Naomi SI., Nascimento F., Nässig
583 WA., Neifar L., Netto-Ferreira AL., Niamir A., Nielsen S V., Nihei SS., Nistri A.,
584 Oceguera-Figueroa A., Odierna G., Ohler A., Ojanguren-Affilastro AA., De Oliveira FF.,
585 De Oliveira ML., De Oliveira OMP., Oliveira SS., Olson LE., Ong'ondo GO., Orlov N.,
586 Ornelas-García CP., Ortega H., Ortega-Andrade M., Ota H., Pariselle A., Passos P., Pastana
587 MNL., Patterson BD., Patitucci LD., Patton JL., Pavan AC., Pavan SE., Pavia M., Peloso
588 PLV., Pelzer A., Pereyra MO., Perez-Gonzalez A., Pérez-Luz B., Pérez CHF., Peterhans
589 JK., Peterson AT., Pétillon J., Philips TK., Picariello O., Pie MR., Pikart TG., Pine RH.,
590 Pinheiro U., Pinho LC., Pinto ÂP., Costa LP., Poggi R., Pombal JP., Prabhu M., Prendini
591 E., Prendini L., Purushothaman J., Pyron RA., Quintela-Alonso P., Quinteros AS., Quiroga-
592 Carmona M., Rabitsch W., Raffaëlli J., Rage JC., Rajaei H., Ramírez MJ., Raposo MA., Py-
593 Daniel LHR., Rasplus JY., Ratcliffe BC., Reddy S., Reis RE., Remsen J V., Richards LR.,
594 Richling I., Robillard T., Rocha MS., Rocha RM., Rödder D., Rödel MO., Rodrigues FP.,
595 Rodriguez E., Rogers DS., Rojas-Runjaic FJM., Röll B., Rosenberger AL., Rowley J., Roza
596 AS., Ruedi M., Salazar-Bravo J., Salcedo NJ., Samyn Y., Santana SE., Santoferrara L.,
597 Santos BF., Santos CMD., Santos JC., Santos MPD., Sargis EJ., Schargel WE., Schätti B.,
598 Scherz MD., Schlick-Steiner BC., Schmidt RC., Schmitt T., Schodde R., Schoeman CS.,
599 Schweiger S., Schwertner CF., Seamark ECJ., Semedo TBF., Shin MK., Siler CD., Silveira
600 LF., Simison WB., Simões M., Sites JW., Smith BT., Smith KT., Song W., Soulier-Perkins
601 A., Sousa LM., Sparks JS., Stampar SN., Steiner FM., Steyer JS., Stiassny MLJ., Stoeck T.,

602 Stopiglia R., Streicher JW., Sturaro MJ., Stys P., Swierk L., Taeger A., Takiya DM.,
603 Taphorn DC., Tavares M., Tavares VDC., Taylor PJ., Tello JG., Teta P., Tillack F., Timm
604 RM., Tokaryk T., Tominaga A., Tonini JFR., Tornabene L., Torres-Carvajal O., Townsend
605 J., Trape JF., Rodrigues MT., Trusch R., Tschopp E., Uhl D., Upham NS., Vacher JP.,
606 Valdesalici S., Van Bocxlaer B., Van Cakenberghe V., Van De Kamp T., Van De Velde I.,
607 Van Den Spiegel D., Vanhove MPM., Vasudevan K., Veerappan D., Velazco PM., Verdade
608 VK., Verheyen E., Vieira LM., Victoriano PF., Vitt LJ., Wagner P., Watkins-Colwell GJ.,
609 Weisse T., Werneck FP., Wheeler WC., Wilson DE., Valero KCW., Wood PL., Woodman
610 N., Quetzalli HDY., Yoshikawa N., Zaher H., Ziegler T., Zima J., Zink RM., Zug G. 2016.
611 Photography-based taxonomy is inadequate, unnecessary, and potentially harmful for
612 biological sciences. *Zootaxa* 4196:435–445. DOI: 10.11646/zootaxa.4196.3.9.

613 Chazot N., Wahlberg N., Freitas AVL., Mitter C., Labandeira C., Sohn J-C., Sahoo RK.,
614 Seraphim N., de Jong R., Heikkilä M. 2018. The trials and tribulations of priors and
615 posteriors in Bayesian timing of divergence analyses: the age of butterflies revisited.
616 *bioRxiv* 259184. DOI: 10.1104/pp.110.169599.

617 Cicconardi F., Borges PAV., Strasberg D., Oromí P., López H., Pérez-Delgado AJ., Casquet J.,
618 Caujapé-Castells J., Fernández-Palacios JM., Thébaud C., Emerson BC. 2017. MtDNA
619 metagenomics reveals large-scale invasion of belowground arthropod communities by
620 introduced species. *Molecular Ecology* 26:3104–3115. DOI: 10.1111/mec.14037.

621 Corthals A., Desalle R. 2005. An application of tissue and DNA banking for genomics and
622 conservation: the Ambrose Monell Cryo-Collection (AMCC). *Systematic Biology* 54:819–
623 823. DOI: 10.1080/10635150590950353.

624 Cox-Foster DL., Conlan S., Holmes EC., Palacios G., Evans JD., Moran NA., Quan P-L., Briese
625 T., Hornig M., Geiser DM., Martinson V., Kalkstein AL., Drysdale A., Hui J., Zhai J., Cui
626 L., Hutchison SK., Simons JF., Egholm M., Pettis JS., Lipkin WI. 2007. A metagenomic
627 survey of microbes in honey bee Colony Collapse Disorder. *Science* 318:283–288.

628 Crampton-Platt A., Timmermans MJTN., Gimmel ML., Kutty SN., Cockerill TD., Khen CV.,
629 Vogler AP. 2015. Soup to tree: the phylogeny of beetles inferred by mitochondrial
630 metagenomics of a Bornean rainforest sample. *Molecular Biology and Evolution* 32:2302–
631 2316. DOI: 10.1093/molbev/msv111.

632 Crampton-Platt A., Yu DW., Zhou X., Vogler AP. 2016. Mitochondrial metagenomics: letting
633 the genes out of the bottle. *GigaScience* 5:15. DOI: 10.1186/s13742-016-0120-y.

634 Deagle BE., Jarman SN., Coissac E., Pompanon F., Taberlet P. 2014. DNA metabarcoding and
635 the cytochrome c oxidase subunit I marker: not a perfect match. *Biology Letters*
636 10:20140562. DOI: 10.1098/rsbl.2014.0562.

637 Dillon N., Austin AD., Bartowsky E. 1996. Comparison of preservation techniques for DNA
638 extraction from hymenopterous insects. *Insect Molecular Biology* 5:21–24. DOI:
639 10.1111/j.1365-2583.1996.tb00036.x.

640 Dowle EJ., Pochon X., C Banks J., Shearer K., Wood SA. 2016. Targeted gene enrichment and
641 high-throughput sequencing for environmental biomonitoring: a case study using freshwater
642 macroinvertebrates. *Molecular Ecology Resources* 16:1240–1254. DOI: 10.1111/1755-
643 0998.12488.

644 Dubois A. 2011. The *International Code of Zoological Nomenclature* must be drastically

improved before it is too late. *Bionomina* 2:1–104. DOI: 10.11646/bionomina.2.1.1.

Ebach MC., Valdecasas AG., Wheeler QD. 2011. Impediments to taxonomy and users of taxonomy: accessibility and impact evaluation. *Cladistics* 27:550–557. DOI: 10.1111/j.1096-0031.2011.00348.x.

Faircloth BC. 2016. Identifying conserved genomic elements and designing universal probe sets to enrich them. *bioRxiv*:77172. DOI: 10.1101/077172.

Faircloth BC., Branstetter MG., White ND., Brady SG. 2015. Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among Hymenoptera. *Molecular Ecology Resources* 15:489–501. DOI: 10.1111/1755-0998.12328.

Ferro ML., Park J-S. 2013. Effect of propylene glycol concentration on mid-term DNA preservation of Coleoptera. *The Coleopterists Bulletin* 67:581–586. DOI: 10.1649/0010-065X-67.4.581.

Frøslev TG., Kjøller R., Bruun HH., Ejrnæs R., Brunbjerg AK., Pietroni C., Hansen AJ. 2017. Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. *Nature Communications* 8:1188. DOI: 10.1038/s41467-017-01312-x.

García-Robledo C., Erickson DL., Staines CL., Erwin TL., Kress WJ. 2013. Tropical plant-herbivore networks: reconstructing species interactions using DNA barcodes. *PLoS ONE* 8:e52967. DOI: 10.1371/journal.pone.0052967.

Garnett ST., Christidis L. 2017. Taxonomy anarchy hampers conservation. *Nature* 546:25–27. DOI: 10.1038/546025a.

666 Gilbert MTP., Moore W., Melchior L., Worebey M. 2007a. DNA extraction from dry museum
667 beetles without conferring external morphological damage. *PLoS ONE* 2:e272. DOI:
668 10.1371/journal.pone.0000272.

669 Gilbert MTP., Sanchez JJ., Haselkorn T., Jewell LD., Lucas SB., Van Marck E., Børsting C.,
670 Morling N., Worobey M. 2007b. Multiplex PCR with minisequencing as an effective high-
671 throughput SNP typing method for formalin-fixed tissue. *Electrophoresis* 28:2361–2367.
672 DOI: 10.1002/elps.200600589.

673 Gilbert MTP., Tomsho LP., Rendulic S., Packard M., Drautz DI., Sher A., Tikhonov A., Dalén
674 L., Kuznetsova T., Kosintsev P., Campos PF., Higham T., Collins MJ., Wilson AS.,
675 Shidlovskiy F., Buigues B., Ericson PGP., Germonpré M., Götherström A., Iacumin P.,
676 Nikolaev V., Nowak-Kemp M., Willerslev E., Knight JR., Irzyk GP., Perbost CS.,
677 Fredrikson KM., Harkins TT., Sheridan S., Miller W., Schuster SC. 2007c. Whole-genome
678 shotgun sequencing of mitochondria from ancient hair shafts. *Science* 317:1927–1930. DOI:
679 10.1126/science.1146971.

680 Gómez-Rodríguez C., Timmermans MJTN., Crampton-Platt A., Vogler AP. 2017. Intraspecific
681 genetic variation in complex assemblages from mitochondrial metagenomics: comparison
682 with DNA barcodes. *Methods in Ecology and Evolution* 8:248–256. DOI: 10.1111/2041-
683 210X.12667.

684 Gómez IC., Sääksjärvi IE., Mayhew PJ., Pollet M., Rey del Castillo C., Nieves-Aldrey JL.,
685 Broad GR., Roininen H., Tuomisto H. 2018. Variation in the species richness of parasitoid
686 wasps (Ichneumonidae: Pimplinae and Rhyssinae) across sites on different continents.
687 *Insect Conservation and Diversity*:In Press. DOI: 10.1111/icad.12281.

688 Grimaldi D., Engel MS. 2005. *Evolution of the Insects*. New York, USA: Cambridge University
689 Press.

690 Hamilton AJ., Basset Y., Benke KK., Grimbacher PS., Miller SE., Novotný V., Samuelson GA.,
691 Stork NE., Weiblen GD., Yen JDL. 2010. Quantifying uncertainty in estimation of tropical
692 arthropod species richness. *The American Naturalist* 176:90–95. DOI: 10.1086/652998.

693 Hammer TJ., Janzen DH., Hallwachs W., Jaffe SP., Fierer N. 2017. Caterpillars lack a resident
694 gut microbiome. *Proceedings of the National Academy of Sciences* 114:9641–9646. DOI:
695 10.1073/pnas.1707186114.

696 Heintzman PD., Elias SA., Moore K., Paszkiewicz K., Barnes I. 2014. Characterizing DNA
697 preservation in degraded specimens of *Amara alpina* (Carabidae: Coleoptera). *Molecular*
698 *Ecology Resources* 14:606–615. DOI: 10.1111/1755-0998.12205.

699 Hendrich L., Morinière J., Haszprunar G., Hebert PDN., Hausmann A., Köhler F., Balke M.
700 2015. A comprehensive DNA barcode database for Central European beetles with a focus
701 on Germany: adding more than 3500 identified species to BOLD. *Molecular Ecology*
702 *Resources* 15:795–818. DOI: 10.1111/1755-0998.12354.

703 Höfer H., Astrin J., Holstein J., Spelda J., Meyer F., Zarte N. 2015. Propylene glycol – a useful
704 capture preservative for spiders for DNA barcoding. *Arachnologische Mitteilungen* 50:30–
705 36. DOI: 10.5431/aramit5005.

706 Huse SM., Welch DM., Morrison HG., Sogin ML. 2010. Ironing out the wrinkles in the rare
707 biosphere through improved OTU clustering. *Environmental Microbiology* 12:1889–1898.
708 DOI: 10.1111/j.1462-2920.2010.02193.x.

709 Ji Y., Ashton L., Pedley SM., Edwards DP., Tang Y., Nakamura A., Kitching R., Dolman PM.,
710 Woodcock P., Edwards FA., Larsen TH., Hsu WW., Benedick S., Hamer KC., Wilcove
711 DS., Bruce C., Wang X., Levi T., Lott M., Emerson BC., Yu DW. 2013. Reliable, verifiable
712 and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters* 16:1245–1257.
713 DOI: 10.1111/ele.12162.

714 Jiggins CD. 2016. *The Ecology and Evolution of Heliconius Butterflies*. Oxford, U.K.: Oxford
715 University Press. DOI: 10.1093/acprof:oso/9780199566570.001.0001.

716 Johnson C., Agosti D., Delabie JH., Dumpert K., Williams DJ., Von Tschirnhaus M., Maschwitz
717 U. 2001. Acropyga and Azteca ants (Hymenoptera: Formicidae) with scale insects
718 (Sternorrhyncha: Coccoidea): 20 million years of intimate symbiosis. *American Museum*
719 *Novitates* 3335:1–18.

720 Kanda K., Pflug JM., Sproul JS., Dasenko MA., Maddison DR. 2015. Successful recovery of
721 nuclear protein-coding genes from small insects in museums using Illumina sequencing.
722 *PLoS ONE* 10:e0143929. DOI: 10.1371/journal.pone.0143929.

723 Kelley JL., Peyton JT., Fiston-Lavier AS., Teets NM., Yee MC., Johnston JS., Bustamante CD.,
724 Lee RE., Denlinger DL. 2014. Compact genome of the Antarctic midge is likely an
725 adaptation to an extreme environment. *Nature Communications* 5:4611. DOI:
726 10.1038/ncomms5611.

727 Knölke S., Erlacher S., Hausmann A., Miller MA., Segerer AH. 2005. A procedure for combined
728 genitalia dissection and DNA extraction in Lepidoptera. *Insect Systematics and Evolution*
729 35:401–409. DOI: 10.1163/187631204788912463.

730 Koebler J. 2013. Earth's life-forms collected to aid in genetic research. *National Geographic*
731 *News*:USA.

732 Labandeira CC., Sepkoski J. J. 1993. Insect diversity in the fossil record. *Science* 261:310–315.

733 Lamarre GPA., Molto Q., Fine PVA., Baraloto C. 2012. A comparison of two common flight
734 interception traps to survey tropical arthropods. *ZooKeys* 216:43–55. DOI:
735 10.3897/zookeys.216.3332.

736 Larsen TH. 2016. *Core standardized methods for rapid biological field assessment*. Arlington,
737 VA: Conservation International. DOI: 10.1017/CBO9780511525438.

738 Lemmon EM., Lemmon AR. 2013. High-throughput genomic data in systematics and
739 phylogenetics. *Annual Review of Ecology, Evolution, and Systematics* 44:99–121. DOI:
740 10.1146/annurev-ecolsys-110512-135822.

741 Lim GS., Balke M., Meier R. 2012. Determining species boundaries in a world full of rarity:
742 Singletons, species delimitation methods. *Systematic Biology* 61:165–169. DOI:
743 10.1093/sysbio/syr030.

744 Linard B., Arribas P., Andújar C., Crampton-Platt A., Vogler AP. 2016. Lessons from genome
745 skimming of arthropod-preserving ethanol. *Molecular Ecology Resources* 16:1365–1377.
746 DOI: 10.1111/1755-0998.12539.

747 Liu S., Wang X., Xie L., Tan M., Li Z., Su X., Zhang H., Misof B., Kjer KM., Tang M., Niehuis
748 O., Jiang H., Zhou X. 2016. Mitochondrial capture enriches mito-DNA 100 fold, enabling
749 PCR-free mitogenomics biodiversity analysis. *Molecular Ecology Resources* 16:470–479.
750 DOI: 10.1111/1755-0998.12472.

751 Lugg WH., Griffiths J., van Rooyen AR., Weeks AR., Tingley R. 2018. Optimal survey designs
752 for environmental DNA sampling. *Methods in Ecology and Evolution*:In Press. DOI:
753 10.1111/ijlh.12426.

754 Maddison DR., Cooper KW. 2014. Species delimitation in the ground beetle subgenus
755 *Liocosmius* (Coleoptera: Carabidae: *Bembidion*), including standard and next-generation
756 sequencing of museum specimens. *Zoological Journal of the Linnean Society* 172:741–770.
757 DOI: 10.1111/zoj.12188.

758 Mamanova L., Coffey AJ., Scott CE., Kozarewa I., Turner EH., Kumar A., Howard E., Shendure
759 J., Turner DJ. 2010. Target-enrichment strategies for next-generation sequencing. *Nature*
760 *Methods* 7:111–118. DOI: 10.1038/nmeth.1419.

761 Mardis ER. 2017. DNA sequencing technologies: 2006-2016. *Nature Protocols* 12:213–218.
762 DOI: 10.1038/nprot.2016.182.

763 Marshall CR. 2017. Five palaeobiological laws needed to understand the evolution of the living
764 biota. *Nature Ecology and Evolution* 1:165. DOI: 10.1038/s41559-017-0165.

765 McMurdie PJ., Holmes S. 2014. Waste not, want not: why rarefying microbiome data is
766 inadmissible. *PLoS Computational Biology* 10:e1003531. DOI:
767 10.1371/journal.pcbi.1003531.

768 Metzker ML. 2010. Sequencing technologies - the next generation. *Nature Reviews Genetics*
769 11:31–46. DOI: 10.1038/nrg2626.

770 Meyer M., Arsuaga JL., De Filippo C., Nagel S., Aximu-Petri A., Nickel B., Martínez I., Gracia
771 A., De Castro JMB., Carbonell E., Viola B., Kelso J., Prüfer K., Pääbo S. 2016. Nuclear

772 DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature*
773 531:504–507. DOI: 10.1038/nature17405.

774 Misof B., Liu S., Meusemann K., Peters RS., Donath A., Mayer C., Frandsen PB., Ware J.,
775 Flouri T., Beutel RG., Niehuis O., Petersen M., Izquierdo-Carrasco F., Wappler T., Rust J.,
776 Aberer AJ., Aspöck U., Aspöck H., Bartel D., Blanke A., Berger S., Böhm A., Buckley TR.,
777 Calcott B., Chen J., Friedrich F., Fukui M., Fujita M., Greve C., Grobe P., Gu S., Huang Y.,
778 Jermiin LS., Kawahara AY., Krogmann L., Kubiak M., Lanfear R., Letsch H., Li Y., Li Z.,
779 Li J., Lu H., Machida R., Mashimo Y., Kapli P., McKenna DD., Meng G., Nakagaki Y.,
780 Navarrete-Heredia JL., Ott M., Ou Y., Pass G., Podsiadlowski L., Pohl H., von Reumont
781 BM., Schütte K., Sekiya K., Shimizu S., Slipinski A., Stamatakis A., Song W., Su X.,
782 Szucsich NU., Tan M., Tan X., Tang M., Tang J., Timelthaler G., Tomizuka S., Trautwein
783 M., Tong X., Uchifume T., Walzl MG., Wiegmann BM., Wilbrandt J., Wipfler B., Wong
784 TKF., Wu Q., Wu G., Xie Y., Yang S., Yang Q., Yeates DK., Yoshizawa K., Zhang Q.,
785 Zhang R., Zhang W., Zhang Y., Zhao J., Zhou C., Zhou L., Ziesmann T., Zou S., Li Y., Xu
786 X., Zhang Y., Yang H., Wang J., Wang J., Kjer KM., Zhou X. 2014a. Phylogenomics
787 resolves the timing and pattern of insect evolution. *Science* 346:763–767. DOI:
788 10.1126/science.1257570.

789 Misof B., Meusemann K., von Reumont BM., Kück P., Prohaska SJ., Stadler PF. 2014b. *A priori*
790 assessment of data quality in molecular phylogenetics. *Algorithms for Molecular Biology*
791 9:22. DOI: 10.1186/s13015-014-0022-4.

792 Mora C., Tittensor DP., Adl S., Simpson AGB., Worm B. 2011. How many species are there on
793 earth and in the ocean? *PLoS Biology* 9:e1001127. DOI: 10.1371/journal.pbio.1001127.

794 Morinière J., Cancian de Araujo B., Lam AW., Hausmann A., Balke M., Schmidt S., Hendrich
795 L., Doczkal D., Fartmann B., Arvidsson S., Haszprunar G. 2016. Species identification in
796 Malaise trap samples by DNA barcoding based on NGS technologies and a scoring matrix.
797 *Plos ONE* 11:e0155497. DOI: 10.1371/journal.pone.0155497.

798 Nieman CC., Yamasaki Y., Collier TC., Lee Y. 2015. A DNA extraction protocol for improved
799 DNA yield from individual mosquitoes. *F1000Research* 4:1314. DOI:
800 10.12688/f1000research.7413.1.

801 Noyes JS. 1989. A study of five methods of sampling Hymenoptera (Insecta) in a tropical
802 rainforest, with special reference to the Parasitica. *Journal of Natural History* 23:285–298.
803 DOI: 10.1080/00222938900770181.

804 Oliver AK., Brown SP., Callaham MA., Jumpponen A. 2015. Polymerase matters: non-
805 proofreading enzymes inflate fungal community richness estimates by up to 15 %. *Fungal*
806 *Ecology* 15:86–89. DOI: 10.1016/j.funeco.2015.03.003.

807 Pawluczyk M., Weiss J., Links MG., Egaña Aranguren M., Wilkinson MD., Egea-Cortines M.
808 2015. Quantitative evaluation of bias in PCR amplification and next-generation sequencing
809 derived from metabarcoding samples. *Analytical and Bioanalytical Chemistry* 407:1841–
810 1848. DOI: 10.1007/s00216-014-8435-y.

811 Pitteloud C., Arrigo N., Suchan T., Mastretta-Yanes A., Vila R., Dinca V., Hernandez-Roldan J.,
812 Brockmann E., Chittaro Y., Kleckova I., Fumagalli L., Buerki S., Alvarez N. 2017. Climatic
813 niche evolution is faster in sympatric than allopatric lineages of the butterfly genus *Pyrgus*.
814 *Proceedings of the Royal Society B: Biological Sciences* 284:20170208.

Pompanon F., Deagle BE., Symondson WOC., Brown DS., Jarman SN., Taberlet P. 2012. Who is eating what: diet assessment using next generation sequencing. *Molecular Ecology* 21:1931–1950. DOI: 10.1111/j.1365-294X.2011.05403.x.

Prendini L., Hanner R., DeSalle R. 2002. Obtaining, storing and archiving specimens and tissue samples for use in molecular studies. In: DeSalle R, Giribet G, Wheeler W eds. *Techniques in Molecular Systematics and Evolution*. Basel, Switzerland: Birkhäuser Verlag, 176–248. DOI: 10.1007/978-3-0348-8125-8_11.

Raposo MA., Stopiglia R., Brito GRR., Bockmann FA., Kirwan GM., Gayon J., Dubois A. 2017. What really hampers taxonomy and conservation? A riposte to Garnett and Christidis (2017). *Zootaxa* 4317:179–184. DOI: 10.11646/zootaxa.4317.1.10.

Ratnasingham S., Hebert PDN. 2007. BOLD: the Barcode of Life Data System (www.barcodinglife.org). *Molecular Ecology Notes* 7:355–364. DOI: 10.1111/j.1471-8286.2006.01678.x.

Reiss RA., Schwert DP., Ashworth AC. 1995. Field preservation of Coleoptera for molecular genetic analyses. *Environmental Entomology* 24:716–719.

Richards S., Murali SC. 2015. Best practices in insect genome sequencing: what works and what doesn't. *Current Opinion in Insect Science* 7:1–7. DOI: 10.1016/j.cois.2015.02.013.

Ritter CD., McCrate G., Nilsson RH., Fearnside PM., Palme U., Antonelli A. 2017. Environmental impact assessment in Brazilian Amazonia: challenges and prospects to assess biodiversity. *Biological Conservation* 206:161–168. DOI: 10.1016/j.biocon.2016.12.031.

836 Ruane S., Austin CC. 2017. Phylogenomics using formalin-fixed and 100+ year old intractable
837 natural history specimens. *Molecular Ecology Resources* 17:1003–1008. DOI:
838 10.1111/ijlh.12426.

839 Scheffers BR., Joppa LN., Pimm SL., Laurance WF. 2012. What we know and don't know about
840 Earth's missing biodiversity. *Trends in Ecology and Evolution* 27:501–510. DOI:
841 10.1016/j.tree.2012.05.008.

842 Schloss PD., Westcott SL., Ryabin T., Hall JR., Hartmann M., Hollister EB., Lesniewski RA.,
843 Oakley BB., Parks DH., Robinson CJ., Sahl JW., Stres B., Thallinger GG., Van Horn DJ.,
844 Weber CF. 2009. Introducing mothur: open-source, platform-independent, community-
845 supported software for describing and comparing microbial communities. *Applied and*
846 *Environmental Microbiology* 75:7537–7541. DOI: 10.1128/AEM.01541-09.

847 Schoonvaere K., De Smet L., Smagghe G., Vierstraete A., Braeckman BP., De Graaf DC. 2016.
848 Unbiased RNA shotgun metagenomics in social and solitary wild bees detects associations
849 with eukaryote parasites and new viruses. *PLoS ONE* 11:e0168456. DOI:
850 10.1371/journal.pone.0168456.

851 Shendure J., Balasubramanian S., Church GM., Gilbert W., Rogers J., Schloss JA., Waterston
852 RH. 2017. DNA sequencing at 40: past, present and future. *Nature* 550:345–353. DOI:
853 10.1038/nature24286.

854 Shi W., Xie S., Chen X., Sun S., Zhou X., Liu L., Gao P., Kyrpides NC., No EG., Yuan JS. 2013.
855 Comparative genomic analysis of the endosymbionts of herbivorous insects reveals eco-
856 environmental adaptations: biotechnology applications. *PLoS Genetics* 9:e1003131. DOI:
857 10.1371/journal.pgen.1003131.

Shokralla S., Gibson JF., King I., Baird DJ., Janzen DH., Hallwachs W., Hajibabaei M. 2016. Environmental DNA barcode sequence capture: targeted, PCR-free sequence capture for biodiversity analysis from bulk environmental samples. *bioRxiv*:87437. DOI: 10.1101/087437.

Shokralla S., Spall JL., Gibson JF., Hajibabaei M. 2012. Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology* 21:1794–1805. DOI: 10.1111/j.1365-294X.2012.05538.x.

Short AEZ., Dikow T., Moreau CS. 2018. Entomological collections in the Age of Big Data. *Annual Review of Entomology* 63:513–530. DOI: 10.1146/annurev-ento-031616-035536.

Srivathsan A., Ang A., Vogler AP., Meier R. 2016. Fecal metagenomics for the simultaneous assessment of diet, parasites, and population genetics of an understudied primate. *Frontiers in Zoology* 13:17. DOI: 10.1186/s12983-016-0150-4.

Staats M., Erkens RHJ., van de Vossenberg B., Wieringa JJ., Kraaijeveld K., Stielow B., Geml J., Richardson JE., Bakker FT. 2013. Genomic treasure troves: complete genome sequencing of herbarium and insect museum specimens. *PLoS ONE* 8:e69189. DOI: 10.1371/journal.pone.0069189.

Stork NE. 2018. How many species of insects and other terrestrial arthropods are there on Earth? *Annual Review of Entomology* 63:31–45. DOI: 10.1146/annurev-ento-020117-043348.

Stork NE., McBroom J., Gely C., Hamilton AJ. 2015. New approaches narrow global species estimates for beetles, insects, and terrestrial arthropods. *Proceedings of the National Academy of Sciences of the United States of America* 112:7519–7523. DOI:

879 10.1073/pnas.1502408112.

880 Suarez A V., Tsutsui ND. 2004. The value of museum collections for research and society.
881 *BioScience* 54:66–74. DOI: 10.1641/0006-3568(2004)054[0066:TVOMCF]2.0.CO;2.

882 Suchan T., Pitteloud C., Gerasimova NS., Kostikova A., Schmid S., Arrigo N., Pajkovic M.,
883 Ronikier M., Alvarez N. 2016. Hybridization capture using RAD probes (hyRAD), a new
884 tool for performing genomic analyses on collection specimens. *PLoS ONE* 11:e0151651.
885 DOI: 10.1371/journal.pone.0151651.

886 Suen G., Scott JJ., Aylward FO., Adams SM., Tringe SG., Pinto-Tomás AA., Foster CE., Pauly
887 M., Weimer PJ., Barry KW., Goodwin LA., Bouffard P., Li L., Osterberger J., Harkins TT.,
888 Slater SC., Donohue TJ., Currie CR. 2010. An insect herbivore microbiome with high plant
889 biomass-degrading capacity. *PLoS Genetics* 6:e1001129. DOI:
890 10.1371/journal.pgen.1001129.

891 Taberlet P., Coissac E., Pompanon F., Brochmann C., Willerslev E. 2012. Towards next-
892 generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*
893 21:2045–2050. DOI: 10.1111/j.1365-294X.2012.05470.x.

894 Tagliavia M., Massa B., Albanese I., La Farina M. 2011. DNA extraction from Orthoptera
895 museum specimens. *Analytical Letters* 44:1058–1062. DOI:
896 10.1080/00032719.2010.506939.

897 Tang M., Tan M., Meng G., Yang S., Su X., Liu S., Song W., Li Y., Wu Q., Zhang A., Zhou X.
898 2014. Multiplex sequencing of pooled mitochondrial genomes - a crucial step toward
899 biodiversity analysis using mito-metagenomics. *Nucleic Acids Research* 42:e166. DOI:

10.1093/nar/gku917.

Tedersoo L., Anslan S., Bahram M., Põlme S., Riit T., Liiv I., Kõljalg U., Kisand V., Nilsson RH., Hildebrand F., Bork P., Abarenkov K. 2015. Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi. *MycoKeys* 10:1–43. DOI: 10.3897/mycokeys.10.4852.

Thompson JF., Milos PM. 2011. The properties and applications of single-molecule DNA sequencing. *Genome Biology* 12:217. DOI: 10.1186/GB-2011-12-2-217.

Thomsen PF., Elias S., Gilbert MTP., Haile J., Munch K., Kuzmina S., Froese DG., Sher A., Holdaway RN., Willerslev E. 2009. Non-destructive sampling of ancient insect DNA. *PloS one* 4:e5048. DOI: 10.1371/journal.pone.0005048.

Thorpe SE. 2017. Is photography-based taxonomy really inadequate, unnecessary, and potentially harmful for biological sciences? A reply to Ceríaco *et al*. (2016). *Zootaxa* 4226:449–450. DOI: 10.11646/zootaxa.4226.3.9.

Timmermans MJTN., Viberg C., Martin G., Hopkins K., Vogler AP. 2016. Rapid assembly of taxonomically validated mitochondrial genomes from historical insect collections. *Biological Journal of the Linnean Society* 117:83–95.

Tin MM-Y., Economo EP., Mikheyev AS. 2014. Sequencing degraded DNA from non-destructively sampled museum specimens for RAD-tagging and low-coverage shotgun phylogenetics. *PLoS ONE* 9:e96793. DOI: 10.1371/journal.pone.0096793.

Toju H. 2015. High-throughput DNA barcoding for ecological network studies. *Population Ecology* 57:37–51. DOI: 10.1007/s10144-014-0472-z.

Veijalainen A., Wahlberg N., Broad GR., Erwin TL., Longino JT., Sääksjärvi IE. 2012. Unprecedented ichneumonid parasitoid wasp diversity in tropical forests. *Proceedings of the Royal Society B: Biological Sciences* 279:4694–4698. DOI: 10.1098/rspb.2012.1664.

Vellend M. 2017. The biodiversity conservation paradox. *American Scientist* 105:94–101. DOI: 10.1511/2011.89.106.

Vesterinen EJ., Ruokolainen L., Wahlberg N., Peña C., Roslin T., Laine VN., Vasko V., Sääksjärvi IE., Norrdahl K., Lilley TM. 2016. What you need is what you eat? Prey selection by the bat *Myotis daubentonii*. *Molecular Ecology* 25:1581–1594. DOI: 10.1111/mec.13564.

Vogel G. 2017. Where have all the insects gone? *Science* 356:576–579.

Wachi N., Matsubayashi KW., Maeto K. 2018. Application of next-generation sequencing to the study of non-model insects. *Entomological Science* 21:3–11. DOI: 10.1111/ens.12281.

Wheat CW., Wahlberg N. 2013. Critiquing blind dating: the dangers of over-confident date estimates in comparative genomics. *Trends in Ecology & Evolution* 28:636–642. DOI: 10.1016/j.tree.2013.07.007.

Wheeler QD., Raven PH., Wilson EO. 2004. Taxonomy: impediment or expedient? *Science* 303:285. DOI: 10.1126/science.303.5656.285.

Yeates DK., Meusemann K., Trautwein M., Wiegmann B., Zwick A. 2016. Power, resolution and bias: recent advances in insect phylogeny driven by the genomic revolution. *Current Opinion in Insect Science* 13:16–23. DOI: 10.1016/j.cois.2015.10.007.

Young AD., Lemmon AR., Skevington JH., Mengual X., Ståhls G., Reemer M., Jordaens K.,

Kelso S., Lemmon EM., Hauser M., De Meyer M., Misof B., Wiegmann BM. 2016. Anchored enrichment dataset for true flies (order Diptera) reveals insights into the phylogeny of flower flies (family Syrphidae). *BMC Evolutionary Biology* 16:143. DOI: 10.1186/s12862-016-0714-0.

Zhou X., Li Y., Liu S., Yang Q., Su X., Zhou L., Tang M., Fu R., Li J., Huang Q. 2013. Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *GigaScience* 2:4. DOI: 10.1186/2047-217X-2-4.

Funding Statement

Funding to PMM was provided by a Marie Skłodowska-Curie fellowship (project MARIPOSAS-704035). CDR received the support from CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brazil). CJB was funded by the Aage V. Jensen Naturfond of Denmark (1121721001). DM and NW received funding from European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 642241 (BIG4 project). NW received funding from the Swedish Research Council. AA is supported by grant from the Knut and Alice Wallenberg Foundation, the Swedish Research Council (B0569601), the Swedish Foundation for Strategic Research, the Faculty of Sciences at the University of Gothenburg, and the David Rockefeller Center for Latin American Studies at Harvard University.

Author Contributions

963 PMM, CDR and AA conceived and led the workshop; all authors participated in the workshop
964 titled “Biodiversity research through the study of insect genomics” organized by the Gothenburg
965 Global Biodiversity Centre in Gothenburg, Sweden; all authors contributed with discussion and
966 ideas for the paper; PMM, CDR and AA organized the structure of the article, PMM wrote the
967 first draft of the article, and all authors contributed to and approved the final version of the
968 article.

969

970 **Conflict of Interests**

971 The authors declare no conflict of interests.

972

973 **Tables**

974 **Table 1. Representative description of methods for mass sampling of insects and their**
975 **application for NGS.** Note that this is not a comprehensive list and is only aimed at providing
976 an overview of available possibilities of widespread use. In the Costs of equipments, we roughly
977 categorized them as Low (approx. < US $50), Medium (approx. US $50 – $100), High (approx.
978 > US $100).

979 **Table 2: Overview of massively parallel DNA sequencing methods applied to insect**
980 **museum specimens.**

981 **Table 3: Non-comprehensive overview of massively parallel DNA sequencing methods**
982 **applied to insects.**

983

984 **Figures**

985 **Figure 1: Schematic of the spatiotemporal continuum of biodiversity and evolution.**
986 Community ecology research usually focus at local scale (Lx), but may be mislead if not taking
987 into account higher-level patterns such as fluctuation of population/species ranges over time
988 (e.g., distributional change due to environmental shift or biotic interaction) and space (e.g., local
989 pool of species depends on the regional diversity and its evolution). At a higher scale (Vx),
990 inferences on population/species evolutionary dynamics may be mislead if not taking into
991 account reduced-scale patterns such as local extinction (e.g., its rate and magnitude, whether
992 local, regional or global) and biotic interaction (e.g., as in diversity-dependent diversification, or
993 continuous replacement of species with similar ecologies over time). Vx is any area of interest

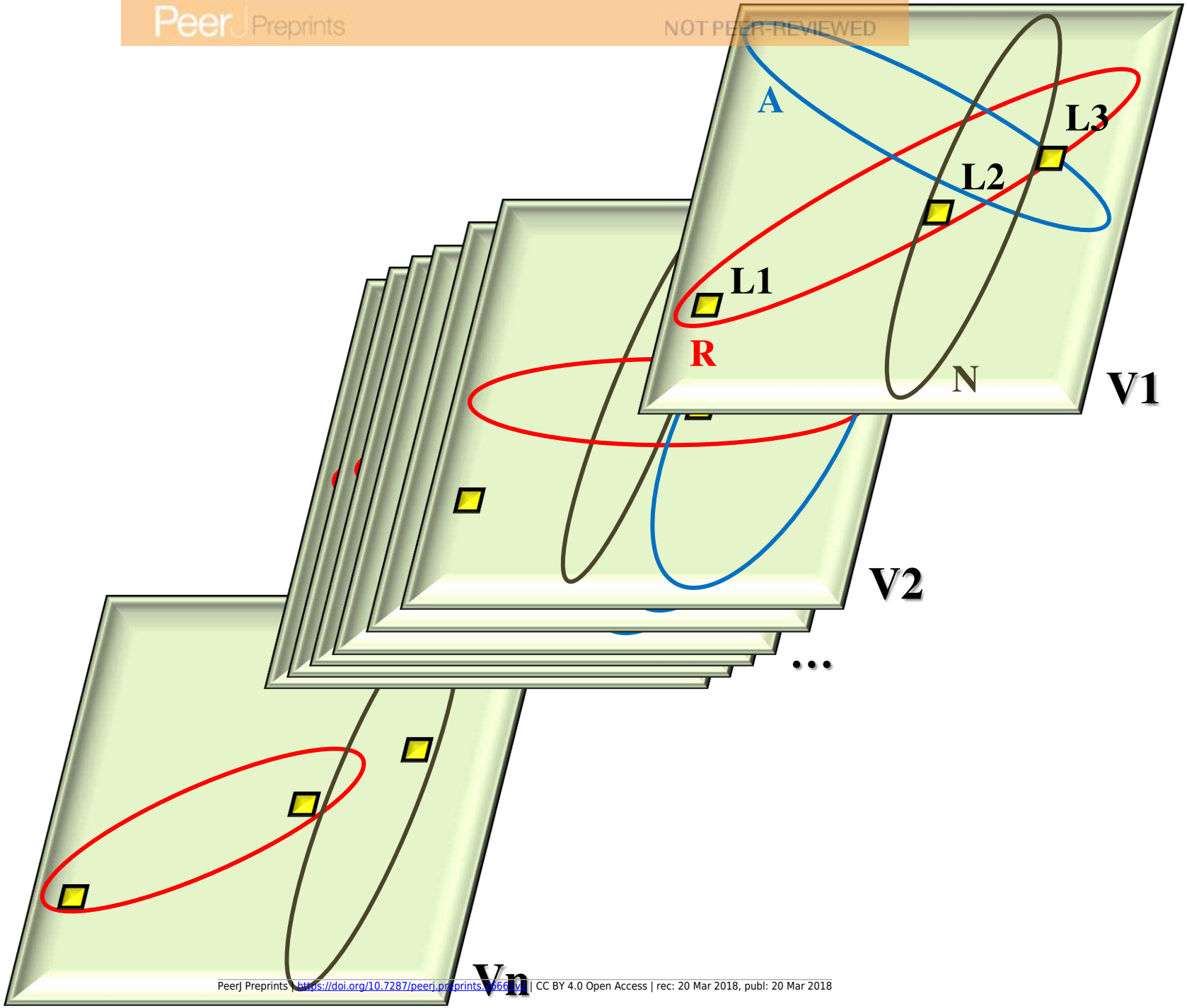
(local, regional, or global) that changes over time (V1, V2 … Vn). A, R, and N are distributional ranges of three different populations/species of interest. Hypothetical sampling sites remain geographically constant over time, whether at geological or ecological scales (L1, L2, and L3).

Figure 2: Flowchart summarizing the 10 rules to study biodiversity through insect genomics and discussed in the paper.

Method	Example	Taxa targeted	Costs	Suitability for genomic research	Sampling effort	Limitations
Trap-sampling	*Van Someren-Rydon*	Fruit-feeding butterflies, from forest floor to canopy	**Low**; negligible if self-built	**Yes**; no killing reagent; baits such as fermented fruit	**Minimum**: 5 traps in forest, 10 traps in open areas; **Collection**: once or twice per day; **Personnel**: 2 people, collection and record; **Complement** with opportunistic hand collection	Need for long-term data because different butterfly communities throughout the year; Other feeding guilts are missing, such as nectar-feeding
Trap-sampling	*Pitfall*	Forest floor insects such as dung beetles, flies, ants	**Low**; negligible if self-built	**Depending on killing reagent**; could be done with detergent and water, propylen glycol; baits such as human dung	**Minimum**: 20 traps per day; linear transect; **Collection**: at least once per day; **Personnel**: 1 person; **Complement** with flight intercept traps	Specimens with a lot of water from pitfall trap, thus a lot of ethanol replaced every week is needed to prevent DNA decay
Leaf-litter collector	*Mini-Winkler*	Leaf-litter and soil insects, such as ants, beetles	**Medium**	**Yes**; 95% EtOH most commonly used as killing reagent	**Minimum**: 20 collectors, each with 1m^2 leaf litter; **Collection**: once, if extraction run in parallel; **Personnel**: 2 people recommended; **Complement** with bait-traps and hand collecting	Limited to forested areas, and not suitable during height of dry or rain season; Not sampling of vegetation-associated, canopy or subterranean insects
Flying-insect collector	*Malaise*	Strong-flying insects, such as Hymenoptera and Diptera	**High**	**Yes**; 95% EtOH most commonly used as killing reagent	**Minimum**: 2 traps for fast surveys; **Collection**: little care, leave in field for 2–4 weeks; **Personnel**: 1 person; **Complement** with flight interception traps	Placement of trap in "likely" flight paths, thus a component of subjectivity is introduced
Flying-insect collector	*Flight interception*	Flying insects, such as beetles, cockroaches, crickets	**Low**; negligible if self-built	**Depending on killing reagent**; could be done in salt-saturated water and detergent, propylen glycol; formaldehyde solutions but in detriment of DNA recovery	**Minimum**: 2 traps for fast surveys; **Collection**: once or twice per day; **Personnel**: 1 person; **Complement** with bait and light traps	Ideal for slow-flying insects, which hit the plastic sheet and fall in the container with killing reagent
Insecticidal knockdown	*Canopy fogging*	Arboreal insect community	**High**	**Yes**; insecticide as killing reagent	**Collection**: laborious and problems with pseudoreplication; **Complement** with canopy light trapping and flight interception traps	Canopy access still limited; High demand on logistics; Environmentally dependent

Authors	Taxon group	Specimens	Sequencing approach	Output
Staats *et al*. (2013)	*Flies and beetles*	**Number**: 3 specimens; **Age**: 1992–1995; **Tissue**: 1–3 legs, thorax, whole specimen (destructive protocol)	**Shotgun** whole genome skimming; Illumina HiSeq 2000	**Read depth**: 3.5x – 146.1x (mt genome); **% Mapping**: 0.002 – 0.82 (mt genome); **Contamination**: 1 specimen extensive bacteriophage & fungal DNA
Tin *et al*. (2014)	*Flies and ants*	**Number**: 11 specimens; **Age**: 1910–1976; **Tissue**: whole specimen (non-destructive protocol)	**Shotgun** whole genome skimming; **RAD-tag**; Illumina MySeq & HiSeq 2500	**Read depth**: 0.08x – 1.0x (whole genome); **% Mapping**: 19 – 76 (whole genome); **Contamination**: not reported
Heintzman *et al*. (2014)	*Beetles*	**Number**: 4 specimens; **Age**: Late Pleistocene (C^{14}), 1875–1950 (museum); **Tissue**: 1 hind leg, pronotum, elytron (destructive protocol)	**Shotgun** whole genome skimming; Illumina HiSeq 2000	**Reads aligned** to reference: 0.009% – 0.225x (mt genome & 5 nuclear loci); **% Insect contigs**: 0.25 – 46.5; **Contamination**: Up to ca. 20% mammalian sequence in contigs
Maddison & Cooper (2014)	*Beetles*	**Number**: 1 specimen; **Age**: 1968; **Tissue**: whole specimen (non-destructive protocol)	**Shotgun** whole genome skimming; Illumina HiSeq 2000	**Read depth**: not reported (8 gene targets); **% Gene length** coverage: 95 – 100 (8 gene targets); **Contamination**: not reported
Kanda *et al*. (2015)	*Beetles*	**Number**: 13 specimens; **Age**: 1929–2010; **Tissue**: whole specimen (non-destructive protocol)	**Shotgun** whole genome skimming; Illumina HiSeq 2000 (2 lanes)	**Read depth**: 0.44x – 4.64x (67 gene targets); **N50**: 280 – 700 (67 gene targets); **Contamination**: Possible in some specimens but not quantified
Timmermans *et al*. (2016)	*Butterflies*	**Number**: 35 specimens; **Age**: 1980–2005; **Tissue**: 1 leg (destructive protocol)	**Shotgun** whole genome skimming; Illumina MySeq (1/3 flow cell)	**% Coverage**: 0 – 100 (mt coding loci); **Contamination**: not reported; **Failure rate**: 4 out of 35 specimens any reads matching mt genomes
Suchan *et al*. (2016)	*Butterflies and grasshoppers*	**Number**: 60 specimens; **Age**: 1908–1997; **Tissue**: legs (destructive protocol)	**Target capture** of RAD probes; Illumina MySeq & HiSeq (one lane each)	**Median depth**: 10x (for each SNP); **% Matrix fullness**: 52 – 72.5 (RAD loci); **Contamination**: ca. 9 % of contigs were of exogenous origin
Blaimer *et al*. (2016)	*Carpenter bees*	**Number**: 51 specimens; **Age**: 1894–2013; **Tissue**: 1 leg (destructive protocol)	**Target capture** of Hymenopteran UCEs; Illumina MySeq	**Average coverage**: 7.4x – 182.4x (UCE loci); **Recovered loci**: 6 – 972 (UCE per sample); **Contamination**: not reported
Pitteloud *et al*. (2017)	*Butterflies*	**Number**: 32 specimens; **Age**: 1929–2012; **Tissue**: legs (destructive protocol)	**PCR Multiplex & Shotgun sequencing**; Illumina MySeq	**Length sequences** (bp): 109 – 7297 (mt and rDNA loci); **Contamination**: not reported

Approach	Case reference	Topic	Taxon group	Impact
Whole-transcriptome shotgun	Misof *et al* . (2014)	*Phylogenomics*	*Class Insecta*	First phylogenomic study to cover all hexapod orders
Whole-genome shotgun	Tang *et al* . (2014)	*Mito-metagenomics*	*Several insect orders*	Pioneering proof-of-concept study to show feasibility of PCR-free mitogenome sequence in bulk samples
RAD-seq	Tin *et al* . (2014)	*Phylogenetics; Museomics*	*Flies and ants*	One of the first insect museomic studies using massive parallel sequencing, and a guideline for non-destructive DNA isolation and library preparation
Target capture	Suchan *et al* . (2016)	*Phylogeography*	*Butterflies and grasshoppers*	New method to target RAD probes (hyRAD). Proof-of-concept using divergent taxa and archived specimens
Target capture	Faircloth *et al* . (2015)	*Phylogenomics*	*Hymenoptera*	Enrichment of Ultraconserved Elements (UCE) of the Hymenoptera order
Single-molecule	Kelley *et al* . (2014)	*Comparative Genomics*	*Antarctic midge*	Single-molecule real time whole-genome sequencing using PacBio



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26661v1 | CC BY 4.0 Open Access | rec: 20 Mar 2018, publ: 20 Mar 2018

1. Define research

2. Collaborate

3. Go to field

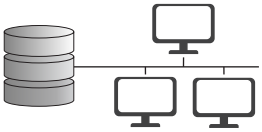
4. Minimize DNA damage

5. Work with taxonomists

6. Isolate DNA

7. Revise sequencing approach

8. Grasp bioinformatics



9. Results: Databases

10. Results: Dissemination