

Reconstructing the Holy Loch Food Web: A simple system for interpreting large numbers of Barcode of Life Database (BOLD) Biodiversity Index Numbers (BINs) for use in taxonomy, natural history and conservation, using Chironomoidea (Diptera) as a model taxon.

Neil Hammatt¹ and the Wellcome Sanger Institute BIOSCAN Collective²

¹Holy Loch Biodiversity Research Laboratory, Honeysuckle Cottage, Clachaig, Dunoon PA23 8RE and ² Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA

Email for correspondence: holylochnaturereserve@gmail.com

Abstract

Mass, low-cost-, non-destructive DNA sequencing/barcoding via UK BIOSCAN, based at the Wellcome Sanger Institute, is now creating hundreds of thousands of invertebrate Cytochrome c Oxidase sub unit 1 (COI) barcodes from specimens collected across the United Kingdom (UK). Holy Loch Nature Reserve (HLNR), in Argyll, Scotland, joined the project in 2024, and in the same year, our first 6350 DNA “barcodes” were generated and uploaded to the Barcode of Life Database (BOLD). Sequences are then organised into statistically distinct clusters which are assigned Biodiversity Index numbers (BINs). The translation of BINs into named species is then mainly dependent on well-curated specimens, with associated barcodes, on BOLD. This paper describes a method to update a locally-held HLNR BIN database, with new sequences added regularly by UK BIOSCAN, and ongoing global taxonomic revisions on BOLD. Many named species at HLNR are not located on the Global Biodiversity Information Facility (GBIF) distribution maps, and, given the complex, multifaceted, nonintegrated nature of the UK biodiversity recording arena, could be new species discoveries in the west of Scotland, and/or Scotland and the UK. Even more unnamed BINs are likely to be new.

Introduction

The world is currently experiencing an accelerating rate of species extinctions driven by human economic activity (IPBES, 2019). Biodiversity monitoring is central to

global conservation efforts, and small community-led projects such as the undisturbed, protected area around the Holy Loch on the Cowal Peninsula, Argyll, Scotland can add to these. Ideally, such monitoring needs a comprehensive baseline audit of all Operational Taxonomic Units (OTUs; Blaxter et al , 2005). Given the close correlation of OTUs to species, this roughly equates to calculating species richness, a standard metric of biodiversity.

In practice, however, generating a complete known species inventory requires a vast, coordinated, and costly input of conventional human expertise. Even if this were achievable, many species, both globally and within the UK remain undescribed by science. Due to a huge lack of global investment in taxonomy and phylogeny, the so-called “taxonomic impediment”(Engel et al, 2021; Wheeler et al. 2024), humanity is realistically decades, perhaps centuries, away from formally documenting every species, at even a small, single site, and many will likely be lost to climate change and other factors (Urban, 2024; Tedesco et al, 2014) before they are named.

Until the recent adoption of DNA sequencing (barcoding) of organisms (Porter & Hajibabaei, 2018), for at least two centuries, traditional taxonomy and phylogeny have relied on the dedication of specialists such as Carl Linnaeus, and taxonomists since, working on individual taxa, and then publishing information on how to identify specimens morphologically as species descriptions and in dichotomous keys.

Identification resources for many UK insect groups do exist, some recent, others several decades old, but all represent snapshots of taxonomic knowledge at the time of publication. Invertebrates are dynamic however: they migrate (either independently or via human transport) and they evolve. Online databases and identification websites can be updated more rapidly, but they also suffer from the taxonomic impediment. As a result, whole families of insects remain difficult for non-specialists, and in many cases, for experts, to identify, and I have dozens of unidentified specimens myself.

DNA barcoding offers a major step forward since it builds on accumulated taxonomic knowledge, enabling rapid and scalable species identification (Antil et al, 2023). UK BIOSCAN combines non-destructive DNA extraction methods (Park et al.) and automation, to generate large numbers of COI barcodes from invertebrates captured

by standard Malaise traps deployed across multiple UK sites. Never before has such a large dataset of barcodes been available for UK biodiversity and ecological studies.

Barcodes are clustered into groups of similar sequences [BINs; Biodiversity Index Numbers; Ratnasingham & Hebert, (2007, 2013)], and then compared with reference library BIN sequences identified by taxonomic experts and stored in the Barcode of Life Database [BOLD; Ratnasingham and Hebert (2007)], to attempt species identification. Similarity thresholds used have been discussed elsewhere (Weigand et al. 2019).

This paper outlines the challenges of working with the large UK BIOSCAN dataset from HLNR, and details a simple method for monitoring taxonomic changes. The Diptera super-family Chironomoidea, containing four families, was chosen for this foundational paper because taxonomic resources vary from a relatively recent identification key (Family: Chironomidae; Langton & Pinder, 2007), to a web-based key for biting midges (Family: Ceratopogonidae; <https://sites.google.com/view/mikes-insect-keys/mikes-insect-keys/keys-for-the-identification-of-british-true-flies-diptera/keys-to-the-british-species-of-family-ceratopogonidae>), and no freely accessible, free-to-use resources (Davies, 1968; 2009) for Blackflies (Family: Simuliidae).

This paper provides a system that allows non-specialists at HLNR to interpret BINs as stable biodiversity units, bypassing the need for morphological expertise.

Methods

Generic methods used for the Holy Loch Food Web Project are deposited on Zenodo (doi: 10.5281/zenodo.18800209).

As monthly UK BIOSCAN malaise trappings never fill up an exact number of 96 well plates, a few additional samples were collected by other means to fill the empty wells. In this study, several larvae were collected from under a piece of wood in a saltmarsh drainage ditch (on 15th September), from a moth pheromone trap containing cider vinegar and six blackberry fruits (*Rubus fruticosus* agg.; also on 15th September). Specimens were washed with phosphate buffer (pH 7) to dilute residual sodium

chloride and acetic acid, prior to transfer to the remainder of 96 well plates containing 100% ethanol.

2024 trapping dates ended on 27th January (carr woodland only), 27th February, 31st March, 27th April, 22nd May, 30th June, 23rd July, 22nd August, 15th September, 21st October, 26th November and 24th December. The spatial layout and marsh trap environment are documented in the aerial photograph (Figs. 1 and 2) and the sampling layout (Fig. 3).

Results

BINs were assigned to the four categories based on the factors listed in Table 1 (doi: 10.5281/zenodo.18094319). Names of these categories are partly based on the Abundance Rarity model of communities described by Magurran & Henderson (2011) and the “rare biosphere” coined by Lynch and Neufeld (2015), where some species are abundant while the majority occur only rarely in populations. BIN allocations are summarised in supplementary tables on Zenodo (<https://doi.org/10.5281/zenodo.18086742>).

Barcoding results

Chironomoidea comprised the greatest number of individuals and BINs recorded at HLNR in 2024. As a result, this superfamily was used as a foundational paper to set the scene for all succeeding papers with other taxa. The majority of chironomids caught at HLNR in 2024 derived from the marsh. Within the family Chironomidae, the genus *Limnophyes* was the most abundant, producing 842 barcodes, of which 348 were assigned to *L. natalensis*. Successful naming of BINs depended on the quality and completeness of the reference library; in several cases, such as BOLD:ACF9576, BINs may contain more than one species, specifically *Culicoides obsoletus* or *C. scoticus*, where morphological differences are not reflected in a 2% DNA sequence divergence. Conversely, some species, such as *Dasyhelea turficola*, spanned multiple BINs.

Within the Chironomidae, twenty-three BINs were assigned to Confirmed Identity (Table 3), while thirty-two BINs occurred only once or a few times and were categorized as Rare Biosphere (Table 4). Among twenty-six BINs with Emerging

Identities assignable only to genus (Table 5), manual BLAST analysis enabled the assignment of up to three possible species. Two relatively frequent BINs assigned to Barcode Silence were identifiable only to the family level (Table 6). For Ceratopogonidae, at least seven species were identified confidently (Table 7), and fifteen rare BINs also had solid identifications. Among the twelve with emerging identities, BLAST analysis revealed that several could not be deciphered between two near-identical species.

Specific attention was given to the Simuliidae, where three larvae collected on September 15th from the saltmarsh field drain were assigned to two *Simulium* BINs, while a total of twelve adults collected in 2024 were allocated to three BINs (Table 8). Notably, BOLD:ACX9864 was captured both as a larva in September and as an adult in October.

Notably, one chironomid BIN appears to be currently unique to the Holy Loch Nature Reserve within BOLD. BIN BOLD:AGX5543 was recorded exclusively within the saltmarsh habitat during April, May, June, August, and September. BLAST analysis of individual barcodes within this BIN revealed a 97.1% similarity to *Limnophyes natalensis*, while maintaining near 100% similarity within the BIN itself, suggesting a potentially distinct or highly localised population.

Discussion

In this initial paper we focus on a single, large superfamily of Diptera, Chironomoidea, which included 878 recognised species on the British list/ United Kingdom Species Inventory (UKSI) in November 2025. All families within this group are notoriously difficult to identify morphologically, but should be abundant in aquatic ecosystems such as at HLNR. By incorporating both named and unnamed BINs, this paper describes the first extensive chironomid fly assemblage from a Scottish coastal site, and far exceeds the one previous, morphologically-identified species of Chironomoidea at HLNR, the infamous Highland Midge (*Culicoides impunctatus*).

These three families are rarely, if ever, represented in wider Diptera surveys because their identification relies on highly specialised morphological expertise and

substantial time investment. Traditional Chironomoidea keys rely heavily on male genitalia (Langton & Pinder, 2007), and females and larvae usually remain unidentifiable. Previous work with Chironomidae, for instance, by Ekrem et al. (2010), has shown that expert morphological approaches regularly encounter large numbers of new, unnamed or provisionally-coded taxa (such as *Limnophyes* sp. 3ES and *L. sp.* 14ES), many waiting years or decades for formal description. The categories assigned in this study (Table 1) allow regular updates of species lists with new batches of sequences from ongoing barcoding.

Twenty-three Chironomidae BINs were assigned to **CONFIRMED IDENTITY**, with near-100% confidence and >98% similarity to multiple library sequences, including *Limnophyes natalensis*. These BINs are long-established, typically defined more than five years ago, and continue to hold as distinct sequence clusters as new sequences are added. They include well-anchored taxa such as *Limnophyes* “3ES” already noted in the literature (Ekrem et al., 2010). Although some long-standing BINs still carry coded names, they remain robustly delimited and require only further taxonomic investigation to formalise species identities (Table 3).

A further 32 BINs occurred only once or a few times in the HLNR dataset (Table 4) and were categorised as **RARE BIOSPHERE**, with one species manually identified. These BINs were represented by three or fewer specimens in 2024 and therefore carry the greatest risk of misidentification. Nonetheless, both BOLD’s Classifier Engine and manual BLAST searches consistently return high-confidence matches to established named BINs. As additional HLNR specimens accumulate, confidence in these names is expected to increase. For both **CONFIRMED IDENTITY** and **RARE BIOSPHERE** categories, presence on the United Kingdom Species Inventory and previous Scottish records on the National Biodiversity Network Atlas provide further support for the identifications.

Twenty-six BINs were placed in **EMERGING IDENTITY**, representing abundant taxa lacking a confirmed species name or associated with only one or a few reference sequences on BOLD (Table 5). BLAST comparison allowed tentative species-level assignment for three of these. This category includes some of the most frequent BINs in the dataset, such as BOLD:ABV35040, provisionally assigned to *Pseudorthocladus curtistylus* based on a single Finnish reference specimen (Roslin

et al., 2022). Many BINs in this group are relatively young, illustrating how UK BIOSCAN is revealing substantial previously undocumented molecular diversity.

Two relatively frequent BINs fell into **BARCODE SILENCE**, as they could not be identified beyond family level (Table 6). These represent cases where no informative reference sequences currently exist, preventing assignment below Chironomidae.

BINs currently unique to HLNR on BOLD warrant targeted investigation to establish whether they represent genuinely distinct lineages or fall within the natural genetic variation of a wider species.

At least seven Ceratopogonidae species were assigned to **CONFIRMED IDENTITY** (Table 7), supported by strong sequence matches and stable BIN concepts. A further 15 rare BINs were placed in the **RARE BIOSPHERE** category, all with solid identifications despite their low specimen numbers. Among the twelve BINs with **EMERGING IDENTITIES**, BLAST analysis showed that several could not be resolved beyond pairs of near- identical species, while four BINs showed close similarity to *Dasyhelea turficola*. Three larvae collected on 15 September from the saltmarsh field drain were assigned to two *Simulium* BINs, and 12 adults collected across 2024 were allocated to three BINs (Table 8). One of these, BOLD:ACX9864, was captured both as a larva on 15 September and as an adult on 21 October, providing a rare direct link between life stages.

In all categories, ongoing global taxonomic work with Chironomoidea will continue to populate BOLD with new reference sequences. As we continually add new barcodes to the HLNR dataset on BOLD, the picture of Chironomoidea community structure at HLNR will become clearer, and some BINs will increase in identification confidence. The non-destructive nature of DNA extraction for this study now enables taxonomists to revisit material after barcoding to further resolve taxonomy and phylogeny. In this data set, some BINs contain more than one species, presumably where morphological differences in adults are not reflected in at least a 2% DNA sequence difference (e.g. BOLD:ACF9576 may be *Culicoides obsoletus* or *C. scoticus*), and some species span many BINs, such as *Dasyhelea turficola*. These

findings may point to cryptic morphological differences, and not necessarily in gross morphology, and/or new genetic lineages and /or evolution at work.

Misidentifications within reference databases are inevitable, just as errors occur in morphological taxonomy. In this investigation, manual BLAST against GENBANK and BOLD, of BINs with genus-only matches, revealed obvious reference misidentifications, and this was not detected by the more conservative Classifier Engine. In several cases, the correct interpretation was clear from sequence similarity patterns, geography, and known ecology.

In spite of being abundant in our study, *Limnophyes* 3ES (Ekrem et al., 2010), remains formally undescribed. These authors also found parthenogenetic, all-female *Smittia* species which are still awaiting formal morphological description via discovery of a male with an identical barcode. Barcoding now allows the inclusion of all-female species in our studies (Ekrem et al., 2010).

In this study, non-destructive DNA extraction and sequencing were extended to larvae and marine/saline locations, allowing capture of further OTU ecological metadata via barcoding.

DNA barcoding promises to add significant numbers of new species to the UK and/or Scottish list of known species. However, a clear picture of whether a species has been recorded before in Scotland is confused by many factors:

- Many taxonomic experts and surveyors do not contribute their records to national repositories such as the National Biodiversity Network (NBN)'s atlas (Harding, 2003).
- Surveyors do not subject their findings to peer review.
- Records are often lost in long-forgotten reports.
- NBN does not record species identified from DNA barcoding.
- Recording is complicated by regular updating of the BOLD sequence library.
- In spite of BINs being workable molecular anchors, only named species are added to the NBN and the GBIF (Global Biodiversity Information Facility).

- The only way to look for whether BINs are new to Scotland is to look at the BOLD distribution map for each BIN.
- Everything added to the NBN atlas is also added to the Global Biodiversity Information Facility, but only by agreement with the recorder which is not always guaranteed.

The result of all of this is a significant inability to assign meaningful, evidence-based conservation categories onto large numbers of (known) UK invertebrates. Species identified via their DNA by BOLD are uploaded to GBIF eventually, as are records from many other sources. We are confident that this study of barcodes from HLNR in 2024, has contributed many newly recorded species of Chironomoidea for Scotland in publicly accessible datasets, including this publication. A simple comparison of the HLNR species found by UK BIOSCAN at HLNR with the distribution map of the UK, and Scotland particularly, on GBIF on December 9th 2025, uncovered in Chironomidae, *Limnophyes pentaplastus*, *Pseudosmittia trilobata*, *Pseudorthocladius cf. curtistylus*, *Micropsectra apposita*, *Micropsectra klinki*, *Tanytarsus ejuncidus* and *Xenopelopia nigricans*. In Ceratopogonidae, *Culicoides comosioculatus* appears to be new to the UK, and rare in most of Europe except Scandinavia. New species of biting midges also include *Dasyhelea arenivaga*, *Forcipomyia monilicornis* and *Atrichopogon cf. hirtidorsum*. Furthermore, all BINs are included in global distribution maps on BOLD.

The majority of new, named species occurred in small numbers; however, *A. cf. hirtidorsum* was represented by eleven insects. It is unclear how species new to the UK will be included in the UKSI. Furthermore, many of the unnamed BINs could also be new species, further emphasising the effect the taxonomic impediment is having on biodiversity recording in the UK. Unfortunately, only publications such as this can make this known publicly at this time. Although for biodiversity monitoring, perfect naming is not essential, for comparisons to previous studies with existing species, a latin binomial is still essential.

Environmental DNA (eDNA) metabarcoding could form a logical expansion of this work. However, eDNA detections are only as reliable as the local reference library they are compared against, and the technology has not been tested against such dense BIN inventories such as this one at the Holy Loch . Every locally obtained BIN acts

as a verified reference against which future eDNA sequences can be interpreted. eDNA may also find new species not attracted to Malaise traps.

Future notes on Chironomoidea at HLNR will include updates on species inventory from 2025 and 2026 for which this paper is the foundation.

Acknowledgements

The Holy Loch Food Web Project is an initiative by Sandbank Community Development Trust, and supported financially by the people of Cowal, and grants from Seachangers, Argyll and Bute Council, the Blodwen Lloyd Binns bequest, Argyll and Bute Climate Action Network, Argyll Bird Club and Naturescot. For generation and release of BIOSCAN data, I thank Mara Lawniczak's team at the Wellcome Sanger Institute, Sanger Scientific Operations, and Sanger Tree of Life, all of whom are funded by the Wellcome Trust award 220540/Z/20/A.

Publication note

This paper was originally submitted to a peer-reviewed journal. The resulting reviews, however, reflected a fundamental disconnect between conventional taxonomic requirements and the mission of UK Bioscan.

The reviewers, operating from a traditional framework, insisted upon anchoring all identifications in the dissection of male genitalia. This critique suggests a lack of engagement with the manuscript's primary objective: addressing the critical shortage of specialists by providing a scalable, modern alternative.

By prioritizing exhaustive morphological dissection over high-throughput methodologies, such conventional critiques overlook the necessity of democratizing taxonomy. UK Bioscan seeks to translate the profound achievements of dipterists, both past and present, into an accessible framework that overcomes the specialized bottlenecks currently hindering large-scale biodiversity monitoring in the UK.

Literature referencing

Species identified in this study for Chironomoidea:

Chironomidae

Brillia bifida

Chaetocladius perennis

Chironomus aprilinus

Chironomus salinarius

Gymnometriocnemus brumalis

Halocladius variabilis

Halocladius varians

Limnophyes edwardsi

Limnophyes habilis

Limnophyes natalensis

Limnophyes pentaplastus

Limnophyes sp. 14ES

Limnophyes sp. 3ES

Metriocnemus albolineatus

Metriocnemus fuscipes

Metriocnemus picipes

Paraphaenocladius impensus

Paraphaenocladius impensus

Pseudorthocladius cf. curtistylus

Pseudorthocladius filiformis

Pseudosmittia trilobata

Smittia leucopogon

Eukiefferiella claripennis

Macropelopia nebulosa

Metriocnemus eurynotus

Metriocnemus sp. 1SW

Micropsectra apposita

Micropsectra klinki

Micropsectra lindrothi

Micropsectra pallidula

Paraphaenocladius impensus

Paraphaenocladius pseudirritus

Paratanytarsus austriacus

Polypedilum uncinatum

Pseudosmittia albipennis

Rheocricotopus atripes

Rheocricotopus sp. 3ES

Smittia cf. *stercoraria*

Smittia cf. *stercoraria*

Smittia paranudipennis

Smittia pratorum

Tanytarsus brundini

Tanytarsus ejuncidus

Tanytarsus signatus

Thienemanniella xena

Xenopelopia nigricans

Limnophyes asquamatus

Ceratopogonidae

Culicoides circumscriptus

Culicoides impunctatus

Culicoides punctatus

Culicoides segnis

Dasyhelea arenivaga

Dasyhelea turficola

Forcipomyia tenuis

Ceratopogon grandiforceps

Culicoides comosioculatus

Culicoides newsteadi

Dasyhelea modesta

Dasyhelea turficola

Dasyhelea turficola

Dasyhelea turficola

Forcipomyia bipunctata

Forcipomyia nigra

Palpomyia flavipes

Palpomyia serripes

Schizohelia leucopeza

Serromyia femorata

Serromyia ledicola

Stilobezzia ochracea

References

Davies, L. (1968) A key to the British species of Simuliidae (Diptera) in the larval, pupal and adult stages. (Scientific Publication No. 24, Freshwater Biological Association, 1–126).

Davies, L. (2009). The taxonomy of British black-flies (Diptera: Simuliidae). Transactions of the Royal Entomological Society of London. 118., 413 - 506. doi: 10.1111/j.1365-2311.1966.tb00832.x.

All other references are archived on Zenodo ([10.5281/zenodo.18086742](https://zenodo.org/doi/10.5281/zenodo.18086742)).

Figure 1.

Aerial photograph of the “Marsh” sampling site at Holy Loch Nature Reserve. The box insert is detailed (with co-ordinates and scale) in Figure 3.

Source: Phill Rawlins



Figure 2.

Layout of the sampling site with approximate descriptions of habitats and electrical conductivity (mS m^{-1}) measurements. Source: Phill Rawlins with annotations by Neil Hammatt

Key

- Holy Loch- a tidal sea loch.

- Upper beach is the highest non-spring tide height.
- The herb zone, gorse (*Ulex europaeus*) scrub and woodland are all on a gravel ridge.
- Shallow flood and upper saltmarsh are only flooded by spring tides.
- The sinuous channel is under investigation, but may be a freshwater seep.
- A 40m linear transect is between points A and B. A is at 55.9945 N, -4.5983; what3words: standard.sunflower.beginning, B is at Grid ref: 55.9941 N, -4.5980; what3words: phantom.overgrown.junction



Figure 3.

The sinuous channel highlighted in Figure 2, located in flower-rich grassland (June 2024). Source: Neil Hammatt



Figure 4.

A malaise trap in the June to March side of the woodland strip (see Figure 3). Source: Neil Hammatt



Figure 5.

General view of the location of the carr woodland malaise trap location, showing the dense, herbaceous understory. Source: Neil Hammatt



Figure 6.

The “Marsh” catch from June 2024. Source: Neil Hammatt

