How to tackle the molecular species inventory for an industrialized nation – lessons from the first phase of the German Barcode of Life initiative GBOL (2012-2015)

Journal:	<i>Genome</i>
Manuscript ID	gen-2015-0185.R1
Manuscript Type:	Article
Date Submitted by the Author:	08-Feb-2016
Complete List of Authors:	<p>Geiger, Matthias; Zoologisches Forschungsmuseum A. Koenig - Leibniz Institut f��r Biodiversit��t der Tiere, Astrin, Jonas; Zoologisches Forschungsmuseum A. Koenig - Leibniz Institut f��r Biodiversit��t der Tiere Borsch, Thomas; BGBM Burkhardt, Ulrich; Senckenberg Museum f��r Naturkunde G��rlitz Grobe, Peter; Zoologisches Forschungsmuseum A. Koenig - Leibniz Institut f��r Biodiversit��t der Tiere Hand, Ralf; BGBM Hausmann, Axel; SNSB-Zoologische Staatssammlung M��nchen Hohberg, Karin; Senckenberg Museum f��r Naturkunde G��rlitz Krogmann, Lars; Staatliches Museum f��r Naturkunde Stuttgart Lutz, Matthias; Institut f��r Evolution und ��kologie, Organismische Botanik Institut f��r Biologie, Universit��t T��bingen Monje, Carlos; Staatliches Museum f��r Naturkunde Stuttgart Misof, Bernhard; Zoologisches Forschungsmuseum A. Koenig - Leibniz Institut f��r Biodiversit��t der Tiere Morini��re, Jerome; SNSB-Zoologische Staatssammlung M��nchen M��ller, Kai; Institute for Evolution and Biodiversity, Westf��lische Wilhelms-Universit��t Pietsch, Stephanie; Zoologisches Forschungsmuseum A. Koenig - Leibniz Institut f��r Biodiversit��t der Tiere Quandt, Dietmar; Nees-Institut f��r Biodiversit��t der Pflanzen Rulik, Bj��rn; Zoologisches Forschungsmuseum A. Koenig - Leibniz Institut f��r Biodiversit��t der Tiere Scholler, Markus; Staatliches Museum f��r Naturkunde Karlsruhe Traunspurger, Walter; Universit��t Bielefeld Haszprunar, Gerhard; SNSB-Zoologische Staatssammlung M��nchen W��gele, Wolfgang; Zoologisches Forschungsmuseum A. Koenig - Leibniz Institut f��r Biodiversit��t der Tiere</p>
Keyword:	Biodiversity assessment, citizen science, DNA barcoding, reference library, Germany

SCHOLARONE™
Manuscripts

Draft

How to tackle the molecular species inventory for an industrialized nation – lessons from the first phase of the German Barcode of Life initiative GBOL (2012-2015)

M. F. Geiger^{*§}, J. J. Astrin^{*}, T. Borsch[§], U. Burkhardt², P. Grobe^{*}, R. Hand[§], A. Hausmann¹, K. Hohberg², L. Krogmann³, M. Lutz⁸, C. Monje³, B. Misof^{*}, J. Morinière¹, K. Müller⁴, S. Pietsch^{*}, D. Quandt⁵, B. Rulik^{*}, M. Scholler⁶, W. Traurnspurger⁷, G. Haszprunar¹, W. Wägele^{*}

^{*}Stiftung Zoologisches Forschungsmuseum Alexander Koenig, Leibniz Institut für Biodiversität der Tiere, Adenauerallee 160, 53113 Bonn

[§]Botanischer Garten und Botanisches Museum Berlin-Dahlem, Freie Universität Berlin, Königin-Luise-Straße 6-8, 14195 Berlin

¹SNSB-Zoologische Staatssammlung München, Münchhausenstraße 21, 81247 München

²Senckenberg Museum für Naturkunde Görlitz, Am Museum 1, 02826 Görlitz

³Staatliches Museum für Naturkunde Stuttgart, Rosenstein 1, 70191 Stuttgart

⁴Institute for Evolution and Biodiversity, Westfälische Wilhelms-Universität, Hüfferstrasse 1, 48149 Münster

⁵Nees-Institut für Biodiversität der Pflanzen, Meckenheimer Allee 170, 53115 Bonn

⁶Staatliches Museum für Naturkunde Karlsruhe, Erbprinzenstr. 13, 76133 Karlsruhe

⁷Universität Bielefeld, Universitätsstraße 25, 33615 Bielefeld

⁸Plant Evolutionary Ecology, Institute of Evolution and Ecology, University of Tübingen, Auf der Morgenstelle 5, 72076 Tübingen[§]corresponding author

Draft

Abstract

Biodiversity loss is mainly driven by human activity. While concern grows over the fate of hot-spots of biodiversity, contemporary species losses still prevail in industrialized nations. Therefore, strategies were formulated to halt or reverse the loss, driven by evidence for its value for ecosystem services. Maintenance of the latter through conservation depends on correctly identified species. To this aim, the German Federal Ministry of Education and Research is funding the GBOL project, a consortium of natural history collections, botanic gardens and universities working on a barcode reference database for the country's fauna and flora. Several noticeable findings could be useful for future campaigns: 1) validating taxon lists to serve as a taxonomic backbone is time-consuming, but without alternative; 2) offering financial incentives to taxonomic experts, often citizen scientists, is indispensable; 3) completion of the libraries for widespread species enables analyses of environmental samples, but the process may not hold pace with technological advancements; 4) discoveries of new species are among the best stories for the media; 5) a commitment to common data standards and repositories is needed, as well as trans-boundary cooperation between nations; 6) after validation, all data should be published online via the BOLD to make them searchable for external users and to allow cross-checking with data from other countries.

Keywords: Biodiversity assessment, citizen science, DNA barcoding, reference library, Germany

Introduction

Recent estimates suggest that there might be 8.7 million species on Earth (Mora et al. 2011), 1.9 million of them being described scientifically and catalogued (Stuart et al. 2010). The discussion about the actual number of species also evoked controversy whether there are fewer or more taxonomists as compared to previous decades (Costello et al. 2013a, 2013b, 2014), tightly linked to the term 'taxonomic impediment' (Crozier 1997; Khuroo et al. 2007). What seems to be common sense is: 1) millions of species remain undescribed, which means hypotheses about interactions on ecosystem-level based on few indicator species might be problematic (Gerlach et al. 2013; Macher et al. 2015), also due to misidentifications (Birk et al. 2012; Balvanera et al. 2014); 2) species determination using exclusively morphological characters is often time-consuming, cost-intensive and too often relies on scarce taxonomic expertise gained only through years of training (Frobel and Schlumprecht 2014); 3) the current rate of species extinction far exceeds the background rate (Pimm et al. 2014), and climate displacement puts additional pressure on many species (Keith et al. 2014); 4) technological advancements exist that could counteract the consequences arising from 1) and 2) (Scheffers et al. 2012; Hebert et al. 2003a). A well-established tool to facilitate the global inventory of biodiversity is DNA barcoding (Hebert et al. 2003b), where short gene segments - unique for each species - have emerged as a standard for fast and reliable identification of specimens to species. Moreover, technological and conceptual advances in the field of taxonomy indicate that multiple loci will in the near future increase identification success and reliability (Korotkova et al. 2011).

The GBOL initiative (German Barcode of Life – www.bolgermany.de) has realized this global vision on a national level and will create the first comprehensive genetic library for species for the whole of Germany. The national network of 17 natural history museums and other biodiversity research institutes has been granted €5.4 million for an initial 3-year period to set up the infrastructure and to start processing the first third of the fauna and flora. This contribution to the special volume in Genome on occasion of the 6th International Barcode of Life Conference presents the major project components that have been established,

elaborates on obstacles overcome and gives insights into first results and an outlook to planned DNA barcoding applications in Germany. Given that several countries and international groups are working on comprehensive barcoding surveys or are in the stage of planning these, our aim is to share our experiences, which could be used to set up similar projects or assist when applying for funding.

Germany has over 80 million inhabitants living on an area of approximately 357,000 km², which makes it one of the most densely populated countries in Europe. Its industry can still become environmentally problematic, and the intensive and highly industrialized agriculture is one of the most severe threats to local biodiversity (Tscharntke, Klein, Kruess, et al. 2005). Thus, greater efforts are needed for environmental protection than in other countries. In the center of Europe, the composition of its flora and fauna was mainly shaped during the last ice age and interglacial periods and contains elements from different biogeographic regions (Schmitt 2007; Taberlet et al. 1998). According to the German Federal Agency for Nature Conservation, about 48,000 animal, 9,500 plant (including different groups of algae) and 14,000 fungal species (likely a gross underestimation) are recorded for continental Germany, and about one fourth is represented in the national Red List for conservation (Bundesamt für Naturschutz 2015).

Generally, Northern temperate ecosystems are projected to experience smaller biodiversity changes in the future, because major land-use alteration has already occurred (Sala et al. 2000). While this certainly holds true also for Germany, within Europe the country is among those with the highest biodiversity threat levels (Federal Ministry for the Environment 2007), with for example 43% of the Red List assessed vertebrates and 27% of flowering plants and ferns being threatened (Bundesamt für Naturschutz 2012). Given that only part of the diversity has been Red List assessed until now (animals: 35%; plants: 50%), it becomes clear that more baseline data are urgently needed. Other compelling arguments why projects like GBOL are needed lie in the agriculture and forestry sectors. Farmland and forests cover about 52% and 30%, respectively, of Germany's surface, and both deliver important

economic and societal services. These sectors would greatly benefit from a faster and more reliable detection and identification of invasive species, new biological pest control applications, or indicator species for monitoring initiatives and habitat quality assessment (Ji et al. 2013).

General structure of GBOL: institutes with taxonomic focus

The participating institutes which applied for funding as a consortium (Table 1) provide their professional taxonomic expertise and existing infrastructure (curators, molecular and morphological collections, databases, bioinformatic platforms and expertise, and laboratories) to comprehensively collect, catalog, describe, and sequence the animal, plant and fungi species from Germany. They also guarantee, after GBOL funding ends, the storage and curation of voucher specimen collections and molecular collections with associated information beyond bare metadata, which has been emphasized as being a crucial aspect of DNA barcoding (Collins and Cruickshank 2013). During the funding period, one central coordinator (employed at ZFMK) covers the reporting duties and general public relations work, organizes workshops, represents the initiative to the outside and guides the development of the web portal. At five major nodes taxonomic coordinators have been employed to take care of the DNA barcoding front-end logistics, often with a special focus on communicating with taxonomic specialists, and to ensure data quality and integrity. The taxonomic coordinators are responsible for the acquisition of suitable material and have expertise in one or more taxonomic groups. They form the backbone of all pre-laboratory workflows and prioritize the collecting and processing of the material. Together with a half-time programmer, five technical assistants and resources for student helpers, GBOL staff salaries make up for 32% of the total budget. In addition, the in-kind contribution of some participating institutions sums up to half of the dedicated cash funding provided through GBOL through the work time of permanent staff (curators, technicians, administration etc.) and infrastructure, reflecting the strong commitment to the program.

Compilation and validation of taxon lists

Although the Federal Agency for Nature Conservation curates and irregularly publishes species lists for Germany, these are neither complete, nor do they allow estimating the distribution of the majority of taxa in Germany. Our strategy was therefore to have expert taxonomists work on particular groups of organisms to generate taxon lists, which serve as a backbone and target lists for the project. While some of these lists existed (e.g., for beetles (Klausnitzer and Köhler 1998), spiders, vascular plants, or collembolans (Schulz et al. 2003, but see <http://goo.gl/wiAUB3>)), or could be built upon using lists from other projects (e.g., the Barcoding Fauna Bavarica Project - www.faunabavarica.de), major efforts were and still are needed in other species-rich groups (e.g., vertebrate parasites or most soil animal taxa e.g., nematodes), where—despite their importance in nutrient cycles—the number of German species can still only be roughly estimated. However, these validated taxon lists are important to (a) check species coverage in GBOL and to focus on gaps and further (b) to check the barcode-based taxonomic units—MOTUs, BINs (Ratnasingham and Hebert 2013), monophyletic entities, etc.—versus the current taxonomic units that were mostly described based on morphology.

It is also evident that we are in a transition phase from alpha-taxonomy based primarily upon morphology to more phylogenetic classification systems in many groups of organisms (Borsch et al. 2015), with the consequence of significant amounts of changing taxon concepts even in comparatively well-known groups, such as flowering plants (Borsch et al. 2015). Therefore, not only the inclusion of taxonomic experts who curate taxon lists is imperative, but also the deposition of DNA extracts in biobanks (Gemeinholzer et al. 2011) is highly advisable to ensure future updating of reference databases with additional markers or even whole genomes.

Most lists are curated in the Diversity Taxon Names (DTN) module of the Diversity Workbench package (www.diversityworkbench.net), where they can be kept up-to-date (<http://goo.gl/wiAUB3>). For some groups of organisms with a large scientific community

working on them, such as flowering plants, we use more inclusive continental-scale syntheses (Euro+Med Plant Base; ww2.bgbm.org/EuroPlusMed/), which provide an overview on currently accepted taxon concepts and corresponding names. These resources are being used to derive continuously updated national checklists (e.g., Buttler and Hand 2008) for the GBOL project. The importance of the target lists grows with the progress in taxonomic coverage achieved, as gaps can be identified and addressed more specifically. As our aim is also to cover the molecular diversity over the range of each species in Germany, we try to include several (2-7) specimens from each of the federal states (animals) or larger geographical regions (plants). This pragmatic approach turns out to be very valuable to quickly query the coverage of particular taxa for a certain region and helps coordinating sampling efforts. Registered users (see below) can use the web portal to retrieve the coverage of a particular taxon for each federal state.

Engagement strategy for citizen scientists

In addition to approx. 70 professional taxonomists (usually museum curators) GBOL depends on the active support of privately acting taxonomic experts. They contribute significantly to the fresh sampling, appropriate preservation and identification of organisms for DNA barcoding, as these selected GBOL partners have excellent knowledge of the morphology, autecology, reproductive mode and life cycle of the taxa they work with, which is indispensable for a correct determination. More than 60% of the new species from Europe have been described by non-professional (i.e. non-hired) taxonomists since the 1950s (Fontaine et al. 2012), but the importance of these specialists for taxonomy is not yet widely acknowledged. Until now, more than 225 external experts have registered (after approval from the taxonomic coordinators) as official taxon-experts through the GBOL web portal. Furthermore, many unregistered experts helped in preparing field trips and locating rare taxa. After registration, they get feedback from a taxonomic coordinator and an introduction to sampling and data recording, as well as assistance in obtaining collecting permits if needed. Given that about 37% of Germanys land area is protected and falls into over 15,000 nature

reserves and similar sanctuaries, we are grateful for a general letter of support from the Federal Government and States Working Group for Nature and Landscape Conservation, and Recreation (LANA). However, due to the federal organization of management and monitoring of biodiversity in Germany, we nevertheless had to apply for numerous exemptions for collecting protected species or in protected areas in all 16 different states.

Once approved, our external experts can order through the web portal different sampling kits (racks with EtOH-prefilled and labelled 2ml or 5ml tubes (fauna), silica bags (flora)) and download the respective data sheet for metadata to be associated with the collected material. One important feature of this system is the initial assignment of GBOL-sample IDs, which remain permanently associated with the specimens: the tubes as well as the data sheet already contain the sample IDs. After collecting, preserving and morphologically identifying the specimen, data collection sheets are populated with relevant additional information (location, date, coordinates, taxonomy notes, habitat type, etc.) and uploaded on the web portal. Alternatively, an online submission form can be used. The sample material will then be shipped to the appropriate GBOL institution for further processing. The benefits for our external experts not only encompass a financial reimbursement (currently usu. €3-5 per identified individual; up to three identified individuals per species and federal state can be contributed in a first round), but also free access to sequences and metadata, and the possibility of co-authorships in joint publications. For many experts the latter arguments were pivotal, and they contributed to GBOL without any financial interest. Visualization of an external taxonomist's contribution on the GBOL-website is a further incentive, but only visible for registered users and only if explicitly chosen.

The sheer amount of material needed – when possible three individuals per species from each of 16 federal states times 48,000 species (= 2,304,000 individuals just for animals) – as well as missing expertise in many taxa makes the inclusion of external taxonomists the most important element of success for GBOL.

There are, however, also groups of organisms for which no external taxonomic specialists would be available, e.g., small dipterans or hymenopterans, soil-dwelling micro- and

mesofauna like nematodes, collembolans and mites, all characterized by small body size and few distinguishing characters. These groups require expensive laboratory facilities (microscopes) and elaborate techniques (live extraction from soil, rearing to empty the intestines as the whole body is used for DNA extraction, special procedures to gain voucher specimens and partially non-destructive DNA extraction protocols). But without doubt, soil animals deserve special effort in developing techniques that facilitate their investigation due to their immense diversity, their high importance for ecosystem services, and the limited number of specialists.

The general utility of DNA barcoding for specimen (re-)identification to species level is no longer a matter of prominent, contentious debate, and the method has become well-established in at least 27 countries worldwide (including GBOL in Germany). Yet, some still see a potential “threat” in DNA barcoding initiatives for their guild and refrain from actively participating. As a consequence, the urgently needed completion of reference databases is hampered. However, judging from the still ongoing gain of new external experts, we witness significant decrease in the reservations towards contributing to GBOL. Often, this is mediated through community dynamic processes (e.g., experiencing how barcoding worked for a colleague, who finally got clues on how to solve the taxonomic riddle in 'his'/'her' group). In the future we aim at setting an additional incentive to engage in GBOL by offering higher reimbursement rates for rare species. We greatly appreciate the opportunity to use the synergies between DNA barcoding and taxonomists, which lay promising foundations for future biodiversity research. Similar and linked to natural history museums archiving snapshots of biodiversity, DNA barcode libraries conserve taxonomic expertise and make it available in a digital form for a plethora of applications. Our experience from workshops and seminars for biology students is that the visual representation of taxonomic conflicts in a dendrogram (i.e. splitting or lumping of taxa) easily attracts the interest of younger biologists, who find an intuitive entry point for deeper taxonomic studies. Fortunately, today also many of the more experienced non-staff taxonomists have realized the value of including DNA barcode data to examine taxonomic issues.

Completion of the reference library

Due to financial restrictions GBOL phase I was designed for a period of initially only three years and excludes microorganisms, most fungi, and marine organisms. For the latter we closely cooperate with an initiative of the German Center for Marine Biodiversity Research (DZMB) focusing on the fauna of the North Sea. We estimated that it would take about 10 years of concerted efforts to cover the complete animal and plant diversity of Germany. Therefore, the main aims for 2012 to 2015 were to generate barcode coverage for about one third of the German fauna and flora and to set up the necessary laboratory and IT infrastructure, as well as all general logistics. After three years, over 225,000 specimens belonging to over 26,000 species have been identified and stored in the collections of the respective GBOL institutes. As of late 2015, we had generated more than 143,000 DNA barcodes from over 20,000 different species (Table 2).

The taxonomic coverage varies among groups, with almost completed reference libraries for Lepidoptera, Aculeata, Ephemeroptera, Trichoptera, some smaller arthropod groups (Neuropterida, Odonata, and Orthoptera), and a few vertebrate groups (amphibians, reptiles, and fishes). Other important and speciose groups, such as Coleoptera and Arachnida, are represented by over 60% of the known German species in the DNA barcode library, including records from the Barcoding Fauna Bavarica project. Diptera, as another major insect order, is possibly the most challenging order globally, and GBOL could not yet cover more than 10% of the known approx. 9,500 species registered for Germany. The progress is similar in non-aculeate Hymenoptera. The soil invertebrate fauna, highly important to the overall performance of terrestrial ecosystems, but also highly complex due to minute size, hidden lifestyle and extreme diversity, is represented with about 4% (nematodes) to 16% (collembolans) of the described German species. While the standard mitochondrial COI barcode marker is widely used for animal taxa, the soil organisms section also included the nuclear 28S rRNA gene to confirm COI results as suggested by, e.g., Porco et al. (2012), and to widen resolution efficiency, as in some taxonomic groups the 28S marker proved

more efficient for separation of closely-related species and for amplification efficiency due to partially poor performance of standard COI primers.

A feature of the GBOL botany section is to provide sequence data for three fast-evolving plastid DNA regions (*trnL-F*, *rpl16*, *trnK/matK*) in combination with the nuclear internal transcribed spacers (ITS). Here we deviate from the officially (CBOL) suggested marker combination (*rbcL* & *matK*) to achieve higher resolution (compare Korotkova et al. 2011). Especially for the nrDNA ITS marker, an increased demand in the future is anticipated, for example in applications that involve the analysis of pollen. Of the flowering plants, about 800 species have been added to the database with DNA barcodes, and a new LIMS/FIMS tool (www.gbol5.de) has been developed to meet the particular requirements of a) distributed work and b) assembly and contig creation of highly complex sequences.

The only group of fungi which is represented in the initial 3.5-year phase of the project is rust fungi (Pucciniales, Basidiomycota). Rust fungi are plant parasites with numerous economically important phytopathogens of cultivated plants. The aim of the rust fungi project is to compile a barcode library for selected taxa in Germany not belonging to *Puccinia* such as species of the *Uromyces pisi* complex. The ITS1 and ITS2 regions of the nuclear rDNA including the 5.8S rDNA and partially also the large subunit of the ribosomal DNA, are suitable common fungal barcode markers (Schoch et al. 2012).

For almost all groups covered in GBOL, more material has been gathered over the first 30 months than could be processed within the time frame of GBOL phase I. This remarkable success was only made possible by activating the taxonomic knowledge outside the natural history collections and universities.

The geographical coverage achieved over Germany varies between taxonomic groups, so some measures had to be taken to compensate for the uneven distribution of collectors (and samples), such as Bioblitz events in underrepresented areas. In the vast majority of species, missing DNA barcodes from distant populations (e.g., from North vs. South Germany) do not hinder unambiguous species identification. However, in a few cases known from well-studied

groups, elevated levels of intraspecific variation had been detected in Germany (Knebelsberger et al. 2015; Hendrich et al. 2015; Hausmann et al. 2011a, 2013; Burkhardt et al. 2014). These might hamper straightforward identification when isolated lineages are not characterized before and would argue for a continued attempt to include material from several distant populations of the same species to be able to estimate levels of genetic diversity. The goal is clearly to have all widespread species being represented in the reference library with multiple representatives in order to enable the analyses of environmental samples to species level. Yet, the process of building reference libraries does currently not keep pace with the technological advancements, and the majority of NGS-derived molecular signatures cannot be identified to species level via reverse taxonomy (see below). It is estimated that about 1/3 of all German species are abundant or very common, 1/3 distributed locally and/or rare, and a last 1/3 very rare, being found only once or twice per 50 years (usually with Red List status). This implies that a DNA barcode reference library with 50-60% species coverage will allow the re-identification of 80-95% of all specimens of a given real-world sample, and thus would be sufficient for the vast majority of applications (while hits for species new to the reference library, e.g., in a metabarcoding application, could prompt search and if necessary protection measures for very rare species).

Highlights/Discoveries of new species

Public relations are crucial to raise funds and help activate external experts. Among the best lead stories for the media are discoveries of new, invasive or other enigmatic species (for example parasitic living lampreys, or the notion of the 'highest DNA barcode' from a moth from the top of Germany's highest mountain). Several scientific publications have resulted from the first phase of GBOL and the Barcoding Fauna Bavarica project, including for example the demonstration of extreme levels (up to 16.7% *p*-distance) of cryptic diversity in the centipede *Stenotaenia linearis* (Wesener et al. 2015) or a widespread freshwater nematode morphospecies (Ristau et al. 2013). Analyzing more than 400 species of true bugs other colleagues showed that DNA barcodes are highly effective in distinguishing over 90%

of Heteroptera species (Raupach et al. 2014), and a similar high level of identification success was observed in neuropterids (Morinière et al. 2014) and Lepidoptera (Hausmann et al. 2011a, 2011b). The release of the first comprehensive and largest DNA barcode reference library for Coleoptera included over 3,500 species (53% of the German fauna), and again more than 90% of the specimens could be unambiguously identified via their DNA barcode (Hendrich et al. 2015). Almost 100% of the German herpetofauna have species-specific DNA barcodes, and the unexpected levels of intraspecific variation in a few members now enable the discrimination of allochthonous lineages and estimation of gene flow (Hawliitschek et al. 2015). Presumably, post-glacially evolved species and incidents of introgressive hybridization resulting from stocking with allochthonous lineages resulted in a lower (78%) efficacy to unambiguously identify freshwater fishes (Knebelsberger et al. 2015). However, what was taken up from the press and brought to attention via several online-channels was the notion of seven 'new', potential invasive fishes, which had been processed within the aim of completing the reference data for all German species.

Commitment to common repositories and international cooperation

Different strategies have been pursued to accumulate GBOL data, which are currently being fused to be available to the user from a single national platform. Museum Koenig (ZFMK) has developed a sample and data processing scheme, which includes a field information management system (FIMS), represented by the Diversity Workbench Framework (developed at SNSB, Munich; <http://diversityworkbench.net/>) and automatic data exchange/annotation with the laboratory information management system (LIMS, Geneious Pro, developed by Biomatters, Auckland) (Fig.1). All specimen and collection-related information along with the sequencing data is transferred to, and displayed on, the continuously enhanced web portal (www.bolgermany.de). The Bavarian State Collection of Zoology in Munich (ZSM) and some other institutes already had substantial pre-GBOL projects or partnerships with the Canadian Centre for DNA Barcoding (CCDB, Guelph), with

direct data flow to and from the Barcode of Life Data Systems (BOLD - www.barcodinglife.org, www.boldsystems.org). As mentioned above, the GBOL botany section (BGBM, IEB, NEES) has developed another web-based application (www.gbol5.de), which is used for sequence assembly and sample management. It is tailored towards the specific needs of the botany section and its targeted markers, and it functions as a bridge to transfer relevant data to the GBOL web portal. Likewise, the Senckenberg Görlitz soil zoology section feeds its taxonomic, sampling and ecological data first to Edaphobase (www.edaphobase.org; Burkhardt et al. 2014), a free online database on soil animals, and subsequently links them to GBOL and BOLD. Together with the BOLD team we are currently adapting the data exchange with the BOLD Web Service Interface (API). The GBOL data scheme complies to the international ABCD-GGBN standard, and using a BioCASE wrapper (www.biocase.org) all metadata, sequences and trace files are provided at a single access point for BOLD. For the data flow from BOLD to GBOL (e.g., ZSM data), the data are harvested at the moment via the BOLD API, specifically querying shared project containers to exclude, for example, non-barcode compliant entries from GenBank.

Additionally, DNA samples are physically deposited in biobanks. GBOL is therefore 1) contributing to the development of collection infrastructures, which support future research and 2) allowing the addition of sequence and multilocus-barcode information to well-documented samples already in the system. By providing access to samples for the scientific community, GBOL further integrates into other biodiversity and taxonomy research programs. The German DNA bank network with its associated institutions (Gemeinholzer et al. 2011), which are all GBOL partners, provides the infrastructure and is integrated at global level into the GGBN Data Portal (Droege et al. 2014).

In January 2015, the GBOL team invited European barcoding groups (incl. ECBOL members) to a workshop in Bonn in order to explore options for more intense cooperation and synergies between the national initiatives. All participants agreed that better trans-boundary cooperation between neighboring nations would both accelerate database completion and increase cost-efficiency. The federalist structure of Europe is not only a key

challenge in politics, but also for the goal of a global inventory of biodiversity. Naturally, species' ranges do not coincide with political borders, and great efforts might be needed to encounter certain taxa, which are only occasionally reported from a given country but are common in other countries. If more national barcoding campaigns would deposit data in repositories with user management that allow joint access to also unpublished data and validated materials, resources could be allocated more efficiently. A common issue when applying for national grants for DNA barcoding campaigns is certainly the possibility of losing politicians' or reviewers' benevolence if the generated data should only be deposited in a 'foreign' data base only, i.e. not situated in the funding nation. In GBOL we follow two strategies, with one institutional node directly cooperating and using all services of the Canadian Centre for DNA Barcoding (CCDB), and the other nodes applying a newly developed work-flow depositing all LIMS and FIMS data into an own system, which exchanges data with BOLD.

DNA barcoding applications

Currently, three different application studies are being conducted within GBOL, either comparing cost efficiency and accuracy of molecular vs. traditional, morphological identification or evaluating the performance of NGS techniques in combination with DNA barcoding in order to process (mixed) environmental samples. Although the results have not been completely evaluated yet, we found a much higher taxonomic resolution at comparable costs when analyzing macrozoobenthic invertebrate samples following the standard DNA barcoding approach (not metabarcoding). Of the approx. 10,000 benthic invertebrates from an EU Water Framework Directive monitoring 53% were identified to species level by a renowned environmental agency based on morphology and following standard procedures for water quality assessments. DNA barcodes confirmed the identity of 73% of the morphologically identified individuals, delivered an incongruent identification in 8%, and 15% of the un-identified specimens could be annotated with a species name by comparison with the reference library. Only 3% of the benthic invertebrates examined could not be identified

by either approach. At the moment we are evaluating whether and how the additional taxonomic information content would affect water quality assessment.

Our first results applying mass sequencing to mixed (environmental) samples from National Park Bayerischer Wald indicate 1) that the technology is well suited to extract species-level information via DNA barcodes, but 2) that the current species coverage in the database allows the unambiguous identification of only 27% of the sequences (from over 25,000 DNA barcodes of several malaise traps), whereas the remaining 73% of successfully amplified COI sequences could be assigned to order, family, or genus level by the automated BIN system via reverse identification of the Barcode of Life database.. The 27% of the data with species name hits belong to almost 7,000 sequences and were assigned to 1,041 different insect species; a great proportion of the non-assigned, remaining sequences belong to Diptera and Hymenoptera species not yet represented in the reference database.

As part of the Global Malaise Program (GMP) of the Biodiversity Institute of Ontario (BIO) and over thirty international contributors, Germany contributed two Malaise traps collecting over a full season each. The GMP involves the standardized large-scale trapping of arthropods across the globe using Malaise traps to obtain tissue material and subsequently determine species diversity using DNA barcoding. One trap was installed in southern Germany in the National Park Bayerischer Wald (conifer-dominated mountain forest), the other in central-western Germany at the Landskrone (Part of the so called Ahr hills with deciduous coppice). Sorting and DNA barcoding of 37,275 specimens from May till September resulted in 5,388 different BINs (species hypothesis, cf. Ratnasingham and Hebert 2013) of which 35% could already be annotated with species-level taxonomy, 7% with genus names, and 26% identified to family level using the BOLD BIN system. The findings illustrate that the main challenge is still to complete the reference database for the German Fauna and Flora. Without this basis most applications will create incomplete species lists and thus unsatisfying results in areas where it is important to assess the presence or absence of certain key taxa or indicator species. However, we embrace the merits of widespread barcoding without *a priori* morphological identification—which is in our view mandatory for the

construction of a reliable reference library—followed by BIN (or other MOTU) analysis, as it is done in the GMP initiative. We explicitly invite the experts to conduct studies—not necessarily only taxonomic works—using the unnamed entities, which would enhance the reference library when names are given to them. As the voucher specimens are available from the GBOL collections and many are already sent out for identification in certain groups, specialists can use both, morphological and molecular data to come to sound conclusions in an integrative context applying the evolutionary species concept. In the meanwhile, the well documented and publically available molecular entities (BINs) allow for the detection of biodiversity shifts over time and site comparisons, possibly much faster than with traditional methods (e.g., Telfer et al. 2015; Wirta et al. 2015).

Outlook GBOL phase II

We argue that funding the creation of reference databases and reference collections is a sustainable investment throughout, as it is required only once for a region and for the existing species. All generated and published data can be used indefinitely and remain valid as long as populations do not drastically change and the respective taxonomic classifications persist (if not, this is where the reference collections come in!). Our prime goals for a continued GBOL project are therefore to extend the DNA barcode reference database and collections, especially to contain all frequent species, as well as important agricultural and forest 'pests', invasive, health-relevant, Red List, FFH, indicator and application-relevant species. Access to these data, together with innovations in the sector of DNA barcoding and mass sample analysis, will give basic and applied science and natural resource management a potent instrument and enables a multitude of economically and socially useful products. The first funding period proved that GBOL is efficient, and the consortium will be able to pursue its goals in a second period of funding just approved.

Acknowledgments

The GBOL project is generously supported by a grant from the German Federal Ministry of Education and Research (FKZ 01LI1101). We thank our numerous colleagues at the involved museums and universities for their time and expertise and are highly indebted to all external taxon experts for collecting and determining tens of thousands of specimens. Claudia Steinke and Sujeevan Ratnasingham with the whole BIO team are thanked for assistance putting GBOL data onto the BOLD platform. We thank our laboratory, technical and IT staff for their valuable work, especially Bruno Cancian de Araujo, Mei-Yu Chen, Dieter Doczkal, Claudia Etzbauer, Helga Fleischer-Notter, Bettina Giesicke, Athena Lam, Claudia Schütte, Jana Thormann and Laura von der Mark. Ludwig Beenken, Wolfgang Maier, Ali Tahir and Marco Thines are thanked for supporting the GBOL fungi section. We also acknowledge all student helpers for labelling, sorting and imaging of specimens.

Draft

References:

- Balvanera, P., Siddique, I., Dee, L., Paquette, A., Isbell, F., Gonzalez, A., Byrnes, J., O'Connor, M.I., Hungate, B.A., and Griffin, J.N. 2014. Linking Biodiversity and Ecosystem Services: Current Uncertainties and the Necessary Next Steps. *Bioscience* 64(1): 49–57. doi:10.1093/biosci/bit003.
- Birk, S., Bonne, W., Borja, A., Brucet, S., Courrat, A., Poikane, S., Solimini, A., van de Bund, W., Zampoukas, N., and Hering, D. 2012. Three hundred ways to assess Europe's surface waters: An almost complete overview of biological methods to implement the Water Framework Directive. *Ecol. Indic.* 18: 31–41. doi.org/10.1016/j.ecolind.2011.10.009.
- Borsch, T., Hernández-Ledesma, P., Berendsohn, W.G., Flores-Olvera, H., Ochoterena, H., Zuloaga, F.O., von Mering, S., and Kilian, N. 2015. An integrative and dynamic approach for monographing species-rich plant groups – Building the global synthesis of the angiosperm order Caryophyllales. *Perspect. Plant Ecol. Evol. Syst.* 17(4): 284–300. Available from <http://linkinghub.elsevier.com/retrieve/pii/S1433831915000451>.
- Bundesamt für Naturschutz. 2015. Artenschutz-Report 2015, Available from https://www.bfn.de/fileadmin/BfN/presse/2015/Dokumente/Artenschutzreport_Download.pdf [accessed 27 September 2015].
- Bundesamt für Naturschutz. 2012. Daten zur Natur 2012, Münster: Landwirtschaftsvlg Münster.
- Burkhardt, U., Russell, D.J., Decker, P., Döhler, M., Höfer, H., Lesch, S., Rick, S., Römbke, J., Trog, C., Vorwald, J., Wurst, E., and Xylander, W.E.R. 2014. The Edaphobase project of GBIF-Germany-A new online soil-zoological data warehouse. *Appl. Soil Ecol.* 83: 3–12. doi.org/10.1016/j.apsoil.2014.03.021.
- Buttler, K.P., and Hand, R. 2008. Liste der Gefäßpflanzen Deutschlands. 107: 1–107.

- Collins, R.A., and Cruickshank, R.H. 2013. The seven deadly sins of DNA barcoding. *Mol. Ecol. Resour.* 13(6): 969–75. Available from <http://www.ncbi.nlm.nih.gov/pubmed/23280099> [accessed 27 September 2013].
- Costello, M.J., Houlding, B., and Joppa, L.N. 2014. Further evidence of more taxonomists discovering new species, and that most species have been named: response to Bebbier et al. (2014). *New Phytol.* 202: 739–740.
- Costello, M.J., May, R.M., and Stork, N.E. 2013. Can We Name Earth's Species Before They Go Extinct? *Science* (80-.). 339(6118): 413–416. doi/10.1126/science.1230318.
- Costello, M.J., Wilson, S., and Houlding, B. 2013. More taxonomists describing significantly fewer species per unit effort may indicate that most species have been discovered. *Syst. Biol.* 62(4): 616–624.
- Crozier, R.H. 1997. Preserving the information content of species: Genetic Diversity, Phylogeny, and Conservation Worth. *Annu. Rev. Ecol. Syst.* 28(1): 243–268.
- Droege, G., Barker, K., Astrin, J.J., Bartels, P., Butler, C., Cantrill, D., Coddington, J., Forest, F., Gemeinholzer, B., Hobern, D., Mackenzie-Dodds, J., O Tuama, E., Petersen, G., Sanjur, O., Schindel, D., and Seberg, O. 2014. The Global Genome Biodiversity Network (GGBN) Data Portal. *Nucleic Acids Res.* 42(D1): D607–D612. doi/10.1093/nar/gkt928.
- Federal Ministry for the Environment, N.C. and N.S. (BMU). 2007. National Strategy on Biological Diversity, Berlin.
- Fontaine, B., van Achterberg, K., Alonso-Zarazaga, M.A., Araujo, R., Asche, M., Aspöck, H., Aspöck, U., Audisio, P., Aukema, B., Bailly, N., Balsamo, M., Bouchet, P., et al. 2012. New Species in the Old World: Europe as a Frontier in Biodiversity Exploration, a Test Bed for 21st Century Taxonomy edited by Schierwater B. *PLoS One* 7(5): e36881. Available from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3359328&tool=pmcentrez&re>

ndertype=abstract [accessed 25 October 2012].

- Frobel, K., and Schlumprecht, H. 2014. Erosion der Artenkenner. Abschlussbericht im Auftrag des BUND Naturschutz in Bayern e.V., Nürnberg.
- Gemeinholzer, B., Dröge, G., Zetzsche, H., Haszprunar, G., Klenk, H.-P., Güntsch, A., Berendsohn, W.G., and Wägele, J.-W. 2011. The DNA Bank Network: The start from a German initiative. *Biopreserv. Biobank*. 9(1): 51–55.
- Gerlach, J., Samways, M., and Pryke, J. 2013. Terrestrial invertebrates as bioindicators: An overview of available taxonomic groups. *J. Insect Conserv.* 17(4): 831–850.
- Hausmann, A., Godfray, H.C.J., Huemer, P., Mutanen, M., Rougerie, R., van Nieukerken, E.J., Ratnasingham, S., and Hebert, P.D.N. 2013. Genetic Patterns in European Geometrid Moths Revealed by the Barcode Index Number (BIN) System. *PLoS One* 8(12): e84518. Available from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3866169&tool=pmcentrez&ndertype=abstract> [accessed 22 January 2014].
- Hausmann, A., Haszprunar, G., and Hebert, P.D.N. 2011. DNA Barcoding the Geometrid Fauna of Bavaria (Lepidoptera): Successes, Surprises, and Questions. *PLoS One* 6(2): 1–9.
- Hausmann, A., Haszprunar, G., Segerer, A.H., Speidel, W., Behounek, G., and Hebert, P.D.N. 2011. Now DNA-barcoded: the butterflies and larger moths of Germany. *Spixiana* 34(1): 47–58.
- Hawltischek, O., Morinière, J., Dunz, A.R., Franzen, M., Rödder, D., Glaw, F., and Haszprunar, G. 2015. Comprehensive DNA barcoding of the herpetofauna of Germany. *Mol. Ecol. Resour.* doi.wiley.com/10.1111/1755-0998.12416.
- Hebert, P.D.N., Cywinska, A., Ball, S.L., and DeWaard, J.R. 2003. Biological identifications through DNA barcodes. *Proc. R. Soc. B Biol. Sci.* 270(1512): 313–321. Available from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1691236&tool=pmcentrez&ndertype=abstract>

ndertype=abstract [accessed 27 October 2012].

Hebert, P.D.N., Ratnasingham, S., and DeWaard, J.R. 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc. R. Soc. B Biol. Sci.* 270(Suppl1): S96–S99. Available from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1698023&tool=pmcentrez&rendertype=abstract> [accessed 28 January 2013].

Hendrich, L., Morinière, J., Haszprunar, G., Hebert, P.D.N., Hausmann, A., Köhler, F., and Balke, M. 2015. A comprehensive DNA barcode database for Central European beetles with a focus on Germany: adding more than 3500 identified species to BOLD. *Mol. Ecol. Resour.* 15(4): 795–818. doi.wiley.com/10.1111/1755-0998.12354.

Ji, Y., Ashton, L., Pedley, S.M., Edwards, D.P., Tang, Y., Nakamura, A., Kitching, R., Dolman, P.M., Woodcock, P., Edwards, F.A., Larsen, T.H., Hsu, W.W., Benedick, S., Hamer, K.C., Wilcove, D.S., Bruce, C., Wang, X., Levi, T., Lott, M., Emerson, B.C., and Yu, D.W. 2013. Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding edited by Holyoak M. *Ecol. Lett.* 16(10): 1245–1257. doi.wiley.com/10.1111/ele.12162 [accessed 19 September 2013].

Keith, D., Mahony, M., Hines, H., Elith, J., Regan, T.J., Baumgartner, J.B., Hunter, D., Heard, G.W., Mitchell, N.J., Parris, K.M., Penman, T., Scheele, B., Simpson, C.C., Tingley, R., Tracy, C.R., West, M., and Akçakaya, H.R. 2014. Detecting extinction risk from climate change by IUCN Red List criteria. *Conserv. Biol.* 28(3): 810–9. doi.wiley.com/10.1111/cobi.12234.

Khuroo, A.A., Dar, G.H., Khan, Z.S., and Malik, A.H. 2007. Exploring an inherent interface between taxonomy and biodiversity: Current problems and future challenges. *J. Nat. Conserv.* 15(4): 256–261. Available from <http://linkinghub.elsevier.com/retrieve/pii/S1617138107000222>.

Klausnitzer, B., and Köhler, F. 1998. Verzeichnis der Käfer Deutschlands, edited by Klausnitzer B., and Köhler F. Dresden: Entomologische Nachrichten und Berichte.

- Kneibelsberger, T., Dunz, A.R., Neumann, D., and Geiger, M.F. 2015. Molecular diversity of Germany's freshwater fishes and lampreys assessed by DNA barcoding. *Mol. Ecol. Resour.* 15(3): 562–572. doi.wiley.com/10.1111/1755-0998.12322.
- Korotkova, N., Borsch, T., Quandt, D., Taylor, N.P., Muller, K.F., and Barthlott, W. 2011. What does it take to resolve relationships and to identify species with molecular markers? An example from the epiphytic Rhipsalideae (Cactaceae). *Am. J. Bot.* 98(9): 1549–1572. Available from <http://www.amjbot.org/cgi/doi/10.3732/ajb.1000502>.
- Macher, J.N., Salis, R.K., Blakemore, K.S., Tollrian, R., Matthaei, C.D., and Leese, F. 2015. Multiple-stressor effects on stream invertebrates: DNA barcoding reveals contrasting responses of cryptic mayfly species. *Ecol. Indic.* doi.org/10.1016/j.ecolind.2015.08.024.
- Mora, C., Tittensor, D.P., Adl, S., Simpson, A.G.B., and Worm, B. 2011. How many species are there on Earth and in the ocean? *PLoS Biol.* 9(8): e1001127. Available from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3160336&tool=pmcentrez&rendertype=abstract> [accessed 25 October 2012].
- Morinière, J., Hendrich, L., Hausmann, A., Hebert, P., Haszprunar, G., and Gruppe, A. 2014. Barcoding Fauna Bavarica: 78% of the Neuropterida fauna barcoded! *PLoS One* 9(10): e109719. Available from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4186837&tool=pmcentrez&rendertype=abstract> [accessed 20 December 2014].
- Pimm, S.L., Jenkins, C.N., Abell, R., Brooks, T.M., Gittleman, J.L., Joppa, L.N., Raven, P.H., Roberts, C.M., and Sexton, J.O. 2014. The biodiversity of species and their rates of extinction, distribution, and protection. *Science* (80-.). 344(6187): 1246752–8. doi/10.1126/science.1246752.
- Porco, D., Potapov, M., Bedos, A., Busmachiu, G., Weiner, W.M., Hamra-Kroua, S., and Deharveng, L. 2012. Cryptic Diversity in the Ubiquist Species *Parisotoma notabilis* (Collembola, Isotomidae): A Long-Used Chimeric Species? *PLoS One* 7(9).

- Ratnasingham, S., and Hebert, P.D.N. 2013. A DNA-Based Registry for All Animal Species: The Barcode Index Number (BIN) System edited by Fontaneto D. PLoS One 8(7): e66213. Available from <http://dx.plos.org/10.1371/journal.pone.0066213>.
- Raupach, M.J., Hendrich, L., Kchler, S.M., Deister, F., Morinire, J., and Gossner, M.M. 2014. Building-Up of a DNA Barcode Library for True Bugs (Insecta: Hemiptera: Heteroptera) of Germany Reveals Taxonomic Uncertainties and Surprises. PLoS One 9(9): e106940. Available from <http://www.ncbi.nlm.nih.gov/pubmed/25203616> [accessed 10 September 2014].
- Ristau, K., Steinfartz, S., and Traunspurger, W. 2013. First evidence of cryptic species diversity and significant population structure in a widespread freshwater nematode morphospecies (*Tobrilus gracilis*). Mol. Ecol. 22(17): 4562–4575. doi.wiley.com/10.1111/mec.12414.
- Sala, O.E., Chapin, F.S., Armesto, J.J., Berlow, E., Bloomfield, J., Dirzo, R., Huber-Sanwald, E., Huenneke, L.F., Jackson, R.B., Kinzig, A., Leemans, R., Lodge, D.M., Mooney, H.A., Oesterheld, M., Poff, N.L., Sykes, M.T., Walker, B.H., Walker, M., and Wall, D.H. 2000. Global biodiversity scenarios for the year 2100. Science 287(5459): 1770–4. doi/10.1126/science.287.5459.1770.
- Scheffers, B.R., Joppa, L.N., Pimm, S.L., and Laurance, W.F. 2012. What we know and don't know about Earth's missing biodiversity. Trends Ecol. Evol. 27(9): 501–510. doi.org/10.1016/j.tree.2012.05.008.
- Schmitt, T. 2007. Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. Front. Zool. 4(1): 11. Available from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1868914&tool=pmcentrez&rendertype=abstract> [accessed 9 January 2014].
- Schoch, C.L., Seifert, K. a., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C. a., Chen, W., Bolchacova, E., Voigt, K., Crous, P.W., Miller, a. N., Schindel, D., et al. 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode

- marker for Fungi. *Proc. Natl. Acad. Sci.* 109(16): 6241–6246.
doi/10.1073/pnas.1117018109.
- Schulz, H.J., Bretfeld, G., and Zimdars, B. 2003. Verzeichnis der Springschwänze (Collembola) Deutschlands. *Entomofauna Ger.* 6: 11–25.
- Stuart, S.N., Wilson, E.O., McNeely, J.A., Mittermeier, R.A., and Rodriguez, J.P. 2010. The Barometer of Life. *Science* (80-.). 328(5975): 177. Available from <http://www.sciencemag.org/cgi/doi/10.1126/science.1188606>.
- Taberlet, P., Fumagalli, L., Wust-Saucy, A.-G., and Cosson, J.-F. 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Mol. Ecol.* 7(4): 453–464.
doi/abs/10.1046/j.1365-294x.1998.00289.x.
- Telfer, A.C., Young, M.R., Quinn, J., Perez, K., Sobel, C.N., Sones, J.E., Levesque-beaudin, V., Derbyshire, R., Fernandez-triana, J., Rougerie, R., Thevanayagam, A., Hebert, P.D.N., et al. 2015. Biodiversity inventories in high gear : DNA barcoding facilitates a rapid biotic survey of a temperate nature reserve. *Biodiversity Data Journal* 3 (August).
doi: 10.3897/BDJ.3.e6313.
- Tscharntke, T., Klein, A.M., Kruess, A., Steffan-Dewenter, I., and Thies, C. 2005. Landscape perspectives on agricultural intensification and biodiversity - Ecosystem service management. *Ecol. Lett.* 8(8): 857–874.
- Wesener, T., Voigtländer, K., Decker, P., Oeyen, J.P., Spelda, J., and Lindner, N. 2015. First results of the German Barcode of Life (GBOL) – Myriapoda project: Cryptic lineages in German *Stenotaenia linearis* (Koch, 1835) (Chilopoda, Geophilomorpha). *Zookeys* 510: 15–29. Available from <http://zookeys.pensoft.net/articles.php?id=5514>.
- Wirta, H., Várkonyi, G., Rasmussen, C., Kaartinen, R., Schmidt, N.M., Hebert, P.D.N., Barták, M., Blagoev, G., Disney, H., Ertl, S., Gjelstrup, P., Gwiazdowicz, D.J., Huldén, L., Ilmonen, J., Jakovlev, J., Jaschhof, M., Kahanpää, J., Kankaanpää, T., Krogh, P.H., Labbee, R., Lettner, C., Michelsen, V., Nielsen, S.A., Nielsen, T.R., Paasivirta, L.,

Pedersen, S., Pohjoismäki, J., Salmela, J., Vilkamaa, P., Väre, H., von Tschirnhaus, M., and Roslin, T. 2015. Establishing a community-wide DNA barcode library as a new tool for arctic research. *Mol. Ecol. Resour.* doi: 10.1111/1755-0998.12489.

Draft

Figure captions:

Figure 1 General data infrastructure set up at ZFMK as FIMS and LIMS system.

Draft

Tables

Table 1: GBOL core institutes by taxonomic focus and node number.

Zoology
SMNK, Staatliches Museum für Naturkunde Karlsruhe (III)
SMNS, Staatliches Museum für Naturkunde Stuttgart (III)
ZSM, Zoologische Staatssammlung München (II)
ZFMK, Zoologisches Forschungsmuseum A. Koenig, Bonn (I, central coordination)
Botany and Mycology
BGBM, Botanischer Garten und Botanisches Museum Berlin, Freie Universität Berlin (V)
SMNK, Staatliches Museum für Naturkunde Karlsruhe (V)
SMNS, Staatliches Museum für Naturkunde Stuttgart (V)
NEES, Nees-Institut, Universität Bonn (V)
IEB, Universität Münster (V)
Soil organisms
SMNG, Senckenberg Museum of Natural History Görlitz (IV)
UB, University Bielefeld (IV)

Draft

Table 2: DNA barcode progress by GBOL nodes and higher taxonomic rank as of late 2015.

node	taxa	barcodes	species
I	Vertebrata	2,430	294
	Arthropoda varia	18,697	2,703
	Diptera	19,958	2,298
	Coleoptera	15,451	3,047
	Total GBOL I:	56,176	8,342
II	various groups	3,747	839
	Lepidoptera	9,866	2,617
	Hymenoptera	16,109	2,557
	Coleoptera	11,173	3,225
	application studies	35,496	3,256
	Total GBOL II:	76,391	12,494
III	Chalcidoidea	2,439	400
	Arachnida	3,771	663
	Pucciniales	256	188
	Total GBOL III:	6,466	1,251
IV	Collembola	461	84
	Gamasina	206	29
	Oribatida	247	46
	Nematoda terrestrial	253	48
	Nematoda limnic	556	49
	Total GBOL IV:	1,723	256
V	Angiosperms	1425	728
	ferns	281	84
	bryophytes	800	310
	Total GBOL V:	2,506	1,122

General data infrastructure set up at ZFMK as FIMS and LIMS system.
520x525mm (100 x 100 DPI)