DNA barcoding for biodiversity

DNA barcoding is a technique used for species-level identification. It can be used to define new species and provide evidence to prosecute for trafficking of endangered species. Molecular geneticist Karen Stephens describes what DNA barcoding is and how students and researchers are using it to investigate insect biodiversity

Karen Stephens

Understanding insect **biodiversity** and the potential impacts of losing **keystone species** are among the big challenges currently faced as part of the **Anthropocene extinction**. Insects are diverse, small, numerous and vital for ecosystems, but are hard to identify, poorly characterised and have rapidly declining populations.

There are thought to be 5.5 million species of insect and they perform vital **ecosystem services**:

- Provisioning services – they are a source of natural resources, such as the stable polymer chitin, which is used in drug delivery systems for controlled release of different formulations.
- Regulating services – they provide benefits through natural processes, such as pollination of crops.
- Cultural services – they provide non-material benefits, such as the beauty of butterflies.
- Supporting services – they are critical for habitat function, such as nutrient cycling aided by invertebrates like earthworms.

A hoverfly (a) and a honeybee (b). Bees and hoverflies both have black and yellow stripy bodies and hover around flowers, feeding on nectar. However, bees have furry bodies and two pairs of wings, whereas hoverflies have one pair of wings and tend to have smooth bodies and larger, compound eyes

However, we have only identified about 10% of these vital animals and have characterised far fewer. Many scientists have expressed concerns that insect decline could impact ecosystem function and human wellbeing. This decline in insect populations has been documented by the Bugs Matter Citizen Science Survey counting 'bug splats' on vehicle number plates in the UK between 2004 and 2023, which estimated a 77.6% decline in insect numbers during this 19-year period.

The magnitude of the threat is being taken seriously, and a host of measures are being implemented to understand and protect our insect biodiversity. These include the Wildlife Trust's 'Action for Insects' campaign to raise awareness and money for conservation, and the Natural Environment Research Council's £2.3 million project to analyse insect populations and model ways to counter population reductions. In 2023 a UK Parliamentary committee was established to consider the effects of insect decline on UK food security.

Improving insect identification

One hurdle is a scarcity of **entomologists** and difficulty in identifying insects based on their appearance, owing to their small size, similarities and different life stages. For example, can you tell a hoverfly from a honeybee?

DNA barcoding offers a different way to collect data on insect biodiversity – to improve identification, to understand ecosystem services provided by a species, and to inform conservation practices and policies. DNA barcoding compares the base sequence of the same region of DNA from different species. The base sequence of this section acts as a 'DNA barcode'. It is the same for members of the same species, but differs between species. The DNA barcode allows a species to be identified, much like the barcodes in shops that allow a product to be identified (see Figure 1).

A DNA barcode

The section of DNA used as a barcode for animals is from the 5′ end of the mitochondrial *cytochrome c oxidase subunit 1* (*CO1*) gene. This sequence is the same within a species but differs between species. The 700-base-pair sequence within the *CO1* gene is a good DNA barcode for two reasons.

Firstly, it is part of the mitochondrial DNA, rather than the nuclear DNA. This is beneficial because there are multiple mitochondria in every eukaryotic cell, each containing a copy of the *CO1* gene. Having more copies of the DNA barcode in a cell makes DNA extraction more successful.

The second benefit is that the *CO1* gene is important for cellular metabolism. The *CO1* gene encodes a protein needed in aerobic respiration, in the electron transport chain. The CO1 protein's important role means that the *CO1* gene evolves relatively slowly, as rapid, extensive mutation would prevent energy transfer as part of cellular metabolism, leading to cell death. This slow evolution helps to explain how mutations accumulate slowly, giving this DNA barcode its ability to remain essentially the same within a species, but to differ between species.

Barcoding biodiversity

Using DNA barcodes in conjunction with collection records helps with identifying the species. Increased knowledge about insect species will allow for more accurate biomonitoring. Examples of increased knowledge include:

- Identifying insect species with the date when they were collected allows us to learn about life cycles – when you are likely to find eggs, larvae, pupae or adults. This can help conservation efforts, such as recommending reduction in pesticide use during specific periods of keystone insect life cycles when the insects are particularly vulnerable.

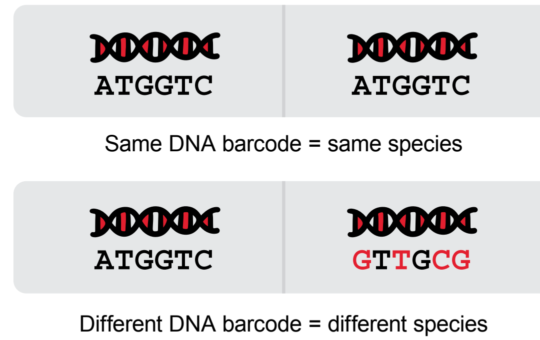

Figure 1 DNA barcoding

- Revealing the time of year at which species are identified helps us to understand seasonal variation in populations. This means we can recognise when climate change or land management have a detrimental effect, e.g. on breeding seasons.
- Combining records on the locations of insect species provides information on that insect's range. Changes in the range help us to understand which elements of an environment are key for that insect's survival.
- Learning about the habitats of different insect species helps us to identify when changes in land use might threaten a species, and the habitat creation that would be beneficial for conservation of key species.

TERMS EXPLAINED

Anthropocene extinction The ongoing extinction event driven primarily by human activities, leading to high rates of species loss.

Biodiversity The variety of species in a particular habitat.

DNA barcoding Using a particular DNA sequence to identify a species.

Ecosystem services The direct and indirect contributions of ecosystems to human wellbeing.

Entomologists Scientists who study insects.

Gel electrophoresis A method of separating DNA by size.

High-throughput sequencing A method that determines the DNA sequence of multiple samples at the same time, allowing rapid DNA sequencing on a large scale.

Keystone species Organisms with a critical role in maintaining the health of their environment. They have a disproportionately large impact on their ecosystem relative to their population size.

Polymerase chain reaction (PCR) A method of making copies of a chosen region of DNA.

Sanger sequencing A method of DNA sequencing that identifies the order of nucleotide bases using modified nucleotides and gel electrophoresis.

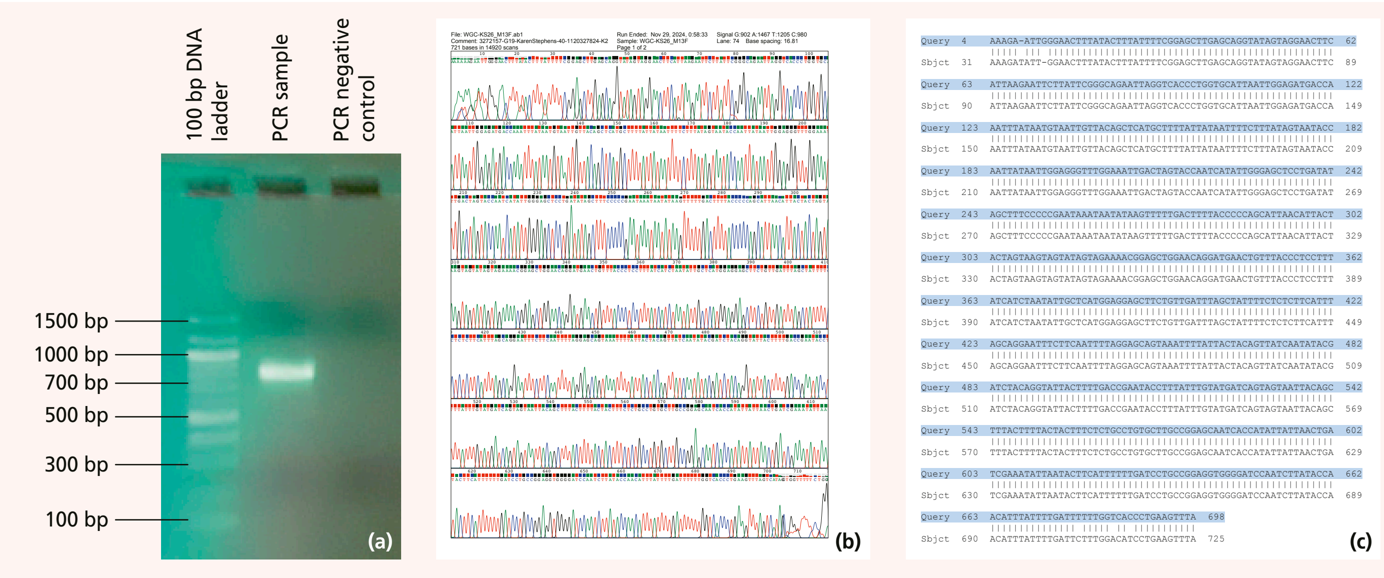

Figure 2 Results from Barcoding for beginners. (a) Agarose gel electrophoresis, showing the PCR sample contains DNA of around the expected 710 base pair DNA barcode, when compared with the DNA ladder. (b) Chromatogram, with clear, non-overlapping peaks showing accurate DNA sequencing. (c) Alignment of the DNA barcode sequence ('query', highlighted in blue), to the DNA sequence of *Sarcophaga carnaria* ('subject', not highlighted), using a BLAST search

Barcoding for beginners

In 2024, the science engagement team at the Wellcome Sanger Institute near Cambridge, supported by a Royal Society Partnership Grant, developed the 'Barcoding for beginners' programme. This upskills educators in DNA barcoding, enabling students to carry out research into invertebrates in their school or college environments.

Following the steps outlined in Box 1, students collect and analyse their own samples. Figure 2a shows an agarose gel with size comparison of the DNA barcode to a 100 base pair DNA ladder. Where a band of the expected ~710 base pair size is present, the DNA barcode, amplified by PCR, is sequenced.

The output of DNA sequencing is a chromatogram (see Figure 2b). Chromatograms provide the order of the nucleotides in the DNA sequence, plotted against the strength of the fluorescent signal from a nucleotide. Students have obtained good quality DNA, shown by clear, evenly spaced peaks for their DNA barcodes.

Students compare a text file containing the sequence of the DNA barcode from the unknown invertebrate to the National Centre for Biological Information database. Alignment of the barcode 'query' to the existing sequence allows identification of the invertebrate. Figure 2c shows how the DNA barcode 'query' sequence from an unidentified invertebrate (shown in blue), aligns to *Sarcophaga carnaria* – the camouflaged flesh fly.

A pioneering cohort of more than 400 students has started biomonitoring of invertebrates in their school environments using DNA barcoding. Among those identified are honeybees, false blister beetles, red mites, Welsh chafers and bean seed flies. We plan to share DNA sequences, collection records and invertebrate identification through the Barcode Of Life Database (BOLD).

Box 1 Finding the DNA barcode for an insect

The technique of DNA barcoding is outlined in Figure 1.1.

1. Collect insect samples. Record the collector, time, location, habitat and a description of the insect.
2. Remove DNA from the insect samples. Disrupt tissues to break the cells apart, lyse (burst) cells to release DNA, and separate from other cell debris.
3. Copy the DNA barcode using the **polymerase chain reaction (PCR)**. After 35 cycles of denaturing, annealing and extending there are more than 34 billion copies.
4. Use **gel electrophoresis** to check that the DNA barcode region has been successfully amplified. Load the PCR sample onto an agarose gel. After separation through the gel by electrophoresis, compare the size of the PCR sample with the DNA ladder to check the DNA is the expected length of the DNA barcode.
5. If the DNA barcode is successfully amplified, sequence using **Sanger sequencing**.
6. Align the sequence of the DNA barcode to an existing database, using bioinformatics to run a Basic Local Alignment Search Tool (BLAST).
7. The DNA barcode either matches an existing entry in the database, allowing rapid identification of a species, or there is no match, suggesting discovery of a new species.

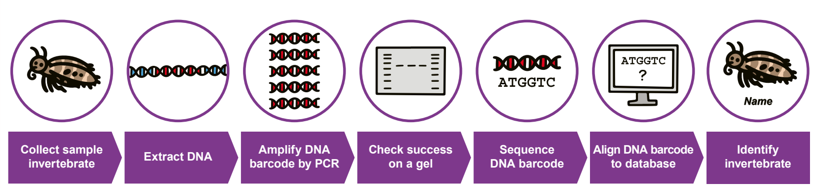

Figure 1.1 An outline of the DNA barcoding technique

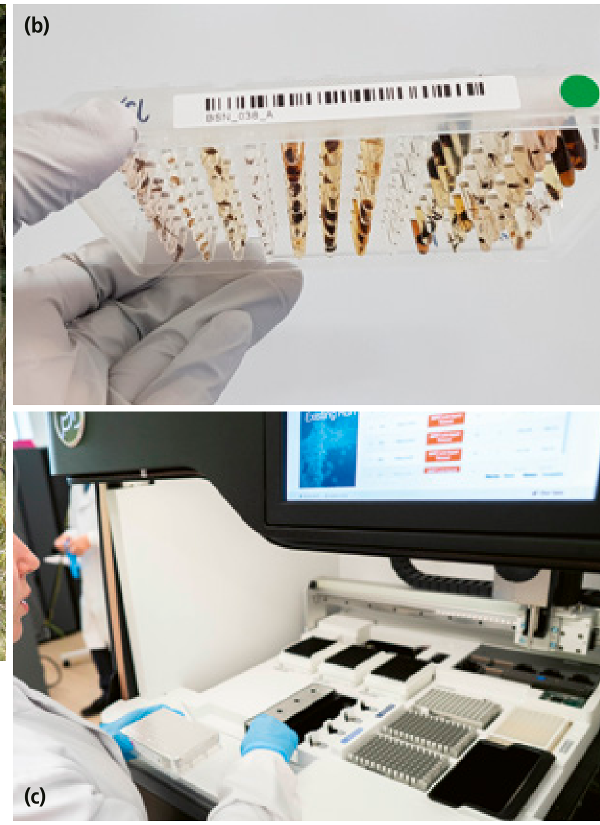

Figure 3 Stages of the BIOSCAN process for DNA barcoding insect biodiversity. (a) A malaise trap used to collect monthly samples for BIOSCAN. The pot of ethanol where samples collect is visible on the top left of the trap. (b) A 96-well plate showing separation of insects into individual wells for DNA extraction and PCR. (c) Massively parallel sequencing of DNA barcodes on a PacBio machine

BIOSCAN

BIOSCAN is a similar research project that operates on a much larger scale. It carries out monthly monitoring of insects collected by many partners, including the National Trust, Wildlife Trusts, Botanic Gardens and zoos, from more than 100 sites around the UK. It aims to DNA barcode 1 million flying insects over 5 years, creating a baseline for UK insect diversity.

The process follows the same outline of DNA barcoding shown in Figure 1.1, but with differences to Barcoding for beginners that mean it can be massively scaled up. These include:

- a monthly collection of flying insects at multiple sites using malaise traps, rather than collection of an individual sample. When insects enter malaise traps, which are essentially tents, they fly upwards and are funnelled into a pot of ethanol to preserve them for analysis (see Figure 3a)
- a high-throughput, automated method, which simultaneously extracts DNA from 96 insect samples in a 96-well plate, instead of one insect at a time (see Figure 3b)
- use of a non-destructive DNA extraction method, which keeps insects intact, allowing entomologists to study and identify insects with a novel DNA barcode. The DNA barcode and insect identification are shared through BOLD
- methods for PCR and quantification of the DNA barcode produced for each sample occur in a 96-well plate in parallel. This means that agarose gel electrophoresis is not needed to examine every sample
- use of **high-throughput sequencing** of DNA (see Figure 3c).

By scaling up insect identification using DNA barcoding, BIOSCAN processes 20 000 samples each month. BIOSCAN shares data through BOLD, including DNA barcodes with collection records and expert identification of insects with previously unknown DNA barcodes. This provides geographical and temporal baseline data for biomonitoring, land management and conservation efforts to benefit insect biodiversity.

Looking to the future

DNA barcoding is providing large-scale data on insect biodiversity. The question now is whether we can apply our knowledge to inform conservation practice and policies before keystone species are lost as part of the Anthropocene extinction.

RESOURCES

Animation of DNA extraction by alkaline lysis:
https://tinyurl.com/your-genome-DNA-extraction

Animation of PCR:
https://tinyurl.com/your-genome-PCR

Animation of gel electrophoresis:
https://tinyurl.com/your-genome-electrophoresis

One method of DNA sequencing:
https://tinyurl.com/DNA-sequencing-3D

Barcoding for beginners:
https://tinyurl.com/barcoding-for-beginners

BIOSCAN website:
https://tinyurl.com/sanger-bioscan

Blog about Barcoding for beginners and BIOSCAN:
https://tinyurl.com/insect-investigators

Dr Karen Stephens was a science engagement manager at the Wellcome Sanger Institute near Cambridge. She started the Barcoding for beginners programme described in this article.