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

View the peer-reviewed version (peerj.com/articles/6727), which is the preferred citable publication unless you specifically need to cite this preprint.

https://doi.org/10.7717/peerj.6727
Biodiversity seen through the perspective of insects: 10 simple rules on methodological choices, common challenges, and experimental design for genomic studies

Pável Matos-Maraví Corresp., 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

1 Department of Biological and Environmental Sciences, University of Gothenburg, Gothenburg, Sweden
2 Gothenburg Global Biodiversity Centre, 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, Copenhagen, Denmark
5 Department of Biology, Lund University, Lund, Sweden
6 Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Stockholm, Sweden
7 Department of Zoology, Stockholm University, Stockholm, Sweden
8 Biodiversity Unit, University of Turku, Turku, Finland
9 Gothenburg Botanical Garden, Gothenburg, Sweden
10 Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, United States

Corresponding Author: Pável Matos-Maraví
Email address: pavelm14@gmail.com

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

Survey Methodology

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

Ten Simple Steps to Study Biodiversity through Insect Genomics

We structure this article in 10 simple rules (Fig. 2) that every biologist should understand to 1)
better interpret the results and conclusions coming from insect biodiversity research, and 2) start
planning a multi-dimensional study of biodiversity using insects as target group and high-
throughput genomic tools. Overall, we briefly review the current state in biodiversity and
evolution research through the study of insect genomics, by revisiting pioneering studies that aim
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. For example, in a recent tropical large-scale species inventory, Borkent & Brown (2015) investigated local species richness of cloud forest Diptera (true flies) by using only two Malaise traps and a one-week intensive “Diptera-Blitz” conducted by several experts. In other cases, such as 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, 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.,


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

Single-molecule sequencing approaches such as PacBio and Oxford Nanopore. The two main advantages are related to a better assembly of genomes with low quantities of poor quality DNA and the portability of some devices (e.g., MinION), which can generate DNA sequences in real-time and in any place in the world, including remote field locations. Moreover, laboratory protocols are simplified and DNA amplification is not at all necessary, which is beneficial for a more accurate quantification of DNA molecules present in the sample pool (Thompson & Milos, 2011). However, technological improvement is needed to reduce the high sequencing error rates (>10 %) (Mardis, 2017; Shendure et al., 2017). Nonetheless, single-molecule sequencing promises to drastically reduce sequencing costs, thus the age where having complete genome sequences for any living insect may be even closer (Kelley et al., 2014). Moreover, the long reads that single-molecule sequencing approaches generate might help resolve long repeat
elements in the genome, thus providing invaluable scaffold for short reads to improve accuracy in assembly and annotation of insect genomes (Richards & Murali, 2015).

Rule 8: Choose the most suitable tools for data analyses

Although genomic sequencing is becoming easier and more affordable, processing the data generated remains a major bottleneck in many projects. Bioinformatic pipelines have been implemented during the past two decades of massive parallel sequencing, thus researchers nowadays count with standard procedures to analyze genomic DNA. However, there exist a number of limitations and challenges that remain to be explored. For example, genomic missing data, as in supermatrices for phylogenomic studies, might hinder statistical power in the inference of species relations, but its effects in systematic bias is yet unclear (Misof et al., 2014a,b). Moreover, taxonomic sampling in phylogenomics is usually lower than in the published Sanger-sequencing work, a fact that might bias systematic inference in insect higher-level phylogenies (Behura, 2015). On the other hand, a number of pipelines have been published for analyzing target-sequencing data from environmental samples (Schloss et al., 2009; Caporaso et al., 2010; Boyer et al., 2016). Such programs provide a delimitation of Operational Taxonomic Units (OTUs), the analogs of species, derived from sequence similarity of typically 97 %. However, assigning thresholds to define analogs of species is problematic because 1) there is a risk to artificially increase or decrease local diversity, 2) inflated OTU richness might be related to sequence chimeras and sequencing errors, and 3) there exist a lack of standardization of threshold values (Huse et al., 2010; Oliver et al., 2015; Alberdi et al., 2018). The shortcomings of using thresholds to define OTUs might even escalate when studying the entomofauna of
hyperdiverse regions such as the tropics. In those cases, there are usually no good estimates of genetic variability between species and a large portion of tropical insects are not represented in reference databases. However, recent advances might alleviate in part the variability caused by sequencing errors by incorporating other observed patterns such as OTUs co-occurrence (Frøslev et al., 2017). In any case, the preservation and morphological study of vouchers are critical to validate taxonomic assignments and thresholds.

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

Rule 9: Make your data and results publicly available
From a practical viewpoint, what is not in a database does not exist (or nearly so). Databases are not only the repositories of genomic information, but also an indispensable tool in the study of biodiversity and evolution. They also allow the reproduction of results and use for other purposes such as in biodiversity assessment. Furthermore, the study of populations, their mechanisms of adaptation and speciation, is an active field given the opportunities that represent the hundreds of insect genome projects published and registered in GenBank (Yeates et al., 2016). Moreover, in the study of species interactions, such as host-parasite and feeding habits, the recovery of cryptic diversity necessitates a reference database because in many cases the identification of taxa through morphological comparison becomes impossible; for example, internal parasites (Schoonvaere et al., 2016), gut microbiota (Hammer et al., 2017), and highly-degraded organic material such as dietary content in gut (Pompanon et al., 2012).

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


Behura SK. 2015. Insect phylogenomics. Insect Molecular Biology 24:403–411. DOI:


Crampton-Platt A., Timmermans MJTN., Gimmel ML., Kutty SN., Cockerill TD., Khen CV.,


Faircloth BC. 2016. Identifying conserved genomic elements and designing universal probe sets


Gilbert MTP., Sanchez JJ., Haselkorn T., Jewell LD., Lucas SB., Van Marck E., Børsting C., Morling N., Worobey M. 2007b. Multiplex PCR with minisequencing as an effective high-


Maddison DR., Cooper KW. 2014. Species delimitation in the ground beetle subgenus *Liocosmius* (Coleoptera: Carabidae: *Bembidion*), including standard and next-generation


Calcott B., Chen J., Friedrich F., Fukui M., Fujita M., Greve C., Grobe P., Gu S., Huang Y.,
Jermiin LS., Kawahara AY., Krogmann L., Kubiak M., Lanfear R., Letsch H., Li Y., Li Z.,
Li J., Lu H., Machida R., Mashimo Y., Kapli P., McKenna DD., Meng G., Nakagaki Y.,
Navarrete-Heredia JL., Ott M., Ou Y., Pass G., Podsiaadlowski L., Pohl H., von Reumont
BM., Schütte K., Sekiya K., Shimizu S., Slipinski A., Stamatakis A., Song W., Su X.,
M., Tong X., Uchifume T., Walzl MG., Wiegmann BM., Wilbrandt J., Wipfler B., Wong
TKF., Wu Q., Wu G., Xie Y., Yang S., Yang Q., Yeates DK., Yoshizawa K., Zhang Q.,
resolves the timing and pattern of insect evolution. Science 346:763–767. DOI:
10.1126/science.1257570.

Misof B., Meusemann K., von Reumont BM., Kück P., Prohaska SJ., Stadler PF. 2014b. A priori
assessment of data quality in molecular phylogenetics. Algorithms for Molecular Biology

Mora C., Tittensor DP., Adl S., Simpson AGB., Worm B. 2011. How many species are there on

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

Nieman CC., Yamasaki Y., Collier TC., Lee Y. 2015. A DNA extraction protocol for improved


Schloss PD., Westcott SL., Ryabin T., Hall JR., Hartmann M., Hollister EB., Lesniewski RA.,
Oakley BB., Parks DH., Robinson CJ., Sahl JW., Stres B., Thallinger GG., Van Horn DJ.,
Weber CF. 2009. Introducing mothur: open-source, platform-independent, community-
supported software for describing and comparing microbial communities. *Applied and

Unbiased RNA shotgun metagenomics in social and solitary wild bees detects associations
with eukaryote parasites and new viruses. *PLoS ONE* 11:e0168456. DOI:
10.1371/journal.pone.0168456.

Shendure J., Balasubramanian S., Church GM., Gilbert W., Rogers J., Schloss JA., Waterston
10.1038/nature24286.

Comparative genomic analysis of the endosymbionts of herbivorous insects reveals eco-
environmental adaptations: biotechnology applications. *PLoS Genetics* 9:e1003131. DOI:
10.1371/journal.pgen.1003131.

Environmental DNA barcode sequence capture: targeted, PCR-free sequence capture for
biodiversity analysis from bulk environmental samples. *bioRxiv*:87437. DOI:
10.1101/087437.


Suchan T., Pitteloud C., Gerasimova NS., Kostikova A., Schmid S., Arrigo N., Pajkovic M.,


**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**

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

Conflict of Interests

The authors declare no conflict of interests.
Tables

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

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

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

Figures

Figure 1: Schematic of the spatiotemporal continuum of biodiversity and evolution. Community ecology research usually focus at local scale (Lx), but may be mislead if not taking into account higher-level patterns such as fluctuation of population/species ranges over time (e.g., distributional change due to environmental shift or biotic interaction) and space (e.g., local pool of species depends on the regional diversity and its evolution). At a higher scale (Vx), inferences on population/species evolutionary dynamics may be mislead if not taking into account reduced-scale patterns such as local extinction (e.g., its rate and magnitude, whether local, regional or global) and biotic interaction (e.g., as in diversity-dependent diversification, or 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.
Table 1 (on next page)

Representative description of methods for mass sampling of insects and their application for NGS

Note that this is not a comprehensive list and is only aimed at providing an overview of available possibilities of widespread use. In the Costs of equipments, we roughly categorized them as Low (approx. < US $50), Medium (approx. US $50 – $100), High (approx. > US $100).
<table>
<thead>
<tr>
<th>Method</th>
<th>Example</th>
<th>Taxa targeted</th>
<th>Costs</th>
<th>Suitability for genomic research</th>
<th>Sampling effort</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trap-sampling</td>
<td>Van Someren-Rydon</td>
<td>Fruit-feeding butterflies, from forest floor to canopy</td>
<td>Low; negligible if self-built</td>
<td>Yes; no killing reagent; baits such as fermented fruit</td>
<td>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</td>
<td>Need for long-term data because different butterfly communities throughout the year; Other feeding guilds are missing, such as nectar-feeding</td>
</tr>
<tr>
<td>Trap-sampling</td>
<td>Pitfall</td>
<td>Forest floor insects such as dung beetles, flies, ants</td>
<td>Low; negligible if self-built</td>
<td>Depending on killing reagent; could be done with detergent and water, propylen glycol; baits such as human dung</td>
<td>Minimum: 20 traps per day; linear transect; Collection: at least once per day; Personnel: 1 person; Complement: with flight intercept traps</td>
<td>Specimens with a lot of water from pitfall trap, thus a lot of ethanol replaced every week is needed to prevent DNA decay</td>
</tr>
<tr>
<td>Leaf-litter collector</td>
<td>Mini-Winkler</td>
<td>Leaf-litter and soil insects, such as ants, beetles</td>
<td>Medium</td>
<td>Yes; 95% EtOH most commonly used as killing reagent</td>
<td>Minimum: 20 collectors, each with 1m² leaf litter; Collection: once, if extraction run in parallel; Personnel: 2 people recommended; Complement: with bait-traps and hand collecting</td>
<td>Limited to forested areas, and not suitable during height of dry or rain season; Not sampling of vegetation-associated, canopy or subterranean insects</td>
</tr>
<tr>
<td>Flying-insect collector</td>
<td>Malaise</td>
<td>Strong-flying insects, such as Hymenoptera and Diptera</td>
<td>High</td>
<td>Yes; 95% EtOH most commonly used as killing reagent</td>
<td>Minimum: 2 traps for fast surveys; Collection: little care, leave in field for 2–4 weeks; Personnel: 1 person; Complement: with flight interception traps</td>
<td>Placement of trap in &quot;likely&quot; flight paths, thus a component of subjectivity is introduced</td>
</tr>
<tr>
<td>Flying-insect collector</td>
<td>Flight interception</td>
<td>Flying insects, such as beetles, cockroaches, crickets</td>
<td>Low; negligible if self-built</td>
<td>Depending on killing reagent; could be done in salt-saturated water and detergent, propylen glycol; formaldehyde solutions but in detriment of DNA recovery</td>
<td>Minimum: 2 traps for fast surveys; Collection: once or twice per day; Personnel: 1 person; Complement: with bait and light traps</td>
<td>Ideal for slow-flying insects, which hit the plastic sheet and fall in the container with killing reagent</td>
</tr>
<tr>
<td>Insecticidal knockdown</td>
<td>Canopy fogging</td>
<td>Arboreal insect community</td>
<td>High</td>
<td>Yes; insecticide as killing reagent</td>
<td>Collection: laborious and problems with pseudoreplication; Complement: with canopy light trapping and flight interception traps</td>
<td>Canopy access still limited; High demand on logistics; Environmentally dependent</td>
</tr>
</tbody>
</table>
Table 2 (on next page)

Overview of massively parallel DNA sequencing methods applied to insect museum specimens
<table>
<thead>
<tr>
<th>Authors</th>
<th>Taxon group</th>
<th>Specimens</th>
<th>Sequencing approach</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Staats <em>et al.</em> (2013)</td>
<td><em>Flies and beetles</em></td>
<td>Number: 3 specimens;</td>
<td>Shotgun whole genome skimming;</td>
<td>Read depth: 3.5x – 146.1x (mt genome);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Age: 1992–1995;</td>
<td>Illumina HiSeq 2000</td>
<td>% Mapping: 0.002 – 0.82 (mt genome);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tissue: 1–3 legs, thorax, whole specimen</td>
<td></td>
<td>Contamination: 1 specimen extensive bacteriophage &amp; fungal DNA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(destructive protocol)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tin <em>et al.</em> (2014)</td>
<td><em>Flies and ants</em></td>
<td>Number: 11 specimens;</td>
<td>Shotgun whole genome skimming;</td>
<td>Read depth: 0.08x – 1.0x (whole genome);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Age: 1910–1976;</td>
<td>RAD-tag;</td>
<td>% Mapping: 19 – 76 (whole genome);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tissue: whole specimen (non-destructive</td>
<td>Illumina MySeq &amp; HiSeq 2500</td>
<td>Contamination: not reported</td>
</tr>
<tr>
<td></td>
<td></td>
<td>protocol)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heintzman <em>et al.</em> (2014)</td>
<td><em>Beetles</em></td>
<td>Number: 4 specimens;</td>
<td>Shotgun whole genome skimming;</td>
<td>Read depth: not reported</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Age: Late Pleistocene (C14), 1875–1950</td>
<td></td>
<td>8 gene targets);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(museum);</td>
<td></td>
<td>% Insect contigs: 0.25 – 46.5;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tissue: 1 hind leg, pronotum, elytron</td>
<td></td>
<td>Contamination: Up to ca. 20% mammalian sequence in contigs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(destructive protocol)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maddison &amp; Cooper (2014)</td>
<td><em>Beetles</em></td>
<td>Number: 1 specimen;</td>
<td>Shotgun whole genome skimming;</td>
<td>Read depth: 0.44x – 4.64x (67 gene targets);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Age: 1968;</td>
<td></td>
<td>N50: 280 – 700 (67 gene targets);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tissue: whole specimen (non-destructive</td>
<td></td>
<td>Contamination: Possible in some specimens but not quantified</td>
</tr>
<tr>
<td></td>
<td></td>
<td>protocol)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kanda <em>et al.</em> (2015)</td>
<td><em>Beetles</em></td>
<td>Number: 13 specimens;</td>
<td>Shotgun whole genome skimming;</td>
<td>Read depth: 0.009% – 0.225x (mt genome &amp; 5 nuclear loci);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Age: 1929–2010;</td>
<td></td>
<td>% Insect contigs: 0.25 – 46.5;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tissue: whole specimen (non-destructive</td>
<td></td>
<td>Contamination: not reported</td>
</tr>
<tr>
<td></td>
<td></td>
<td>protocol)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Timmermans <em>et al.</em> (2016)</td>
<td><em>Butterflies</em></td>
<td>Number: 35 specimens;</td>
<td>Shotgun whole genome skimming;</td>
<td>Read depth: not reported</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Age: 1980–2005;</td>
<td></td>
<td>8 gene targets);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tissue: 1 leg (destructive protocol)</td>
<td></td>
<td>% Gene length coverage: 95 – 100 (8 gene targets);</td>
</tr>
<tr>
<td>Suchan <em>et al.</em> (2016)</td>
<td><em>Butterflies</em></td>
<td>Number: 60 specimens;</td>
<td>Target capture of RAD probes;</td>
<td>Contamination: not reported</td>
</tr>
<tr>
<td></td>
<td>and grasshoppers</td>
<td>Age: 1908–1997;</td>
<td>Illumina MySeq &amp; HiSeq (one lane each)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tissue: legs (destructive protocol)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blaimer <em>et al.</em> (2016)</td>
<td><em>Carpenter bees</em></td>
<td>Number: 51 specimens;</td>
<td>Target capture of Hymenopteran UCEs;</td>
<td>Average coverage: 7.4x – 182.4x (UCE loci);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Age: 1894–2013;</td>
<td>Illumina MySeq</td>
<td>Recovered loci: 6 – 972 (UCE per sample);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tissue: 1 leg (destructive protocol)</td>
<td></td>
<td>Contamination: not reported</td>
</tr>
<tr>
<td>Pitteloud <em>et al.</em> (2017)</td>
<td><em>Butterflies</em></td>
<td>Number: 32 specimens;</td>
<td>PCR Multiplex &amp; Shotgun sequencing;</td>
<td>Length sequences (bp): 109 – 7297 (mt and tDNA loci);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Age: 1929–2012;</td>
<td>Illumina MySeq</td>
<td>Contamination: not reported</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tissue: legs (destructive protocol)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3 (on next page)

Non-comprehensive overview of massively parallel DNA sequencing methods applied to insects
<table>
<thead>
<tr>
<th>Approach</th>
<th>Case reference</th>
<th>Topic</th>
<th>Taxon group</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole-transcriptome shotgun</td>
<td>Misof et al. (2014)</td>
<td><strong>Phylogenomics</strong></td>
<td><strong>Class Insecta</strong></td>
<td>First phylogenomic study to cover all hexapod orders</td>
</tr>
<tr>
<td>Whole-genome shotgun</td>
<td>Tang et al. (2014)</td>
<td><strong>Mito-metagenomics</strong></td>
<td><strong>Several insect orders</strong></td>
<td>Pioneering proof-of-concept study to show feasibility of PCR-free mitogenome sequence in bulk samples</td>
</tr>
<tr>
<td>RAD-seq</td>
<td>Tin et al. (2014)</td>
<td><strong>Phylogenetics; Museomics</strong></td>
<td><strong>Flies and ants</strong></td>
<td>One of the first insect museomic studies using massive parallel sequencing, and a guideline for non-destructive DNA isolation and library preparation</td>
</tr>
<tr>
<td>Target capture</td>
<td>Suchan et al. (2016)</td>
<td><strong>Phylogeography</strong></td>
<td><strong>Butterflies and grasshoppers</strong></td>
<td>New method to target RAD probes (hyRAD). Proof-of-concept using divergent taxa and archived specimens</td>
</tr>
<tr>
<td>Target capture</td>
<td>Faircloth et al. (2015)</td>
<td><strong>Phylogenomics</strong></td>
<td><strong>Hymenoptera</strong></td>
<td>Enrichment of Ultraconserved Elements (UCE) of the Hymenoptera order</td>
</tr>
<tr>
<td>Single-molecule</td>
<td>Kelley et al. (2014)</td>
<td><strong>Comparative Genomics</strong></td>
<td><strong>Antarctic midge</strong></td>
<td>Single-molecule real time whole-genome sequencing using PacBio</td>
</tr>
</tbody>
</table>
Community ecology research usually focus at local scale (Lx), but may be mislead if not taking into account higher-level patterns such as fluctuation of population/species ranges over time (e.g., distributional change due to environmental shift or biotic interaction) and space (e.g., local pool of species depends on the regional diversity and its evolution). At a higher scale (Vx), inferences on population/species evolutionary dynamics may be mislead if not taking into account reduced-scale patterns such as local extinction (e.g., its rate and magnitude, whether local, regional or global) and biotic interaction (e.g., as in diversity-dependent diversification, or 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).
Flowchart summarizing the 10 rules to study biodiversity through insect genomics and discussed in the paper
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