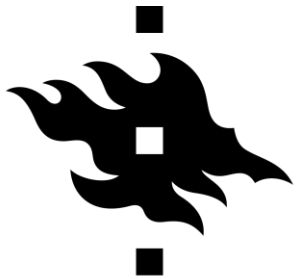
LUOMUS
FINNISH MUSEUM OF NATURAL HISTORY

Use of collections in taxonomic research with a focus on genetic data

Kyung Min Lee

Zoology Unit, LUOMUS

23 August 2022



HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Lecture outline

- DNA barcoding
- Nuclear genes
- Genome assembly
- Genomics / museomics

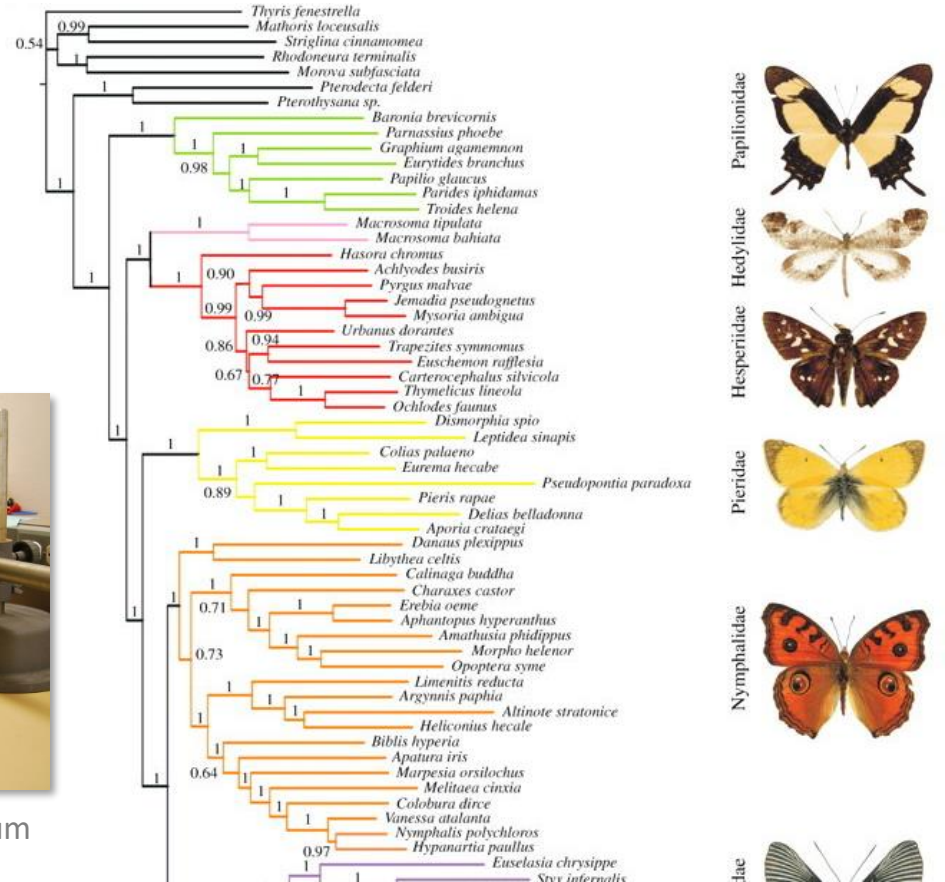
Objectives of the lecture

- To understand the importance of biological collections in taxonomic research
- To understand what type of data can be obtained from collection specimens and samples, with a focus on sequence and genome-level characters
- To learn how the state-of-the-art DNA technology works for museum samples and how it can serve present and future taxonomic research

Taxonomic research

- The science of naming, defining (circumscribing) and classifying groups of biological organisms on the basis of shared characteristics
- Understanding biodiversity
- Order of Evolution
- Many applications (e.g., conservation)

UC Berkely, Jepson Herbarium

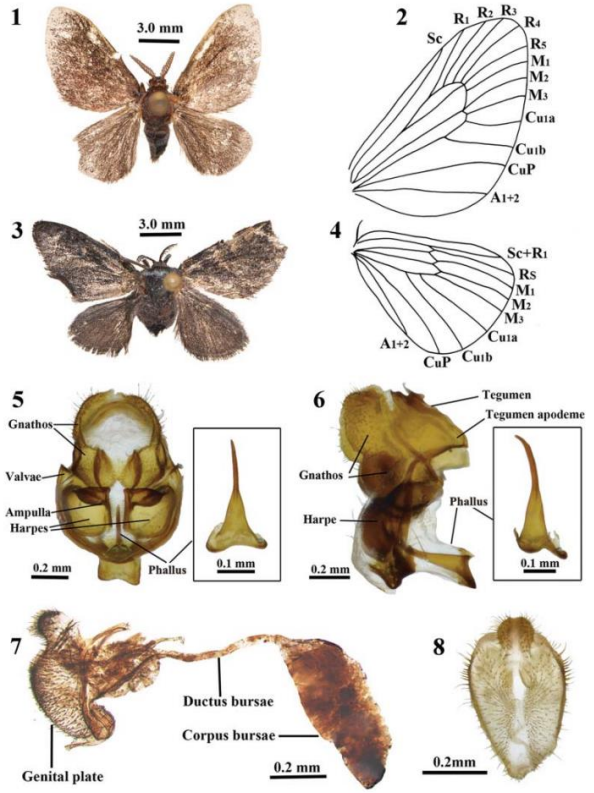


Evolution of DATA used in taxonomic research

Morphology → DNA barcoding → Nuclear genes → High throughput sequencing (HTS)

Wing colour / patterns

Wing venation



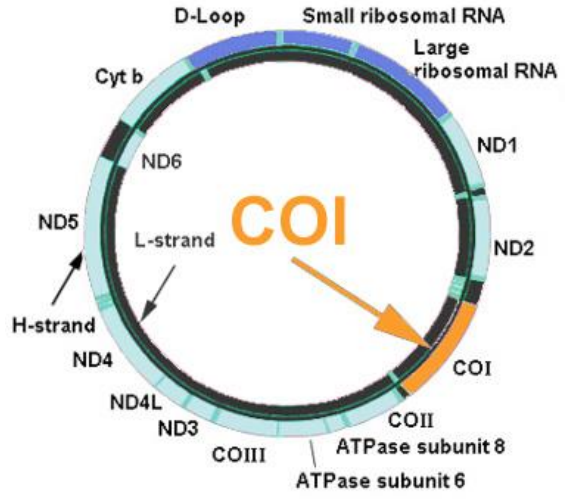
Genitalia of male / female

Single gene approach

maternal inheritance

648 bp

Mitochondria



Multi gene approach

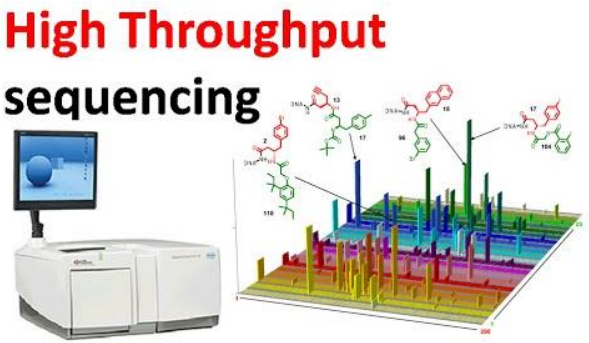
Homologous recombination

Thousands of nucleotides

List of nuclear genes

- ArgK arginine kinase
- CAD carbamoylphosphate synthetase
- RpS5 ribosomal protein S5
- IDH Isocitrate dehydrogenase
- EF1-a Elongation factor 1 alpha
- WGL wingless
- Nex9 sorting nexin-9-like
- MDH cytosolic malate dehydrogenase

Millions of nucleotides



- whole genome sequencing
- reduced representation sequencing (e.g., RAD-seq, anchored enrichment)

Integrative approach

- Morphology
- DNA barcodes
- Genomic data (ddRAD-seq)

ZooKeys 927: 75–97 (2020)
doi: 10.3897/zookeys.927.51142
https://zookeys.pensoft.net

RESEARCH ARTICLE

A peer-reviewed open-access journal
ZooKeys
Launched to accelerate biodiversity research

Revision of the genus *Hoplodrina* Boursin, 1937 (Lepidoptera, Noctuidae, Xyleninae). I. *Hoplodrina octogenaria* (Goeze, 1781) and its sister species *H. alsinides* (Costantini, 1922) sp. rev. in Europe

Peter Huemer[1], Jean Haxaire[2], Kyung Min Lee[3], Marko Mutanen[3], Oleg Pekarsky[4], Stefano Scalercio[5], László Ronkay[6]

1 *Tiroler Landesmuseen Betriebsges.m.b.H., Naturwissenschaftliche Sammlungen, Krajnc-Str. 1, A-6060 Hall, Austria* **2** *Le Roc, 47310 LaPlume, France* **3** *Department of Ecology and Genetics, University of Oulu, PO Box 3000, FI-90014, Oulu, Finland* **4** *Felsőerdősor u. 16-18, H-1068, Budapest, Hungary* **5** *Council for Agriculture Research and Economics, Research Centre for Forestry and Wood, Rende, Italy* **6** *Hungarian Natural History Museum, Budapest, Hungary*

Corresponding author: Peter Huemer (p.huemer@tiroler-landesmuseen.at)

Academic editor: Alberto Zilli | Received 14 February 2020 | Accepted 20 March 2020 | Published 16 April 2020

http://zoobank.org/4908DDE1-C3B5-499E-B003-DFB06A132EE6

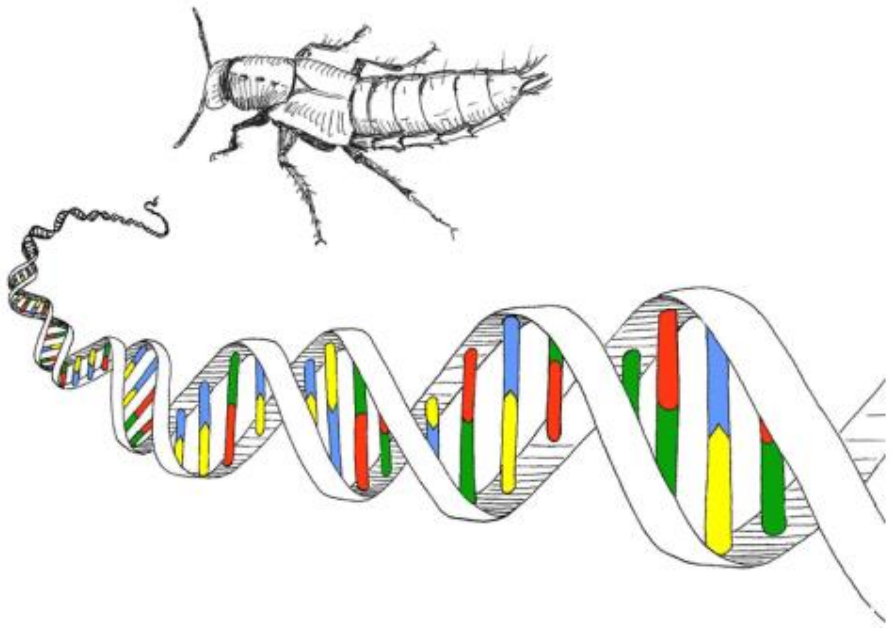
Citation: Huemer P, Haxaire J, Lee KM, Mutanen M, Pekarsky O, Scalercio S, Ronkay L (2020) Revision of the genus *Hoplodrina* Boursin, 1937 (Lepidoptera, Noctuidae, Xyleninae). I. *Hoplodrina octogenaria* (Goeze, 1781) and its sister species *H. alsinides* (Costantini, 1922) sp. rev. in Europe. ZooKeys 927: 75–97. https://doi.org/10.3897/zookeys.927.51142

Abstract

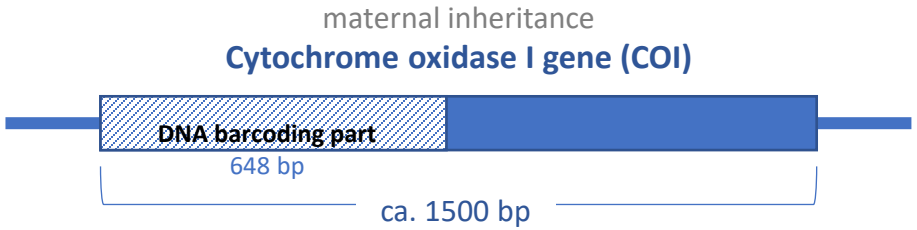
The taxonomic status of the European *Hoplodrina octogenaria* (Goeze, 1781) is discussed and its partly sympatric sister species, *Hoplodrina alsinides* (Costantini, 1922) **sp. rev.**, is separated and re-described based on morphological and molecular taxonomic evidence. The adults and their genitalia are illustrated

DNA barcoding

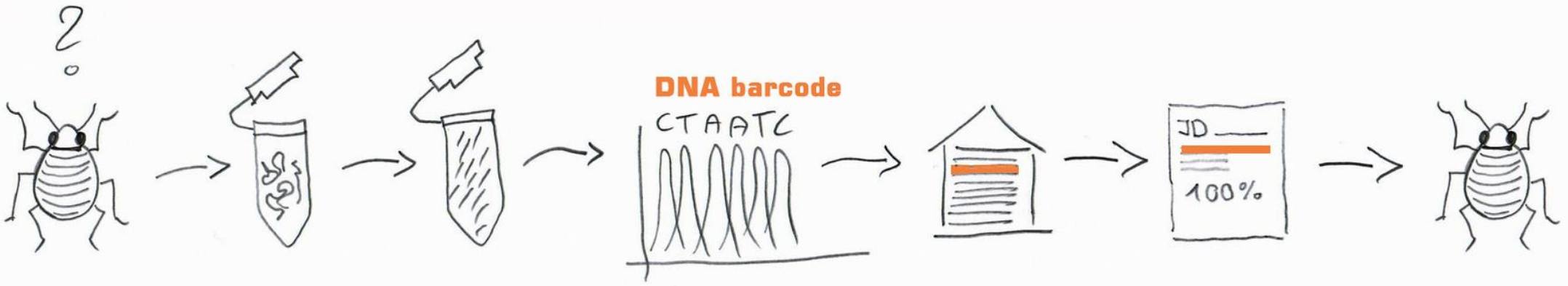
- DNA barcodes is a short sequence that can be used to identify an organism to species
- Standardised DNA region (500-1000 bp)
- Different gene regions are used to identify the different organismal groups (e.g., mitochondrial COI for animals, ITS for fungi, rbcL for plants)



DNA barcoding



How does DNA Barcoding work ?

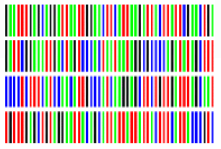


Tissue sampling

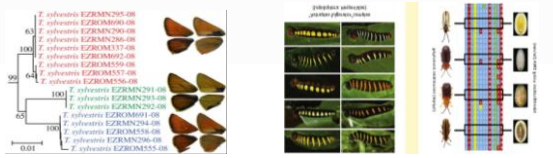
PCR amplification

e.g. CCDB

e.g. BOLD



Species discovery & biodiversity assessment



DNA barcoding

- Inside the core lab facility at the CCDB
- Virtually all automated

DNA barcoding

CCDB
Canadian Centre for DNA Barcoding

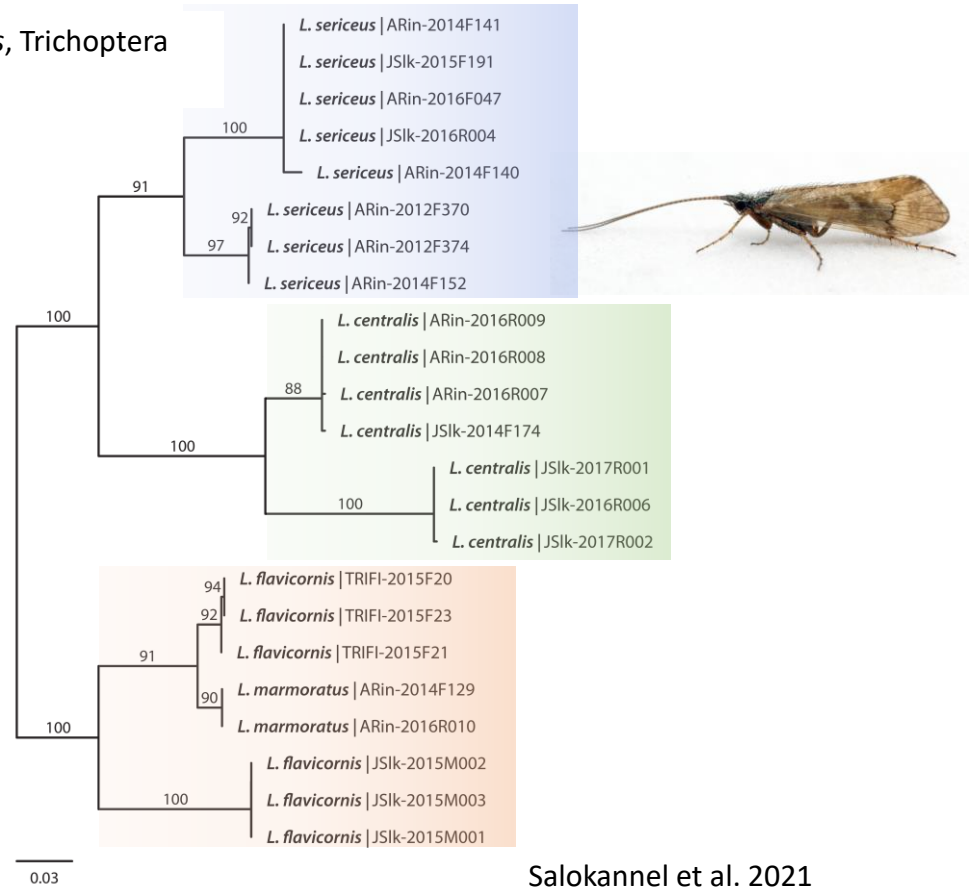
- Data management
  - BOLD system (www.boldsystems.org)
- Workbench for DNA barcoding data
  - Manage
  - Archive
  - Mine
  - Analyse
  - Publish
  - Share

DNA barcoding

- Provides high resolution in shallow relationships (species/population level)

Cryptic (=hidden) diversity / Deep split

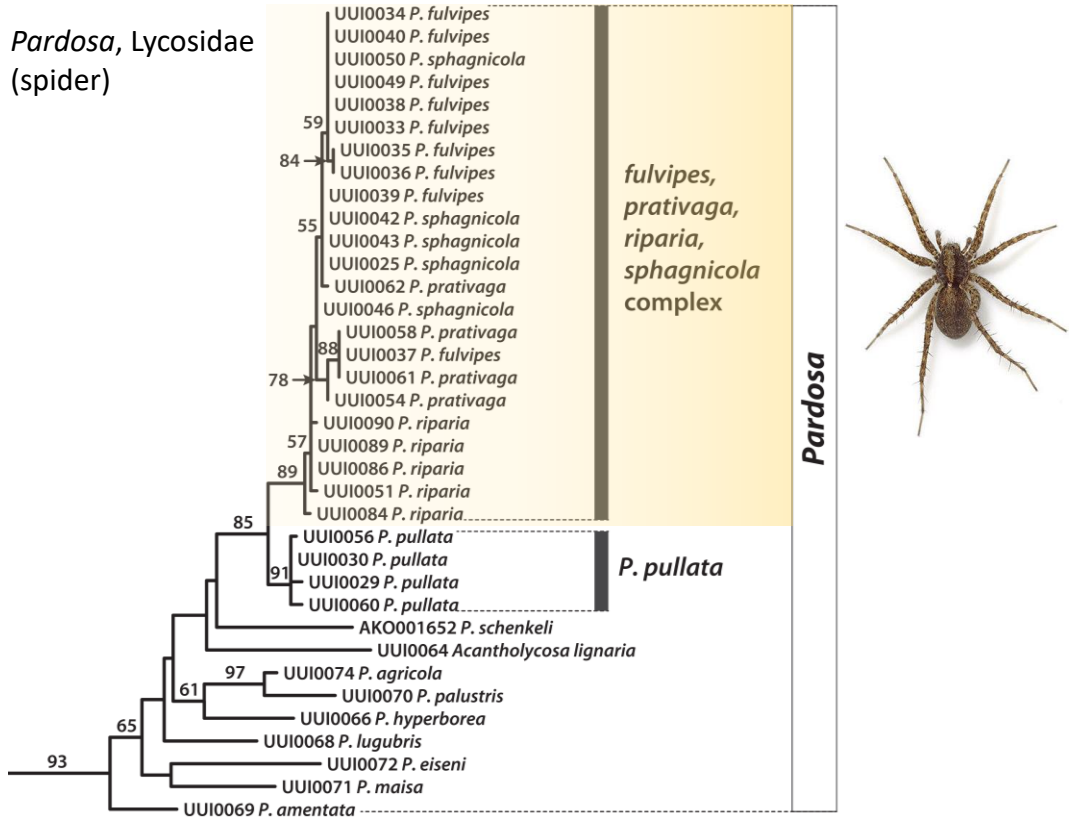
Limnephilus, Trichoptera (caddisfly)



Salokannel et al. 2021

DNA barcode sharing

Pardosa, Lycosidae (spider)

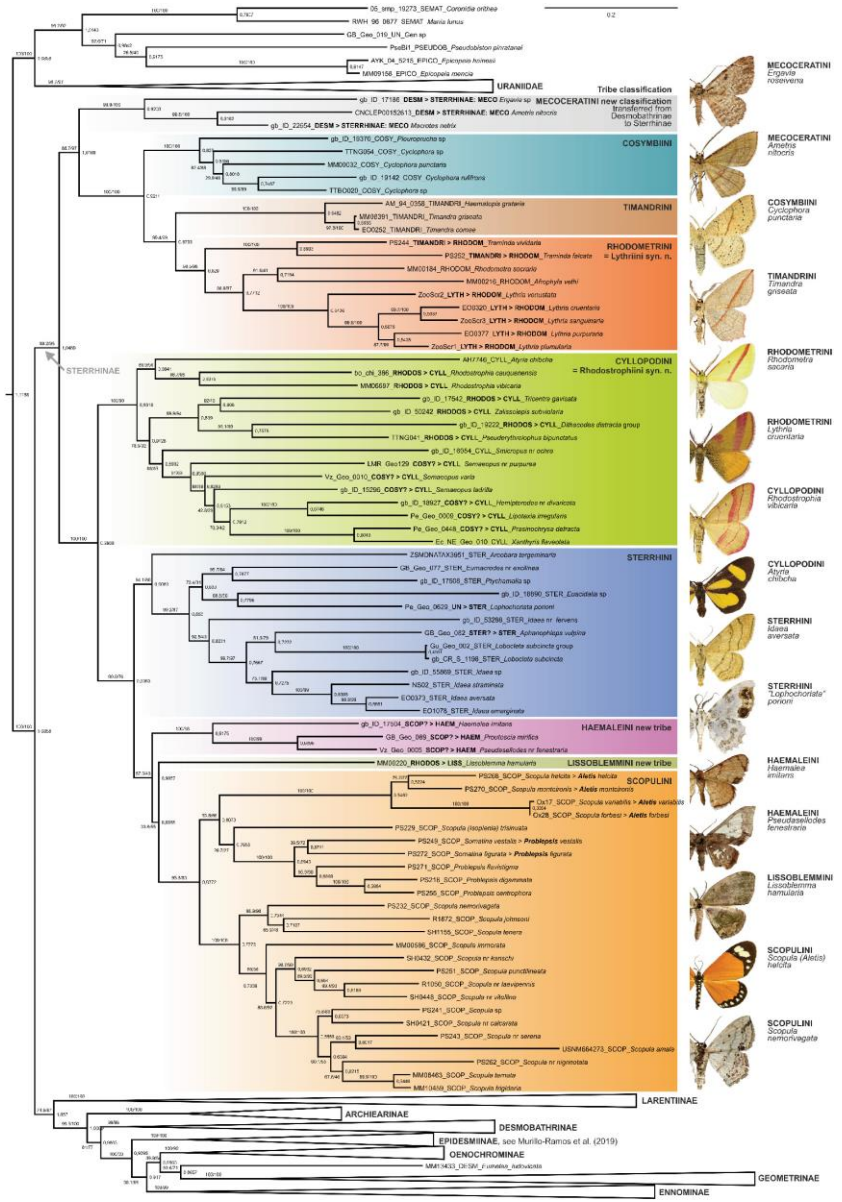


Ivanov et al. 2018

Nuclear genes

- Homologous recombination
- Provide high resolution in deeper relationships (genus, tribe, family, order level)
- List of most commonly used genes

Genes		Length (bp)	Reference
ArgK	arginine kinase	388	Wahlberg et al. 2016
Nex9	sorting nexin-9-like	420	
CAD	carbamoylphosphate synthetase	826	Wahlberg & Wheat 2008
RpS5	ribosomal protein S5	603	
IDH	Isocitrate dehydrogenase	722	
EF1-a	Elongation factor 1 alpha	1047	
WGL	wingless	400	
MDH	cytosolic malate dehydrogenase	407	

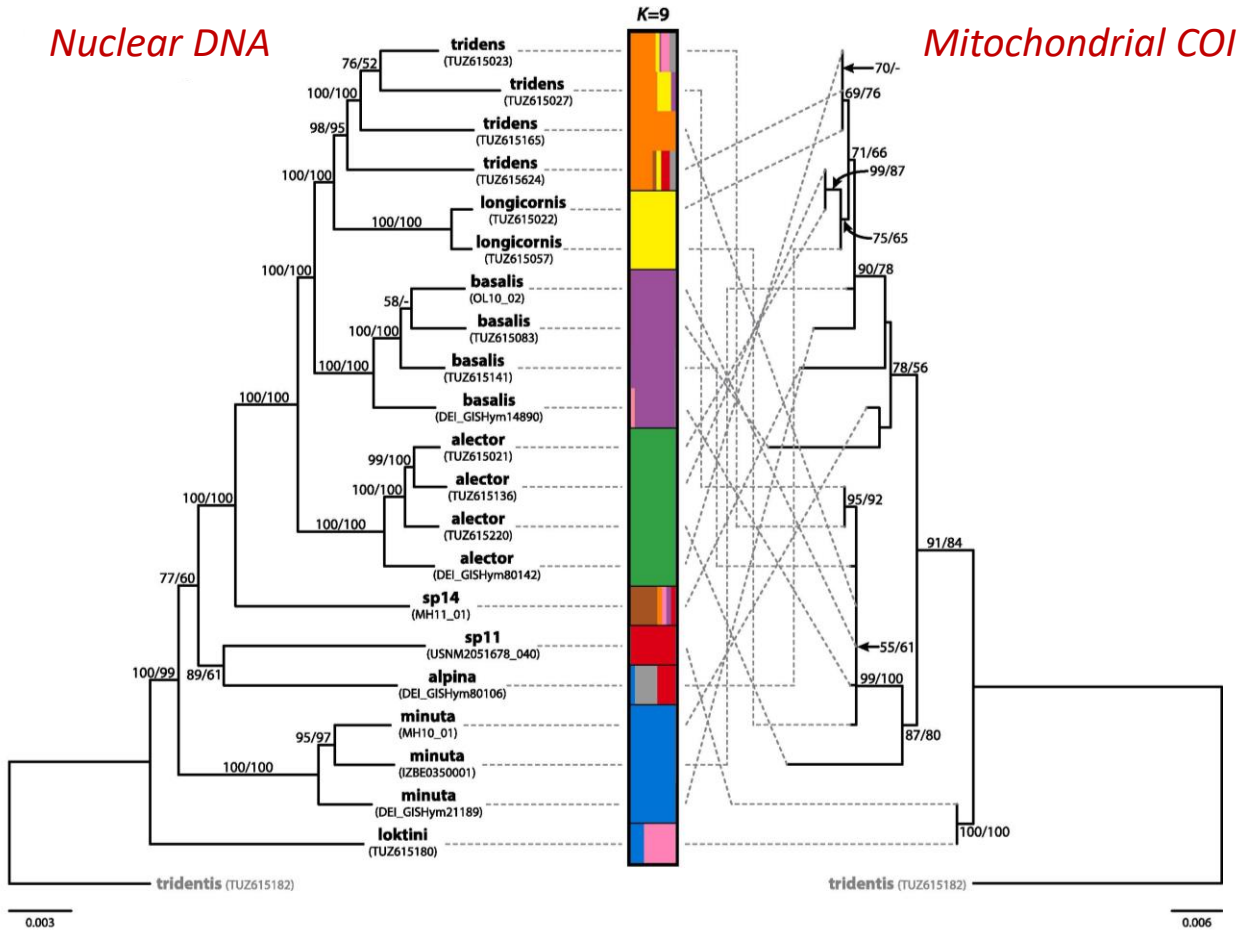


Sihvonen et al. 2020

Nuclear genes

- DNA barcoding VS. nuclear data
- conflict OR congruence?
- Possible reasons behind the inconsistency can be
  - Operational bias (misidentification)
  - Maternally-inherited endosymbiotic bacteria (e.g., *Wolbachia*)
  - Hybridization & incomplete lineage sorting
- Mitochondrial discordance

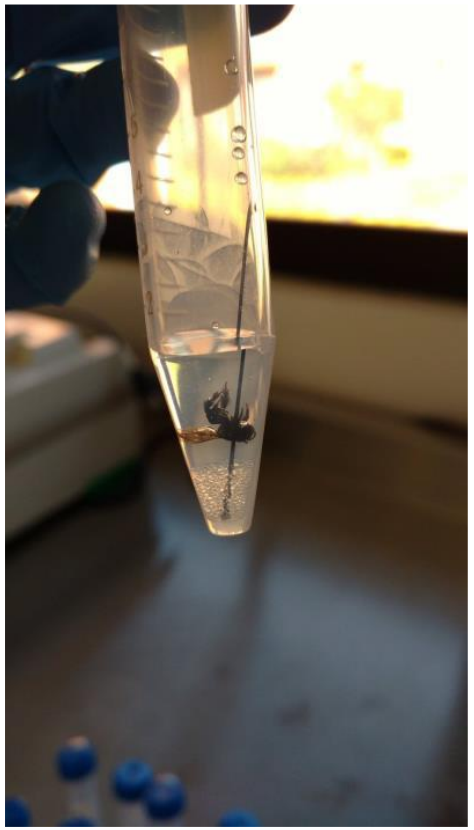
***Empria longicornis* group (sawflies)**



Prous et al. 2020

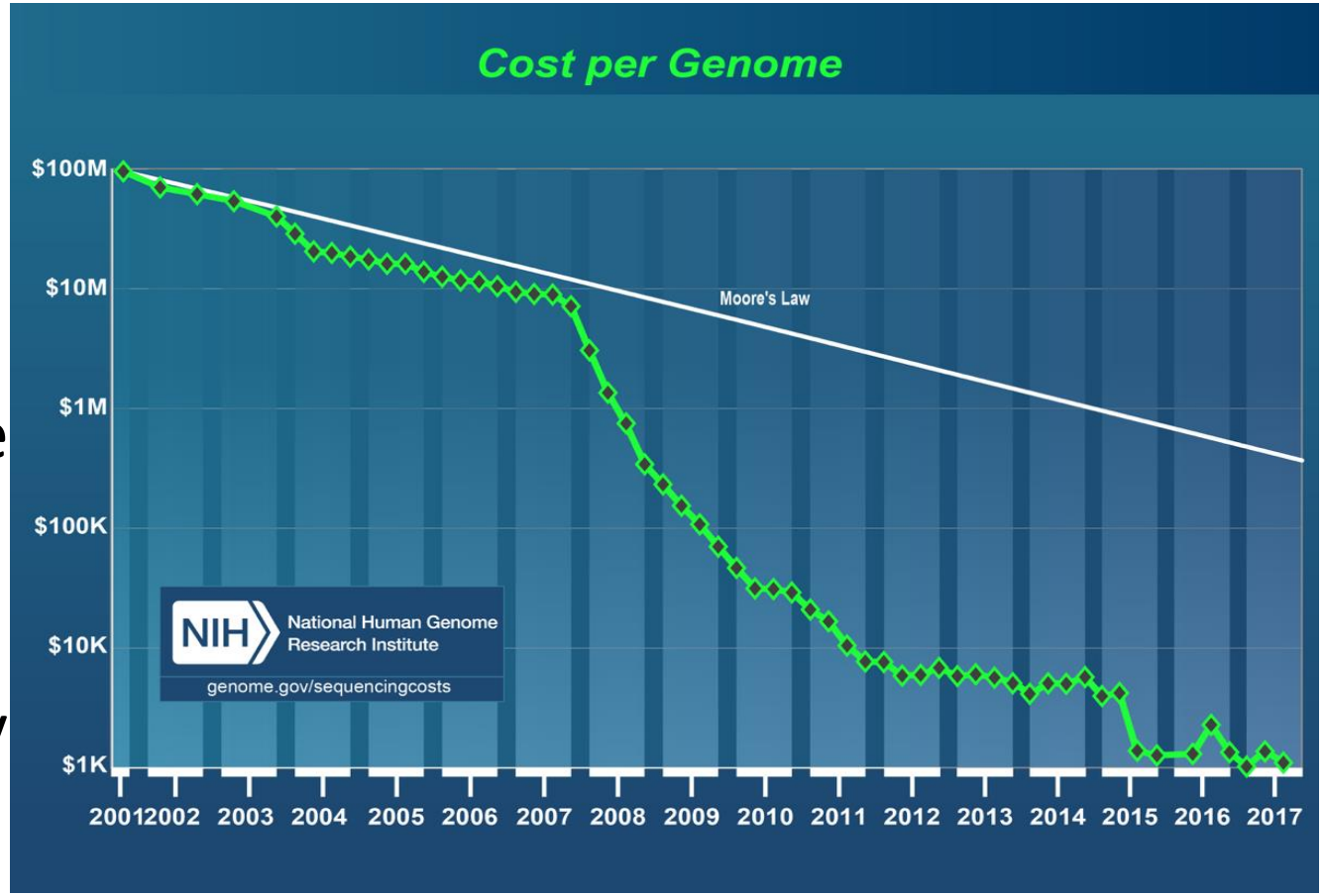
Development of new methods to extract DNA from museum specimens

- e.g. from very old samples or specimens in formaldehyde
- Non-destructive methods (e.g. TYPE specimens)



High-throughput sequencing (HTS)

- Genomics and museomics – also known as next-generation sequencing (NGS)
- Rapid and cost-effective
- HTS techs enable hundre sequenced at a time
- Enable more reliable phy



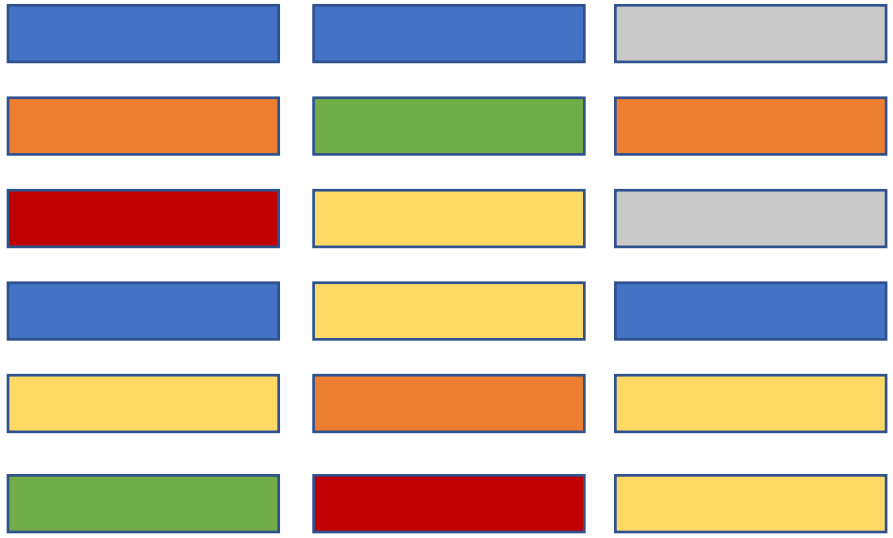
High-throughput sequencing (HTS)

- Whole genome seqeuncing (WGS)
- Reduced representation sequencing
  - Restriction-site associated DNA sequencing (RAD-seq or ddRAD-seq)
  - Target enrichment (TE)

WGS vs Genome reduction method

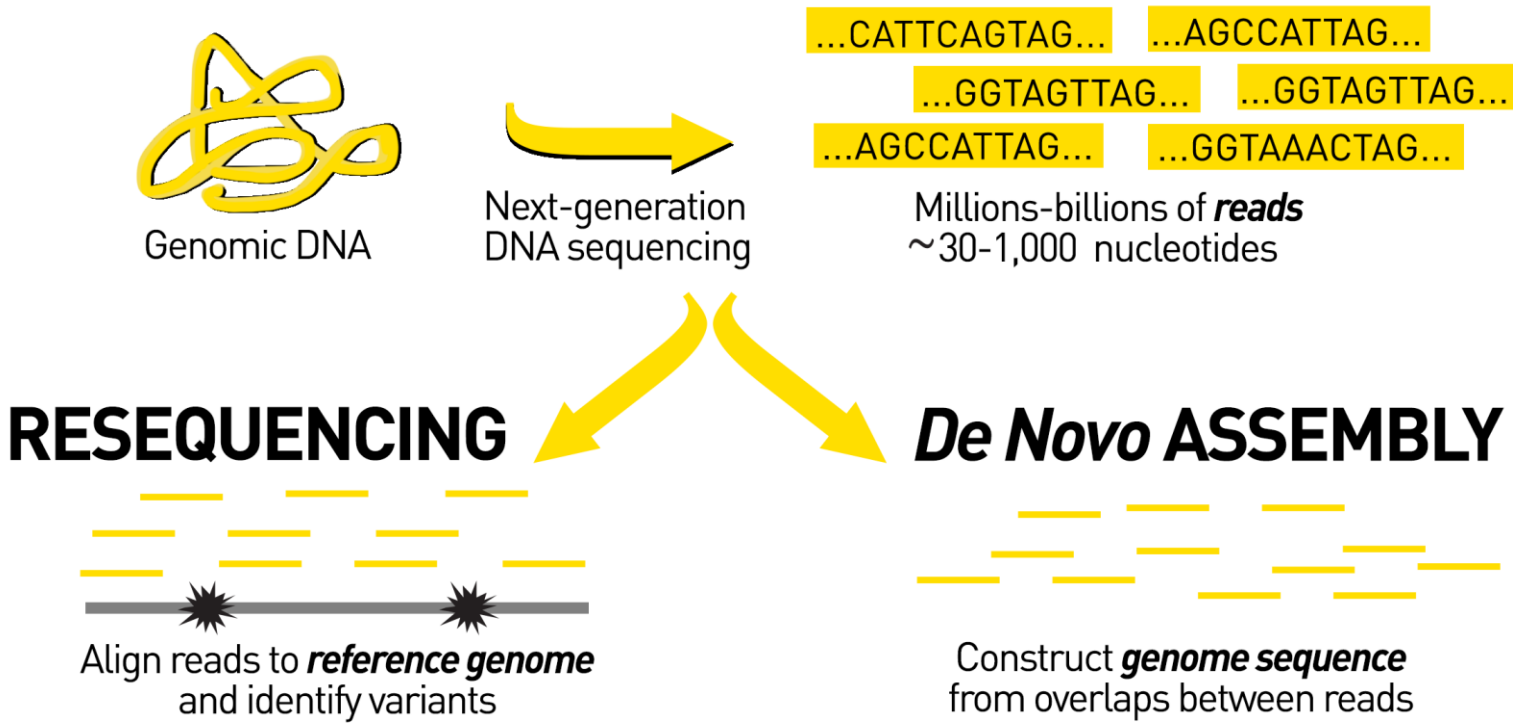
WGS

Genome reduced method

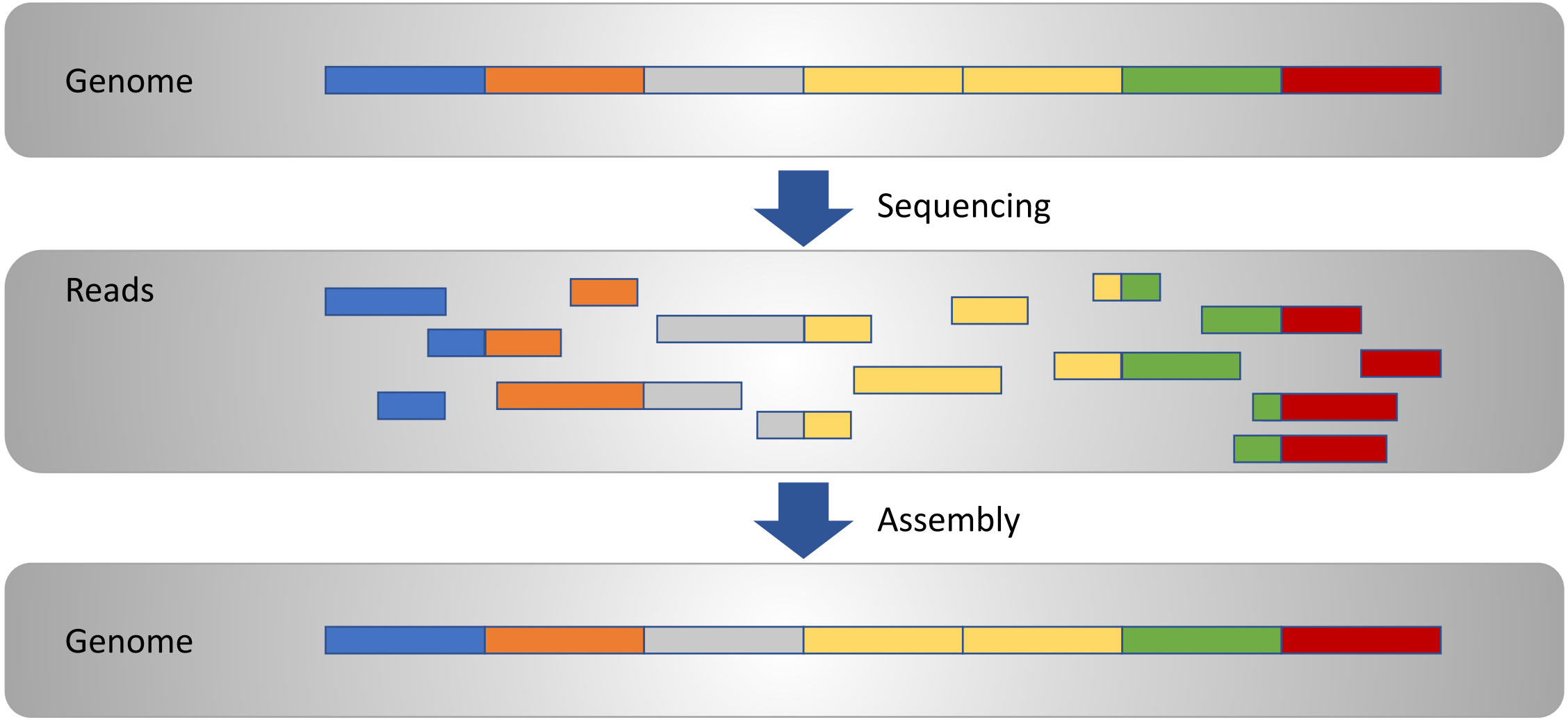


Genome assembly

- De novo assembly – not reference based
- Reference-based assembly
- NCBI (SRA) database



Genome assembly



Genome assembly

- Bioinformatics
- Next generation sequencing analysis
- Visualisation of annotated genomes & assemblies
- SNP variant analysis

Genome assembly

- Bioinformatics
- Assembly & mapping
- Reference mapping with reliable algorithms & de novo assembly

Genome assembly

- Bioinformatics
- Alignment & phylogenetics
- Trusted alignment algorithms, MAFFT & Clustal Omega
- Build phylogenetic trees with RaxML & PAUP*

Museomics

- HTS technologies offer a promise of efficient ways of sequencing degraded DNA
    - HTS involves sequencing of short fragments of DNA, which is characteristic of DNA extracted from old museum specimens, e.g. type specimens
    - Large volumes of sequence data from relatively small amounts of starting material
- Museomics opens up the variety of interesting taxa available to study & the scope of questions that can be investigated in order to further knowledge about biodiversity

"Frankly, I don't see the resemblance."

kaamranhafeez.com

Museomics: placing mysterious genera

Systematic Entomology

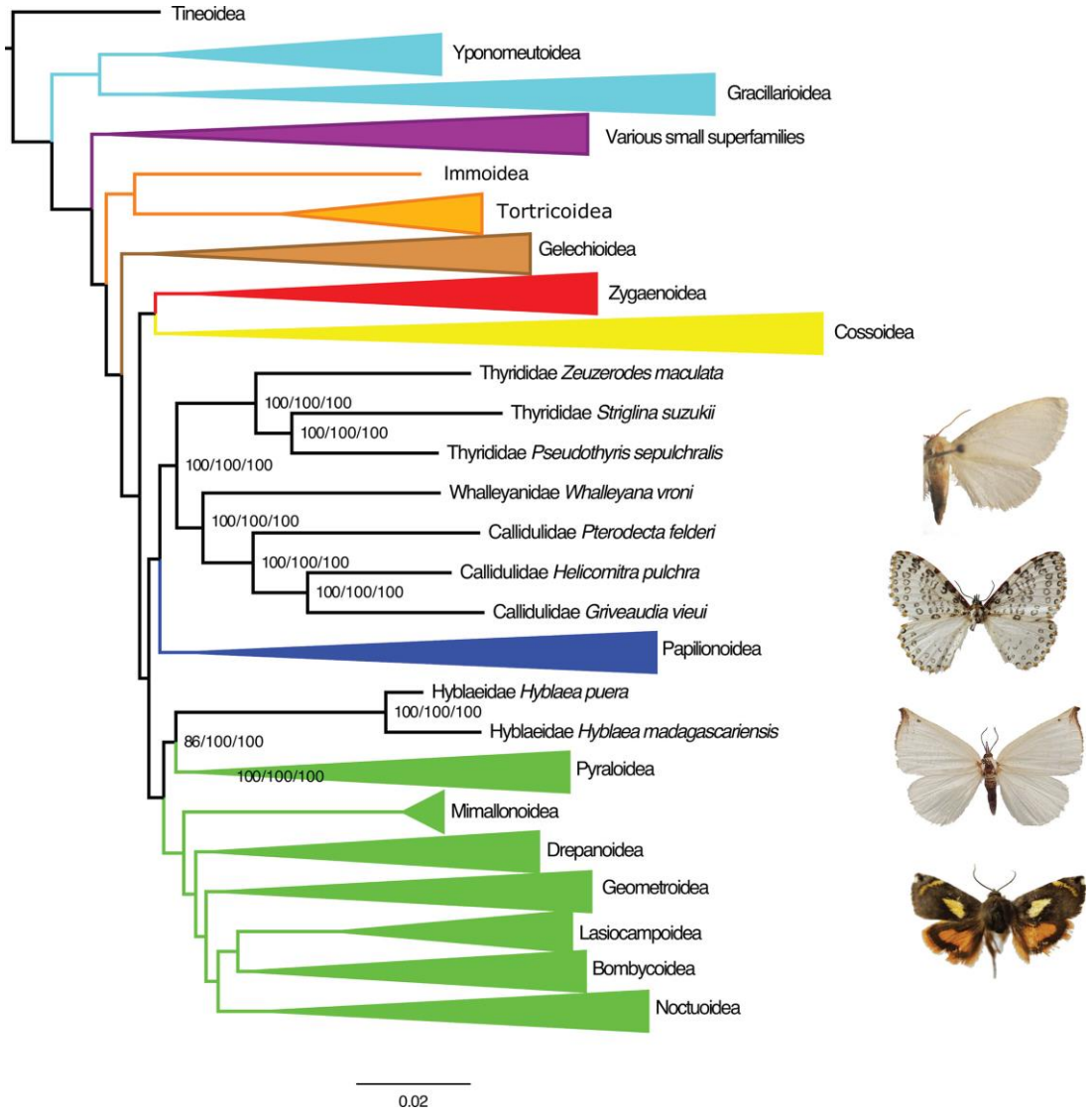
Systematic Entomology (2021), **46**, 926–937 DOI: 10.1111/syen.12503

Museomics of a rare taxon: placing Whalleyanidae in the Lepidoptera Tree of Life

VICTORIA G. TWORT[1,2], JOËL MINET[3], CHRISTOPHER W. WHEAT[4] and NIKLAS WAHLBERG[1]

[1]Department of Biology, Lund University, Lund, Sweden, [2]The Finnish Museum of Natural History Luomus, Zoology Unit, University of Helsinki, Helsinki, Finland, [3]Muséum National d'Histoire Naturelle, ISYEB, Paris, France and [4]Department of Zoology, Stockholm University, Stockholm, Sweden

- Whole genome sequencing (WGS) approach -> 332 genes
- *Whalleyana* species collected between 1969 and 1974
- *De novo* genome assembly



Museomics: exploring the suitability of a genome reduction method on museum specimens

Insect Systematics and Diversity, (2021) 5(2): 6; 1–10
doi: 10.1093/isd/ixaa021
Research

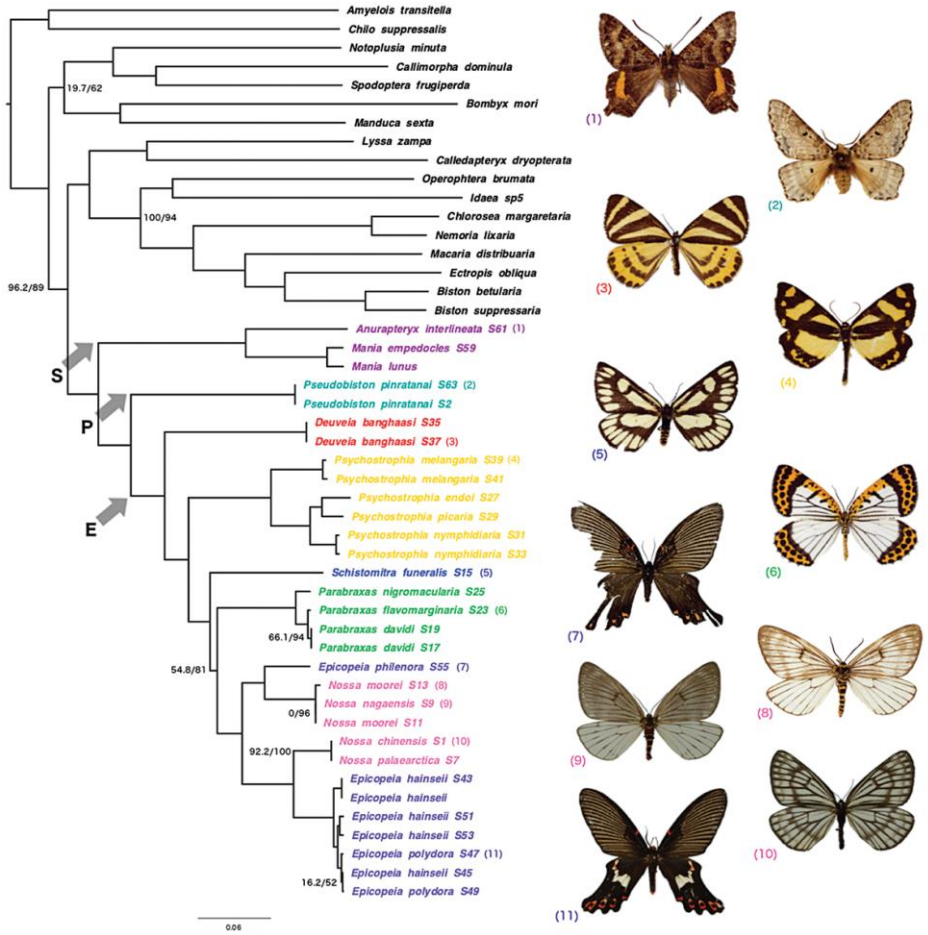
Molecular Phylogenetics, Phylogenomics, and Phylogeography

Museomics: Phylogenomics of the Moth Family Epicopeiidae (Lepidoptera) Using Target Enrichment

Elsa Call,[1,5] Christoph Mayer,[2] Victoria Twort,[1,3] Lars Dietz,[2] Niklas Wahlberg,[1] and Marianne Espeland[4]

[1]Department of Biology, Lund University, 22362 Lund, Sweden, [2]Statistical Phylogenetics and Phylogenomics, Zoological Research Museum Alexander Koenig, 53113 Bonn, Germany, [3]University of Helsinki, Finnish Natural History Museum, Luomus, Helsinki, Finland, [4]Arthropoda Department, Zoological Research Museum Alexander Koenig, 53113 Bonn, Germany, and [5]Corresponding author, e-mail: elsa.call.fr@gmail.com

- Target enrichment (TE) approach
- Museum specimens of Lepidoptera collected between 1892 and 2001
- *De novo* genome assembly



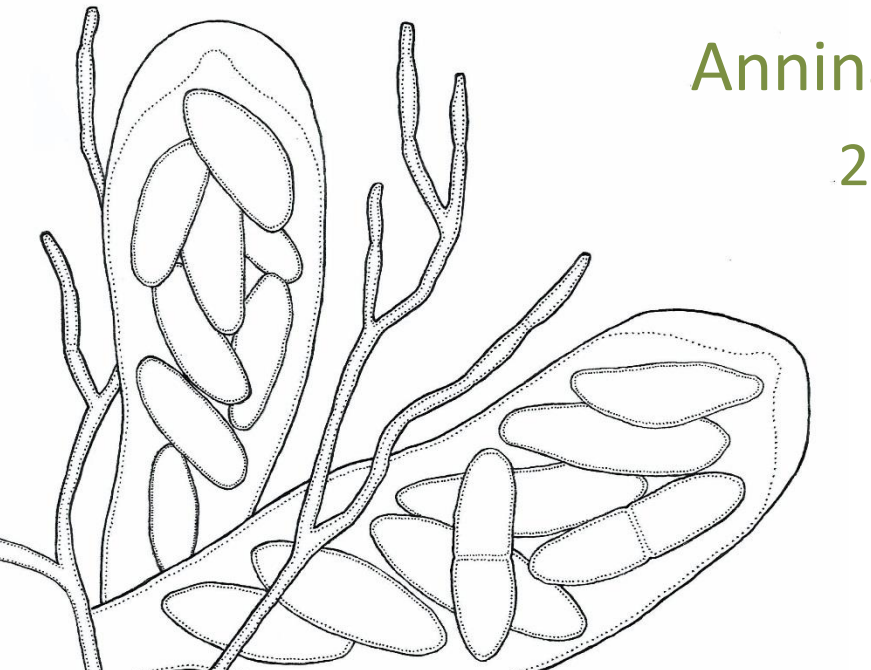
Challenges in museomics

- Development of best-practices in isolating, processing, and analysing historical DNA (hDNA) remain underexplored.
- The quality of hDNA can be largely dependent on preparation types, tissues sources, archival ages, and collecting histories.
- Researchers still face challenges in producing and analysing data.
- Obtaining adequate sequencing coverage, minimizing missing data, correcting for DNA degradation, and removing contaminant DNA are major challenges for genome sequencing of hDNA samples.

Biological collections and genetic data

Annina Kantelinen

23.8.2022



Contents

- Research examples
  - Morphology
  - DNA-methods

Lichens are small ecosystems:
mycobiont & photobiont, but also
other fungi, algae, secondary photobionts,
protozoa and non-photosynthetic bacteria.

Collections

- Natural history museums are diverse biobanks of biodiversity
- Many species are more accessible in collections than in their original habitats
  - remote geographical areas
  - rare or endangered taxa
  - extinct taxa
  - taxa that have not been seen since their initial collection

Type specimens

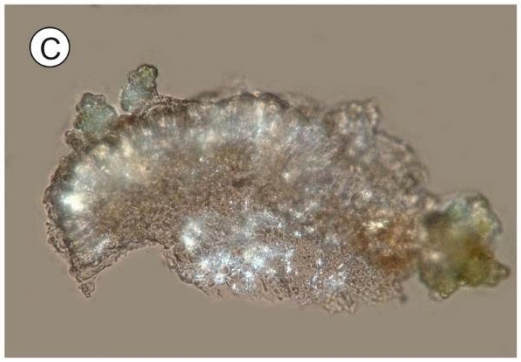
- When a species is first discovered by scientists, a type specimen is nominated. If we are later in doubt about what are the characters of the species, we can check the type.

- ´The Type´ – a song by John Hinton for the Natural History Museum London
  - https://www.youtube.com/watch?v=gfQL7bXwzvM

- Type specimens are often old and DNA-sequencing can be difficult…

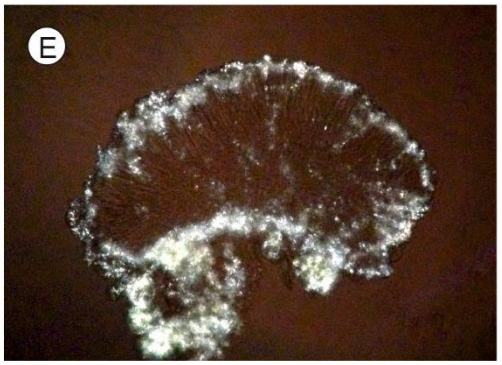
Morphology and old type

- A new character, crystalline granules, was relevant in linking an old type specimen to fresh material
  - The original *M. prasina* type specimen is from 1825
  - Fresh material resolved into three DNA lineages



”*Micarea prasina* 1”

”*M. prasina* 2”



M. prasina s. str.

Launis et al. 2019

Morhology and DNA barcoding

- ”Deficiently known forest lichens – identification through DNA-barcoding” 2011-2012
- Specimens were collected, morphologically identified, sequenced, and deposited in the herbarium and DNA databases
- DNA barcode was created for 108 lichen species
  - Also scientifically new species

PUTTE

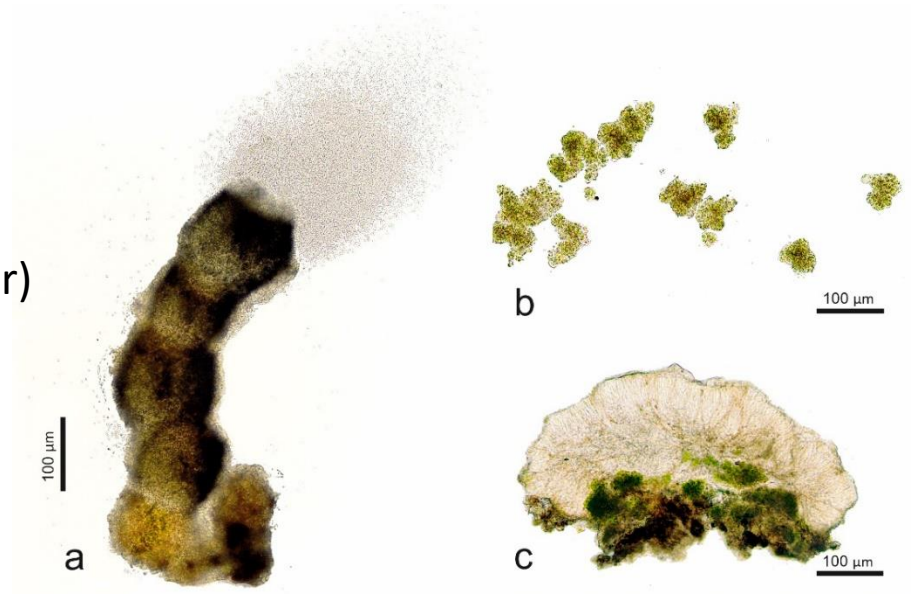
Puutteellisesti tunnettujen ja uhanalaisten metsälajien tutkimus-ohjelma

Morphology and molecular systematics

- Sequencing hundreads of specimens is not always possible or smart

- **New article:** Lichen speciation is sparked by a substrate requirement shift and reproduction mode differentiation. Kantelinen et al. 2022

-516 studied herbarium specimens in Central Europe and Fennoscandia
-Reproduction mode (sexual/asexual)
-Substratum (bark/decaying wood/other)
-3 DNA loci of selected samples



Ancient DNA

- AncientDNA and museomics methods are rapidly evolving
  - Morphology is useful, but requires expertise and time

- Collections are increasingly used in biogeographical, environmental and taxonomic studies

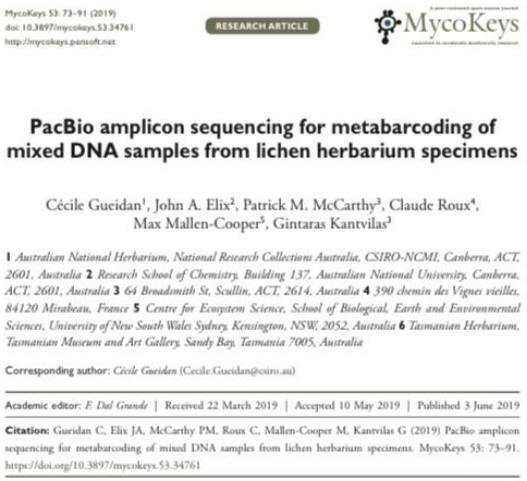
New article: DNA sequencing historical lichen specimens. 2019. Kistenich et al.

-Target sequences (mtSSU)
-Samples from every 25 years from present to 150 years back in time.
-Received satisfactory DNA sequence information for 54 of 56 specimens
-Recovered full-length sequences for several more than 100-years-old specimens!

Metabarcoding

- DNA-based identification meets HTS
  - 500-1000 bp, 16S rDNA
- Great for mixed species samples
  - environmental samples, eg. dead wood
  - community ecology

New article: PacBio amplicon sequencing for metabarcoding of mixed DNA samples from lichen herbarium specimens. 2019. Gueidan, C. et al.



MycoKeys 53: 73–91 (2019)
doi: 10.3897/mycokeys.53.34761
http://mycokeys.pensoft.net

RESEARCH ARTICLE

MycoKeys

PacBio amplicon sequencing for metabarcoding of mixed DNA samples from lichen herbarium specimens

Cécile Gueidan[1], John A. Elix[2], Patrick M. McCarthy[3], Claude Roux[4], Max Mallen-Cooper[5], Gintaras Kantvilas[3]

1 *Australian National Herbarium, National Research Collections Australia, CSIRO-NCMI, Canberra, ACT, 2601, Australia* **2** *Research School of Chemistry, Building 137, Australian National University, Canberra, ACT, 2601, Australia* **3** *64 Broadsmith St, Scullin, ACT, 2614, Australia* **4** *390 chemin des Vignes vieilles, 84120 Mirabeau, France* **5** *Centre for Ecosystem Science, School of Biological, Earth and Environmental Sciences, University of New South Wales Sydney, Kensington, NSW, 2052, Australia* **6** *Tasmanian Herbarium, Tasmanian Museum and Art Gallery, Sandy Bay, Tasmania 7005, Australia*

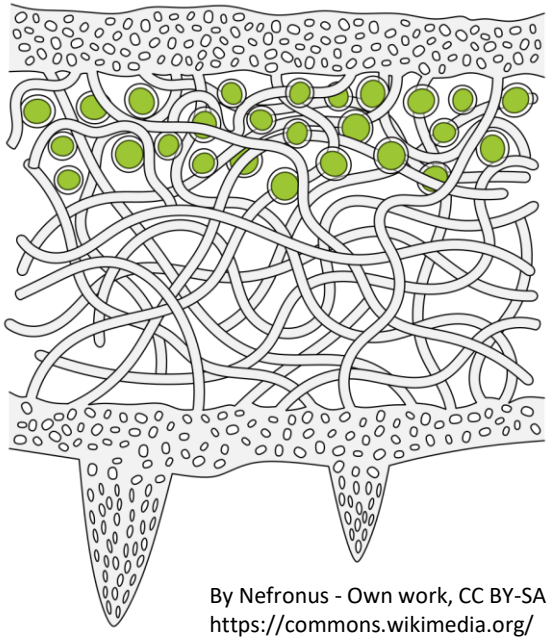
Corresponding author: *Cécile Gueidan* (Cecile.Gueidan@csiro.au)

Academic editor: *F. Dal Grande* | Received 22 March 2019 | Accepted 10 May 2019 | Published 3 June 2019

Citation: Gueidan C, Elix JA, McCarthy PM, Roux C, Mallen-Cooper M, Kantvilas G (2019) PacBio amplicon sequencing for metabarcoding of mixed DNA samples from lichen herbarium specimens. MycoKeys 53: 73–91. https://doi.org/10.3897/mycokeys.53.34761

Genomics & lichens

- Genomic studies in lichenology are considerable delayed due to the symbiotic nature of lichens
- Symbiosis makes it difficult to obtain myco-/photobiont genomes by techniques widely used in other groups of organisms
  - Researchers have tried to culture the mycobiont, but obtaining and maintaining such cultures is difficult and unpredictable



Whole genome sequencing

- More data, wide range of research questions
  - Evolution, adaptation, metabolism, genetics…

New article: The lichen symbiosis re-viewed through the genomes of *Cladonia grayi* and its algal partner *Asterochloris glomerata.* Armaleo, D. et al. 2019.

-The first parallel genomic analysis of lichen symbionts

-From cultures

Metagenomics

- Study of ALL genomes from a mixed community of organisms
  - Environmental samples, eg. microbes
  - Symbiotic organisms, e.g. lichens
- In metagenomics the DNA present in the entire lichen symbiosis is massively sequenced, and the mycobiont part is recovered using computational tools

"*Because of its ability to reveal the previously hidden diversity of microscopic life, metagenomics offers a powerful lens for viewing the microbial world that has revolutionized understanding of the entire living world*"

Marco, D. 2011

Metagenomics & lichens

New(ish) articles:

Fungal Diversity
https://doi.org/10.1007/s13225-018-0407-7

Phylogenomic analysis of 2556 single-copy protein-coding genes resolves most evolutionary relationships for the major clades in the most diverse group of lichen-forming fungi

David Pizarro[1] · Pradeep K. Divakar[1] · Felix Grewe[2] · Steven D. Leavitt[3] · Jen-Pan Huang[2] · Francesco Dal Grande[1,4] · Imke Schmitt[4,5] · Mats Wedin[6] · Ana Crespo[1] · H. Thorsten Lumbsch[2]

Received: 7 December 2017 / Accepted: 1 August 2018

Abstract
Phylogenomic datasets continue to enhance our understanding of evolutionary relationships in many lineages of organisms. However, genome-scale data have not been widely implemented in reconstructing relationships in lichenized fungi. Here we generate a data set comprised of 2556 single-copy protein-coding genes to reconstruct previously unresolved relationships in the most diverse family of lichen-forming fungi, Parmeliaceae. Our sampling included 51 taxa, mainly from the subfamily Parmelioideae, and represented six of the seven previously identified major clades within the family. Our results provided strong support for the monophyly of each of these major clades and most backbone relationships in the topology were recovered with high nodal support based on concatenated dataset and species tree analyses. The alectorioid clade was strongly supported as sister-group to all remaining clades, which were divided into two major sister-groups. In the first major clade the anzioid and usneoid clades formed a strongly supported sister-group relationship with the cetrarioid + hypogymnioid group. The sister-group relationship of *Evernia* with the cetrarioid clade was also strongly supported, whereas that between the anzioid and usneoid clades needs further investigation. In the second major clade *Oropogon* and *Platismatia* were sister to the parmelioid group, while the position of *Omphalora* was not fully resolved. This study demonstrates the power of genome-scale data sets to resolve long-standing, ambiguous phylogenetic relationships of lichen-forming fungi. Furthermore, the topology inferred in this study will provide a valuable framework for better understanding diversification in the most diverse lineage of lichen-forming fungi, Parmeliaceae.

Keywords Fungi · Lecanorales · Lichenized fungi · Parmeliaceae · Parmelioideae · Phylogeny · Systematics

SCIENTIFIC REPORTS

OPEN

Received: 15 June 2017
Accepted: 12 October 2017
Published online: 02 November 2017

Sequencing genomes from mixed DNA samples - evaluating the metagenome skimming approach in lichenized fungi

Anjuli Meiser[1,3], Jürgen Otte[2], Imke Schmitt[1,2] & Francesco Dal Grande[2]

The metagenome skimming approach, i.e. low coverage shotgun sequencing of multi-species assemblages and subsequent reconstruction of individual genomes, is increasingly used for in-depth genomic characterization of ecological communities. This approach is a promising tool for reconstructing genomes of facultative symbionts, such as lichen-forming fungi, from metagenomic reads. However, no study has so far tested accuracy and completeness of assemblies based on metagenomic sequences compared to assemblies based on pure culture strains of lichenized fungi. Here we assembled the genomes of *Evernia prunastri* and *Pseudevernia furfuracea* based on metagenomic sequences derived from whole lichen thalli. We extracted fungal contigs using two different taxonomic binning methods, and performed gene prediction on the fungal contig subsets. We then assessed quality and completeness of the metagenome-based assemblies using genome assemblies as reference which are based on pure culture strains of the two fungal species. Our comparison showed that we were able to reconstruct fungal genomes from uncultured lichen thalli, and also cover most of the gene space (86–90%). Metagenome skimming will facilitate genome mining, comparative (phylo)genomics, and population genetics of lichen-forming fungi by circumventing the time-consuming, sometimes unfeasible, step of aposymbiotic cultivation.

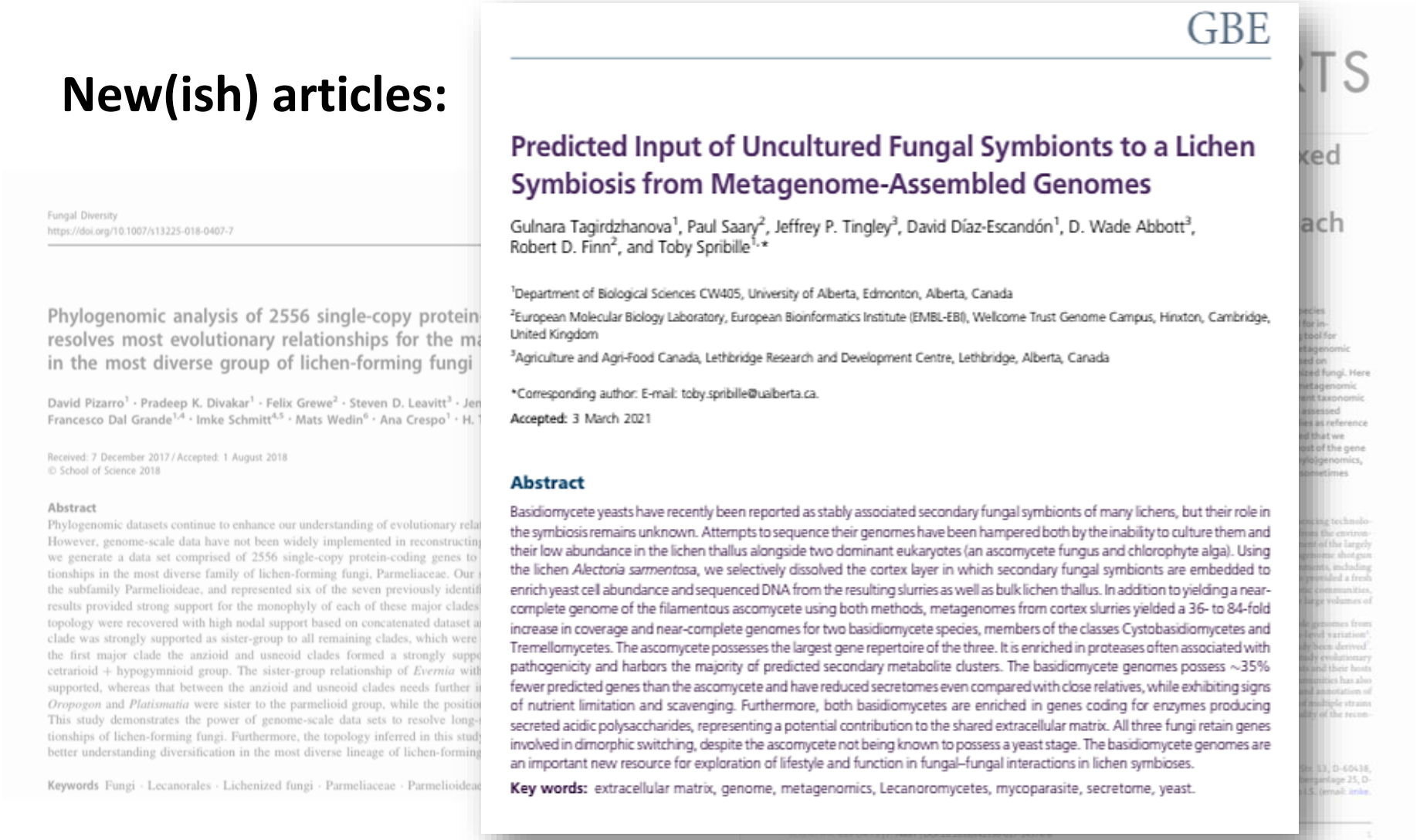
In recent years, the decreasing costs and higher accessibility of high-throughput DNA sequencing technologies have revolutionized microbial ecology research. Direct sequencing of genomic material from the environment, commonly referred to as metagenomics, can provide a cultivation-independent assessment of the largely untapped genetic diversity and functional aspects of microbial communities. Whole-metagenome shotgun sequencing has been applied to study diverse microbiomes, spanning a range of natural environments, including the human body[1–3]. Metagenomics has not only been used to catalogue diversity, but it has also provided a fresh perspective on our understanding of the intricate, multi-species interactions driving symbiotic communities, and how these interactions influence ecosystems[4]. On the other hand, the conversion of these large volumes of sequencing data to biologically useful information remains a major challenge[5].

With the improvement of bioinformatics tools, it is increasingly possible to assemble whole genomes from environmental communities of both prokaryotes and eukaryotes, and analyse their strain-level variation[6]. Although research on metagenomic assembly is still in its infancy, valuable insights have already been derived[7]. The annotation of metagenomic contigs from multi-species communities has proven useful to study evolutionary patterns, metabolic complementation, genetic exchange and/or modification between symbionts and their hosts in several symbiotic systems. The reconstruction of individual genomes from multi-species communities has also been used to isolate genes associated with the biosynthesis of novel biomolecules[8]. Assembly and annotation of sequencing data, however, pose several analytical challenges[9]. In particular, the co-occurrence of multiple strains or similar species – sometimes present at highly uneven ratios – may drastically reduce the quality of the reconstructed genomes[10].

[1]Institute of Ecology, Evolution and Diversity, Goethe University Frankfurt, Max-von-Laue Str. 13, D-60438, Frankfurt, Germany. [2]Senckenberg Biodiversity and Climate Research Centre (SBiK-F), Senckenberganlage 25, D-60486, Frankfurt, Germany. Correspondence and requests for materials should be addressed to I.S. (email: imke.schmitt@senckenberg.de) or F.D.G. (email: francesco.dalgrande@senckenberg.de)

Metagenomics & lichens

New(ish) articles:



Fungal Diversity
https://doi.org/10.1007/s13225-018-0407-7

Phylogenomic analysis of 2556 single-copy protein
resolves most evolutionary relationships for the ma
in the most diverse group of lichen-forming fungi

David Pizarro[1] · Pradeep K. Divakar[1] · Felix Grewe[2] · Steven D. Leavitt[3] · Jen
Francesco Dal Grande[1,4] · Imke Schmitt[4,5] · Mats Wedin[6] · Ana Crespo[1] · H.

Received: 7 December 2017 / Accepted: 1 August 2018
© School of Science 2018

Abstract
Phylogenomic datasets continue to enhance our understanding of evolutionary rela
However, genome-scale data have not been widely implemented in reconstructin
we generate a data set comprised of 2556 single-copy protein-coding genes to
tionships in the most diverse family of lichen-forming fungi, Parmeliaceae. Our
the subfamily Parmelioideae, and represented six of the seven previously identif
results provided strong support for the monophyly of each of these major clades
topology were recovered with high nodal support based on concatenated dataset a
clade was strongly supported as sister-group to all remaining clades, which were
the first major clade the anzioid and usneoid clades formed a strongly suppo
cetrarioid + hypogymnioid group. The sister-group relationship of *Evernia* with
supported, whereas that between the anzioid and usneoid clades needs further i
Oropogon and *Platismatia* were sister to the parmelioid group, while the positio
This study demonstrates the power of genome-scale data sets to resolve long-
tionships of lichen-forming fungi. Furthermore, the topology inferred in this study
better understanding diversification in the most diverse lineage of lichen-forming

Keywords Fungi · Lecanorales · Lichenized fungi · Parmeliaceae · Parmelioidea

GBE

Predicted Input of Uncultured Fungal Symbionts to a Lichen Symbiosis from Metagenome-Assembled Genomes

Gulnara Tagirdzhanova[1], Paul Saary[2], Jeffrey P. Tingley[3], David Díaz-Escandón[1], D. Wade Abbott[3], Robert D. Finn[2], and Toby Spribille[1,*]

[1]Department of Biological Sciences CW405, University of Alberta, Edmonton, Alberta, Canada

[2]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom

[3]Agriculture and Agri-Food Canada, Lethbridge Research and Development Centre, Lethbridge, Alberta, Canada

*Corresponding author: E-mail: toby.spribille@ualberta.ca.

Accepted: 3 March 2021

Abstract

Basidiomycete yeasts have recently been reported as stably associated secondary fungal symbionts of many lichens, but their role in the symbiosis remains unknown. Attempts to sequence their genomes have been hampered both by the inability to culture them and their low abundance in the lichen thallus alongside two dominant eukaryotes (an ascomycete fungus and chlorophyte alga). Using the lichen *Alectoria sarmentosa*, we selectively dissolved the cortex layer in which secondary fungal symbionts are embedded to enrich yeast cell abundance and sequenced DNA from the resulting slurries as well as bulk lichen thallus. In addition to yielding a near-complete genome of the filamentous ascomycete using both methods, metagenomes from cortex slurries yielded a 36- to 84-fold increase in coverage and near-complete genomes for two basidiomycete species, members of the classes Cystobasidiomycetes and Tremellomycetes. The ascomycete possesses the largest gene repertoire of the three. It is enriched in proteases often associated with pathogenicity and harbors the majority of predicted secondary metabolite clusters. The basidiomycete genomes possess ~35% fewer predicted genes than the ascomycete and have reduced secretomes even compared with close relatives, while exhibiting signs of nutrient limitation and scavenging. Furthermore, both basidiomycetes are enriched in genes coding for enzymes producing secreted acidic polysaccharides, representing a potential contribution to the shared extracellular matrix. All three fungi retain genes involved in dimorphic switching, despite the ascomycete not being known to possess a yeast stage. The basidiomycete genomes are an important new resource for exploration of lifestyle and function in fungal–fungal interactions in lichen symbioses.

Key words: extracellular matrix, genome, metagenomics, Lecanoromycetes, mycoparasite, secretome, yeast.

Challenges and future aspects

- HTS challenges:
  - De novo assembly, because no reference data available
  - Short reads are bioinformatically demanding
  - Low recovery
- HTS future:
  - Short reads are becoming longer (Illumina 200 bp -> PacBio 15000 bp)
- Not just what you CAN do, but what you WANT to do
  - What is your research question?
  - What is your research question 5 years from now?

More info about the methods

- A lecture “Fundamentals of Genome Assembly” by Jared Simpson (Ontario Institute for Cancer Research) https://www.youtube.com/watch?v=5wvGapmA5zM

- Workshop on Genomics, Cesky Krumlov (www.evomics.org)

References

1. Annina Launis, Jiri Malicek , Måns Svensson , Andrei Tsurykau , Emmanuel Serusiaux & Leena Myllys. (2019). Sharpening species boundaries in the *M. prasina* group with a new circumscription of the type species *M. prasina*. Mycologia 111: 574-592.

2. Annina Kantelinen, Christian Printzen, Péter Poczai & Leena Myllys. (2022). Lichen speciation is sparked by a substrate requirement shift and reproduction mode differentiation. Scientific Reports 12:11048.

3. Victoria G. Twort, Joël Minet, Christopher W. Wheat and Niklas Wahlberg. (2020). Museomics of a rare taxon: placing Whalleyanidae in the Lepidoptera Tree of Life bioRxiv preprint doi: https://doi.org/10.1101/2020.08.18.255182.

4. Daniele Armaleo, Olaf Müller, François Lutzoni, Ólafur S. Andrésson, Guillaume Blanc, Helge B. Bode, Frank R. Collart, Francesco Dal Grande, Fred Dietrich, Igor V. Grigoriev, Suzanne Joneson, Alan Kuo, Peter E. Larsen, John M. Logsdon Jr, David Lopez, Francis Martin, Susan P. May, Tami R. McDonald, Sabeeha S. Merchant, Vivian Miao, Emmanuelle Morin, Ryoko Oono, Matteo Pellegrini, Nimrod Rubinstein, Maria Virginia Sanchez-Puerta, Elizabeth Savelkoul, Imke Schmitt, Jason C. Slot, Darren Soanes, Péter Szövényi, Nicholas J. Talbot, Claire Veneault-Fourrey & Basil B. Xavier. (2019). The lichen symbiosis re-viewed through the genomes of *Cladonia grayi* and its algal partner *Asterochloris glomerata.* BMC Genomics volume 20, Article number: 605.

5. Marco, D. ed. (2011). Metagenomics: current innovations and future trends. Caister Academic Press. ISBN 978-1-904455-87-5.

6. Anjuli Meiser, Jürgen Otte, Imke Schmitt & Francesco Dal Grande. (2017). Sequencing genomes from mixed DNA samples - evaluating the metagenome skimming approach in lichenized fungi. Scientific Reports 7: 14881.

7. David Pizarro, Pradeep Divakar, Felix Grewe, Steven Leavitt, Jen-Pan Huang, Francesco Dal Grande, Imke Schmitt, Mats Wedin, Ana Crespo, Thorsten Lumbsch. 2018. Phylogenomic analysis of 2556 single-copy protein-coding genes resolves most evolutionary relationships for the major clades in the most diverse group of lichen-forming fungi. Fungal Diversity.

8. Sonja Kistenich, Rune Halvorsen, Audun Schrøder-Nielsen, Lisbeth Thorbek, Einar Timdal and Mika Bendiksby. (2019). DNA Sequencing Historical Lichen Specimens., Front. Ecol. Evol., 23.

Today´s group work

- Form groups of 2-3 persons

- Examine a specimen and discuss how it could be used in research
  - Are DNA-studies possible?
  - What might be the challenges?
  - What kind of research questions would you like to ask?

Literature

- Call, E., Mayer, C., Twort, V., Dietz, L., Wahlberg, N., & Espeland, M. (2021). Museomics: phylogenomics of the moth family Epicopeiidae (Lepidoptera) using target enrichment. *Insect Systematics and Diversity*. 5:6. https://doi.org/10.1093/isd/ixaa021

- Huemer P., Haxaire J., Lee K.M., Mutanen M., Pekarsky O., Scalercio S., Ronkay L. 2020. Revision of the genus *Hoplodrina* Boursin, 1937 (Lepidoptera, Noctuidae, Xyleninae). I. *Hoplodrina octogenaria* (Goeze, 1781) and its sister species *H. alsinides* (Costantini, 1922) sp. rev. in Europe. *Zookeys*. 927:75–97. https://doi.org/10.3897/zookeys.927.51142

- Ivanov, V., Lee, K. M., & Mutanen, M. 2018. Mitonuclear discordance in wolf spiders: Genomic evidence for species integrity and introgression. *Molecular Ecology*. 27:1681-1695. https://doi.org/10.1111/mec.14564

- Raxworthy, C.J. and Smith, B.T., 2021. Mining museums for historical DNA: advances and challenges in museomics. *Trends in Ecology & Evolution.* 36:1049-1060. https://doi.org/10.1016/j.tree.2021.07.009

- Twort, V.G., Minet J., Wheat C.W., Wahlberg N. 2021. Museomics of a rare taxon: placing Whalleyanidae in the Lepidoptera Tree of Life. *Systematic Entomology*. 46:926–937. https://doi.org/10.1111/syen.12503

- Wahlberg, N., & Wheat, C. W. 2008. Genomic outposts serve the phylogenomic pioneers: designing novel nuclear markers for genomic DNA extractions of Lepidoptera. *Systematic biology*. 57:231-242. https://doi.org/10.1080/10635150802033006

- Wahlberg, N., Peña, C., Ahola, M., Wheat, C. W., & Rota, J. 2016. PCR primers for 30 novel gene regions in the nuclear genomes of Lepidoptera. *ZooKeys.* 596:129. https://doi.org/10.3897/zookeys.596.8399