

Masterarbeit: Methodenentwicklung zum automatischen Abgleich von morphologischen Artenlisten und DNA-Daten

Betreuende: Brandon Seah, Wiebke SICKEL

Rückfragen / Interesse an: Brandon Seah (brandon.seah@thuenen.de)

Start: zum nächstmöglichen Zeitpunkt

Ziel:

Entwicklung von Methoden zum automatisierten Abgleich morphologischer und genetischer Daten unter Berücksichtigung von Unterschieden in taxonomischer Auflösung und Erfassungsebenen.

Diese Methoden sollen Auswertungen ermöglichen hinsichtlich folgender Fragestellungen:

- In wie vielen Fällen (=Proben) sind sich Morphologie und DNA einig?
- In wie vielen Fällen und für welche Taxa konnte Morphologie oder DNA eine höhere taxonomische Auflösung erreichen / den jeweils anderen Datensatz erweitern?
- In wie vielen Fällen und für welche Taxa liefern Morphologie und DNA Uneinigheiten?
- Inwiefern können solche Uneinigheiten automatisch korrigiert werden?

Hintergrund:

Die Erfassung von Arten in einer Probe, einem Ökosystem (=Biodiversitätserfassung) erlebt seit einigen Jahren eine Revolution hinsichtlich der angewandten Methodik. Man beobachtet eine Verschiebung von klassischen, morphologischen Erfassungsmethoden hin zur automatisierten Erfassung mittels Hochdurchsatz-DNA-Sequenzierung, insbesondere Metabarcoding und eDNA. Beide Methoden werden aktuell in vielen Fällen parallel angewandt, meistens zur Validierung der DNA-basierten Ergebnisse. Der Vergleich der Methoden findet allerdings meistens auf der Gesamtliste der detektierten Taxa statt, obwohl ein proben-spezifischer Vergleich aussagekräftiger wäre.

Im speziellen Fall des Wildbienen-Monitorings wird ein solcher Vergleich regelmäßig notwendig, da die tötungsfreie Erfassung von hohlraumnestenden Wildbienen mittels Fotos und eDNA erfolgt. Die Erfassungen erfolgen allerdings auf unterschiedlichen Aggregationsebenen (Niströhren im Fall der Fotos, Nisthilfenpaare im Fall der eDNA). Hier muss ein automatischer Abgleich der erfassten Daten stattfinden, der auch auf kritische Mismatches hinweist.

Was macht es kompliziert?

- Unterschiedliche Namen und Taxonomien werden in verschiedenen Datenbanken verwendet
- Unsicherheit bei der morphologischen Bestimmung (z.B. Bestimmung nur auf Gattung-Niveau)
- Ggf. begrenzte Auflösung der genetischen Marker

Vorgehensweise:

- Auswahl der Test-Datensätze (siehe unten)
- Bewertung der Referenzdaten:
 - Inwiefern sind Namen in verschiedene Taxonomien einheitlich (z.B. Synonyme)?
 - Sind alle Arten die wir untersuchen möchten in den Datenbanken repräsentiert?
 - Gibt es genug genetische Auflösung, um die zu unterscheiden?
- Aufreinigung der Daten – z.B. sind Proben einheitlich benannt im morpho und DNA-Teil; gibt es Typos in der Taxonomie, etc.
- Inwiefern können phylogenetische Verwandtschaften berücksichtigt werden?
- Wie geht man mit Diskrepanzen mit der taxonomischen Auflösung um, z.B. morphologisch nur bis Familie bestimmt, mittels DNA bis Art (und vice versa)?

Ihr Profil:

mind. 1 Kriterium muss erfüllt sein

- Grundkenntnis der Taxonomie
- Grundkenntnis molekulare Ökologie
- Kenntnis einer Programmiersprache, z.B. Python oder R

Literatur / mögliche Test-Datensätze:

Terrestrische eDNA:

<https://onlinelibrary.wiley.com/doi/full/10.1111/1755-0998.13148>

Pollen-Analyse (nur zwei willkürliche Beispiele):

<https://academic.oup.com/ee/article/46/1/38/2845897>

<https://onlinelibrary.wiley.com/doi/full/10.1111/mec.15675>

Licht-Falle:

<https://onlinelibrary.wiley.com/doi/full/10.1002/edn3.125>