

Reliability of iNaturalist for Insect Identification Compared with DNA Barcoding: A Bioblitz Study in Bailu Park, Suzhou, China

Author: Mutong (Katherine) Wang
Mentors: David Micklos and Xinyue Wang
Cold Spring Harbor Asia DNA Learning Center

Abstract

AI-identification for insects and other organisms based on photographs is becoming increasingly accessible through applications such as iNaturalist, but certain species can be difficult to distinguish. This study evaluated the reliability of iNaturalist identifications by comparing them with DNA barcoding results from insects collected during a Bioblitz in Bailu Park, Suzhou, on August 7, 2026. I collected nine samples and first identified them using iNaturalist. Next, the insects' DNA was extracted using the Chelex method, amplified through PCR using COI primers, and underwent gel electrophoresis. Afterwards, the DNA was sequenced, then analyzed on DNA Subway. Four out of nine iNaturalist identifications corresponds with the barcoding identification at species level, while most correspond at the genus level. The results generally support the hypothesis that iNaturalist is reliable as an preliminary insect identification tool, although it has limited accuracy for insects that are visually similar, poorly photographed, or unrecorded in NCBI Genebank.).

Introduction

Accurate insect identification is important because it can largely influence the effect of insect conservation, biodiversity research, control of invasive species, and so on. Traditional identification often relies on the morphology of insects, such as the shape, color, wing patterns, and other visible characteristics (Indriati and Hidayat). However, morphological identification can be difficult in many situations, such as when samples are damaged, recorded poorly, visually similar or closely related to other species ("Morphological ID: Is It Really Worth the Disadvantages?").

iNaturalist is an online biodiversity platform that allows users to upload photographs of plants or insects and receive identification suggestions from AI models and other users. In order to assign observations to a taxon, iNaturalist uses artificial-intelligence-aided identification through computer vision models (iNaturalist). According to existing research on iNaturalist (López-Guillén et al.), the identification quality is influenced by identifier experience and by the quality and completeness of the photographic evidence .

DNA barcoding provides an alternative method for species identification. It tests a small DNA fragment that varies between species but remains consistent within a species, and then checks that fragment against databases to find the most accurate match (CD Genomics) . For insects, a region of the mitochondrial cytochrome c oxidase subunit I (COI) gene is commonly used as the barcode region.

The objective of this study was to evaluate to what extent iNaturalist identifications corresponds with COI DNA barcoding identifications for insects collected in Bailu Park, Suzhou. The research question was: How reliable is iNaturalist for identifying insects when compared with COI DNA barcoding results? My hypothesis is that iNaturalist would provide generally reliable preliminary identifications, particularly at the genus level. I also hypothesize that differences between the iNaturalist and DNA barcoding results would occur for low-quality photographs and damaged samples.

Materials & Methods

1.Sampling (Fig. 1 and 2)

The samples were collected during a Bioblitz conducted from 10:00 a.m. to 11:30 a.m. on August 7, 2026, in Bailu Park, Suzhou, China. Bailu Park is chosen because it has several different habitats, including ponds, a lake, grassland, and woods, which can support a variety of insects and other organisms. Samples were collected using an insect net. The net was repeatedly swept through grass and tree branches to collect insects that touches the net. Each collected organism was first photographed using the camera in the iNaturalist application and identified by iNaturalist. Then, the samples were transferred into tubes containing 95% ethanol that preserved the insects bodies.

Fig 1: Sampling locations in Bailu Park, Suzhou (Barcode Sample Database)



Fig 2: Photographs of samples 1-01 to 1-09

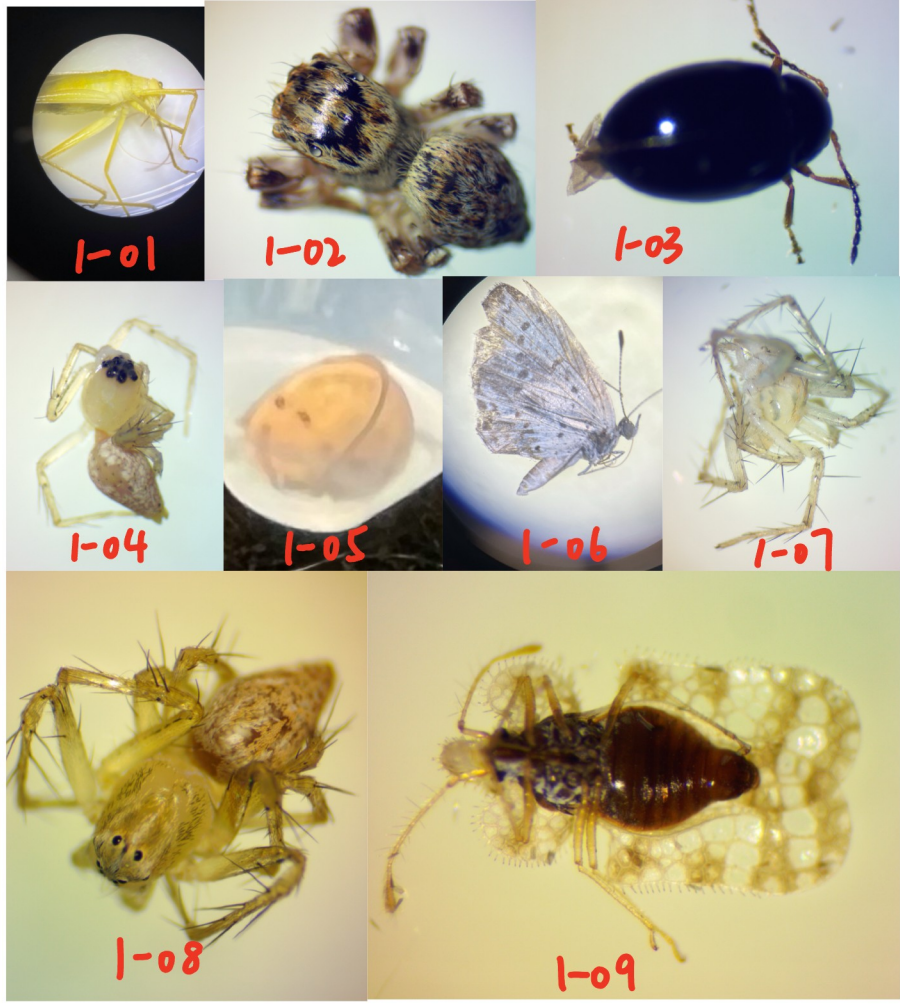


Fig 3: Gel electrophoresis results

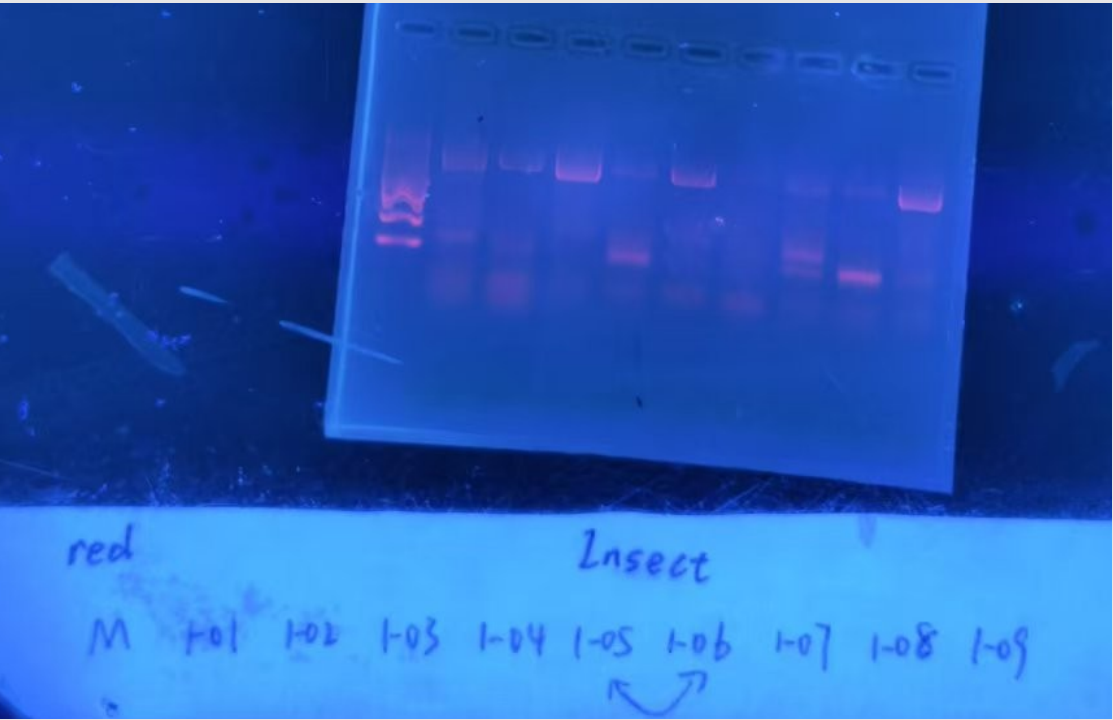
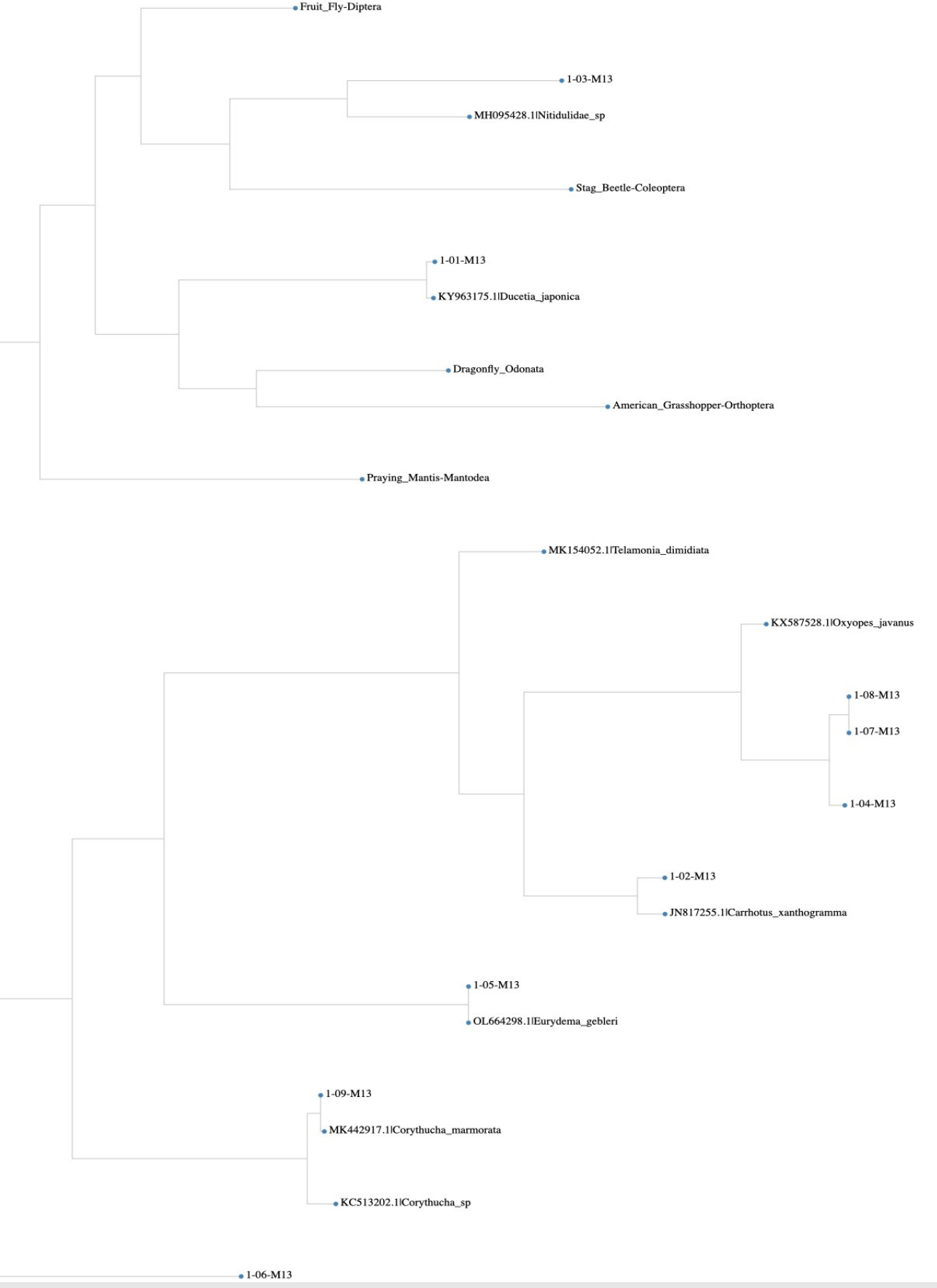


Table 1: Comparison between iNaturalist and DNA barcoding identifications (iNaturalist and DNA Subway 2.0)

	BLAST results	iNaturalist results
1-01	Bush katydid / Ducetia japonica	Bush katydid / Ducetia antipoda
1-02	Carrhotus xanthogramma	Carrhotus xanthogramma
1-03	Sap beetle / Nitidulidae	Cabbage-stem flea beetle / Psylliodes chrysocephala
1-04	Oxyopes sertatus	Zoosection Rta clade
1-05	Ladybug / Coccinellidae	Ladybug / Coccinellidae
1-06	Pale grass blue / Pseudozizeeria maha	Pale grass blue / Pseudozizeeria maha
1-07	Oxyopes sertatus	Oxyopes lineatus
1-08	Oxyopes sertatus	Oxyopes lineatus
1-09	Chrysanthemum lace bug / Corythucha marmorata	Chrysanthemum lace bug / Corythucha marmorata

Fig 4: Phylogenetic tree of the samples and other references (DNA Subway 2.0)



2.DNA extraction and PCR

DNA was extracted from the insect or arthropod tissue using the Chelex method. The extracted DNA was then amplified through PCR. A region of the mitochondrial COI gene was amplified using COI primers. Next, the PCR products were analyzed using gel electrophoresis.

3.DNA sequencing and BLAST analysis on DNA subway

The PCR products were sent to a local DNA sequencing company. Then, resulting sequences were analyzed through the DNA Subway website and compared with sequences in NCBL Genebank using BLAST. The closest matches were recorded as the barcoding results for the samples.

Results

All nine samples produced relatively clear PCR bands after gel electrophoresis, indicating that DNA extraction and COI amplification were successful for all samples (Fig.3).

The DNA Subway BLAST results and the preliminary iNaturalist identifications are shown in Table 1. At the species level, the two methods had the exactly same identification for four of the nine samples: 1-02, 1-05, 1-06, 1-09. At genus level, samples 1-01, 1-07, 1-08 shared the same identification, but iNaturalist assigned the samples to different species. For, sample 1-03, both the genus and species given by the two methods are different. Among BLAST results for sample 1-03, the identification with fewest mismatches has a high number of 58 mismatches, suggesting that 1-03 is likely to be a new species in the Nitidulidae genus that has not been recorded in GENEbank before.

Discussion

1.Analysis of the results

The results generally support the hypothesis that iNaturalist is reliable for preliminary insect identification, although it has some limitations. The application produced results that correspond with the DNA barcoding results at species level for 44.4% of all samples and at genus level for 60% of the remaining samples. This suggests that iNaturalist was often able to recognize the general morphological group, though it sometimes fails to distinguish between closely related species like Oxyopes sertatus and Oxyopes lineatus, shown in sample 1-07 and 1-08, and it can make mistakes when the photo of the sample is low-quality (1-04).

This result can be explained by the intended role of iNaturalist. iNaturalist is designed to provide quick identification based mainly on visual similarity, and is therefore useful for people who are not professional taxonomists, especially when the organism is photographed clearly and belongs to a commonly seen species. However, iNaturalist's own website states that computer-vision suggestions should not be treated as the only evidence for an identification (iNaturalist). The results shows that iNaturalist can be a first step in a biodiversity investigation, while DNA barcoding can provide additional evidence to complement iNaturalist's identification.

The disagreements in iNaturalist and DNA barcoding results of some of the samples also demonstrate why identification based on genotype can be reliable. DNA barcoding examines a sequence in the sample itself and therefore does not depend on whether a particular external feature is visible, avoiding mistaking a sample for another visually similar species. Nevertheless, there are situations where neither iNaturalist nor DNA barcoding are able to give reliable identifications, such as when the barcode of an organism like 1-09 is missing from the reference database. My original hypothesis did not fully account for this possibility. In this case, a reliable conclusion requires additional evidence or examination by specialists.

2.Limitations and improvements

a. This investigation has a small sample size, because it was conducted as a short Bioblitz, where sampling lasted only about 90 minute. Also, the sampling took place on a hot summer morning when many insects could have been hiding. Therefore, the results should be interpreted as a comparison of identification methods for these particular samples, not as a measurement of iNaturalist's accuracy for all organisms. To improve, one can repeat the sampling at different times of day and under different weather conditions.

b.Some samples were small or partly damaged after collection. This could have reduced the performance of iNaturalist's AI-identification. To improve, one should try to better preserve samples and take higher quality photos.

To conclude, this investigation demonstrates that applications based on phenotype identification and molecular methods can complement one another. Therefore, the implications for future biodiversity investigations is that combining both aforementioned methods can improve the reliability of species identification and help students participate in meaningful research.

Acknowledgements

I would like to thank my mentors, David Micklos and Xinyue Wang, for their guidance throughout the whole barcoding camp experience. I also thank Nancy, the teaching assistant for her support. I am also grateful to the other students in the class for collaborating during the DNA extraction. This investigation was supported by the Cold Spring Harbor Laboratory Asia DNA Learning Center, which I also would like to thank.

References

Barcode Sample Database. "Barcode Project Sample Database." Dnalc.Org. 2026. <https://sampledb.dnalc.org/barcoding/project/3383/sample/MZT-107>.
CD Genomics. "DNA Barcoding 101: Definition, Steps, and How It Works." 2022.
iNaturalist. "iNaturalist.Org." iNaturalist. 2025.