UK Barcode of Life: 2024 project

DEFRA Centre of Excellence for DNA Methods

October 2024

Natural England Commissioned Report NECR572

www.gov.uk/natural-england

About Natural England

Natural England is here to secure a healthy natural environment for people to enjoy, where wildlife is protected and England's traditional landscapes are safeguarded for future generations.

Further Information

This report can be downloaded from the Natural England Access to Evidence Catalogue. For information on Natural England publications or if you require an alternative format, please contact the Natural England Enquiry Service on 0300 060 3900 or email enquiries@naturalengland.org.uk.

Copyright

Catalogue code: NECR572

Report details

Authors

Ben Price, Joana Cristovao, Olga Sivell, Daniel W. Hall, Jordan Beasley, Alyssa Paul, Owain Powell, Oliver W. White and William Askew

Natural England Project manager

William Askew

Contractor

The Natural History Museum, Cromwell Road, London, SW7 5BD https://nhm.ac.uk/

Keywords

DNA barcoding, genome skimming, UK DNA reference library.

Acknowledgements

The authors thank the following NHM curators for providing specimens for genome skimming: Jan Beccaloni, Max Barclay, Michael Geiser, Duncan Sivell and Sue Ryder.

Citation

Price, B., Cristovao, J., Sivell, O., Hall, D.W., Beasley, J., Paul A., Powell O., White, O.W., and Askew, W. (2024) UK Barcode of Life: 2024 project update. DEFRA Centre of Excellence for DNA Methods. *Natural England Commissioned Report* NECR572

Foreword

Natural England is part of the Defra DNA Centre of Excellence, which champions the uptake of DNA based tools for monitoring the environment to inform its management and regulation. Natural England commissioned this report to report on the progress of the UK Barcode of Life project which has received funding from the Defra DNA Centre of Excellence in 21/22 and 20/21 and Natural England in 22/23 and 23/24.

Natural England commission a range of reports from external contractors to provide evidence and advice to assist us in delivering our duties. The views in this report are those of the authors and do not necessarily represent those of Natural England or the DNA Centre of Excellence.

Executive summary

DNA-based methods offer a significant opportunity to monitor individual species and species assemblages where appropriate, for example those that may be difficult to monitor using traditional methods. However, with the exception of some individual species such as the great crested newt, there is still much development of these techniques required before they can be used in routine monitoring. Natural England has been developing the use of DNA-based methods for monitoring for several years and is a founding member of the Defra DNA Centre of Excellence, which was set up to encourage collaboration across the Defra group to progress the use of DNA based methods by tackling cross-cutting barriers.

Gaps in DNA reference libraries of UK species were identified by the Defra DNA Centre of Excellence Working Group as one of the main barriers preventing the further uptake of DNA based methods for environmental species monitoring. This report details the continuation of the UK Barcode of Life (UKBOL) project and progress in barcoding priority species. In particular, an additional 2,862 specimens of 1,515 species were collected and databased, with 1,171 specimens of 750 species sequenced and added to BOLD. The project website (including a data portal), was developed to facilitate a continuous gap analysis of species coverage.

Key points:

- There are approximately 76,000 eukaryote species recognised in the UK, the majority of which are poorly known.
- DNA barcoding uses a short, standardised segment of an organism's genome for identification by comparison to a reference library.
- A Defra funded gap analysis highlighted that almost half the known UK species lack DNA barcode data (see report: Price and others, 2020).
- To rectify this a steering group was formed to initiate a UK Barcode of Life (UKBOL) project, and begin sequencing priority species (see report: Price and others, 2022)
- This report provides an overview of progress towards three main goals (a) coordinating a steering group, (b) sequencing priority fresh/museum specimens, (c) planning future fieldwork, and (d) developing the website and data portal.
- In the past year a further 2,862 specimens of 1,515 species were databased and 1,171 specimens of 750 species were sequenced and made public on BOLD.
- The data portal was updated to combine the authoritative list of all UK species and existing data on BOLD.
- The public UKBOL dataset now includes 23,795 specimens of 7,935 species.

Contents

Introduction

There are approximately 76,000 eukaryote species recognised in the UK, however the majority are poorly known. For those few species with sufficient data, their abundance and distribution have, on average, declined since 1970 and of the 8,431 species that have been formally assessed, 15% are threatened with extinction and 133 species are already extinct in the UK (Hayhow and others, 2019).

DNA barcoding uses a short, standardized segment of an organism's genome for identification, much like the barcodes found on commercial products (Hebert and others, 2003). These DNA-based identifications require comparison to reference libraries of DNA barcodes sequenced from identified individuals, typically deposited in natural history collections. See Price and others (2020, 2022a, 2022b) for additional background information.

A gap analysis of public DNA data in major DNA reference libraries highlighted that almost half the known UK species lack DNA barcode data, and that quality assurance is variable for those with data (Price and others, 2020). The previous report recommended the formation of a steering group to initiate a UK Barcode of Life (UKBOL) project to coordinate UK barcoding and begin sequencing priority species (Price and others, 2020).

The UKBOL steering group was initiated in 2021 and represents government agencies and national repositories from the devolved nations, national sequencing centres and organisations representing citizen science networks (Price and others, 2022a). A live priority list checks the barcode of life database (BOLD) weekly to assess coverage for priority species.

The aims of the project in 2023/24 were:

1. Continue to coordinate the UKBOL steering group;
2. Identify priority species critical for Natural England monitoring;
3. Genome skim museum specimens or barcode fresh specimens;
4. Plan collaborative field work;
5. Maintain and develop a project website and online data portal.

Progress towards objectives

Objective 1: Co-ordination and management of UKBOL steering group

The steering group membership and Terms of Reference are provided in Price and others (2022a). A steering group meeting was held online in January 2024, and further meetings are planned to continue quarterly. Several members of the steering group are partners in the Biodiversity Genomics Europe project (https://biodiversitygenomics.eu/) which includes a strand developing European DNA barcode reference libraries for three groups: (a) pollinators, and species monitored in (b) freshwater and (c) marine habitats across Europe. The steering group will work closely with the BGE project to ensure synergy.

Objective 2: Priority species barcoding from existing collections

Priority species, critical for Natural England monitoring on sensitive sites such as SSSI's were identified based on assemblages within the Pantheon database (Webb and others, 2018):

- Bare Sand and Chalk inverts (F111)
- Open Short Sward Inverts (F112)
- Scrub Edge Inverts (F001)
- Bark and Sapwood Decay Inverts (A212)
- Heartwood Decay Inverts (A211)
- Fungi (C14)

These species were compared with existing data on BOLD to identify species needing additional data (e.g., species without data, with less than 3 public barcodes, or with multiple DNA clusters requiring clarification), resulting in a list of 208 species (128 Arthropoda and 80 Fungi) requiring sequence data.

The arthropods were checked for representatives in the NHM collection, with 108 species present with exemplar specimens selected for genome skimming. Genome skimming approaches were used as amplicon sequencing of older museum specimens (i.e., collected over 10-20 years ago) is error prone and prohibitively expensive due to the fragmentation of DNA over time (see Mullin and others, 2022; Price and others, 2022b). The fungal gap list was provided to Kew with samples in preparation for genome skimming. For species in hand the specimens were databased and imaged, before tissue samples (e.g., a leg) were removed for genome skimming.

DNA was extracted using a modified ancient DNA protocol (Hall and others, 2023) after which the tissue (e.g., a leg) was returned to the original voucher and 35uL of the DNA

extract was archived at -80°C. Libraries were prepared with unique dual indexing for each specimen, using a reduced volume Santa Cruz Reaction (Kapp and others, 2021). Libraries were pooled targeting 5-10M paired-end reads per specimen and sequenced on a single NovaSeq XPlus 2*150bp lane through an external provider. The barcode fragment of COI was recovered using the skim2mt pipeline (White and others, 2024). In addition to the barcode data being uploaded to BOLD, the entire genome skim sequence data for each specimen has been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB51803, adding to previous UKBOL skim data. (https://www.ebi.ac.uk/ena/browser/view/PRJEB51803).

Objective 3: Planning future field collection of priority species

The gap analysis and subsequent search of museum collections (Objective 2) left twenty arthropod species targets for the 2024 field season (Table 1). These species will be targeted in collaboration with Natural England Field Unit, national conservation schemes and conservation societies.

Table 1: Priority species for collection in 2024: Bare Sand and Chalk inverts (F111)

Phylum	Class	Species
Arthropoda	Arachnida	*Argenna subnigra* (O.Pickard-Cambridge, 1861)
Arthropoda	Arachnida	*Gnaphosa lugubris* (C.L.Koch, 1839)
Arthropoda	Arachnida	*Baryphyma maritimum* (Crocker & Parker, 1970)
Arthropoda	Arachnida	*Jacksonella falconeri* (Jackson, 1908)
Arthropoda	Arachnida	*Lessertia dentichelis* (Simon, 1884)
Arthropoda	Arachnida	*Micrargus laudatus* (O.Pickard-Cambridge, 1881)
Arthropoda	Arachnida	*Parapelecopsis nemoralioides* (O.Pickard-Cambridge, 1884)
Arthropoda	Arachnida	*Porrhomma microphthalmum* (O.Pickard-Cambridge, 1871)
Arthropoda	Arachnida	*Agroeca lusatica* (L.Koch, 1875)
Arthropoda	Arachnida	*Alopecosa barbipes* (Sundevall, 1833)
Arthropoda	Arachnida	*Arctosa perita* (Latreille, 1799)
Arthropoda	Arachnida	*Neon pictus* Kulczynski, 1891
Arthropoda	Arachnida	*Enoplognatha oelandica (Thorell, 1875)*
Arthropoda	Arachnida	*Xysticus sabulosus* (Hahn, 1832)
Arthropoda	Insecta	*Euheptaulacus villosus* (Gyllenhal, 1806)
Arthropoda	Insecta	*Trixoscelis similis* Hackman, 1970
Arthropoda	Insecta	*Anoscopus duffieldi* (Le Quesne, 1964)
Arthropoda	Insecta	*Psammotettix maritimus* (Perris, 1857)
Arthropoda	Malacostraca	*Buddelundiella cataractae* Verhoeff, 1930
Arthropoda	Insecta	*Wesmaelius balticus* (Tjeder, 1931)
Mollusca	Gastropoda	*Vitrea subrimata (*Reinhardt, 1871)
Mollusca	Gastropoda	*Pyramidula umbilicata* (Montagu, 1803)
Mollusca	Gastropoda	*Truncatellina cylindrica* (J.B.Férussac, 1807)

Table 2: Priority species for collection in 2024: Scrub Edge Inverts (F001)

Phylum	Class	Species
Arthropoda	Insecta	*Homoneura biumbrata* (Loew, 1847)

Table 3: Priority species for collection in 2024: Open Short Sward Inverts (F112)

Phylum	Class	Species
Arthropoda	Insecta	*Chlorops dasycerus* Loew, 1866
Mollusca	Gastropoda	*Truncatellina callicratis* (Scacchi, 1833)

Objective 4: Overall numbers of specimens collected.

This objective continues the barcode sequencing of priority species identified by Price and others (2020) and sequencing started by Price and others (2022a, 2022b, 2023). Natural England has estimated 700 species require barcoding to make substantial progress in providing more complete reference libraries of priority taxa. The objective for this contract was to sequence 1,000 fresh specimens with data uploaded to BOLD.

Specimens were provided via several sources, but primarily through the UK recording community, a bioblitz at Thorne and Hatfield Moors in June 2023, and Natural England and Natural History Museum staff throughout the year.

A summary of the additional specimens added to the UKBOL project in 2023-24 is provided (Table 2), with specimen details provided in Price (2024). In total 2,862 specimens of 1,515 species were collected and databased in the UKBOL project, with 1,171 specimens of 750 species sequenced and uploaded to BOLD.

Table 4 Current summary of UKBOL projects on BOLD and new records added in 2023-24.

Project	Description	Specimens added 2023-24	Specimen Total	Species added 2023-24	Species Total
ANBIO	"Ainsdale BioBlitz": Natural England and NHM BioBlitz held at Ainsdale NNR July 2019.	0	625	0	309
BURE	Bure Marshes Bioblitz, June 2022	378	378	96	96
THORN	Thorne & Hatfield Marshes NNR Bioblitz, July 2023	237	237	164	164
NMS	National Museums Scotland	0	475	0	399
UKAN	"UK Barcoding – Animals": General project for UK animal barcoding.	2,519	4,880	1,367	2,525
UKPL	"UK Barcoding – Plants": General project for UK plant barcoding.	0	50	0	27
FBUK	FreshBase. Freshwater invertebrate barcoding.	60	490	10	431
FPUK	Flowering Plants UK	0	4,750	0	1,455
TOTAL		2,922	14,783	760	5,315

In addition to the UKBOL projects in Table 2, the Darwin Tree of Life has contributed 9,096 specimens of 4,816 species to BOLD since initiation in 2020 which have been added to the UKBOL dataset, bringing the dataset total to 23,795 specimens of 7,935 species().

Samples were databased, imaged and then dissected / tissue sampled as appropriate for the taxon group (eg, leg(s) removed). In the case of very small specimens the whole body was used for DNA extraction, with the voucher recovered after the extraction was completed and subsequently stored in 80% ethanol at -20°C until further processed (eg mounted as a museum voucher).

Samples were extracted using either the KAPA Express Extraction Kit following the kit protocol, or overnight at 56°C using Lysis C (Korlevic and Lawniczak, 2023). Lysate is stored at -80°C with an aliquot diluted 1:10 with PCR water and then amplified using a one-step PCR with uniquely indexed LCO and HCO primers (Folmer and others, 1994) using the KAPA 3G PCR kit following the kit protocol. Resulting amplicons were sequenced on a GridION, using either MinION R10 or Flongle R10 flowcells (Oxford Nanopore), and processed using ONTbarcoder (Srivathsan and others, 2021).

Voucher barcode sequences were checked against BOLD using BOLDigger (Buchner and Leese, 2020) and any unexpected matches were inspected manually. Only a small proportion of unexpected matches occurred, resulting from (a) poor sample preservation leading to contaminant bacterial / fungal / human sequences being recovered, or (b) misidentified species already on BOLD. For the former additional specimens were sought and for the latter the identification was confirmed before upload to BOLD.

Objective 5: Develop the UKBOL website

The UKBOL website continues to be developed in consultation with the steering group. The site includes several static pages for general project information and the priority species checklist. The priority checklist queries the BOLD database for DNA barcode data available for each priority species, either without a geographic restriction, or restricted to specimens collected in the UK. The UKBOL data portal uses the UK species inventory (UKSI) as the taxonomic backbone and monthly public data snapshots provided by BOLD.

The existing website is built on Drupal with a MySQL database, with one server which hosts both services. As such, these tools do not have the functionality required so the site is being re-developed with new infrastructure consisting of four main parts:

- Main app
    - Python based API (Flask)
    - Frontend (Vue)
- Elasticsearch cluster(s) to contain the bulk of the data imported from other services.
- Postgres database to contain other UKBoL-specific data such as species tags.
- Queue (Celery) to manage large requests such as downloads.

These updates will allow the data portal to update weekly from public data on BOLD and are anticipated to be live from the end of May 2024.

Next Steps

As monitoring programmes begin to incorporate more DNA-based techniques to identify species, filling the gaps in UK DNA reference libraries becomes more important to the success of these programmes.

The next steps for the project are to (1) continue advertising the project with relevant organisations, including national recording schemes who may be able to contribute expertly identified specimens to the project; (2) continue to develop the data portal in response to user needs; (3) continue to deliver DNA barcode sequences of priority species; (4) secure large-scale funding to fill the remaining gaps, enabling comprehensive DNA-based biodiversity exploration and monitoring through such initiatives as the Natural Capital and Ecosystem Assessment programme; (5) incorporate several projects based in UK overseas territories; and (6) coordinate with the Biodiversity Genomics Europe (BGE) project team to ensure complementary taxon sampling and data portal development.

References

Buchner D., Leese F. (2020) BOLDigger – a Python package to identify and organise sequences with the Barcode of Life Data systems. *Metabarcoding and Metagenomics* 4: e53535. https://doi.org/10.3897/mbmg.4.53535

Folmer O., Black M., Hoeh W., Lutz R., Vrijenhoek R. (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol*, 3(5):294-9.

Kapp J.D., Green R.E., Shapiro B. (2021) A Fast and Efficient Single-Stranded Genomic Library Preparation Method Optimized for Ancient DNA. *Journal of Heredity* 112(3): 241–49. https://doi.org/10.1093/jhered/esab012.

Korlević P., Lawniczak M. (2023) SOP - Lysis C plate based DNA extraction. protocols.io https://dx.doi.org/10.17504/protocols.io.ewov1o3molr2/v1

Hayhow D.B., Eaton M.A., Stanbury A.J., Burns F, Kirby W.B., Bailey, N. (2019) The State of Nature 2019. The State of Nature partnership.

Hebert P.D., Cywinska A., Ball S.L., & Dewaard J.R. (2003) Biological identifications through DNA barcodes. *Proceedings. Biological sciences*, 270 (1512): 313–321.

Mullin V.E., Stephen W., Arce A.N., Nash W., Raine C., Notton D.G., Whiffin A., Blagderov V., Gharbi K., Hogan J., Hunter T., Irish N., Jackson S., Judd S., Watkins C., Haerty W., Ollerton J., Brace S., Gill R. J., & Barnes I. (2022). First large-scale quantification study of DNA preservation in insects from natural history collections using genome-wide sequencing. *Methods in Ecology and Evolution*, 00, 1– 12. https://doi.org/10.1111/2041-210X.13945

Price B.W., Briscoe A., Misra R., Broad G.R. (2020) DEFRA Centre of Excellence for DNA Methods: Evaluation of DNA barcode libraries used in the UK and developing an action plan to fill priority gaps. *Natural England Joint Publication* JP035. http://hdl.handle.net/10141/622857

Price B.W., Misra R., Broad G.R., Clark K. (2022a) Initiation of UK Barcode of Life and Filling Priority Gaps: DEFRA Centre of Excellence for DNA Methods. *Natural England Joint Publication* JP041.

Price B.W., Cristovao J., Hall A.C., Cuber P., Geeves C., Salatino S., White O., and others (2022b) UK Barcode of Life: 2022 Project Update. *Natural England Commissioned Report* NECR457.

Price B., Cristovao J., Sivell O., Hall A., Hall D.W., White O., Wilbraham J., Brodie J., Blagoderov V.A., Whiffin A., Cuber P., Davis R., Misra R., Leatherland D., Clark K. (2023) UK Barcode of Life: 2023 project update. DEFRA Centre of Excellence for DNA Methods. *Natural England Commissioned Report* NECR497.

Price, B. (2024). UKBOL project samples for the 2023 - 2024 reporting period [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10870001

Srivathsan A., Lee L., Katoh K., Hartop E., Kutty S.N., Wong J., Yeo D., Meier R. (2021) MinION barcodes: biodiversity discovery and identification by everyone, for everyone. bioRxiv.:2021.03.09.434692. https://doi.org/10.1101/2021.03.09.434692.

Webb J., Heaver D., Lott D., Dean H.J., Van Breda J., Curson J., Harvey M.C., Gurney M., Roy D.B., Van Breda A., Drake M., Alexander K.N.A., Foster G. (2018) Pantheon - database version 3.7.6. https://pantheon.brc.ac.uk/

White O., Hall A., Price B.W., Williams S., Clark M.D. (2024) A snakemake toolkit for the batch assembly, annotation, and phylogenetic analysis of mitochondrial genomes and ribosomal genes from genome skims of museum collections. bioRxiv: 2023.08.11.552985. https://doi.org/10.1101/2023.08.11.552985

www.gov.uk/natural-england