

Suggestions for a molecular biodiversity assessment of South East Asian freshwater invertebrates. Lessons from the megadiverse beetles (Coleoptera)

Michael BALKE,^{1,2*} Lars HENDRICH,¹ Emmanuel F.A. TOUSSAINT,^{1,2} Xin ZHOU,³ Thomas von RINTELEN,⁴ Mark DE BRUYN⁵

¹Zoological State Collection, Münchhausenstrasse 21, 81247 Munich; ²GeoBioCenter, Ludwig-Maximilians-University, Geschwister-Scholl-Platz 1, Munich, Germany; ³China National GeneBank, BGI-Shenzhen, 518083 Shenzhen, China; ⁴Museum für Naturkunde, Leibniz-Institut für Evolutions und Biodiversitätsforschung an der Humboldt-Universität zu Berlin, Invalidenstrasse 43, 10115 Berlin, Germany; ⁵Molecular Ecology and Fisheries Genetics Laboratory, School of Biological Sciences, Environment Centre for Wales, University of Bangor, Deiniol Road, LL57 2UW Bangor, UK

*Corresponding author: Coleoptera-ZSM@zsm.mwn.de

ABSTRACT

The beginning of the 21st century was characterised by an unprecedented human-mediated loss of biodiversity, with an astonishing number of undescribed species disappearing from the Earth. To counter this major erosion of biodiversity, we need to describe and monitor what we want to preserve. Unfortunately, the velocity of deforestation, ecosystem degradation and the escalating threat to the last wilderness areas on the planet overwhelm traditional taxonomists in their bid to describe all of Earth's biodiversity. Based on empirical studies on weevils and diving beetles (Coleoptera), we show that biodiversity assessments based on *cox1* DNA sequence data deliver comparably accurate estimates of species diversity, even using a simple clustering method with a preset threshold. The method works best for large datasets, where lineage idiosyncratic errors such as species lumping or splitting compensate each other. *Cox1* clusters cannot be translated into formal species per se, but can help taxonomists to accelerate their work. We suggest that large-scale sequencing campaigns for the Asian freshwater fauna will reveal patterns relevant for conservation priority setting, and enhance our understanding of macroevolutionary processes that have shaped current biodiversity in the region. Along with next generation sequencing approaches, we also suggest that our understanding of alpha taxonomy will benefit.

Key words: South East Asia, macrozoobenthos, monitoring and inventories, DNA barcoding, next generation sequencing.

Received: January 2013. Accepted: June 2013.

INTRODUCTION

This paper was presented as a keynote at the *Freshwater Invertebrates of Southeast Asia: Biodiversity and Origin* (FISA) workshop (Maha Sarakham, Thailand, 2012) by M. Balke and has subsequently been modified to address issues raised by different speakers at the workshop.

With the dramatic increase of biodiversity erosion during the past decades, one of the greatest challenges in biodiversity research is biodiversity. How many species are there? How to best identify already described species? These are questions immediately relevant to paramount topics such as conservation of biodiversity, pest control and ultimately human welfare (Dincă *et al.*, 2011). However, with millions of undescribed species on Earth (Mora *et al.*, 2011), and most of the *ca.* 1,500,000 already described species (Wilson, 1988) rather difficult to identify, there is a need for reliable and fast tools to measure and describe this biodiversity (House of Lords, 2008). The common currency for biologists is the species entity, and for easiest communication and comparison of data, species require a formally assigned name based on one physically available voucher specimen, the holotype. In a series of publications, Hebert

et al. (2003a, 2003b, 2004a, 2004b) advocated the use of a DNA barcode, a term coined to describe the use of a standard DNA fragment to build a truly global species database linking DNA sequence data with actual species names and thus morphology, ethology and *e.g.* species ecology. They suggested using a 658 bp fragment of the mitochondrial gene encoding for cytochrome *c* oxidase I (COI) (*cox1*). This marker can usually be amplified using universal primers and is informative at the population and species level (Hebert *et al.*, 2003a, 2003b, 2004a, 2004b), but this can also often be, but not always, for group species according to their higher classification (Hendrich *et al.*, 2010; Rach *et al.*, 2008). In the past decade, and in the wake of the Consortium for the Barcode of Life (CBOL) (<http://barcoding.si.edu>), more than 2 million specimens representing about 180,000 species have been sequenced and added to a growing central database [*Barcode Of Life Data systems* (BOLD)] (<http://www.boldsystems.org>). With this online library of DNA data, it is possible to use an identification interface to paste a query sequence and explore if there are close matches (Ratnasingham and Hebert, 2007). Species in this database typically have their own page, with se-

quence data and most notably voucher information, distribution maps and voucher photographs. While this system may not be perfect, it does illustrate very well how taxonomic data can be implemented and managed.

Many studies have shown that this DNA barcoding fragment can successfully be used to identify species (once a reference database has been created) (Janzen *et al.*, 2009; Dincă *et al.*, 2011), and potentially detect overlooked species even in well-studied faunas (Herbert 2004a, 2004b; Hausmann *et al.*, 2011). It was suggested that the genetic divergence among species is usually larger than within species, which should allow for species identification/delineation using simple preset divergence thresholds (Hebert *et al.*, 2003b), or using more elaborate methods employing statistical techniques to describe population genetic and speciation processes (Monaghan *et al.*, 2005, 2009; Pons *et al.*, 2006). However, several studies urge for caution (Meier *et al.*, 2006), as identification success might be low in certain cases, especially when moving from local to regional sampling levels where increasing numbers of closely related species are included (Bergsten *et al.*, 2012b; Hendrich *et al.*, 2010), or where interspecific DNA divergence levels between some species of one genus can be less than within other species of the same genus (Hendrich *et al.*, 2009). In young species, the genetic signal of *cox1* can be unstructured between species (Hendrich *et al.*, 2010), only allowing for delineation of a species-complex rather than individual species.

Taking these potential pitfalls into account (Moritz and Cicero, 2004; Will *et al.*, 2005), however, DNA sequence data can be a useful addition to the biologist's toolkit (Hebert, 2004a; Goldstein and DeSalle, 2011). An example of such an approach is the Barcoding Fauna Bavarica (Germany) project, which assembled data for more than 10,000 animal species, *all identified by experts*, and all vouchered (www.faanabavarica.de). This was possible because existing taxonomic knowledge was good, and experts for a range of taxa were relatively abundant. Even among the supposedly best-characterised species, overlooked sibling species were identified [*e.g.* documented by Bergsten *et al.* (2012a)], and taxonomic revision for the whole dataset has recently begun.

In most tropical regions, however, the situation is more challenging, mainly due to a lack of taxonomic expertise and a very high proportion of undescribed species (May, 2010; Riedel *et al.*, 2013a). This also became evident during the FISA workshop, for a range of taxa from rotifers to small Crustacea and aquatic insects. There is no doubt that a solid taxonomic framework is the most desirable foundation for any kind of biodiversity assessment. Proposed tools to accelerate taxonomy are manifold (Mallet and Willmott, 2003; Dayrat, 2005; Padial *et al.*, 2010; Riedel *et al.*, 2013a), and a review is not within the scope of this paper. Here, we rather focus on an alternative

approach, and suggest the use of large scale DNA sequencing and objective clustering of data (Meier *et al.*, 2006) to provide first insights into large scale patterns of aquatic invertebrate diversity across South East Asia. Specifically, we suggest using standardised protocols based on the DNA barcoding method outlined above (Hebert *et al.*, 2003a), to create an objective, scientific basis to better understand freshwater diversity in Asia, to provide publicly accessible data for conservation and the setting of research priorities, and, for example, biogeographic analysis and community ecology. This is not to rival established taxonomic expertise and formal species description, but rather a supporting measure to these approaches (Riedel *et al.*, 2013b). We refer to our campaign as *molecular biodiversity assessment* (MBA), and this can be further developed into metabarcoding (see below).

METHODS

The DNA laboratory and analytical procedures required to obtain barcodes have been described in detail in previous open access publications (Hendrich *et al.*, 2010; Hawlitschek *et al.*, 2012; Tänzler *et al.*, 2012). Detailed information can also be found in our laboratory wiki at: http://zsm-entomology.de/wiki/The_Beetle_D_N_A_Lab. While museum specimens may deliver DNA adequate for barcoding purposes (Hebert *et al.*, 2013), the use of freshly collected specimens and/or specimens stored in 96% ethanol at -80°C is strongly advocated. Whenever possible, non-destructive DNA extractions (*e.g.* using the Qiagen™ column kit; Qiagen, Venlo, The Netherlands) should be performed using whole specimens for small taxa, and leg, thoracic, abdominal or antennae tissues for larger taxa. The preservation of specimen integrity is of great importance, since they will be retained in a public natural history collection as a physical voucher (Riedel *et al.*, 2013a). All vouchers are then digitised and the pictures uploaded to the database for further ease of identification and referencing. The DNA fragment we use is either the so-called barcoding fragment, the 5' end of cytochrome COI (Hebert *et al.*, 2003a) or the 3' end of *cox1* (for diving beetles, because we traditionally focused on that fragment; Hendrich *et al.*, 2010). Once a specimen has been processed through this pipeline, the corresponding sequence, picture and exhaustive information (classification, collector, locality) should be uploaded to complete the global database.

For our preliminary studies (Hendrich *et al.*, 2010; Tänzler *et al.*, 2012), we used the SpeciesIdentifier module of TaxonDNA software (Meier *et al.*, 2006). This module allows clustering sequence data at different thresholds, *e.g.* ranging 1-10% of sequence divergence using uncorrected *p*-distances (Fig. 1). SpeciesIdentifier accounts for threshold violations according to the *triangle inequity* (*i.e.* when the divergence between A-B and B-C is 3% or less, but A-C exceeds 3%, then A, B and C would still be

grouped into one 3% cluster by Taxon DNA) (Fig. 1). In our case, a maximum threshold of 10% was likely to capture any kind of lineage idiosyncrasy. It is very important to recall that such distances between species should not be seen as fixed values, as they depend on the lineage or even the area under consideration. Although a generalised cut-off level for the SE Asian freshwater beetles here is suggested to be *ca.* >3% (but see below), one should take into account that biodiversity assessments using molecular screening should always be interpreted within context. Hebert *et al.* (2003b), for example, show that in Cnidaria, *cox1* sequences between species tend to be extremely low (<1%), while in certain Crustacea, they can be very high (>20%), as we will discuss below for *Trigonopterus* weevils from New Guinea. This illustrates that the context should indeed be taken into account, but also that a combination of more than one technique (*e.g.* morphology+genetics) is highly important especially to test molecular methods for potential focal taxa and of course for proper taxonomic work.

SpeciesIdentifier recognises *a priori* delineated species from sequence names as long as the name follows the format *genus species*, *i.e.* *Rhantus suturalis*, or *Exocelina australiaone* MB1307. The output summarises the number of different species names in the dataset, the number of clusters found under the preset threshold (*e.g.* 1, 2, 3%), the number of clusters containing only one species name, and the number of *perfect clusters* (those that contain all individuals under one species name and only those individuals, *i.e.* monophyly). Therefore, we can calculate the number of split clusters (one species split into more than one cluster, *i.e.* paraphyly) and lumped clusters

(more than one species name in a cluster). SpeciesIdentifier was also used for species richness estimation, with clusters taken as species surrogates. For any clustering threshold (*e.g.* at 1, 2, 3%), two values were reported.

The first of these values was the number of clusters found relative to the number of morphology-based species names in the dataset (*agreement* hereafter). This is illustrated in Fig. 2, left part (same height for columns means a perfect match); Fig. 3 shows the clustering performance at different preset thresholds. For example, a dataset with a hundred species names and a threshold clustering at 25% divergence would likely reveal only one cluster. Thus, our species richness estimation would amount to a meager 1% (*agreement*) of the true species, as delineated by morphology. Second, and more importantly, we report *taxonomic accuracy*, which was calculated as the number of perfect clusters (*i.e.* clusters containing all sequences of a morphology based species and only those sequences) relative to the number of species in the dataset. The number of perfect clusters can increase when the existing taxonomy is revised to accommodate cryptic or overlooked species. A 100% *accuracy* means that all clusters perfectly mirror putative species based on morphology (Fig. 2, right part – above circle means full congruence, middle circles are values in between, and lower circles mean full incongruence; Fig. 3). Higher numbers indicate species number overestimation/oversplitting, lower numbers indicate species number underestimation/lumping.

It is important to use species names, either formally identified described species, or operational expert taxonomic units (*e.g.* *Exocelina australiaone* MB1307) to be able to evaluate the performance of the method.

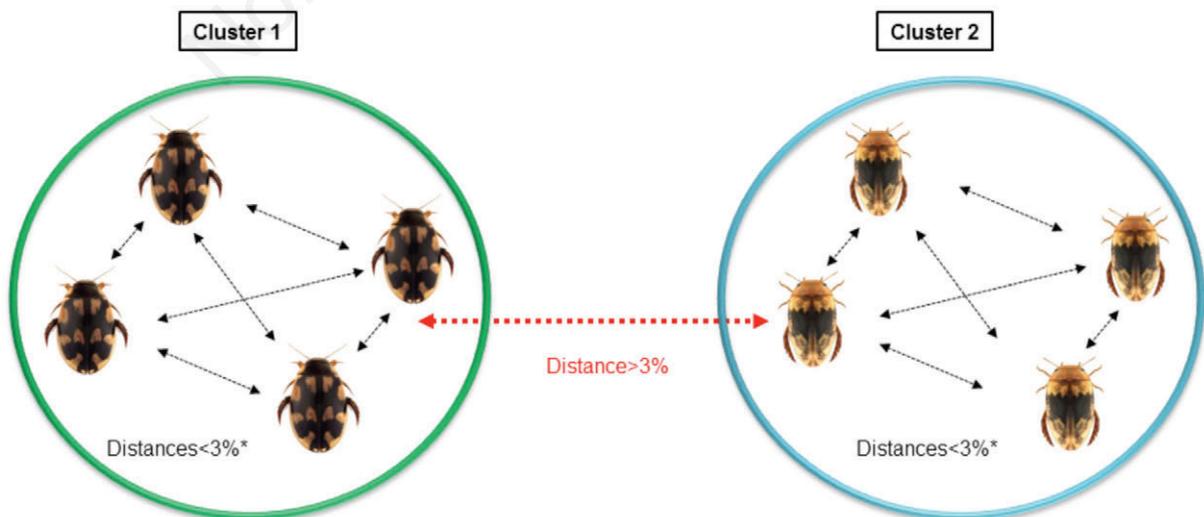


Fig. 1. Principle of clustering DNA sequence data. *The distance between specimens within a 3% cluster can exceed 3% because of the *triangle inequality*.

For the proposed molecular biodiversity assessment campaign, SpeciesIdentifier can and will have to use sequence names such as *Ephemeroptera sample MB7673* and so forth until further refinements are made based on taxonomists' feedback.

RESULTS AND DISCUSSION

Molecular biodiversity assessment

In preliminary studies we have asked: how well does the *number* of clusters agree with the *number* of species in the dataset (quantitative assessment) (Fig. 2)?

We also asked: how well do the *contents* of clusters agree with *contents* of species in the dataset? In other words: is there a method that detects species from our sequence matrix with 100% accuracy? (qualitative assessment) (Fig. 2).

For the entire Australian diving beetle (Coleoptera: Dytiscidae) fauna, with >270 sequenced species, we have empirically shown that the best results occur at cut-off thresholds between 2 and 3%. At 2-3%, we find a 100%

correct estimation of the species number in a sample (quantity), however the quality was lower than expected, only *ca.* 80% of the clusters could be translated directly into formal species (Fig. 3) (Hendrich *et al.*, 2010). In a dataset of 279 species of New Guinean *Trigonopterus* weevils (Curculionidae) (Tänzler *et al.*, 2012), the quality of the clustering was 86% at a 3% cut-off, but the species number was 16% higher than that based on morphologically described species. Because of their high interspecific divergences, performance for *Trigonopterus* was best at 8% cut-off. The surprising result was that for both groups, utility of a standard threshold at *e.g.* 3% delivers error margins for taxonomic accuracy of <30% (diving beetles) (mean error for diving beetles individual genera was 21%) and <14% (*Trigonopterus* weevils), which were lower than in many studies using morphospecies sorting, where error could be up to 80% (Krell, 2004).

Clustering of *cox1* sequences can therefore provide a first insight into local species diversity (but of course not replace proper taxonomic revision). However, it has also

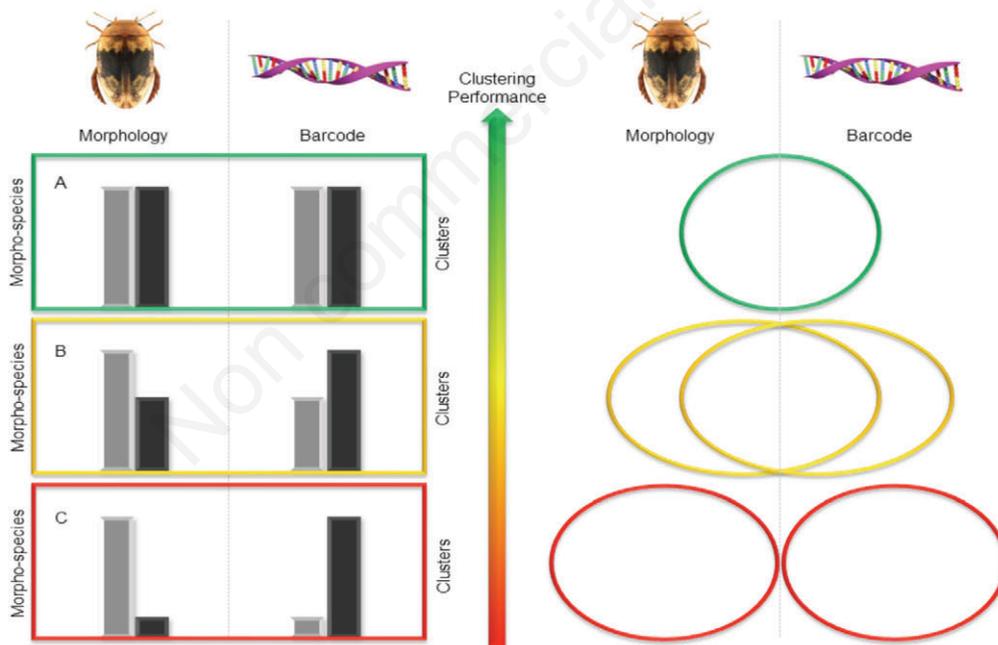


Fig. 2. Different cases of clustering performances and summary of the issue of lineage idiosyncrasy. A) The numbers of both morphospecies and genetic clusters are highly congruent. In such a case, the clustering performance is optimal, meaning that the agreement between morphologically identified species and genetic clusters based on the barcode fragment is excellent as represented by broadly overlapping circles. B) The number of morphospecies and genetic clusters are different, with two possibilities, *i.e.* the clustering based on barcodes recovers less putative species than the number of species based on morphology, or the opposite. In this case, morphology and barcoding are not in full agreement, and the clustering performance is then moderate. C) The number of species recovered by either morphology or barcoding is completely different, and the clustering performance is poor, as represented by non-overlapping circles between morphology-based and barcode-based identifications. It is noteworthy that the efficiency of clustering is highly dependent on lineage idiosyncrasy, and that despite this pitfall, the performance can be improved by tuning the threshold of clustering.

been shown that datasets should be large and ideally cover many lineages for such an MBA approach. Performance is highly lineage idiosyncratic, and errors for individual lineages can be significantly higher (Hendrich *et al.*, 2010). Studies on additional taxa are certainly needed. However, we feel confident that this approach can be widely applied to little-known communities to gain first insights into diversity patterns. Over time, and with improving taxonomic knowledge and databases, cluster-by-cluster sorting of lineages will be named, and their contents evaluated.

Molecular biodiversity assessment data for conservation, community ecology and biogeography

Haplotype clusters can be used to calculate similarity indices, such as the Sorenson index. Also, results for larger datasets were consistent between different analytical methods, *e.g.* traditional species delineation *vs* DNA entities, delineated by clustering or more sophisticated general mixed Yule coalescent (GMYC) analysis (Hendrich *et al.*, 2010; Tänzler *et al.*, 2012). Sequence data can moreover be used to calculate richness accumulation curves, using either haplotype or phylogenetic diversity. For regional samples of aquatic macroinvertebrates from Canada, a high correlation with morphologically delineated entities was found (Zhou *et al.*, 2009). Sequence data help to assign those lifestages to species names for which there is otherwise little hope to achieve reliable identification. This greatly aids ecological surveys and environmental impact assessments, especially when immature stages and adults occupy very different habitats, *e.g.* in the Trichoptera and Ephemeroptera, and it is crucial for conservation priority setting to understand where certain species actually breed (Ruiter *et al.*, 2013; Zhou *et al.*, 2007, 2010). Zhou *et al.* (2013) provide references on national biomonitoring programmes and possible sequencing-based approaches to aid these ecological surveys.

Limitations and pitfalls

DNA sequence data and MBA cannot automatically be translated into species entities, and does not provide a substitute for traditional taxonomy. Yet, the method can be a proxy for *species* sorting, and such *species* can be cross-checked with morphological or other evidence to accurately delineate species. This notwithstanding, problems associated with incompatible molecular *vs* morphological evidence are well established (Hendrich *et al.*, 2010; Hawlitschek *et al.*, 2012). Causes include, for example, very recent origins of species with associated incomplete lineage sorting and introgression (Monaghan *et al.*, 2006; Hawlitschek *et al.*, 2012), and this problem may occur particularly when a study covers species along with their closest relatives, *e.g.* densely sampled radiations or

densely sampled faunas (Hendrich *et al.*, 2010; Bergsten *et al.*, 2012a). There is also not necessarily a straightforward approach when nuclear DNA markers are included (Skale *et al.*, 2012). However, these are not major problems as long as a study is carefully designed, taking these pitfalls into consideration. In doing so, a lot can be gained: MBA does not substitute carefully conducted taxonomic studies, it is really only a very first step. Nevertheless, MBA can be very useful as a framework for establishing further, integrative studies (Bergsten *et al.*, 2012a, 2012b).

A vision for a molecular biodiversity assessment campaign of Asian macrozoobenthos

Asian wetlands and streams are incredibly diverse biologically, yet we are only now beginning to comprehend the true magnitude of their species diversity, and the patterns of this biodiversity between sites. Concurrently, Asian wetlands, in particular, are under enormous anthropogenic pressure, and conservation measures to maintain their ecosystem services are of utmost importance.

The setting of conservation priorities is hampered by a lack of taxonomic knowledge, and the inability to place re-

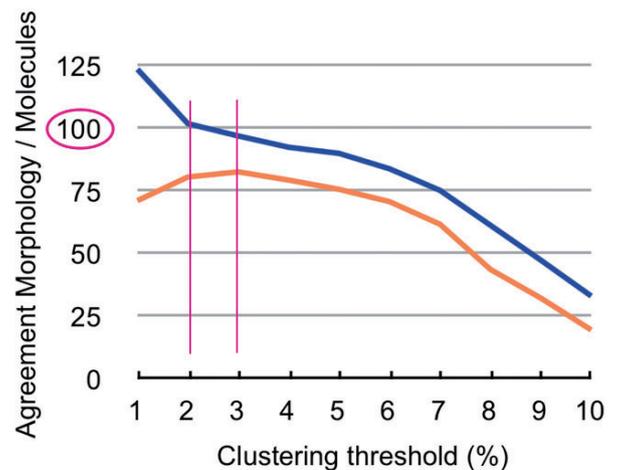


Fig. 3. Agreement and taxonomic accuracy of molecular clusters estimated at thresholds of 1-10% sequence divergence. Y-axis represents the relation of morphological entities *vs* molecular ones in % (*e.g.* if the *a priori* accepted morphological species number is 100, and we find 125 clusters using a 1% threshold, the value is 125%, meaning 25% overestimation of species number). Upper line is the agreement of species diversity estimation (percentage of clusters relative to the number of morphological species) while the lower line represents the taxonomic accuracy of clusters (number of clusters containing all sequences of a named species). At 100, there would be perfect agreement between both methods (modified from Hendrich *et al.*, 2010).

sults from local species inventories into a regional or global context. It has been shown that the full geographic range of a species and their genetic diversity has to be carefully considered (Bálint *et al.*, 2011), and the same is true for levels of endemism (species turnover or β -diversity).

We suggest the creation of a globally accessible database that has the power to strongly improve knowledge on Asian macrozoobenthos (and beyond), and begin to remedy the current problems associated with insufficient taxonomic knowledge.

The basic workflow is that of a typical DNA barcoding campaign, bringing together expert taxonomists, field biologists and molecular taxonomists. The major theme is sustainability and accountability, with all vouchers digitised and deposited in public natural history collections; examples can be seen at the *All Lepidoptera* website (<http://www.lepbarcoding.org>), or the already mentioned BOLD website (<http://www.boldsystems.org>). Pfrender *et al.* (2010) provide a review on project strategies that could also be adopted for work in SE Asia.

We envision a planning workshop at the next FISA meeting, to discuss an initial funding and sampling strategy. The obvious initial focus might be directed to lowland wetlands, which are under extreme threat. To gain first insights into diversity patterns, we suggest initial sampling at 200 localities across Southern and SE Asia west of Wallace line, covering the various biogeographical and ecological regions.

The DNA sequence data can serve as an immediate proxy for formally named species, to satisfy the requirement for a preliminary assessment of species diversity and diversity patterns using the MBA outlined above. Such a system would also support taxonomy, as data and vouchers will be freely available for experts, and monographic work should be strongly encouraged and financially supported, wherever possible. The preliminary data would certainly highlight areas of greatest potential species richness, which should incentivise taxonomists to prioritise efforts in these geographic areas of highest biodiversity.

A suggestion for taxon selection

We feel that a targeted approach is the key to success, and taxon selection should be based on discussions within the community to identify taxa where there is current taxonomic expertise to rapidly facilitate project outcomes, and also the selection of taxa where Asian projects can feed directly into existing global initiatives for added value. Examples are the international Barcode of Life's (iBOL) working group 1.7. Freshwater Bio-Surveillance which covers the groups commonly used for environmental impact assessments (EIAs) such as Ephemeroptera, Plecoptera, Trichoptera (EPT) and Odonata, or the Trichoptera barcode of life (<http://trichopteraabol.org/>). Oth-

ers include different groups of Coleoptera, Crustacea and Mollusca where local (national) and international working groups already study Asian representative taxa. Groups should in general have been shown to add relevant information for environmental impact assessment, and should be comparably easy to handle and curate. The Barcoding Fauna Bavarica project (<http://www.faubavarica.de/>) is an example where taxon selection was focused that way, funding and involving existing taxonomic expertise. After three years, more than 10,000 species have been collected, identified and sequenced. The setting for taxon identification is different in the widely understudied Asian fauna, but the basic framework is the same.

New technologies

There seems to exist a misconception as to the strength of DNA barcoding. It lies in the use of a standard marker that can link genotype to species. This requires a carefully well-finished and taxonomically revised database as basic knowledge of which haplotype belongs to which species, or which group of similar haplotypes represents which lineage of closely related young species (Hawlitschek *et al.*, 2012). Such a database forms the reference system for future mass sequencing technologies – no matter what percentage of a genome is being sequenced – so that matching the genetic data with taxonomy will only require a *cox1* database search in most cases.

Next generation sequencing

DNA barcoding can be coupled with next-generation sequencing (NGS) and this method is known as metabarcoding (Taberlet *et al.*, 2012a, 2012b). The latest DNA sequencing technologies can provide incredible amounts of information economically and more rapidly than ever envisioned before. Next-generation sequencing makes it possible to sequence large amounts of specimens at the same time. The length of sequence reads is increasing as technologies advance, and this may allow efficient sequencing of bulk samples, an approach referred to as environmental sequencing or metabarcoding (Hajibabaei *et al.*, 2011; Thomsen *et al.*, 2012; Yu *et al.*, 2012).

Hajibabaei *et al.* (2011) suggested that this is feasible for bulk samples of aquatic insects. Thomsen *et al.* (2012) showed that it is possible to detect (large) target species from water samples alone. That means that NGS methods offer an opportunity to monitor rare and threatened animal species from DNA traces in their freshwater environments, and even potential for identifying new species. The new DNA-based method is effective even in locations where the animals are extremely rare. The method if further developed might also show some correlation between the amount of DNA in the environment and the density of individuals, meaning that the DNA detection method can even be used

to estimate population sizes. This is crucial in the monitoring of rare animals, where one often wants to know whether the population is large or small. The United Nations (UN) have agreed to halt the decline of biodiversity, but a prerequisite to do so is that we are capable of accurately documenting the status of threatened species. This new approach is a fundamental step forward making it cheaper and faster to monitor endangered species, and thus prioritise efforts to the benefit of biodiversity at a broad scale. Recent advances in NGS technologies enable such rapid quantification of these parameters directly from environmental water samples. Yu *et al.* (2012) show that NGS provides robust alpha and beta biodiversity assessments even in the absence of existing taxonomy, essentially taking the molecular biodiversity assessment methods suggested by Hendrich *et al.* (2010) and Tänzler *et al.* (2012) to the ultrasequencing level.

Finally, Zhou *et al.* (2013) describe an ultra-deep sequencing pipeline, *i.e.* NGS sequencing of environmental samples using a sequencing platform that produces extremely high amounts of sequence reads, which does not rely on pre-sequencing polymerase chain reaction (PCR) amplification. This means that the sample will go through a mitochondrial DNA enrichment step (centrifugation), and the DNA extraction step, followed by NGS sequencing. Such a pipeline, once optimised, would allow the simultaneous sequencing of mixtures of phylogenetically rather heterogeneous samples, avoiding PCR primer mismatch problems. Also, without PCR amplification, there is some hope to be able to reveal relative abundance patterns in the sample.

CONCLUSIONS

Molecular biodiversity assessment might be a good alternative to tackle morphotype sorting, as it is more objective and relies on more standardised protocols especially in the context of larger-scale projects with diverse personnel.

Conducted carefully, MBA for the first time in history: i) helps to gather scientific data where there is little or no hope for other information; ii) enables meaningful assessment of diversity without guessing how many *morphospecies* there are; and iii) most importantly, it enables student/local participation in an objective framework.

New sequencing technologies will also enable us to conduct large surveys or monitoring for conservation purposes based on objective scientific data.

Understanding how ecosystems function, however, relies on knowledge of species and their biotic and abiotic interactions, so taxonomists and field biologists/ecologists will always play a key role. The approach outlined here should facilitate targeted taxonomic work, and free taxonomists from routine work that can certainly be standardised and automated more efficiently and effectively.

ACKNOWLEDGMENTS

Our sincere thanks to Dr. Maria K. Hołyńska and Prof. La-orsri Sanoamuang for organising the *Freshwater Invertebrates of Southeast Asia: Biodiversity and Origin* (FISA) workshop in Maha Sarakham, Thailand, 30th July to 4th August, 2012. This work was supported by DFG grant BA2152/11-1 to Dr. Michael Balke.

REFERENCES

- Bálint M, Domisch S, Engelhardt CHM, Haase P, Lehrian S, Sauer J, Theissinger KT, Pauls SU, Nowak C, 2011. Cryptic biodiversity loss linked to global climate change. *Nat. Clim. Chang.* 1:313-318.
- Bergsten J, Bilton DT, Fujisawa T, Elliott M, Monaghan M, Balke M, Hendrich L, Geijer J, Herrmann J, Foster G, Ribera I, Nilsson AN, Barraclough TG, Vogler AP, 2012a. The effect of geographical scale of sampling on DNA barcoding. *Syst. Biol.* 61:851-869.
- Bergsten J, Brilmyer G, Crampton-Platt A, Nilsson AN, 2012b. Sympatry and colour variation disguised well-differentiated sister species: *sphrodytes* revised with integrative taxonomy including 5 kbp of housekeeping genes (Coleoptera: Dytiscidae). *DNA Barcodes (Berlin)* 1:1-18.
- Dayrat B, 2005. Towards integrative taxonomy. *Biol. J. Linn. Soc.* 85:407-415.
- Dincă V, Zakharov EV, Hebert PD, Vila R, 2011. Complete DNA barcode reference library for a country's butterfly fauna reveals high performance for temperate Europe. *P. Roy. Soc. B-Biol. Sci.* 278:347-355.
- Goldstein PZ, DeSalle R, 2011. Integrating DNA barcode data and taxonomic practice: determination, discovery, and description. *Bioessays* 33:135-147.
- Hajibabaei M, Shokralla S, Zhou X, Singer GAC, Baird DJ, 2011. Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS ONE* 6:e17497.
- Hausmann A, Haszprunar G, Hebert PDN, 2011. DNA barcoding the geometrid fauna of Bavaria (Lepidoptera): successes, surprises, and questions. *PLoS ONE* 6:e17134.
- Hawltischek O, Hendrich L, Espeland M, Toussaint EFA, Genner MJ, Balke M, 2012. Pleistocene climate change promoted rapid diversification of aquatic invertebrates in Southeast Australia. *BMC Evol. Biol.* 12:142.
- Hebert PDN, Cywinska A, Ball SL, deWaard JR, 2003a. Biological identifications through DNA barcodes. *P. Roy. Soc. Lond. B Bio.* 270:313-321.
- Hebert PDN, Penton EH, Burns J, Janzen DH, Hallwachs W, 2004a. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly, *Astraptes fulgerator*. *P. Natl. Acad. Sci. USA* 101:14812-14817.
- Hebert PDN, Ratnasingham S, deWaard JR, 2003b. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *P. Roy. Soc. Lond. B Bio.* 270(Suppl.1): S96-99.
- Hebert PDN, Stoeckle MY, Zemplak TS, Francis CM, 2004b. Identification of birds through DNA barcodes. *PLoS Biol.* 2:1657-1663.
- Hebert PDN, Zakharov EV, Prosser SW, Sones JE, McKeown JT, Mantle B, La Salle J, 2013. A DNA 'Barcode Blitz': rapid digitization and sequencing of a natural history collection. *PLoS ONE* 8:e68535.

- Hendrich L, Hawlitschek O, Balke M, 2009. A revision of Australasian *Neobidessodes Sphrodytes* gen.n. diving beetles - integrating morphology, cybertaxonomy, DNA taxonomy and phylogeny (Coleoptera: Dytiscidae, Bidessini). *Zootaxa* 2288:1-41.
- Hendrich L, Pons J, Ribera I, Balke M, 2010. Mitochondrial *cox1* sequence data reliably uncover patterns of insect diversity but suffer from high lineage-idiosyncratic error rates. *PLoS ONE* 5:e14448.
- House of Lords, 2008. Systematics and taxonomy: follow-up. House of Lords ed., London: 330 pp. Available from: <http://www.publications.parliament.uk/pa/ld200708/ldselect/ldscitech/162/162.pdf>
- Janzen DH, Hallwachs W, Blandin P, Burns JM, Cadiou J-M, Chacon I, Dapkey T, Deans AR, Epstein ME, Espinoza B, Franclemont JG, Haber WA, Hajibabaei M, Hall JPW, Hebert PDN, Gauld ID, Harvey DJ, Hausmann A, Kitching IJ, Lafontaine D, Landry J-F, Lemaire C, Miller JY, Miller JS, Miller L, Miller SE, Montero J, Munroe E, Rab Green S, Ratnasingham S, Rawlins JE, Robbins RK, Rodriguez JJ, Rougerie R, Sharkey MJ, Smith MA, Solis MA, Sullivan JB, Thiaucourt P, Wahl DB, Weller SJ, Whitfield JB, Willmott KR, Wood DM, Woodley NE, Wilson JJ, 2009. Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. *Mol. Ecol. Resour.* 9:1-26.
- Krell F-T, 2004. Parataxonomy vs. taxonomy in biodiversity studies - pitfalls and applicability of 'morphospecies' sorting. *Biodivers. Conserv.* 13:795-812.
- Mallet J, Willmott K, 2003. Taxonomy: renaissance or tower of Babel? *Trends Ecol. Evol.* 18:57-59.
- May RM, 2010. Tropical arthropod species, more or less? *Science* 329:41-42.
- Meier R, Shiyang K, Vaidya G, Ng PKL, 2006. DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Syst. Biol.* 55:715-728.
- Monaghan MT, Balke M, Gregory TR, Vogler AP, 2005. DNA-based species delineation in tropical beetles using mitochondrial and nuclear markers. *Philos. T. Roy. Soc. B* 360: 1925-1933.
- Monaghan MT, Balke M, Pons J, Vogler AP, 2006. Beyond barcodes. Complex DNA taxonomy of a South Pacific Island radiation. *P. Roy. Soc. Lond. B Bio.* 273:887-893.
- Monaghan MT, Wild R, Elliot M, Fujisawa T, Balke M, Inward DJ, Lees DC, Ranaivosolo R, Eggleton P, Barraclough TG, Vogler AP, 2009. Accelerated species inventory on Madagascar using coalescent-based models of species delimitation. *Syst. Biol.* 58:298-311.
- Mora C, Tittensor DP, Adl S, Simpson AG, Worm B, 2011. How many species are there on Earth and in the ocean? *PLoS Biol.* 9:e1001127.
- Moritz C, Cicero C, 2004. DNA barcoding: promise and pitfalls. *PLoS Biol.* 2:e354.
- Padial JM, Miralles A, De la Riva I, Vences M, 2010. The integrative future of taxonomy. *Front. Zool.* 7:16.
- Pfrender ME, Hawkins CP, Bagley M, Courtney GW, Creutzburg BR, Epler JH, Fend S, Schindel D, Ferrington LC Jr, Hartzell PL, Jackson S, Larsen DP, Lévesque A, Morse JC, Petersen MJ, Ruiter D, Whiting M, 2010. Assessing macroinvertebrate biodiversity in freshwater ecosystems: advances and challenges in DNA-based approaches. *Q. Rev. Biol.* 85:319-340.
- Pons J, Barraclough TG, Gomez-Zurita J, Cardoso A, Duran DP, Hazell S, Kamoun S, Sumlin WD, Vogler AP, 2006. Sequence-based species delimitation 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Syst. Biol.* 55:595-609.
- Rach J, DeSalle R, Sarkar IN, Schierwater B, Hadrys H, 2008. Character-based DNA barcoding allows discrimination of genera, species and populations in Odonata. *P. Roy. Soc. Lond. B Bio.* 275:237-247.
- Ratnasingham S, Hebert PDN, 2007. BOLD: the barcode of life data system. *Mol. Ecol. Notes* 7:355-364.
- Riedel A, Sagata K, Suhardjono YR, Tänzler R, Balke M, 2013a. Integrative taxonomy on the fast track - towards more sustainability in biodiversity research. *Front. Zool.* 10:15.
- Riedel A, Sagata K, Surbakti S, Tänzler R, Balke M, 2013b. One hundred new species of *Trigonopterus* weevils from New Guinea. *Zookeys* 280:1-150.
- Ruiter DE, Boyle EE, Zhou X, 2013. DNA barcoding facilitates associations and diagnoses for Trichoptera larvae of the Churchill (Manitoba, Canada) area. *BMC Ecol.* 13:5.
- Skale A, Tänzler R, Hendrich L, Balke M, 2012. High mitochondrial DNA sequence divergence in New Guinea *Carabdytes* stream beetles and the taxonomist's dilemma when other evidence is kind of subtle (and collecting localities are far far away). *Zookeys* 247:31-43.
- Taberlet P, Coissac E, Hajibabaei MM, Rieseberg LH, 2012a. Environmental DNA. *Mol. Ecol.* 21:1789-1793.
- Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E, 2012b. Towards next generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.* 21:2045-2050.
- Tänzler R, Sagata K, Surbakti S, Balke M, Riedel A, 2012. DNA barcoding for community ecology. How to tackle a hyperdiverse, mostly undescribed Melanesian fauna. *PLoS ONE* 7:e28832.
- Thomsen PF, Kielgast J, Iversen LL, Wiuf C, Rasmussen M, Gilbert MT, Orlando L, Willerslev E, 2012. Monitoring endangered freshwater biodiversity using environmental DNA. *Mol. Ecol.* 2:2565-2573.
- Will KW, Mishler BD, Wheeler QD, 2005. The perils of DNA barcoding and the need for integrative taxonomy. *Syst. Biol.* 54:844-851.
- Wilson EO, 1988. Biodiversity. National Academies Press, Washington, DC: 538 pp.
- Yu DW, Ji Y, Emerson BC, Wang X, Ye C, Yang C, Ding Z, 2012. Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods Ecol. Evol.* 3:613-623.
- Zhou X, Adamowicz S, Jacobus LM, DeWalt RE, Hebert PDN, 2009. Towards a comprehensive barcode library for arctic life. Ephemeroptera, Plecoptera, and Trichoptera of Churchill, Manitoba, Canada. *Front. Zool.* 6:30.
- Zhou X, Jacobus LM, DeWalt RE, Adamowicz S, Hebert PDN, 2010. Ephemeroptera, Plecoptera, and Trichoptera fauna of Churchill (Manitoba, Canada): insights into biodiversity patterns from DNA barcoding. *J. N. Am. Benthol. Soc.* 29:814-837.
- Zhou X, Kjer KM, Morse JC, 2007. Associating larvae and adults of Chinese Hydropsychidae (Insecta: Hydropsychidae) using DNA sequences. *J. N. Am. Benthol. Soc.* 26:718-741.
- 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.