

UITOTO:

a new R package and Shiny app for simplifying optimal molecular diagnoses in high-throughput taxonomy

Ambrosio Torres^a, Leshon Lee^b,
Amrita Srivathsan^a, and Rudolf Meier^a

^aCenter for Integrative Biodiversity Discovery, Museum für Naturkunde, Germany

^bDepartment of Biological Sciences, National University of Singapore, Singapore

Presentation for the:



July 29, 2024



① Introduction

② The UITOTO R package

③ Testing UITOTO

④ Discussion and conclusions

⑤ Acknowledgements

⑥ References

Descriptive taxonomy (or alpha taxonomy)

- Forms the basis of biological classification by identifying distinct biological units (species).

(Llorente-Bousquets, 1990; Renner, 2016; Braby *et al.*, 2024)

*These hypotheses are independent of phylogenetic trees, unlike secondary homologies (Richter, 2005; Agnarsson & Coddington, 2008)

Descriptive taxonomy (or alpha taxonomy)

- Forms the basis of biological classification by identifying distinct biological units (species).
 - These basic units are used across various studies (e.g., ecological, evolutionary).

(Llorente-Bousquets, 1990; Renner, 2016; Braby *et al.*, 2024)

*These hypotheses are independent of phylogenetic trees, unlike secondary homologies (Richter, 2005; Agnarsson & Coddington, 2008)

Descriptive taxonomy (or alpha taxonomy)

- Forms the basis of biological classification by identifying distinct biological units (species).
 - These basic units are used across various studies (e.g., ecological, evolutionary).
 - Classification principles have evolved to integrate new:

(Llorente-Bousquets, 1990; Renner, 2016; Braby *et al.*, 2024)

*These hypotheses are independent of phylogenetic trees, unlike secondary homologies (Richter, 2005; Agnarsson & Coddington, 2008)

Descriptive taxonomy (or alpha taxonomy)

- Forms the basis of biological classification by identifying distinct biological units (species).
 - These basic units are used across various studies (e.g., ecological, evolutionary).
 - Classification principles have evolved to integrate new:
 - Data

(Llorente-Bousquets, 1990; Renner, 2016; Braby *et al.*, 2024)

*These hypotheses are independent of phylogenetic trees, unlike secondary homologies (Richter, 2005; Agnarsson & Coddington, 2008)

Descriptive taxonomy (or alpha taxonomy)

- Forms the basis of biological classification by identifying distinct biological units (species).
 - These basic units are used across various studies (e.g., ecological, evolutionary).
 - Classification principles have evolved to integrate new:
 - Data
 - Methods

(Llorente-Bousquets, 1990; Renner, 2016; Braby *et al.*, 2024)

*These hypotheses are independent of phylogenetic trees, unlike secondary homologies (Richter, 2005; Agnarsson & Coddington, 2008)

Descriptive taxonomy (or alpha taxonomy)

- Forms the basis of biological classification by identifying distinct biological units (species).
 - These basic units are used across various studies (e.g., ecological, evolutionary).
 - Classification principles have evolved to integrate new:
 - Data
 - Methods
 - Regulations for standardized biodiversity communication

(Llorente-Bousquets, 1990; Renner, 2016; Braby *et al.*, 2024)

*These hypotheses are independent of phylogenetic trees, unlike secondary homologies (Richter, 2005; Agnarsson & Coddington, 2008)

Descriptive taxonomy (or alpha taxonomy)

- Forms the basis of biological classification by identifying distinct biological units (species).
 - These basic units are used across various studies (e.g., ecological, evolutionary).
 - Classification principles have evolved to integrate new:
 - Data
 - Methods
 - Regulations for standardized biodiversity communication
 - Since Linnaeus (1758), alpha taxonomy relies on diagnostic features, traditionally morphological (e.g., primary homology hypotheses to distinguish species*).

(Llorente-Bousquets, 1990; Renner, 2016; Braby *et al.*, 2024)

*These hypotheses are independent of phylogenetic trees, unlike secondary homologies (Richter, 2005; Agnarsson & Coddington, 2008)

Integrative taxonomy and molecular diagnoses

- Integrative taxonomy incorporates diverse information sources (e.g., morphology, DNA sequencing, chromosome mapping), providing a comprehensive understanding of species boundaries.

(Brower, 2010; Renner, 2016; Meier *et al.*, 2022; Ahrens, 2024)

*International Commission on Zoological Nomenclature (Rheindt et al., 2023)

Integrative taxonomy and molecular diagnoses

- Integrative taxonomy incorporates diverse information sources (e.g., morphology, DNA sequencing, chromosome mapping), providing a comprehensive understanding of species boundaries.
 - Despite the potential of DNA data to enhance diagnostic precision, incorporating Diagnostic Molecular Combinations (DMCs) into species descriptions remains uncommon.

(Brower, 2010; Renner, 2016; Meier *et al.*, 2022; Ahrens, 2024)

*International Commission on Zoological Nomenclature (Rheindt et al., 2023)

Integrative taxonomy and molecular diagnoses

- Integrative taxonomy incorporates diverse information sources (e.g., morphology, DNA sequencing, chromosome mapping), providing a comprehensive understanding of species boundaries.
 - Despite the potential of DNA data to enhance diagnostic precision, incorporating Diagnostic Molecular Combinations (DMCs) into species descriptions remains uncommon.
 - Moreover, the ICBN* has recently recommended including specific molecular sites in species diagnoses to improve standards.

(Brower, 2010; Renner, 2016; Meier *et al.*, 2022; Ahrens, 2024)

*International Commission on Zoological Nomenclature (Rheindt et al., 2023)

Towards the gold standard in molecular diagnoses

- The gold standard for diagnoses includes:

(Fedosov *et al.*, 2022; Meier *et al.*, 2022; Rheindt *et al.*, 2023)

*Used by Sharkey *et al.*, (2021)

Towards the gold standard in molecular diagnoses

- The gold standard for diagnoses includes:
 - Contrastiveness (differentiation based on comparisons with related species)

(Fedosov *et al.*, 2022; Meier *et al.*, 2022; Rheindt *et al.*, 2023)
 *Used by Sharkey *et al.*, (2021)

Towards the gold standard in molecular diagnoses

- The gold standard for diagnoses includes:
 - Contrastiveness (differentiation based on comparisons with related species)
 - State-specificity (identification of unique molecular sites or combinations of sites)

(Fedosov *et al.*, 2022; Meier *et al.*, 2022; Rheindt *et al.*, 2023)

*Used by Sharkey *et al.*, (2021)

Towards the gold standard in molecular diagnoses

- The gold standard for diagnoses includes:
 - Contrastiveness (differentiation based on comparisons with related species)
 - State-specificity (identification of unique molecular sites or combinations of sites)
 - However, some molecular diagnostic methods fail to meet these standards, resulting in inaccurate species delimitation (e.g., consensus barcodes* lack explicit distinguishing molecular sites).

(Fedosov *et al.*, 2022; Meier *et al.*, 2022; Rheindt *et al.*, 2023)

*Used by Sharkey *et al.*, (2021)

Towards the gold standard in molecular diagnoses

- The gold standard for diagnoses includes:
 - Contrastiveness (differentiation based on comparisons with related species)
 - State-specificity (identification of unique molecular sites or combinations of sites)
 - However, some molecular diagnostic methods fail to meet these standards, resulting in inaccurate species delimitation (e.g., consensus barcodes* lack explicit distinguishing molecular sites).
 - Effective methods to integrate molecular and morphological data are needed!

(Fedosov *et al.*, 2022; Meier *et al.*, 2022; Rheindt *et al.*, 2023)

*Used by Sharkey *et al.*, (2021)

Available tools for molecular diagnoses

- Cladistic Haplotype Aggregation (CHA. Brower, 1999)
- CAOS-R (Sarkar *et al.*, 2008; Bergmann, 2024)
- Spider (Brown *et al.*, 2012)
- FASTACHAR (Merckelbach & Borges, 2020)
- QUIDDICH (Kühn & Hasse, 2020)
- DeSignate (Hütter *et al.*, 2020)
- MOLD (Fedosov *et al.*, 2022)

Received: 8 April 2021

Revised: 19 January 2022

Accepted: 21 January 2022

DOI: 10.1111/1755-0998.13590

RESOURCE ARTICLE

MOLECULAR ECOLOGY
RESOURCES WILEY

MOLD, a novel software to compile accurate and reliable DNA diagnoses for taxonomic descriptions

Alexander Fedosov^{1,2}  | Guillaume Achaz^{2,3,4} | Andrey Gontchar⁵ | Nicolas Puillandre² 



Challenges in molecular diagnoses

- However, the current software tools exhibit some of the following drawbacks:

Fedorov *et al.* (2022) compared MOLD with other contemporary options
Brower & DeSalle (2024) discussed and listed methods for incorporating DNA into taxonomic descriptions.

Challenges in molecular diagnoses

- However, the current software tools exhibit some of the following drawbacks:
 - **Not user-friendly:** non-intuitive interfaces and occasional subjective decision-making.

Fedorov *et al.* (2022) compared MOLD with other contemporary options
Brower & DeSalle (2024) discussed and listed methods for incorporating DNA into taxonomic descriptions.

Challenges in molecular diagnoses

- However, the current software tools exhibit some of the following drawbacks:
 - **Not user-friendly:** non-intuitive interfaces and occasional subjective decision-making.
 - **Significant time consumption:** high-throughput taxonomy analyzes thousands of sequences and hundreds of new species.

Fedorov *et al.* (2022) compared MOLD with other contemporary options
Brower & DeSalle (2024) discussed and listed methods for incorporating DNA into taxonomic descriptions.

Challenges in molecular diagnoses

- However, the current software tools exhibit some of the following drawbacks:
 - **Not user-friendly:** non-intuitive interfaces and occasional subjective decision-making.
 - **Significant time consumption:** high-throughput taxonomy analyzes thousands of sequences and hundreds of new species.
 - **Lack of integrated verification tools:** needed for validating Diagnostic Molecular Combinations (DMCs) and specimen identification.

Fedorov *et al.* (2022) compared MOLD with other contemporary options
Brower & DeSalle (2024) discussed and listed methods for incorporating DNA into taxonomic descriptions.

Challenges in molecular diagnoses

- However, the current software tools exhibit some of the following drawbacks:
 - **Not user-friendly:** non-intuitive interfaces and occasional subjective decision-making.
 - **Significant time consumption:** high-throughput taxonomy analyzes thousands of sequences and hundreds of new species.
 - **Lack of integrated verification tools:** needed for validating Diagnostic Molecular Combinations (DMCs) and specimen identification.
 - **Unreliable DMCs:** DMCs of limited practical use, affecting diagnostic efficiency.

Fedorov *et al.* (2022) compared MOLD with other contemporary options
Brower & DeSalle (2024) discussed and listed methods for incorporating DNA into taxonomic descriptions.

Specificity vs. inclusivity

- The trade-off between specificity and inclusivity is crucial for all diagnosis types (morphological, molecular, etc.).

Specificity vs. inclusivity

- The trade-off between specificity and inclusivity is crucial for all diagnosis types (morphological, molecular, etc.).
- Consequences of imbalances in diagnoses:

Specificity vs. inclusivity

- The trade-off between specificity and inclusivity is crucial for all diagnosis types (morphological, molecular, etc.).
- Consequences of imbalances in diagnoses:
 - **Too specific (i.e., very long DMCs):** may fail to include all specimens of a species, especially those from different geographical regions or collected at different times.

Specificity vs. inclusivity

- The trade-off between specificity and inclusivity is crucial for all diagnosis types (morphological, molecular, etc.).
- Consequences of imbalances in diagnoses:
 - **Too specific (i.e., very long DMCs):** may fail to include all specimens of a species, especially those from different geographical regions or collected at different times.
 - **Too general (i.e., very short DMCs):** may inaccurately group multiple species together, leading to false positives.

Specificity vs. inclusivity

- The trade-off between specificity and inclusivity is crucial for all diagnosis types (morphological, molecular, etc.).
- Consequences of imbalances in diagnoses:
 - **Too specific (i.e., very long DMCs):** may fail to include all specimens of a species, especially those from different geographical regions or collected at different times.
 - **Too general (i.e., very short DMCs):** may inaccurately group multiple species together, leading to false positives.
- Tools like MOLD use improved algorithms to address these issues, though challenges remain with high-throughput taxonomy datasets (see below).

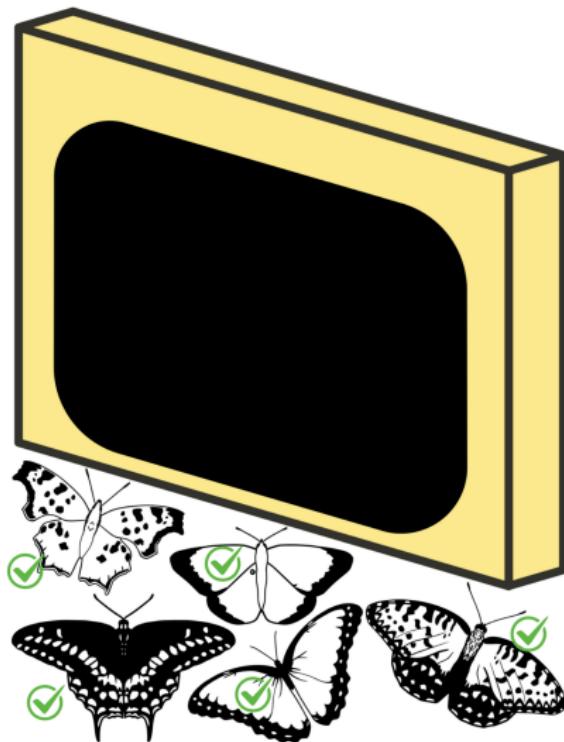
The goal of diagnoses (The "Hole in the Wall" game)



Lovely (unrealistic?) diagnoses



Diagnoses (too general / very short DMCs)



Diagnoses (too specific / very long DMCs)



In this presentation we:

- Introduce UITOTO: a user-friendly R package (including a Shiny app) designed to address key challenges of molecular diagnoses in high-throughput taxonomy.

In this presentation we:

- Introduce UITOTO: a user-friendly R package (including a Shiny app) designed to address key challenges of molecular diagnoses in high-throughput taxonomy.
- Provide an empirical evaluation of UITOTO using three datasets: Megaselidae, Mycetophylidae, and European butterflies.

In this presentation we:

- Introduce UITOTO: a user-friendly R package (including a Shiny app) designed to address key challenges of molecular diagnoses in high-throughput taxonomy.
- Provide an empirical evaluation of UITOTO using three datasets: Megaselidae, Mycetophylidae, and European butterflies.
- Discuss efficient molecular diagnoses and offer detailed guidelines for achieving optimal results, particularly with UITOTO.

1 Introduction

2 The UITOTO R package

3 Testing UITOTO

4 Discussion and conclusions

5 Acknowledgements

6 References

UITOTO



Home

Find DMCs

Tax. Verif./Ident. with DMCs

Visualizer

UITOTO: User InTerface for Optimal molecular diagnoses in high-throughput TaxOnomy

Center for Integrative Biodiversity Discovery (CIBD)

Museum für Naturkunde - Leibniz Institute for Evolution and Biodiversity Science

Find Diagnostic Molecular Combinations (DMCs)

Taxonomic verification and identification using DMCs

Visualize DMCs

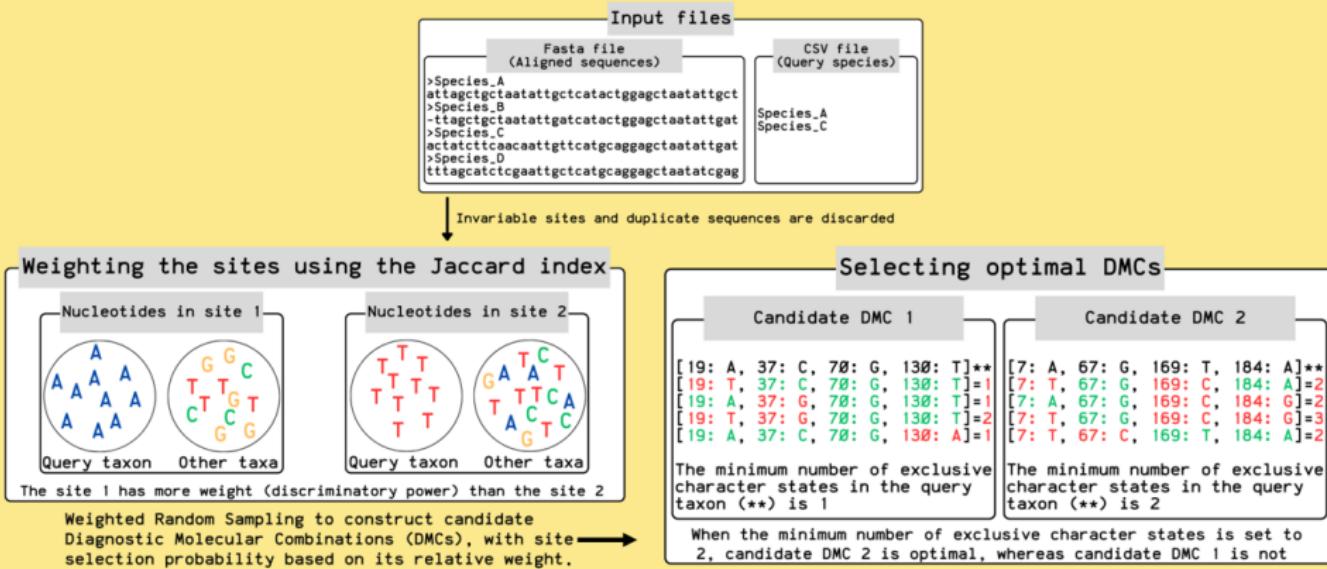


Author: Ambrosio Torres G. - Code:

UITOTO GitHub Repository: <https://github.com/atorresgalvis/UITOTO>
UITOTO Shiny App: <https://atorresgalvis.shinyapps.io/MolecularDiagnoses/>



Identification of Diagnostic Molecular Combinations (DMCs)



- Then, it selects the most frequent sites in the optimal DMCs to build one final DMC.

Taxonomic Verification and Identification Using DMCs

- UITOTO provides robust tools for analyzing both aligned and unaligned sequences through three functions:

Taxonomic Verification and Identification Using DMCs

- UITOTO provides robust tools for analyzing both aligned and unaligned sequences through three functions:
 - **Identifier:** matches unknown specimens based on the provided DMC patterns.

Taxonomic Verification and Identification Using DMCs

- UITOTO provides robust tools for analyzing both aligned and unaligned sequences through three functions:
 - **Identifier:** matches unknown specimens based on the provided DMC patterns.
 - **ALnID:** align the unknown sequences and then operates similarly to Identifier.

Taxonomic Verification and Identification Using DMCs

- UITOTO provides robust tools for analyzing both aligned and unaligned sequences through three functions:
 - **Identifier:** matches unknown specimens based on the provided DMC patterns.
 - **ALnID:** align the unknown sequences and then operates similarly to Identifier.
 - **IdentifierU:** employs an alignment-free approach with a dynamic sliding window.

Taxonomic Verification and Identification Using DMCs

- UITOTO provides robust tools for analyzing both aligned and unaligned sequences through three functions:
 - Identifier: matches unknown specimens based on the provided DMC patterns.
 - ALnID: align the unknown sequences and then operates similarly to Identifier.
 - IdentifierU: employs an alignment-free approach with a dynamic sliding window.
- Users can set permissible mismatches to enhance flexibility in taxonomic identification.

Taxonomic Verification and Identification Using DMCs

- UITOTO provides robust tools for analyzing both aligned and unaligned sequences through three functions:
 - **Identifier:** matches unknown specimens based on the provided DMC patterns.
 - **ALnID:** align the unknown sequences and then operates similarly to Identifier.
 - **IdentifierU:** employs an alignment-free approach with a dynamic sliding window.
- Users can set permissible mismatches to enhance flexibility in taxonomic identification.
- Results are useful to validate and refine DMCs through cross-validation with known taxonomic assignments.

Visualization of DMCs

- Generate customizable publication-quality DMC comparisons and visualizations.

File with the available DMCs (e.g., Output from the OpDMC approach) ①:

DMCs.csv

Upload complete

Hex color code for sequence names:

Hex color code for the sites:

Hex color code for DMC's sites:

Fasta file with the sequences to draw the DMCs of the species included in the previous file ②:

sequences.txt

Upload complete

Font size for sequence titles (approx. pt):

Font size for sequence sites (approx. pt):

Font size for DMC's sites (approx. pt):

Use the sequences containing the string/word!!! ③:

Letter case:

Font type for sequence names:

Font type for sequence sites:

Output format:

Include DMCs after sequence names Sequence names in bold

Sequence names in italic Sequence sites in bold

Nucleotides color (available for the Format 3):

Palette color 1: A C G T

Palette color 2: A C G T

Palette color 3: A C G T

Palette color 4: A C G T

④ Download visualization here

Species_A_holotype [19: A, 37: C, 46: A, 56: C, 70: G]

Species_B_holotype [22: C, 34: T, 128: G, 129: T, 140: C]

Species_C_holotype [11: T, 31: A, 125: T, 127: C, 145: T]

Note: This tool is available only in the Shiny version of UITOTO.



1 Introduction

2 The UITOTO R package

3 Testing UITOTO

4 Discussion and conclusions

5 Acknowledgements

6 References

Data Acquisition (*Cytochrome c Oxidase I* sequences)

- Genus *Megaselia* (Diptera: Phoridae)
 - Seq. DMC searching: 2,229 (313 bp) from Singapore, representing 69 new species (Lee *et al.*, *In prep*).
 - Seq. DMC testing: 30,289 (313 bp) from the Indo-Pacific region (Lee *et al.*, *In prep*).

Data Acquisition (*Cytochrome c Oxidase I* sequences)

- Genus *Megaselia* (Diptera: Phoridae)
 - Seq. DMC searching: 2,229 (313 bp) from Singapore, representing 69 new species (Lee *et al.*, *In prep*).
 - Seq. DMC testing: 30,289 (313 bp) from the Indo-Pacific region (Lee *et al.*, *In prep*).
- Mycetophilidae (Diptera: Bibionomorpha)
 - Seq. DMC searching: 1,456 (313 bp) from Singapore, representing 115 new species (Amorin *et al.*, 2024).
 - Seq. DMC testing: 60,349 from BOLD (498-658 bp, Mean = 605 bp) (Ratnasingham & Hebert, 2007).

Data Acquisition (*Cytochrome c Oxidase I* sequences)

- Genus *Megaselia* (Diptera: Phoridae)
 - Seq. DMC searching: 2,229 (313 bp) from Singapore, representing 69 new species (Lee *et al.*, *In prep*).
 - Seq. DMC testing: 30,289 (313 bp) from the Indo-Pacific region (Lee *et al.*, *In prep*).
- Mycetophilidae (Diptera: Bibionomorpha)
 - Seq. DMC searching: 1,456 (313 bp) from Singapore, representing 115 new species (Amorin *et al.*, 2024).
 - Seq. DMC testing: 60,349 from BOLD (498-658 bp, Mean = 605 bp) (Ratnasingham & Hebert, 2007).
- European Butterflies (Lepidoptera)
 - Seq. DMC searching: 591 (658 bp) from the UK, representing 49 species (Dincă *et al.*, 2021).
 - Seq. DMC testing: 21,483 from the rest of Europe (113-1,498 bp, Mean = 656 bp) (Dincă *et al.*, 2021).

Search for Diagnostic Molecular Combinations (DMCs)

- We performed seven different runs using the OpDMD function of UITOTO with varying settings*:
 - Three searches with fixed DMC minimum length ($MnLen = 6$) and varying exclusive (2, 3, 4)
 - Three searches with free DMC minimum length**, varying exclusive (2, 3, 4)
 - One search with fixed $MnLen = 10$ and exclusive = 4
- We also conducted a DMC search using MOLD (Fedosov *et al.*, 2022) with default settings for comparison.

*Note: all searches involved 50,000 iterations with a refinement strength of 0.33, treating gaps as missing data.

**Example: A DMC with exclusive = 2 has a minimum possible length of two (2).

Assessing the Reliability of DMCs (cross-validation)

- We utilized two different functions of UITOTO for matching 'unknown' sequences against the provided DMC patterns:
 - Identifier: ***Megaselia*** (aligned sequences for testing)
 - ALnID: **Mycetophilidae** and **European butterflies** (unaligned sequences for testing)

Assessing the Reliability of DMCs (cross-validation)

- We utilized two different functions of UITOTO for matching 'unknown' sequences against the provided DMC patterns:
 - Identifier: ***Megaselia*** (aligned sequences for testing)
 - ALnID: **Mycetophilidae** and **European butterflies** (unaligned sequences for testing)
- We used two different tolerances for mismatches between the DMCs and the sequences:
 - Allowing up to one mismatch
 - Zero tolerance for mismatches

Assessing the Reliability of DMCs (cross-validation)

- We utilized two different functions of UITOTO for matching 'unknown' sequences against the provided DMC patterns:
 - Identifier: **Megaselia** (aligned sequences for testing)
 - ALnID: **Mycetophilidae** and **European butterflies** (unaligned sequences for testing)
- We used two different tolerances for mismatches between the DMCs and the sequences:
 - Allowing up to one mismatch
 - Zero tolerance for mismatches
- We developed the function **ConfuClass** to generate:
 - Confusion matrices
 - Classification statistics (e.g., precision, recall, specificity, accuracy, F1 score)

Classification statistics I

- **Recall:** measures the DMC ability to correctly recognize the species of interest.

$$\text{Recall} = \frac{\text{TruePos}}{\text{TruePos} + \text{FalseNeg}}$$

- **Precision:** measures how often the identified species are truly the target species.

$$\text{Precision} = \frac{\text{TruePos}}{\text{TruePos} + \text{FalsePos}}$$

- **Specificity:** measures the DMC ability to correctly recognize species that are not of interest.

$$\text{Specificity} = \frac{\text{TrueNeg}}{\text{TrueNeg} + \text{FalsePos}}$$

Classification statistics II

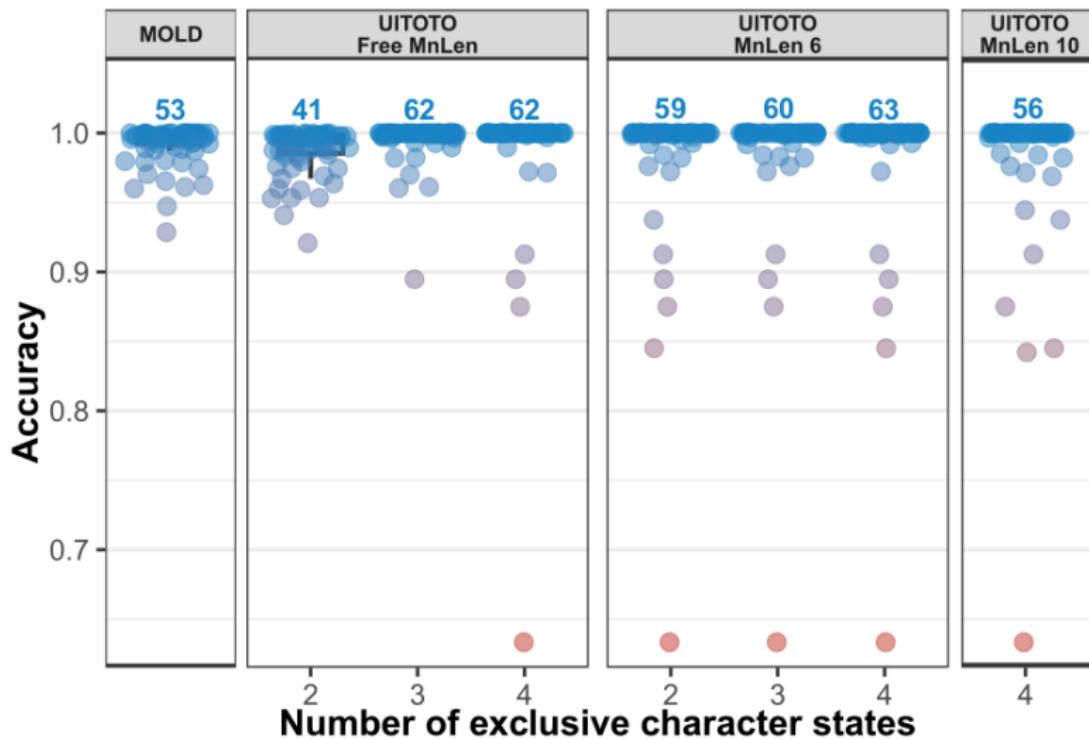
- **Accuracy:** measures the average of recall and specificity to provide a balanced view of DMC performance.

$$\text{Accuracy} = \frac{\text{Recall} + \text{Specificity}}{2}$$

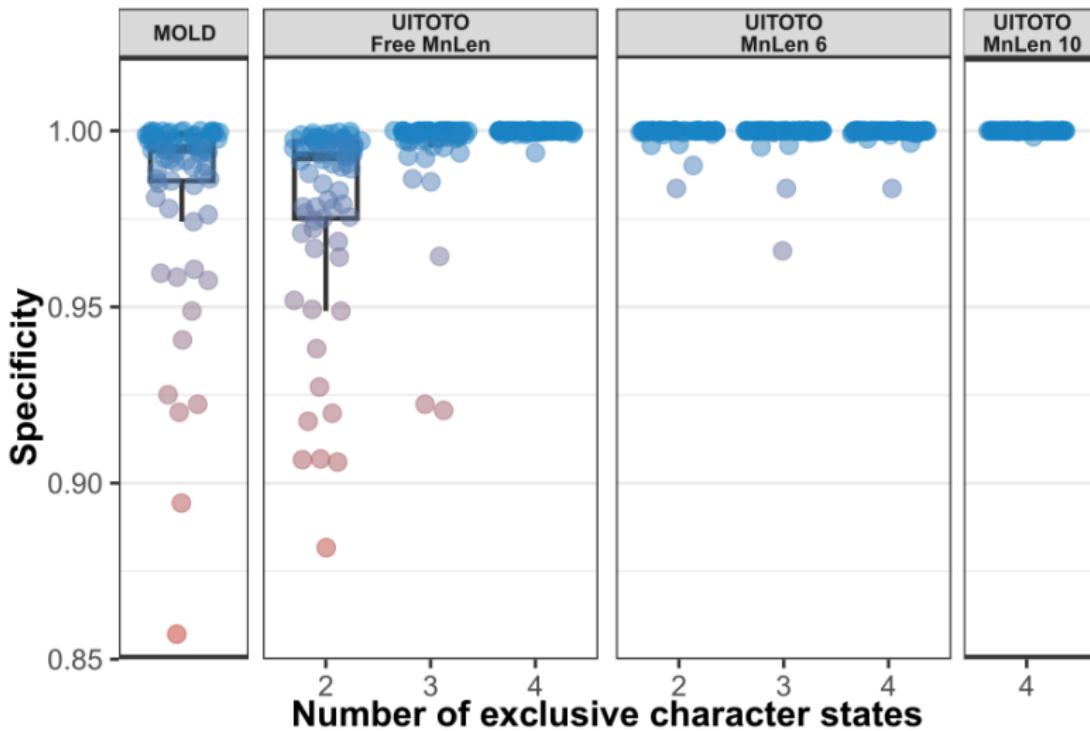
- **F1 Score:** provides a single metric that balances both the ability to find the target species and the Precision in doing so.

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

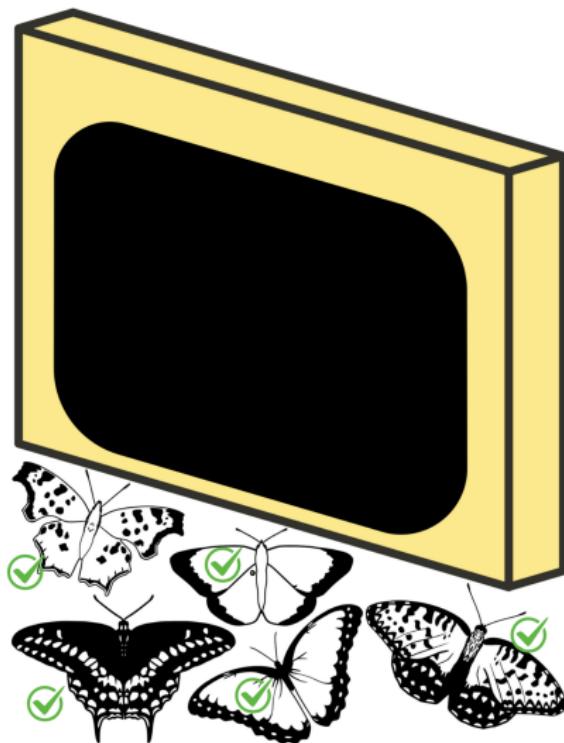
DMC Accuracy (*Megaselia*)



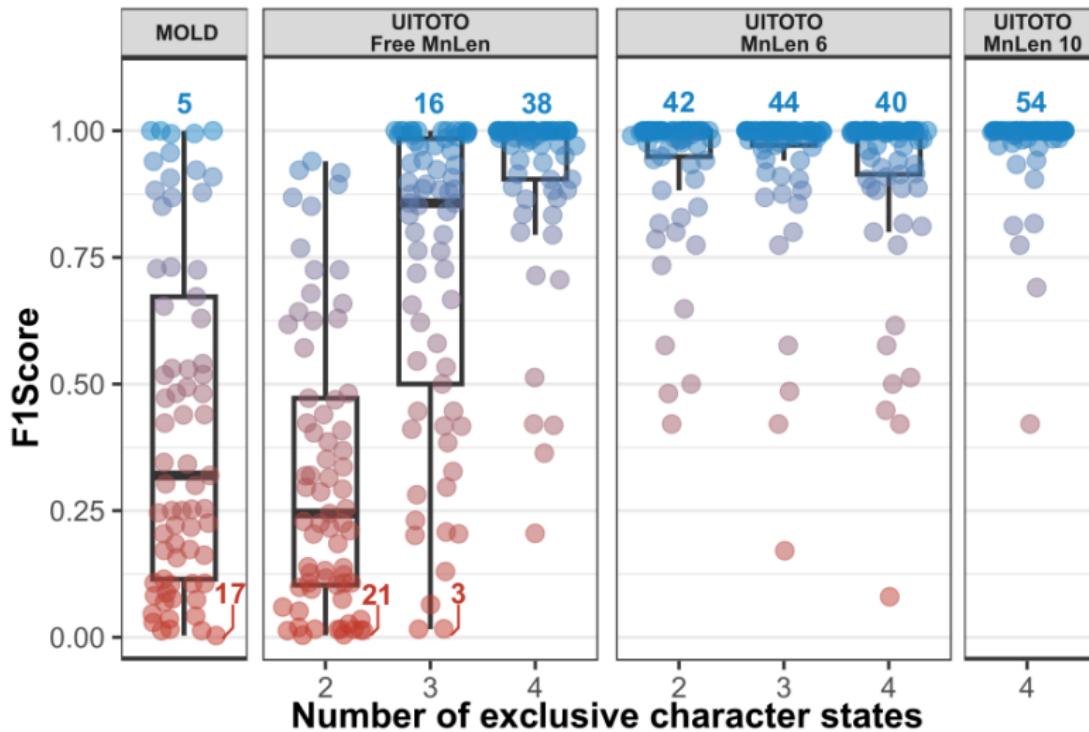
DMC Specificity (*Megaselia*)



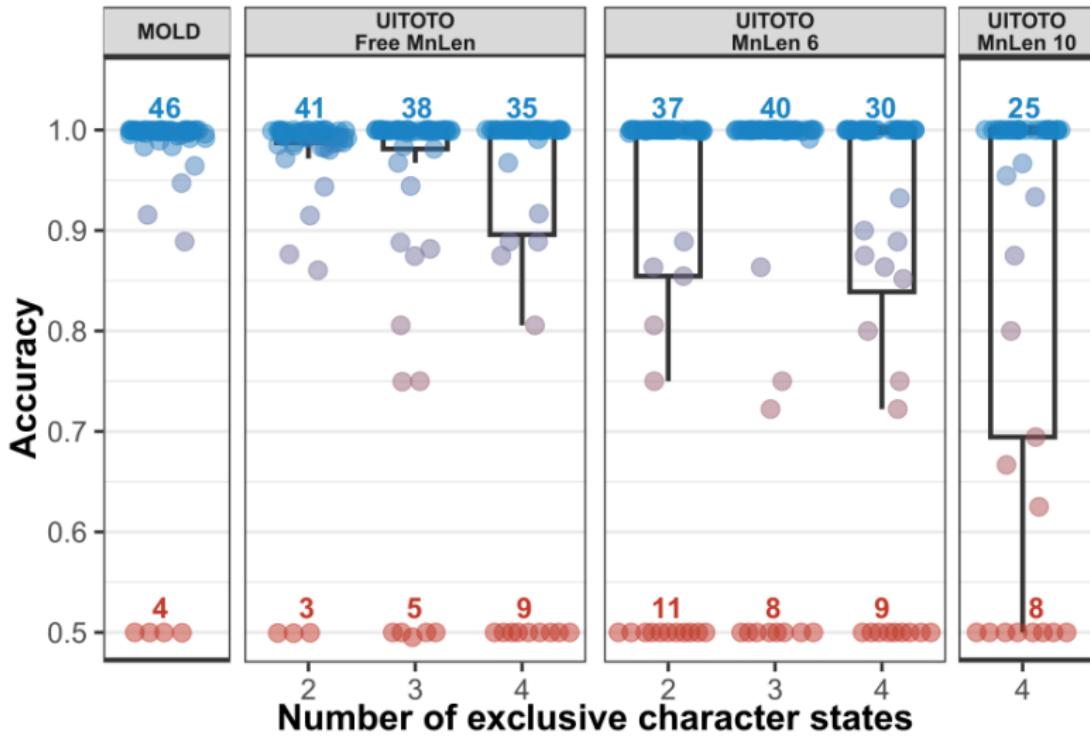
So...



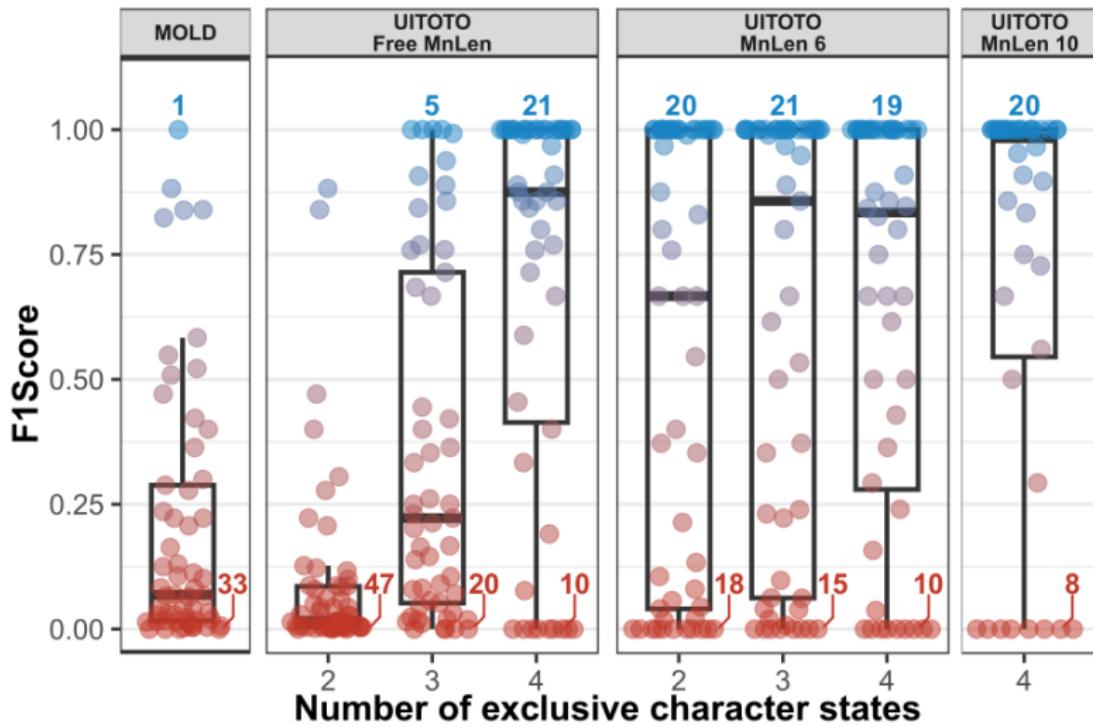
DMC F1 Score (*Megaselia*)



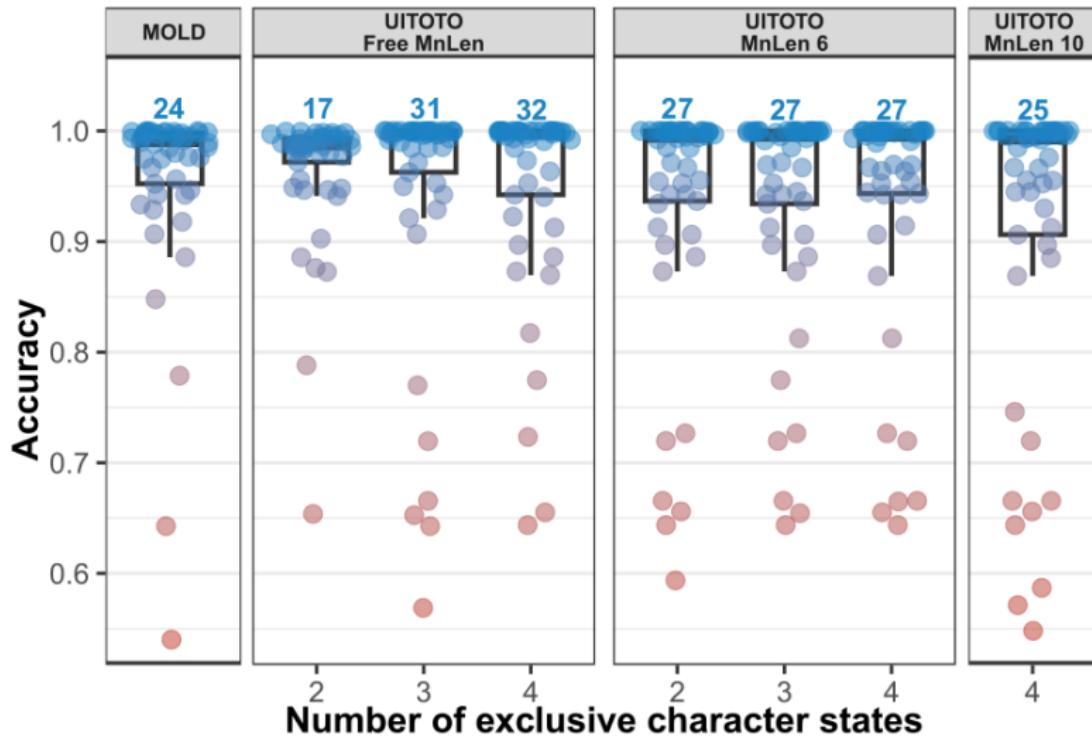
DMC Accuracy (Mycetophilidae)



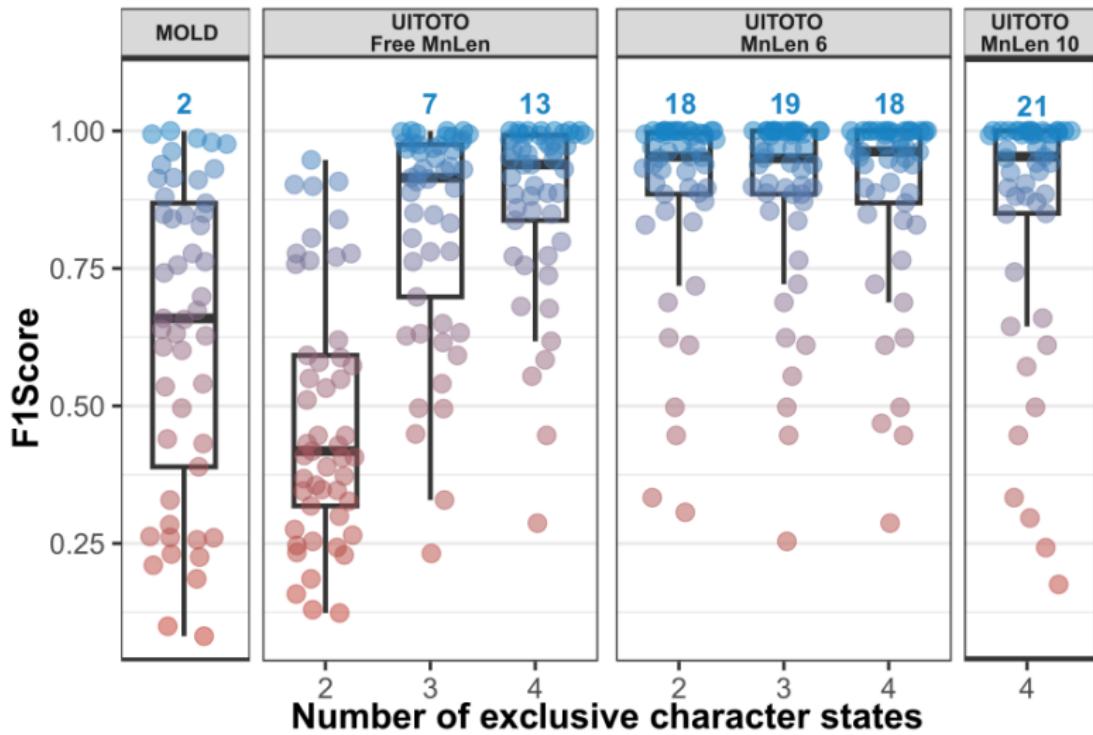
DMC F1 Score (Mycetophilidae)



DMC Accuracy (European butterflies)



DMC F1 Score (European butterflies)



① Introduction

② The UITOTO R package

③ Testing UITOTO

④ Discussion and conclusions

⑤ Acknowledgements

⑥ References

Then, what should we do?

- Use different settings for performing the DMC searches.
Or...you can start with 3 (or 4) exclusive character states for the DMCs and let the DMC length vary freely.

Then, what should we do?

- Use different settings for performing the DMC searches.
Or...you can start with 3 (or 4) exclusive character states for the DMCs and let the DMC length vary freely.
- Use appropriate pipelines for testing the reliability of the DMCs! This includes using high numbers of sequences for your cross-validation analyses.

Then, what should we do?

- Use different settings for performing the DMC searches.
Or...you can start with 3 (or 4) exclusive character states for the DMCs and let the DMC length vary freely.
- Use appropriate pipelines for testing the reliability of the DMCs! This includes using high numbers of sequences for your cross-validation analyses.
- Take into account that the settings for obtaining optimal DMCs may vary among species.

Then, what should we do?

- Use different settings for performing the DMC searches.
Or...you can start with 3 (or 4) exclusive character states for the DMCs and let the DMC length vary freely.
- Use appropriate pipelines for testing the reliability of the DMCs! This includes using high numbers of sequences for your cross-validation analyses.
- Take into account that the settings for obtaining optimal DMCs may vary among species.
- Consider the taxonomic level of your study when identifying and testing DMCs.

Conclusions

- The UITOTO R package provides a robust and intuitive platform for molecular diagnoses, balancing DMC length with specificity (reliable DMCs).

Conclusions

- The UITOTO R package provides a robust and intuitive platform for molecular diagnoses, balancing DMC length with specificity (reliable DMCs).
- Empirical evaluations using three datasets covering different insect groups and taxonomic levels (genus, family, and order) demonstrated the good performance of UITOTO, also outperforming the software MOLD.

Conclusions

- The UITOTO R package provides a robust and intuitive platform for molecular diagnoses, balancing DMC length with specificity (reliable DMCs).
- Empirical evaluations using three datasets covering different insect groups and taxonomic levels (genus, family, and order) demonstrated the good performance of UITOTO, also outperforming the software MOLD.
- It is highly recommended to use different settings for the DMC searches, as not all groups exhibit the same levels of intraspecific and interspecific variation (this is also observed in morphology).

① Introduction

② The UITOTO R package

③ Testing UITOTO

④ Discussion and conclusions

⑤ Acknowledgements

⑥ References

Thanks For Your Attention!



Contact: atorresgalvis@gmail.com
Ambrosio.TorresGalvis@mfn.berlin

UITOTO: <https://atorresgalvis.shinyapps.io/MolecularDiagnoses/>

① Introduction

② The UITOTO R package

③ Testing UITOTO

④ Discussion and conclusions

⑤ Acknowledgements

⑥ References

References

- Agnarsson, I., Coddington, J.A. (2008). Cladistics, 24, 5161.
- Ahrens, D. (2024). In: DeSalle, R. (eds) DNA barcoding. Methods in Molecular Biology, vol 2744. Humana, New York, NY.
- Amorim, D.S., et al. (2024). Integr. Syst., in press.
- Bergmann, T. (2024). In: DeSalle, R. (eds) DNA Barcoding. Methods in Molecular Biology, vol 2744. Humana, New York, NY.
- Brower, A.V.Z. (2010). Syst. Biodivers., 8(4), 485491.
- Brower, A.V.Z., DeSalle, R. (2024). In: DeSalle, R. (eds) DNA Barcoding. Methods in Molecular Biology, vol 2744. Humana, New York, NY.
- Brown, S.D.J., et al. (2012). Mol. Ecol. Resour., 12, 562565.
- Dincă, V., et al. (2021). Commun. Biol., 4, 315.
- Fedosov, A., et al. (2022). Mol. Ecol. Resour., 22, 20382053.
- Hütter, T., et al. (2020). BMC Bioinformatics, 21, 151.
- Jaccard, P. (1908). Bull. Soc. Vaudoise Sci. Nat., 44, 223270.

References

- Kühn, A.L., Haase, M. (2020). J. Zool. Syst. Evol. Res., 58, 2226.
- Lee, L., et al.. *In prep.*
- Llorente-Bousquets, J. (1990). Fondo de Cultura Económica, México, D.F.
- Meier, R., et al. (2022). Cladistics, 38, 264275.
- Merckelbach, L.M., Borges, L.M.S. (2020). Mol. Ecol. Resour., 20, 17611768.
- Renner, S.S. (2016). Syst. Biol., 65(6), 10851095.
- Richter, S. (2005). Theory Biosci., 124, 105120.
- Rheindt, F.E., et al. (2023). PLoS Biol., 21(8), e3002251.
- Sarkar, I.N., et al. (2008). Mol. Ecol. Resour., 8, 12561259.
- Sharkey, M.J., et al. (2021). Zool. J. Linnean Soc., 192, 430445.
- Ratnasingham, S., Hebert, P.D. (2007). Mol. Ecol. Notes, 7(3), 355364.