D N A  L i b r a r y  o f  L i f e

Toward the DNA Library of Life

Line LE GALL 1, Frédéric DELSUC 2, Stéphane HOURDEZ 3, Guillaume LECOINTRE 4 & Jean-Yves RASPLUS 5

2 Institut des Sciences de l’Évolution, UMR 5554, CNRS, IRD, EPHE, Université de Montpellier, Montpellier, France.
3 Station Biologique de Roscoff, CNRS – UPMC, UMR 7144, Adaptation et Biologie des Invertebrés en Conditions Extrêmes, 29680 Roscoff, France.
5 INRA, UMR 1062, CBGP Centre de Biologie pour la Gestion des Populations, 34988 Montferrier-sur-Lez, France.

Email: 1 legall@mnhn.fr 2 Frederic.Delsuc@umontpellier.fr 3 hourdez@sb-roscoff.fr 4 guillaume.lecointre@mnhn.fr 5 jean-yves.rasplus@inra.fr

Abstract. The special set of papers entitled “DNA Library of Life” constitutes an outcome of the project “Bibliothèque du vivant” (BdV), which aims to promote the molecular taxonomy of eukaryotes by offering research teams the possibility to produce and manage a molecular library linked with specimens deposited in natural history museums. The project was funded by three French institutions (the CNRS, INRA and MNHN), and provided access to the sequencing power offered by the Genoscope for 105 teams between 2011 and 2013. It was subsequently supported by the CNRS through the “Groupement de Recherche Génomique Environnementale”. The scientific objectives of this programme were threefold: 1) species delimitation among species complexes; 2) phylogenetic reconstruction (including phylogenomics); and 3) metabarcoding and improving NGS methods for systematic purposes. Within the present collection, 19 papers contribute to these objectives across a large taxonomic range and a worldwide geographic coverage. These papers propose taxonomic novelties (22 new species and 3 new genera) in both animal and plant taxa.

Keywords. Census of Life, eukaryotic biodiversity, molecular assisted systematics, Museum life, species description.

A brief history of the DNA Library of Life

Long before the discovery of DNA (Watson & Crick 1953), humans had already started to distinguish, name and classify organisms. The first inventories could be found within the written texts left by ancient Greek philosophers, especially Aristotle and his pupil Theophrastus (4th–3rd century AD) who greatly contributed to organizing a census of existing species, classifying them into “animal” for organisms with the ability to move and “vegetal” for “unanimated” ones (for an in-depth review, see Mayr 1982). While the usage of some organisms (especially plants) had been well documented in the Middle Ages, almost no progress was made on the classification of organisms before the Renaissance and its technological advents. The construction of the microscope by A. van Leeuwenhoek (Hamarneh 1960) fostered the observation of tiny organisms and unveiled a previously invisible world that accommodates a wealth of species. A formal system of naming species was subsequently proposed by Carl von Linné in the *Species Plantarum* (Linnaeus 1753). Swiftly, the binomial system gained recognition in the scientific world and almost two million species have since been described, which nevertheless may represent only half (Costello *et al.* 2013) – or even as little as one fifth (Mora *et al.* 2011) – of the overall eukaryotic diversity on Earth. Whereas only morpho-anatomical characters were initially used to describe species, more recently the use of molecular characters has allowed biologists to estimate phylogenetic relationships more reliably and to discriminate between morphologically cryptic species. In providing a comprehensive phylogenetic framework, molecular systematics has enhanced our understanding of species diversity by unravelling cryptic diversity, morphological convergence, heteromorphic life cycles, sexual dimorphism and other evolutionary phenomena.

Molecular data consist of universal and quantifiable characters, which present the advantage of being both objective and abundant compared to morphological ones. The comparison of homologous characters facilitates the reconstruction of historical relationships among biological entities, as well as the estimation of their divergence times. Even if molecular characters are not completely devoid of pitfalls, especially when inferring species relationships from gene genealogies (e.g., Collins & Cruickshank 2012), they have proven to be amenable in establishing libraries of life for prokaryotes and eukaryotes (e.g., Hebert *et al.* 2003; Blaxter 2004; Quast *et al.* 2013), albeit with a few limitations.

Identifying what level of molecular difference defines a species is also a challenging task, as it refers to the debated plethora of species concepts and ultimately relates to underlying speciation mechanisms (Roux *et al.* 2016). Nonetheless, several methods for species delimitation using DNA barcoding data have been proposed (Pons *et al.* 2006; Puillandre *et al.* 2012; Zhang *et al.* 2013). Moreover, the addition of novel data has in some cases challenged previous species delimitations, highlighting the need for automated and time-proficient algorithms to delimit species (Ratnasingham & Hebert 2013), ideally relying on the analysis of multiple loci in a coalescent framework (Yang & Rannala 2014). Despite the apparent simplicity of using DNA sequences to build a Library of Life, the number of species collected in the core database is far from representative of the diversity on Earth. For instance, on the GenBank taxonomy page (Anonymous 2017), it is stated that all the organisms in the public sequence databases currently represent no more than 10% of the total number of described species.

In the context of the on-going 6th global biodiversity crisis (mostly referred to as a mass extinction), the major challenge that taxonomists are currently facing is to describe species diversity before it actually becomes extinct. Whereas some taxa and some regions on Earth have been thoroughly studied (i.e., more than 70% of plants are presumed to have been described already), the biodiversity of certain communities (e.g., soil) and some of the smallest organisms (e.g., microbial eukaryotes (de Vargas *et al.* 2015), fungi (Hawksworth 2001) and small inconspicuous animals (Appeltans *et al.* 2012)) remain poorly known, rendering the conservation status of those species impossible to assess (Régnier *et al.* 2015). DNA-assisted taxonomy is often seen as the holy grail in accelerated species discovery and description, as molecular data are abundant and amenable to quantitative statistical treatment. For instance, in the
marine realm, assuming that the current rhythm of species description is maintained, most of the species should be discovered before the end of this century (Appeltans et al. 2012). Nevertheless, the time-lapse between species discovery and species description is still 21 years on average (Fontaine et al. 2012), which reminds us that the process leading to a formal description is long and leaves a gap between uncovered species and named species (Yahr et al. 2016). The objective of this special set of papers in the European Journal of Taxonomy is to make a contribution, no matter how slim, toward the construction of a comprehensive DNA Library of Life.

Contributions in this special set of papers to the DNA Library of Life

This special collection of articles is comprised of the present introduction, 18 research papers spanning the fields of biogeography, phylogeny and molecular systematics, as well as an opinion paper on the pitfalls in supermatrix phylogenomics (Philippe et al. in press). The geographic coverage of the sampling conducted for the 18 research papers spans the entire planet and includes organisms from marine (Castelin et al. in press a, Castelin et al. in press b, Fedosov et al. in press, Galindo et al. in press, Manghisi et al. in press, Rousseau et al. in press, Sabroux et al. in press) and terrestrial biomes (Azofeifa-Bolaños et al. in press, Galkowski et al. in press, Legendre et al. in press, Le Ru et al. in press, Ohler & Nicolas in press, Prigent et al. in press, Rabeau et al. in press, Ramage et al. in press, Soldati et al. in press, Tu et al. in press, Veron et al. in press). It is noteworthy that the freshwater biome is not covered by any of these studies, except for the paper by Ohler & Nicolas (in press) focusing on frogs that occupy freshwater habitats for at least part of their life cycle.

Interestingly, although the funding intended for this project originates from French institutions, only a single study (Galkowski et al. in press) is limited to metropolitan France. Most of the sampling for this project was performed in tropical areas, including some overseas French territories (Fedosov et al. in press, Rousseau et al. in press, Ramage et al. in press, Castelin et al. in press b, Sabroux et al. in press, Legendre et al. in press). Last but not least, the majority of the papers deal with animals; the remainder deals with plants. There is no contribution on fungi or on unicellular eukaryote lineages.

Within this special set of papers for the DNA Library of Life, ten papers propose taxonomic novelties, consisting of 22 new species (Fedosov et al. in press, Le Ru et al. in press, Tu et al. in press, Galindo et al. in press, Soldati et al. in press, Azofeifa-Bolaños et al. in press, Galkowski et al. in press) and three new genera (Rousseau et al. in press, Rabeau et al. in press, Sabroux et al. in press).

The European Journal of Taxonomy “Library of Life” initiative includes the following papers (in order of publication, which will take place from the 30th of January to the 3rd of March, 2017):

• Le Gall L., Delsuc F., Hourdez S., Lecointre G. & Rasplus J.-Y. Toward the DNA Library of Life.
• Fedosov A.E., Stahlschmidt P., Puillandre N., Aznar-Corman L. & Bouchet P. Not all spotted cats are leopards: evidence for a Hemilienardia ocellata species complex (Gastropoda: Conoidea: Raphitomidae).
• Rousseau F., Gey D., Kurihara A., Maggs C.A., Martin-Lescanne J., Payri C., Reviers B. de, Sherwood A.R. & Le Gall L. Molecular phylogenies support taxonomic revision of three species of Laurencia (Rhodomelaceae, Rhodophyta), with the description of a new genus.
• Le Ru B., Capdevielle-Dulac C., Musyoka B.K., Pallangyo B., Njaku M., Mubenga O., Chipabika G., Ndemah R., Bani G., Molo R. Ong’amo G. & Kergoat G.J. Phylogenetic analysis and systematics of
the *Acrapex unicolora* Hampson species complex (Lepidoptera, Noctuidae, Noctuinae, Apameini), with the description of five new species from the Afrotropics.

- **Ohler A. & Nicolas V.** Which frog’s legs do *froggies* eat? The use of DNA barcoding for identification of deep frozen frog legs (Dicroglossidae, Amphibia) commercialized in France.


- **Tu V.T., Csorba G., Ruedi M., Furey N.M., Son N.T., Thong V.D., Bonillo C. & Hassanin A.** Comparative phylogeography of bamboo bats of the genus *Tylonycteris* (Chiroptera, Vespertilionidae) in Southeast Asia.


- **Soldati L., Condamine F.L., Clamens A.-L. & Kergoat G.J.** Documenting tenebrionid diversity: progress on *Blaps* Fabricius (Coleoptera, Tenebrionidae, Tenebrioninae, Blaptini) systematics, with the description of five new species.

- **Philippe H., de Vienne D.M., Ranwez V., Roure B., Baurain D. & Delsuc F.** Pitfalls in supermatrix phylogenomics.


- **Pringent S.R., Suwalski A. & Veuille M.** Connecting systematic and ecological studies using DNA barcoding in a population survey of Drosophilidae (Diptera) from Mt Oku (Cameroon).


- **Castelin M., de Mazancourt V., Marquet G., Zimmerman G. & Keith P.** Genetic and morphological evidence for cryptic species in *Macrobrachium australis* and resurrection of *M. ustulatum* (Crustacea, Palaemonidae).


- **Legendre F., Grandcolas P. & Thouzé F.** Molecular phylogeny of Blaberidae (Dictyoptera, Blattodea) with implications for taxonomy and evolutionary scenarios.
Perspectives toward a comprehensive DNA Library of Life

High-throughput DNA sequencing has widened the field of possibilities for automated and accelerated taxon identification (Coissac et al. 2016). This technological advancement has enhanced the use of molecular data to study functional ecology, organism interactions, community ecology and biodiversity, plus the dynamics and evolution of past and present interactions. Nevertheless, further efforts are needed to aggregate a comprehensive DNA Library of Life. The well-known taxonomic impediment is of course relevant, even if some authors have questioned it (Appeltans et al. 2012). In addition, one of the major current challenges consists of improving our sampling methods and instrumentation to discover organisms that live in extreme environments. Finally, species description is governed by the International Code for Zoological Nomenclature (ICZN 1999) and the International Code of Nomenclature for Algae, Fungi, and Plants (McNeill & International Association for Plant Taxonomy 2012), which leaves many microbial eukaryotic lineages orphaned and results in species belonging to the same phylogenetic lineage being described according to both codes (Yilmaz et al. 2014).

In order to perform effective biodiversity monitoring, upon which conservation depends, we are counting upon new technologies to further our understanding of biodiversity as a whole. This, however, will not be possible without a proper and well-curated taxonomic framework. We therefore urge our founding agencies to continue their financial support to taxonomy, with the aim of achieving a comprehensive DNA Library of Life.

Acknowledgments

We would like to express our sincere thanks to Koen Martens for accepting to host this special set of papers in the European Journal of Taxonomy. We are indebted to Charlotte Thionois and Danny Eibye-Jacobsen who took care of most of the production process. The entire editorial team of European Journal of Taxonomy has always been most helpful. This special set of papers would not have been possible without the encouragement, the scientific support and the financial contributions of the Institut Ecologie et Environnement (CNRS), the Muséum national d’Histoire naturelle, and the Institut national de Recherche agronomique. The Sanger sequencing has been performed with the operational support of the Commissariat à l’Energie Atomique at Genoscope. This is publication ISEM 2017-014 of the Institut des Sciences de l’Evolution.

References


Toward the DNA Library of Life


Ratnasingham S. & Hebert P.D.N. 2013. A DNA-based registry for all animal species: the Barcode Index Number (BIN) system. *PLoS ONE* 8: e66213. [http://dx.doi.org/10.1371/journal.pone.0066213](http://dx.doi.org/10.1371/journal.pone.0066213)


Watson J.D. & Crick F.H.C. 1953. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature* 171: 737–738. [http://dx.doi.org/10.1038/171737a0](http://dx.doi.org/10.1038/171737a0)


*Manuscript received: 12 January 2017*
*Manuscript accepted: 20 January 2017*
*Published on: 30 January 2017*
*Guest editors: Line Le Gall, Frédéric Delsuc, Stéphane Hourdez, Guillaume Lecointre and Jean-Yves Rasplus*
*Handling editor: Koen Martens*
*Desk editor: Danny Eibye-Jacobsen*

Printed versions of all papers are also deposited in the libraries of the institutes that are members of the *EJT* consortium: Muséum national d’Histoire naturelle, Paris, France; Botanic Garden Meise, Belgium; Royal Museum for Central Africa, Tervuren, Belgium; Natural History Museum, London, United Kingdom; Royal Belgian Institute of Natural Sciences, Brussels, Belgium; Natural History Museum of Denmark, Copenhagen, Denmark; Naturalis Biodiversity Center, Leiden, the Netherlands.