SPECIES DETERMINATION INCL. DNA METHODOLOGY

Paul Henning Krogh
THE TRAITS – ECOSYSTEM SERVICE CONCEPT

CMW: Community Weighted trait Means
SPG: Specific environmental Protection Goals
DEFINITIONS

DNA barcoding

- Molecular species identification with the help of ‘barcoding’ genes, which are short sequences of DNA that vary greatly between species.
- Uses Sanger sequencing.

DNA metabarcoding

- Community DNA: Species identification of a mixture of species from a mixture of their DNA, “DNA soup” subject to amplification of barcodes.
- Soil eDNA: Species identification DNA extracted from soil subject to amplification of barcodes.
- Sequencing by NGS.
eDNA WORKFLOW
METABARCODING OF AMPLIFIED MARKER GENES

Sampling → DNA extraction → PCR → Sequencing → Bioinformatics → Stats
COMPARISON BETWEEN WORKFLOWS FOR ONE TAXON

<table>
<thead>
<tr>
<th></th>
<th>Classical</th>
<th>Metabarcoding</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Community DNA</td>
<td>Soil eDNA</td>
</tr>
<tr>
<td><strong>Sampling</strong></td>
<td>20 soil cores</td>
<td>20 soil cores</td>
</tr>
<tr>
<td><strong>Heat extraction</strong></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td><strong>Sieving</strong></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td><strong>Identification and counting</strong></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td><strong>DNA extraction</strong></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td><strong>PCR#1&amp;2 – NGS</strong></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td><strong>Bioinformatics</strong></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td><strong>Statistics</strong></td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
REFERENCE DATABASES

Where do we find proper DNA reference databases?
NCBI/GENBANK and BOLD
They “harvest” from each other
DANISH COLLEMBOLA
COI AND 16S COVERAGE BY REFERENCE DATABASES

Sources:
BOLD boldsystems.org
Allearter allearter-databasen.dk
GERMAN COLLEMBOLA
COI AND 16S COVERAGE BY REFERENCE DATABASES

Sources:
BOLD boldsystems.org
portal.edaphobase.org
Results Summary

Found 137,398 records, with 137,398 records with sequences, forming 5,671 BINs (clusters), with specimens from 81 countries, deposited in 58 institutions.

Of these records, 57,056 have species names, and represent 992 species.

True number of species

True number after curation 461
TWO MORPHOLOGICALLY DISTINGUISHABLE COLLEMBOLANS

Isotomurus palustris

Isotomurus pseudopalustris
INCONSISTENCY BETWEEN BARCODES AND CLASSICAL BINARY SPECIES NAMES

Same name but different barcode on BOLD

**Isotomurus palustris**

Found 26 published records, with 26 records with sequences, forming 4 BINs (clusters), with specimens from 4 countries, deposited in 4 institutions.

Of these records, 26 have species names, and represent 1 species.

**Isotomurus pseudopalustris**

Found 23 published records, with 23 records with sequences, forming 2 BINs (clusters), with specimens from 4 countries, deposited in 2 institutions.

Of these records, 23 have species names, and represent 1 species.
NCBI BLAST SEARCH RESULT OF A 158 BP MINIBARCODE

<table>
<thead>
<tr>
<th>Description</th>
<th>Max Score</th>
<th>Total Score</th>
<th>Query Cover</th>
<th>E value</th>
<th>Per. Identi</th>
<th>Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isotomidae sp. BIOUG21966-G08 cytochrome oxidase subunit I (COI) gene, partial cds; mito</td>
<td>252</td>
<td>252</td>
<td>95%</td>
<td>1e-66</td>
<td>96.69%</td>
<td>MG038791.1</td>
</tr>
<tr>
<td>Seira delamarei isolate Sd_7_4 cytochrome c oxidase subunit I (COI) gene, partial cds; mito</td>
<td>167</td>
<td>167</td>
<td>98%</td>
<td>5e-41</td>
<td>85.90%</td>
<td>MH579456.1</td>
</tr>
<tr>
<td>Seira delamarei isolate Sd_7_3 cytochrome c oxidase subunit I (COI) gene, partial cds; mito</td>
<td>167</td>
<td>167</td>
<td>98%</td>
<td>5e-41</td>
<td>85.90%</td>
<td>MH579456.1</td>
</tr>
<tr>
<td>Seira delamarei isolate Sd_7_2 cytochrome c oxidase subunit I (COI) gene, partial cds; mito</td>
<td>167</td>
<td>167</td>
<td>98%</td>
<td>5e-41</td>
<td>85.90%</td>
<td>MH579454.1</td>
</tr>
<tr>
<td>Seira delamarei isolate Sd_7_1 cytochrome c oxidase subunit I (COI) gene, partial cds; mito</td>
<td>167</td>
<td>167</td>
<td>98%</td>
<td>5e-41</td>
<td>85.90%</td>
<td>MH579453.1</td>
</tr>
<tr>
<td>Seira delamarei cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial</td>
<td>167</td>
<td>167</td>
<td>98%</td>
<td>5e-41</td>
<td>85.90%</td>
<td>KM978370.1</td>
</tr>
<tr>
<td>Desoria sp. 2 JRD-2015 voucher BIOUG06014-F07 cytochrome oxidase subunit I (COI) gene</td>
<td>161</td>
<td>161</td>
<td>98%</td>
<td>2e-39</td>
<td>85.26%</td>
<td>KM622345.1</td>
</tr>
<tr>
<td>Desoria sp. 2 JRD-2015 voucher BIOUG06737-D10 cytochrome oxidase subunit I (COI) gene</td>
<td>161</td>
<td>161</td>
<td>98%</td>
<td>2e-39</td>
<td>85.26%</td>
<td>KM621117.1</td>
</tr>
<tr>
<td>Desoria sp. 2 JRD-2015 voucher BIOUG05179-H09 cytochrome oxidase subunit I (COI) gene</td>
<td>161</td>
<td>161</td>
<td>98%</td>
<td>2e-39</td>
<td>85.26%</td>
<td>KM617258.1</td>
</tr>
<tr>
<td>Desoria sp. 2 JRD-2015 voucher BIOUG04852-H03 cytochrome oxidase subunit I (COI) gene</td>
<td>161</td>
<td>161</td>
<td>98%</td>
<td>2e-39</td>
<td>85.26%</td>
<td>KM617233.1</td>
</tr>
</tbody>
</table>
### BOLD SEARCH RESULT OF A 158 BP MINIBARCODE

<table>
<thead>
<tr>
<th>Kingdom</th>
<th>Phylum</th>
<th>Class</th>
<th>Order</th>
<th>Species</th>
<th>Similarity</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arthropoda</td>
<td>Collembola</td>
<td>Entomobryomorpha Isotomidae</td>
<td>Isotomus pseudopalustris</td>
<td>98.67</td>
<td>Published</td>
<td></td>
</tr>
<tr>
<td>Arthropoda</td>
<td>Collembola</td>
<td>Entomobryomorpha Isotomidae</td>
<td>Isotomus pseudopalustris</td>
<td>98.67</td>
<td>Published</td>
<td></td>
</tr>
<tr>
<td>Arthropoda</td>
<td>Collembola</td>
<td>Entomobryomorpha Isotomidae</td>
<td>Isotomus pseudopalustris</td>
<td>98.67</td>
<td>Published</td>
<td></td>
</tr>
<tr>
<td>Arthropoda</td>
<td>Collembola</td>
<td>Entomobryomorpha Isotomidae</td>
<td>Isotomus pseudopalustris</td>
<td>98.67</td>
<td>Published</td>
<td></td>
</tr>
<tr>
<td>Arthropoda</td>
<td>Collembola</td>
<td>Entomobryomorpha Isotomidae</td>
<td>Isotomus pseudopalustris</td>
<td>98.67</td>
<td>Published</td>
<td></td>
</tr>
<tr>
<td>Arthropoda</td>
<td>Collembola</td>
<td>Entomobryomorpha Isotomidae</td>
<td>Isotomus pseudopalustris</td>
<td>98.67</td>
<td>Published</td>
<td></td>
</tr>
<tr>
<td>Arthropoda</td>
<td>Collembola</td>
<td>Entomobryomorpha Isotomidae</td>
<td>Isotomus pseudopalustris</td>
<td>98.67</td>
<td>Published</td>
<td></td>
</tr>
<tr>
<td>Arthropoda</td>
<td>Collembola</td>
<td>Entomobryomorpha Isotomidae</td>
<td>Isotomus pseudopalustris</td>
<td>98.65</td>
<td>Early-Release</td>
<td></td>
</tr>
<tr>
<td>Arthropoda</td>
<td>Collembola</td>
<td>Entomobryomorpha Isotomidae</td>
<td>Isotomus pseudopalustris</td>
<td>98</td>
<td>Private</td>
<td></td>
</tr>
<tr>
<td>Arthropoda</td>
<td>Collembola</td>
<td>Entomobryomorpha Isotomidae</td>
<td>Isotomus pseudopalustris</td>
<td>98</td>
<td>Private</td>
<td></td>
</tr>
<tr>
<td>Arthropoda</td>
<td>Collembola</td>
<td>Entomobryomorpha Isotomidae</td>
<td>Isotomus pseudopalustris</td>
<td>98</td>
<td>Early-Release</td>
<td></td>
</tr>
<tr>
<td>Arthropoda</td>
<td>Collembola</td>
<td>Entomobryomorpha Isotomidae</td>
<td>Isotomus pseudopalustris</td>
<td>96.67</td>
<td>Early-Release</td>
<td></td>
</tr>
<tr>
<td>Arthropoda</td>
<td>Collembola</td>
<td>Entomobryomorpha Isotomidae</td>
<td>Isotomus pseudopalustris</td>
<td>96.67</td>
<td>Early-Release</td>
<td></td>
</tr>
</tbody>
</table>
Lumbricus terrestris and Lumbricus herculeus

James et al. 2010
L. herculeus DISTRIBUTION BY BOLD
L. terrestris DISTRIBUTION BY BOLD
NO-TILL VS TILLAGE FIELD STUDY
PCA of the log transformed earthworm fresh mass showing the first (PC1) and second principal component scores (PC2) representing 44% of the variability.
A multivariate analysis (PCA) of earthworm conventionally estimated biomass and metabarcoding DNA reads from Illumina NGS. Samples came from two locations, Aulum and Jerslev, with contrasting farming tillage practices.

CT: Conventional Tillage; RT: Reduced Tillage; CA: Conservation Agriculture
NMDS OF EARTHWORM eDNA PRESENCE-ABSENCE DATA FROM A STUDY OF NO-TILL VS TILL

Aulum

Jerslev

Dim1

Dim2

No-tillage

Tillage

0

0.5

1

1.5

2

2.5

0

0.5

1

1.5

2

2.5

-2

-1

0

1

2

-2

-1

0

1

2
SPECIES LISTS OF COLLEMBOLA OBTAINED BY METABARCODING AND BY MORPHOLOGICAL IDENTIFICATION (AULUM 2017-2018)
SPECIES LISTS OF COLLEMBOLA OBTAINED BY METABARCoding AND BY MORPHOLOGICAL IDENTIFICATION (JERSLEV 2017-2018)
SPECIES LISTS OF EARTHWORMS OBTAINED BY METABARCODING AND BY MORPHOLOGICAL IDENTIFICATION (JERSLEV 2017-2018)
SPECIES LISTS OF EARTHWORMS OBTAINED BY METABARCoding AND BY MORPHOLOGICAL IDENTIFICATION
(AULUM 2017-2018)

<table>
<thead>
<tr>
<th>Metabarcoding</th>
<th>Morphospecies</th>
</tr>
</thead>
<tbody>
<tr>
<td>L. herculeus</td>
<td>L. terrestris</td>
</tr>
<tr>
<td>A. longa</td>
<td>Dendrobaena sp.</td>
</tr>
<tr>
<td>A. rosea</td>
<td>D. rubidus</td>
</tr>
<tr>
<td>A. tuberculata</td>
<td>A. caliginosa</td>
</tr>
<tr>
<td>A. caliginosa</td>
<td>A. tuberculata</td>
</tr>
<tr>
<td>A. chlorotica</td>
<td>A. rosea</td>
</tr>
<tr>
<td>D. octaedra</td>
<td>D. octaedra</td>
</tr>
<tr>
<td>L. rubellus</td>
<td>D. rubidus</td>
</tr>
<tr>
<td>D. rubidus</td>
<td>D. rubidus rubidus</td>
</tr>
<tr>
<td>D. rubidus subrubicundus</td>
<td></td>
</tr>
</tbody>
</table>

Metabarcoding: 7
Morphospecies: 4
Overlap: 2
CONVENTIONAL SAMPLING OF COLLEMBOLA

No. of samples

No. of species

Jerslev

Aulum

No. of samples

No. of species
COLLEMBOLA RAREFACTION OF DNA READS

Jerslev

Aulum

No. of species

samples

No. of species

samples
CONCLUSIONS

- eDNA techniques can detect effects of stressors on soil invertebrate communities
- There is a need to investigate
  - Optimal sample size and number per habitat/treatment
  - Optimal homogenization and subsampling
  - Optimal DNA extraction to address the current actual soil invertebrate biodiversity
- Develop methods to estimate changes in biomass from metabarcoding and metamitogenomes
- The large international DNA databases are goldmines, but you have to curate and make QA to avoid errors