

The age of **MUSEOMICS**

How to get genomic information
from museum specimens of
Lepidoptera

Elsa CALL

CBGP

19th April 2022



LUND
UNIVERSITY



- Natural history museum collections are abundant
 - Over 2.5 – 3 billion specimens



- Natural history museum collections are abundant
 - Over 2.5 – 3 billion specimens
- Value of museum collections
 - Systematics
 - Evolution
 - Biodiversity
 - Habitat loss & Climate change
 - Biological invasion
 - Public health and safety
 - History of infectious diseases



- DNA from these specimens
 - Too degraded to be used
- Next-Generation Sequencing
 - Short fragments
 - Works for extinct taxa
 - Neanderthals, mammoths & cave bears



WHY IS GENOME SEQUENCING DIFFICULT?



Types of ancient DNA damage

- **Strand breaks**

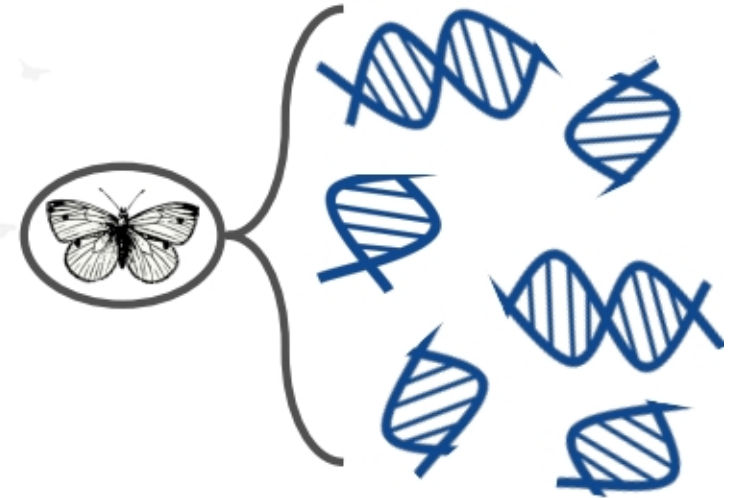
- Natural fragmentation of the DNA

- **Effects**

- Low quantity of DNA
- Short fragment length

- **Solutions**

- Amplify short (<100 – 300 bp)
- Overlapping fragments



Types of ancient DNA damage

- **Miscoding**

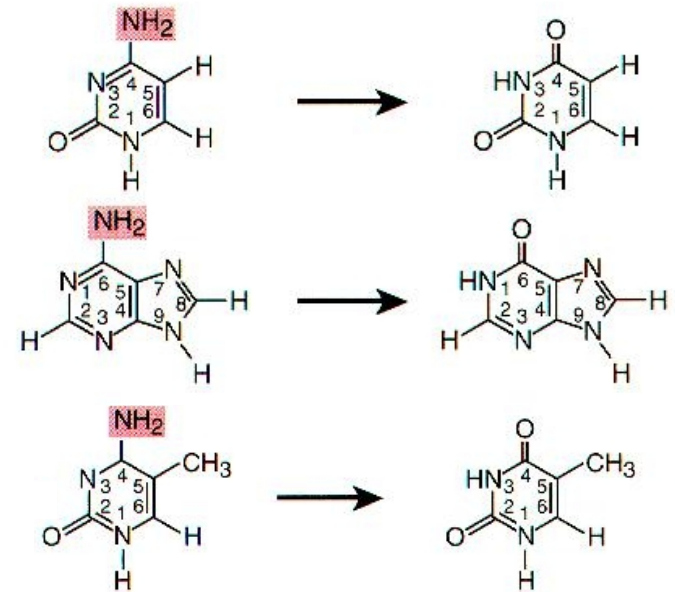
- Chemical degradation of DNA

- Effects: base misincorporation

- *Type 2*: C \rightarrow T (U) and G \rightarrow A
- *Type 1*: A \rightarrow G

- Solutions

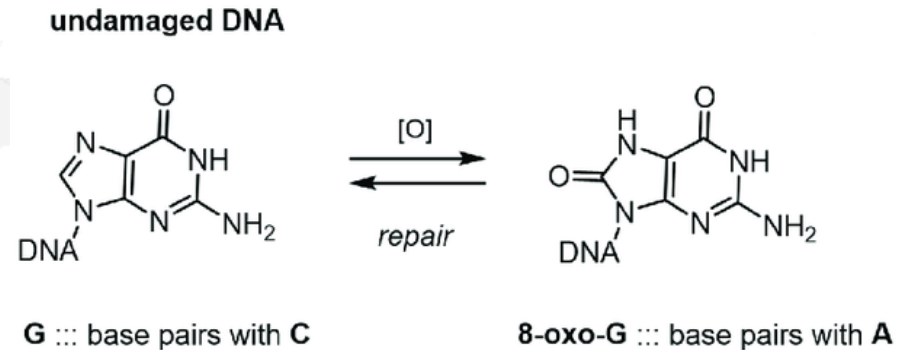
- Multiplex extractions and amplifications
- Remove Uracil





Types of ancient DNA damage

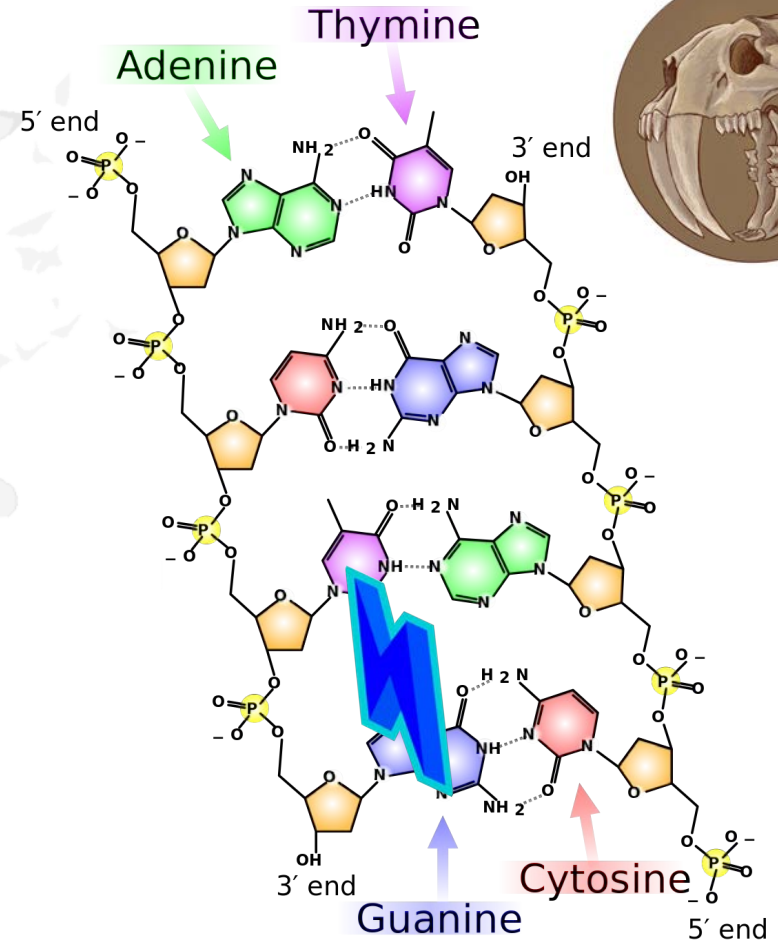
- **Blocking**
 - Chemical modification of DNA
- **Effects**
 - Blocks amplification
 - Base misincorporation $G \rightsquigarrow T$
- **Solutions**
 - Special polymerase
 - Multiple amplifications



Types of ancient DNA damage

● Crosslinks

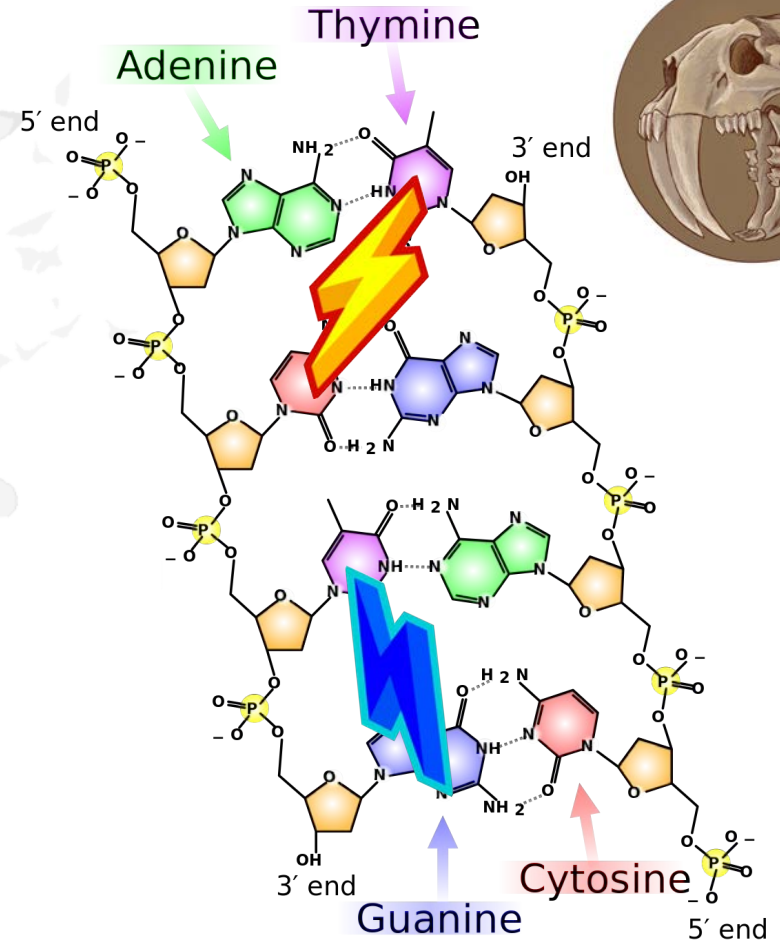
- Linkage between 2 nucleotides
- Within the same strand



Types of ancient DNA damage

● Crosslinks

- Linkage between 2 nucleotides
- Within the same strand
- Or between opposite strands



Types of ancient DNA damage

● Crosslinks

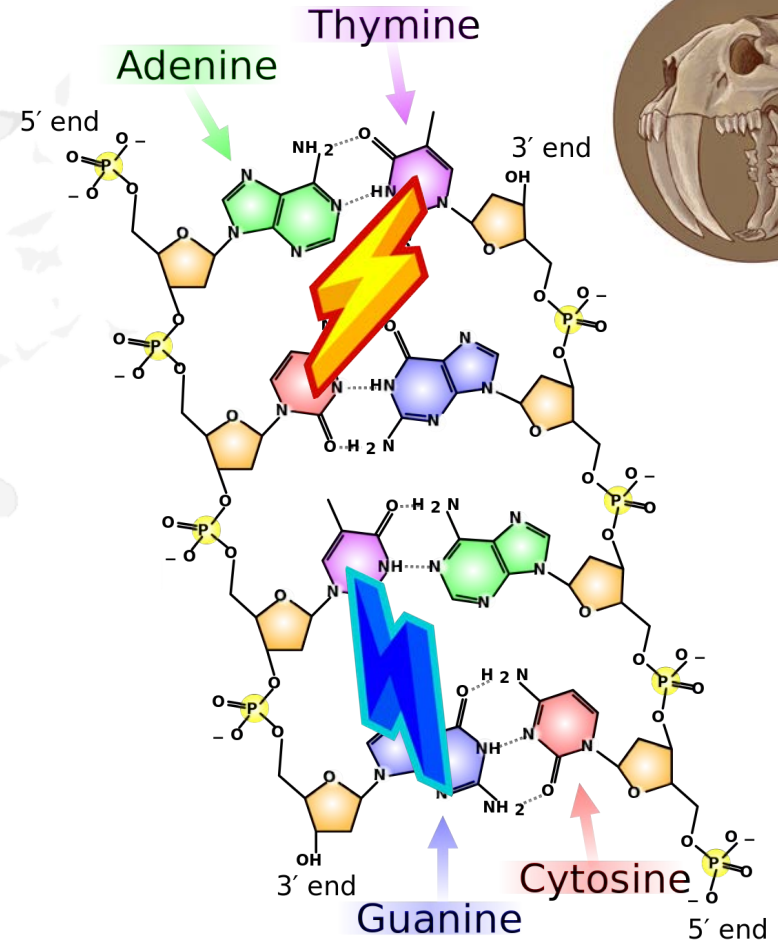
- Linkage between 2 nucleotides
- Within the same strand
- Or between opposite strands

● Effects

- No amplification

● Solutions

- Cleave crosslinks



MUSEOMICS EXAMPLES



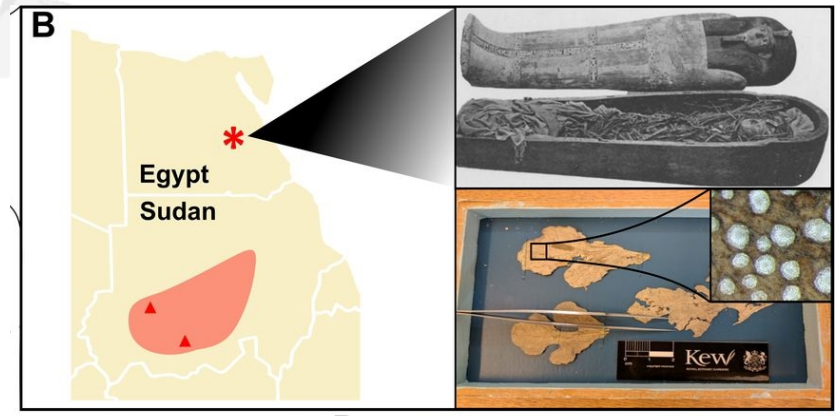
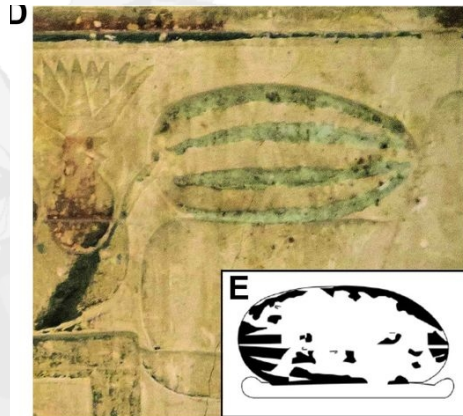
- Quagga, extinct in 1883
- Dry muscles from a museum specimen
- Sanger sequencing
- Sequence of mtDNA
 - 229 nucleotide pairs
 - 12 base substitutions
- Date of the genus *Equus* 3-4 Myr ago



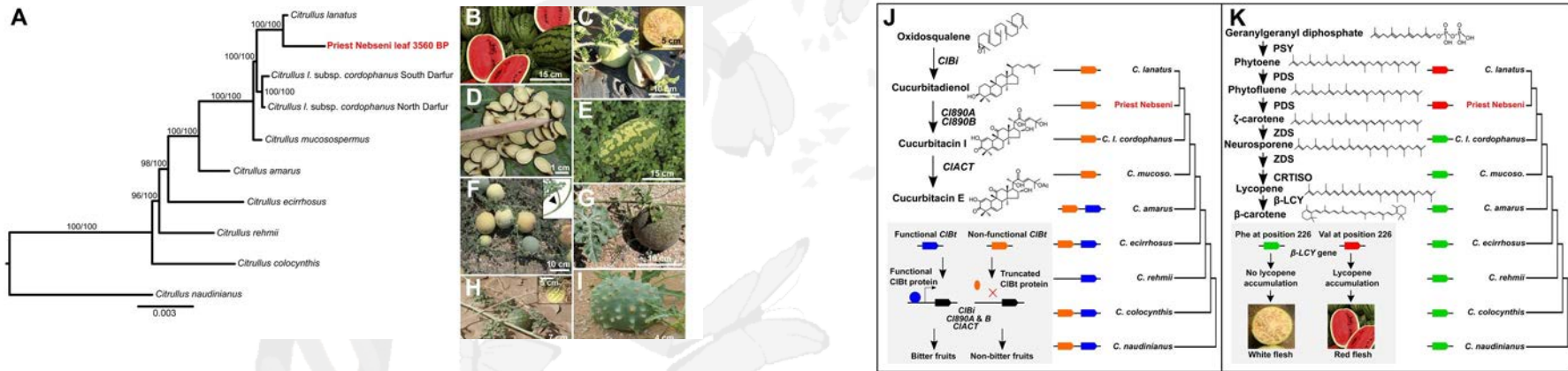
- Flightless NZ bird that went extinct following human settlement in 13th century
- Used NGS to recover
 - Complete MT genome
 - ~900 Mb of Nuclear genome (~75% of Emu genome)
- Identify microsats that can be used to study past population of moa



- Origin & domestication of watermelon is unclear
 - Archeological evidence restricted to Egypt & Libya
 - 7 edible species (sweet red) & 6 inedible species (white pulp)
- A *Citrullus* leaf founded in a tomb



- Genome skimming on the leaf plus all *Citrullus* species
 - Phylogenetic analysis + markers of bitterness vs. sweetness
- ⇒ Confirmation the leaf is sister to domesticated watermelon

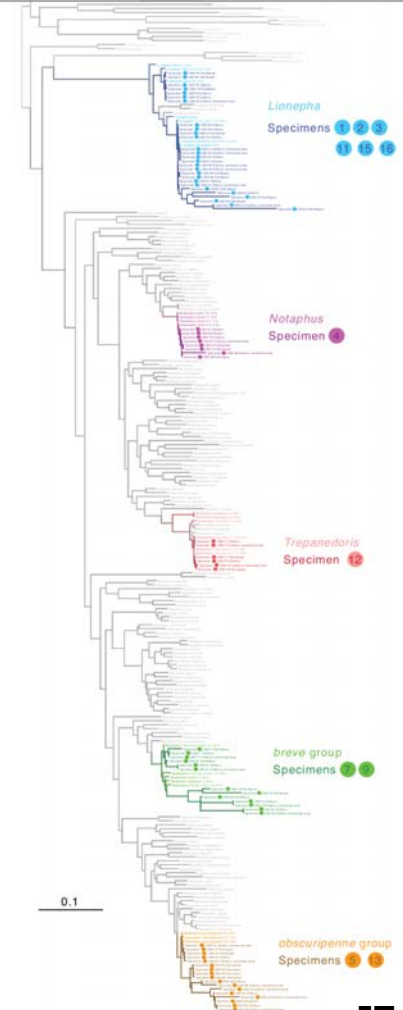


- Highlights that 3,500 years ago Egyptians had cultivated a sweet, red watermelon

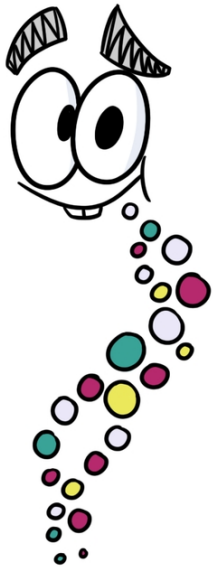
- Sequencing of historical samples
 - Beetles 58-159 years old
- Non-destructive method
 - Low amount of input DNA
- Repaired and un-repaired protocols
- Compare preparation libraries protocols

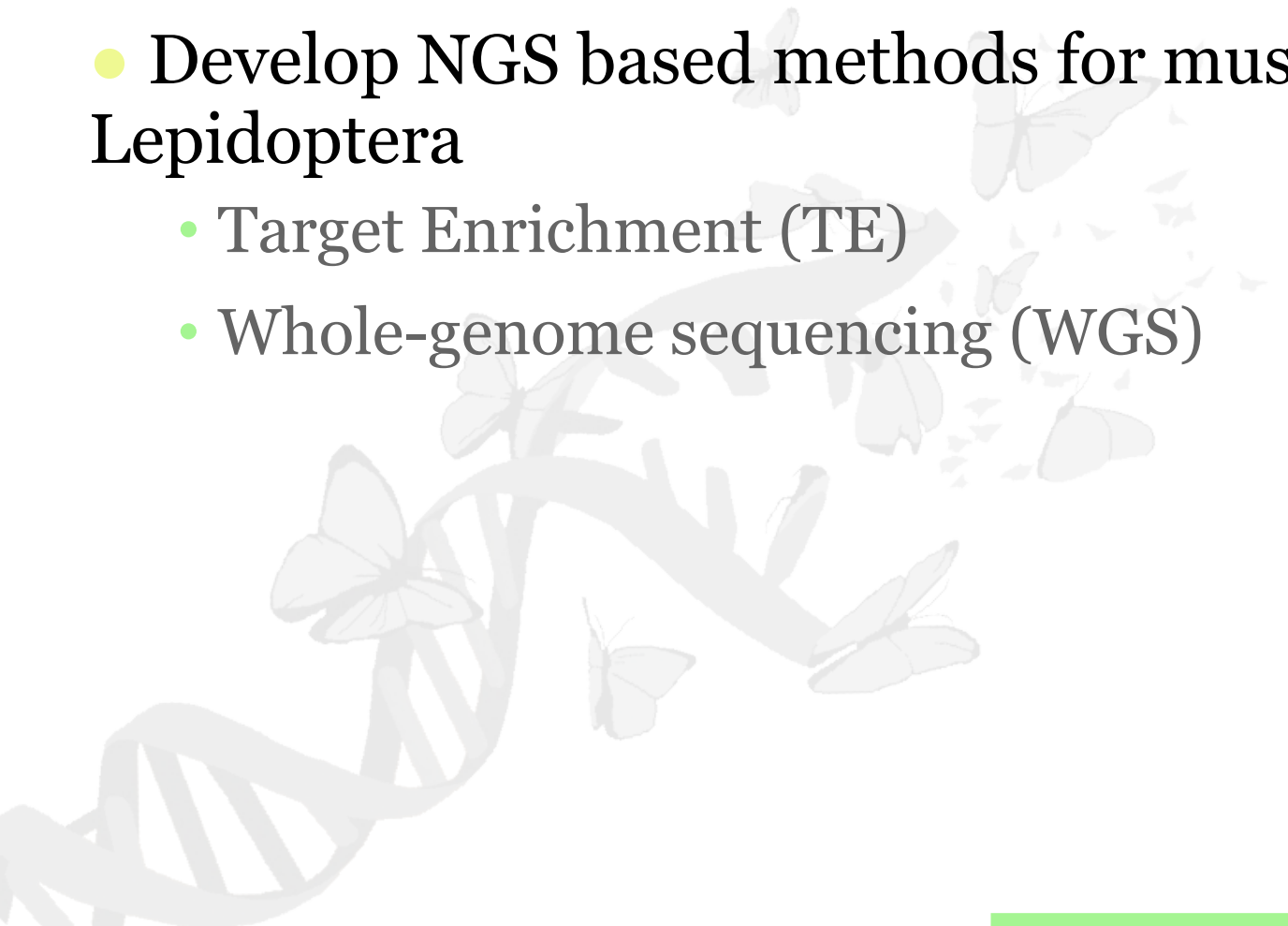


- Recover 85% of the mtGenome
- 67 nuclear protein coding genes
 - Average 0.5-65% recovery
- Compared historical specimens with modern samples
- List of guidelines
 - Cost-effective sample preparation
- Facilitate new museomics projects



HOW TO GET GENOMIC INFORMATION FROM MUSEUM SPECIMENS



- Develop NGS based methods for museum specimens of Lepidoptera
 - Target Enrichment (TE)
 - Whole-genome sequencing (WGS)
- 

- Develop NGS based methods for museum specimens of Lepidoptera
 - Target Enrichment (TE)
 - Whole-genome sequencing (WGS)
- Explore museomics applications
 - Phylogenetic relationships among families of moths
 - Population genetics?

- Develop NGS based methods for museum specimens of Lepidoptera
 - Target Enrichment (TE)
 - Whole-genome sequencing (WGS)
- Explore museomics applications
 - Phylogenetic relationships among families of moths
 - Population genetics?
- Investigate the level of DNA preservation in specimens of various ages

- Epicopeiidae, Sematuridae & Pseudobistonidae
 - Collections in museums (Copenhagen, Bonn, Tokyo, *etc.*)
 - Small families
 - 27 Epicopeidae – Asia
 - 42 Sematuridae – South America
 - 2 Pseudobistonidae – Asia



- *Pieris napi* (Pieridae)
- Common in Sweden
⇒ important collections
- Reference genome



HOW DOES SAMPLE PREPARATION WORK?



METHOD OVERVIEW

DNA extraction



METHOD OVERVIEW

DNA extraction



DNA repair (USER)

METHOD OVERVIEW

DNA extraction

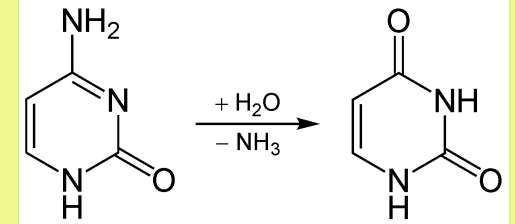
DNA repair (USER)

WHY?

Issue: Cytosine deamination

Cytosine

Uracil



⇒ sequencing errors

C/G ⇒ T/A

HOW?

USER: remove uracil bases



METHOD OVERVIEW

DNA extraction

DNA repair (USER)

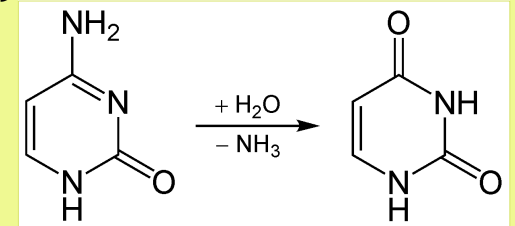
Library preparation

WHY?

Issue: Cytosine deamination

Cytosine

Uracil



⇒ sequencing errors

C/G ⇒ T/A

HOW?

USER: remove uracil bases



METHOD OVERVIEW

DNA extraction

DNA repair (USER)

Sequencing

Library preparation

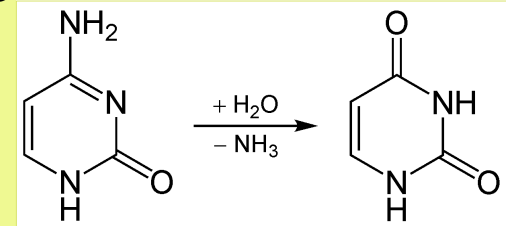


WHY?

Issue: Cytosine deamination

Cytosine

Uracil



⇒ sequencing errors

C/G ⇒ T/A

HOW?

USER: remove uracil bases



METHOD OVERVIEW

DNA extraction

DNA repair (USER)

Sequencing

Library preparation

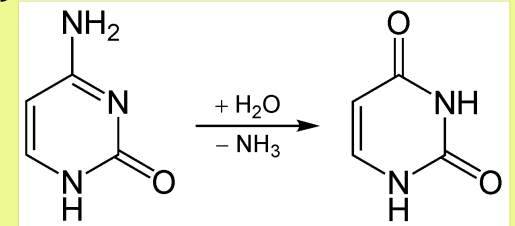
Downstream analysis

WHY?

Issue: Cytosine deamination

Cytosine

Uracil



⇒ sequencing errors

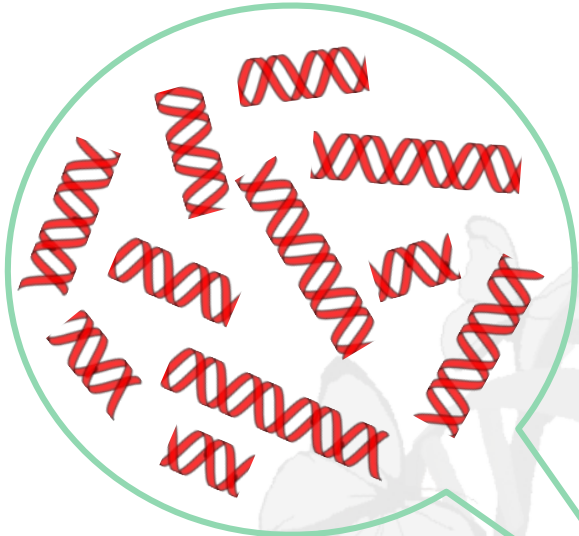
C/G ⇒ T/A

HOW?

USER: remove uracil bases

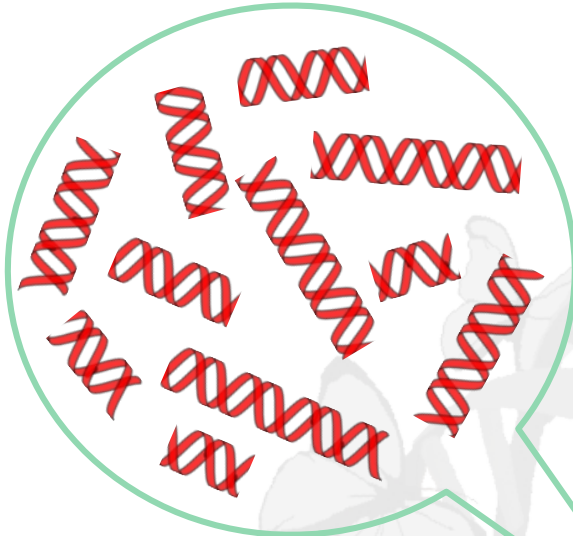


WHOLE GENOME SEQUENCING (WGS)



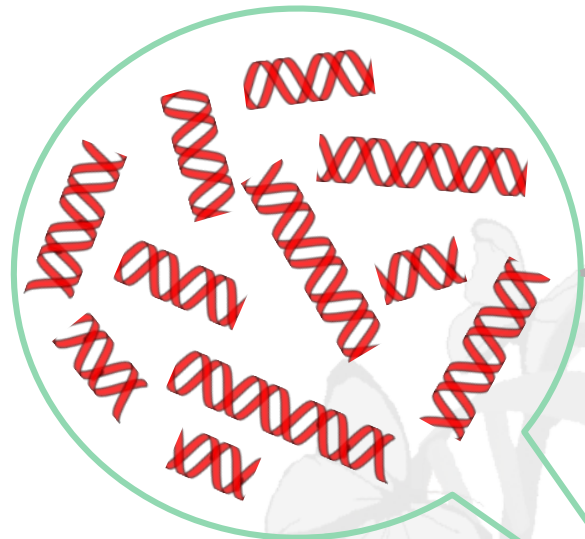
WHOLE GENOME SEQUENCING (WGS)

Reference assembly

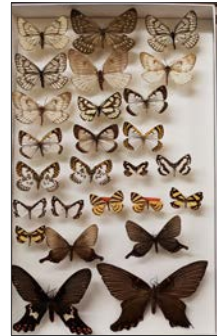
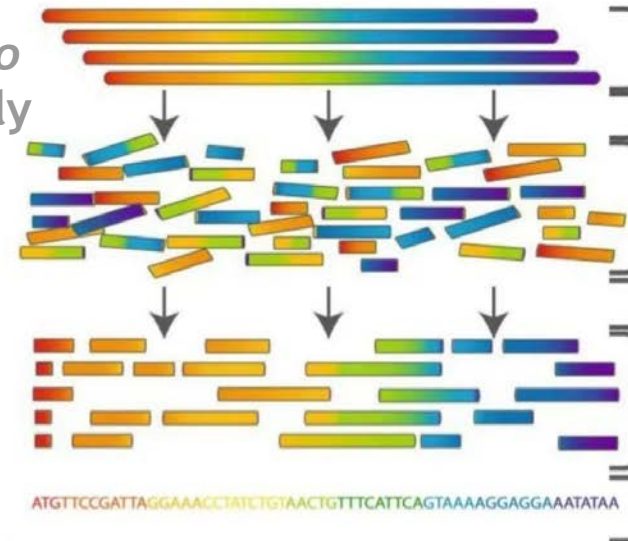


WHOLE GENOME SEQUENCING (WGS)

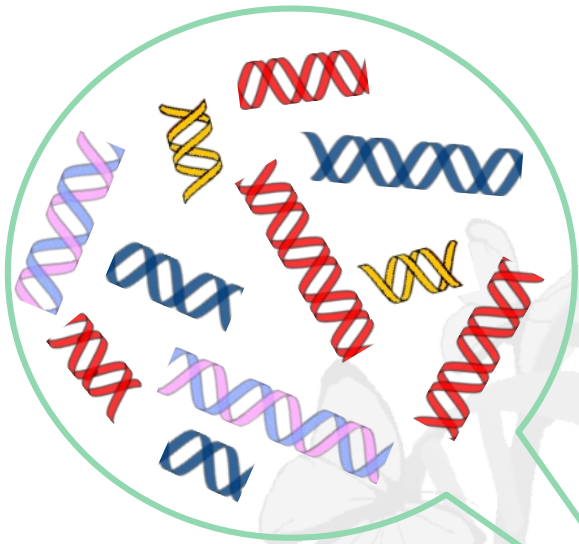
Reference assembly



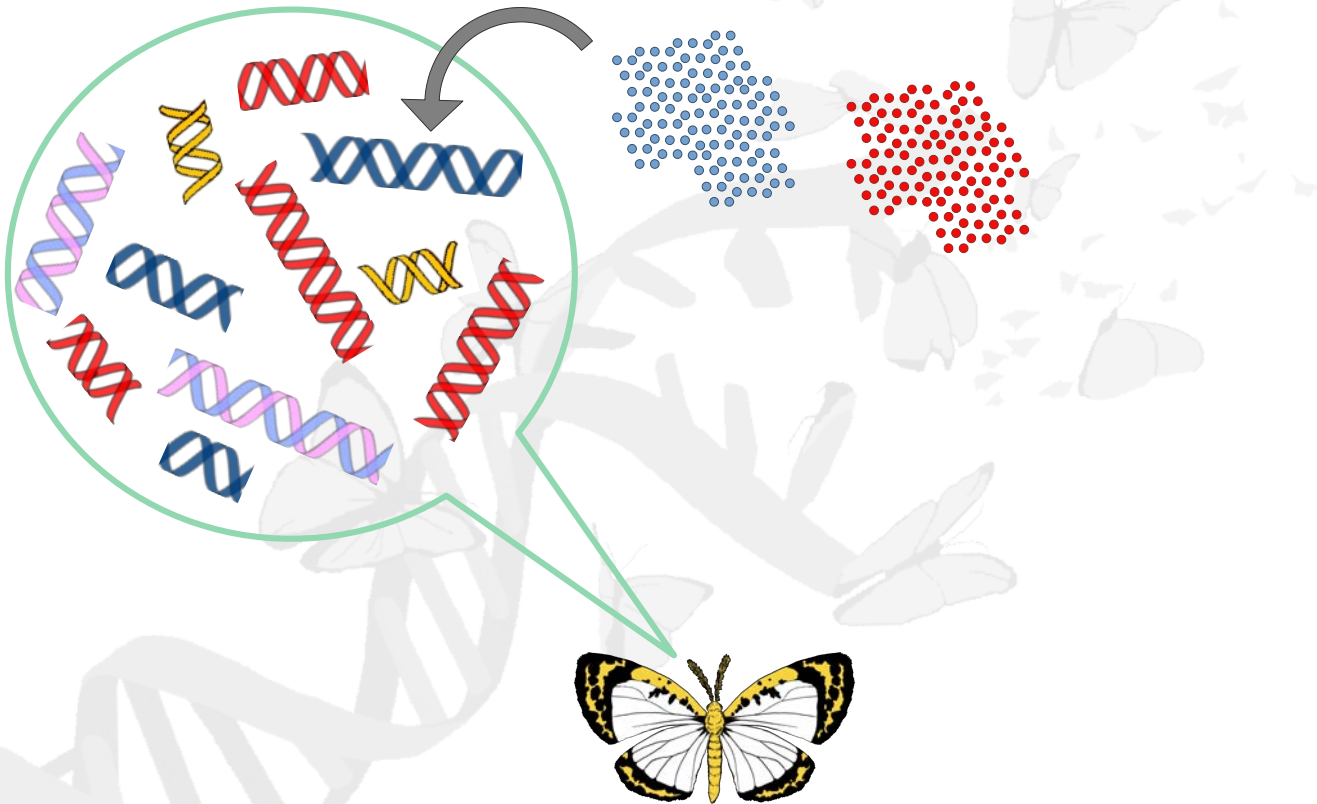
De novo
assembly



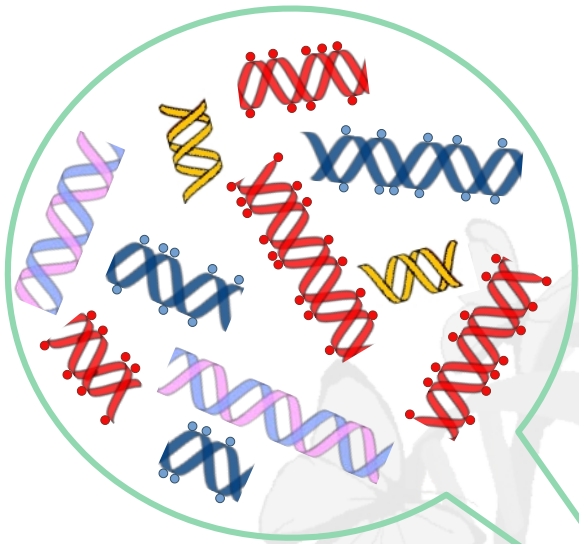
TARGET ENRICHMENT (TE)



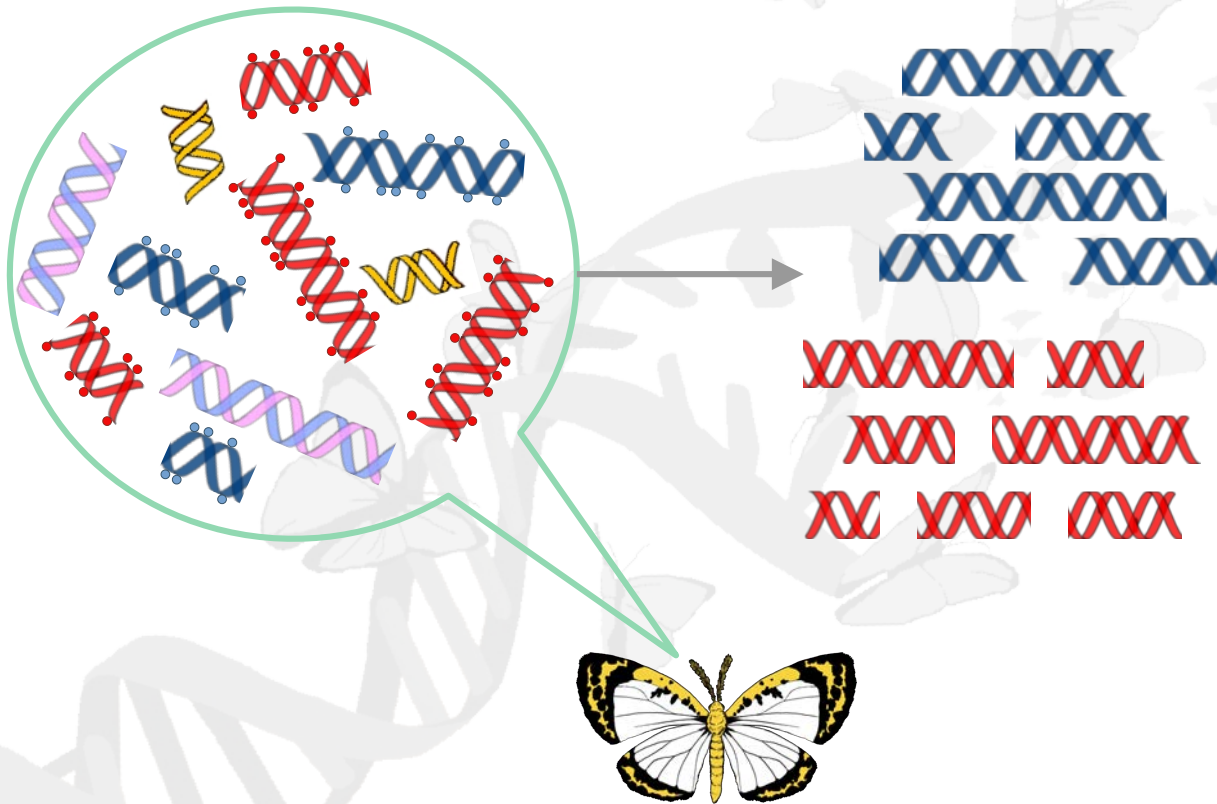
TARGET ENRICHMENT (TE)



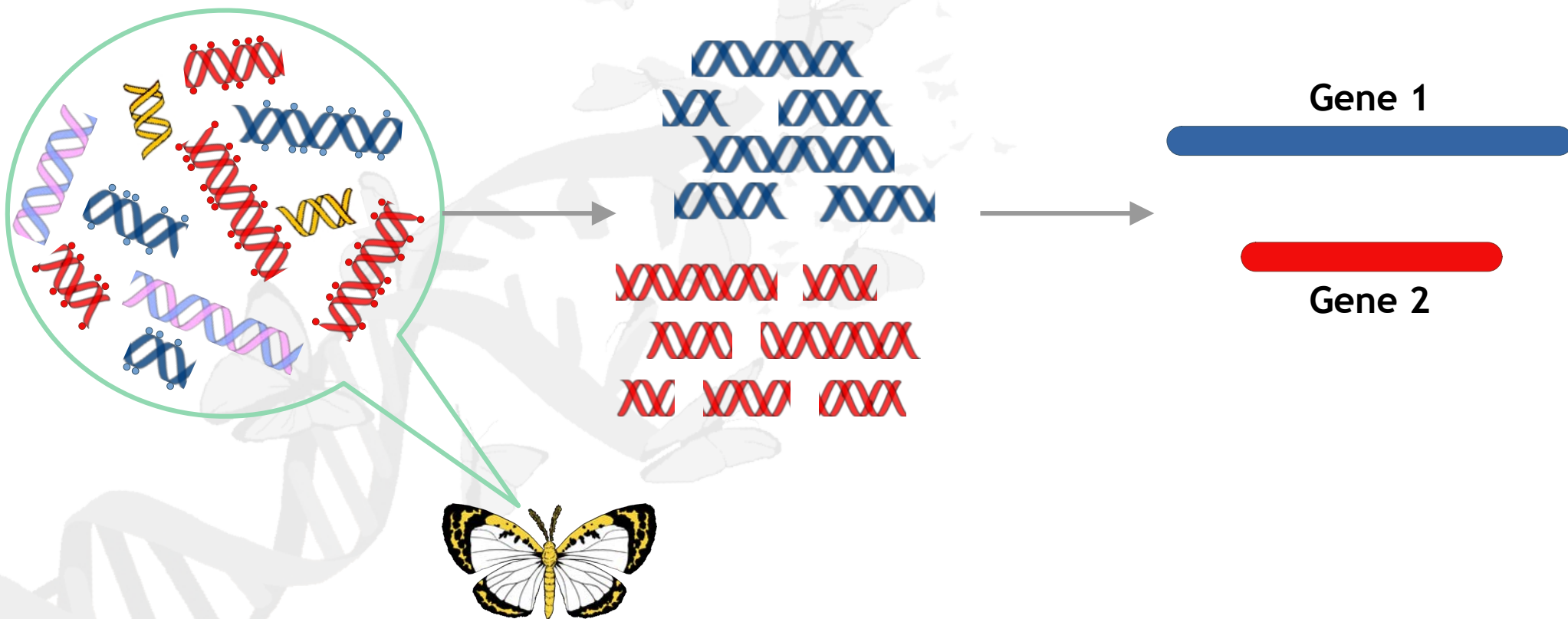
TARGET ENRICHMENT (TE)



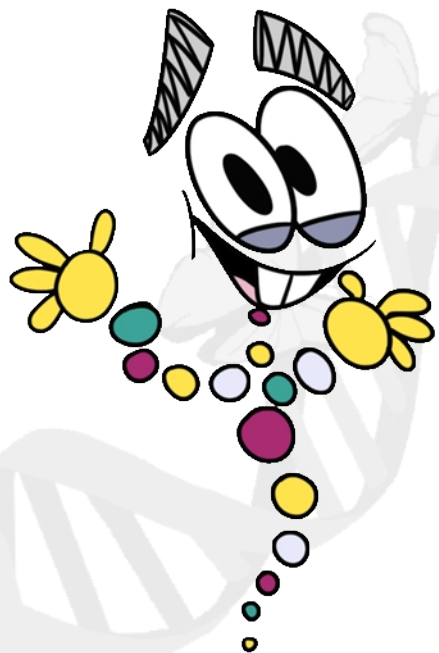
TARGET ENRICHMENT (TE)



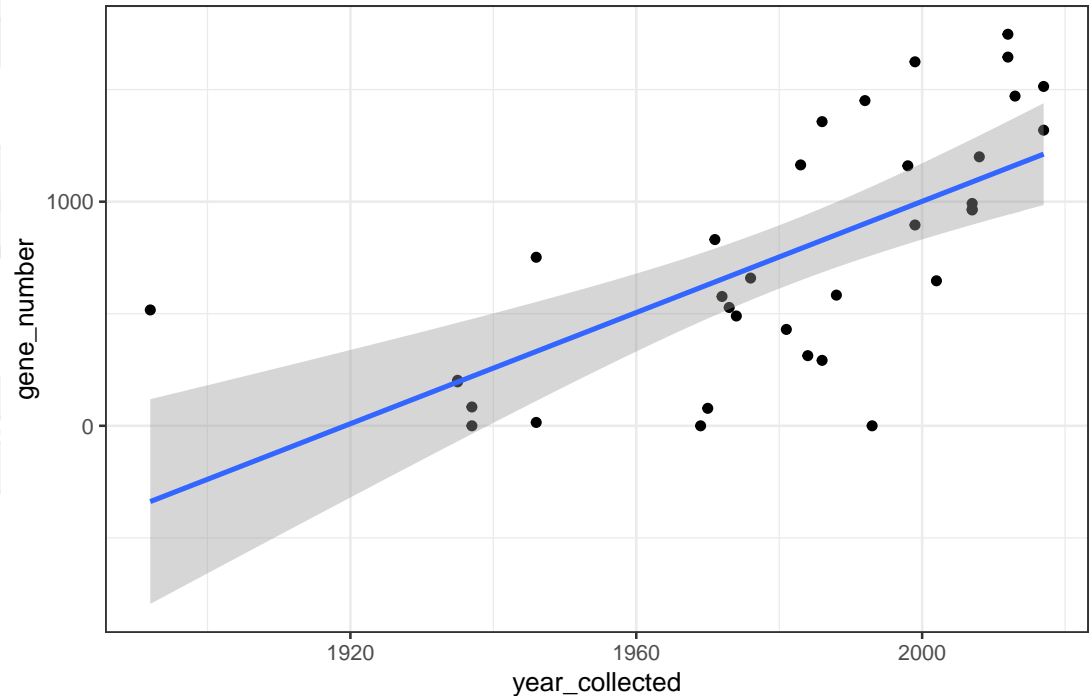
TARGET ENRICHMENT (TE)



TARGET ENRICHMENT RESULTS & DISCUSSION



- Present a TE kit for Lepidoptera
 - 2,953 loci in 1,753 orthologous genes
- Recovered between
 - 78 – 1,747 loci
- Oldest specimen
 - From 1892: 517 loci



Adding leaves to the Lepidoptera tree: capturing hundreds of nuclear genes from old museum specimens

CHRISTOPH MAYER¹, LARS DIETZ², ELSA CALL³,
 SANDRA KUKOWKA⁴, SEBASTIAN MARTIN⁵
 and MARIANNE ESPELAND²

- Same kit, family level
 - Use TE probes set
- Phylogenomics 378 loci; 327 genes

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 Espeland⁴

- Same kit, family level
 - Use TE probes set
- Phylogenomics 378 loci; 327 genes
- Average length 367 bp
 - Total length: 134,881 bp
- Old samples have fewer loci
 - Oldest specimen yielded 516 raw loci

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

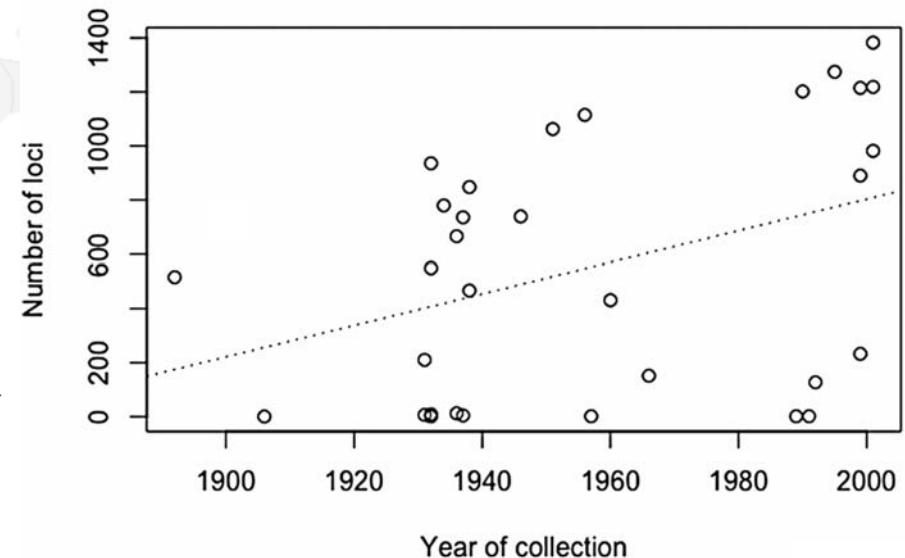
Molecular Phylogenetics, Phylogenomics, and Phylogeography

OXFORD

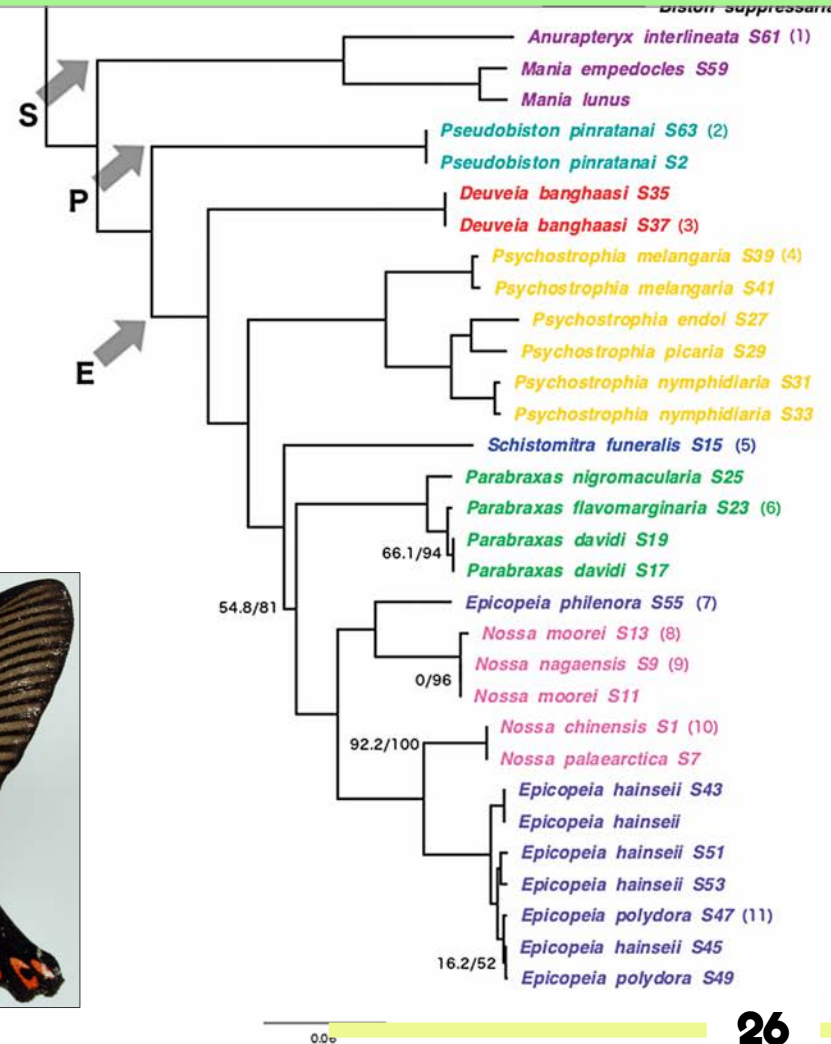
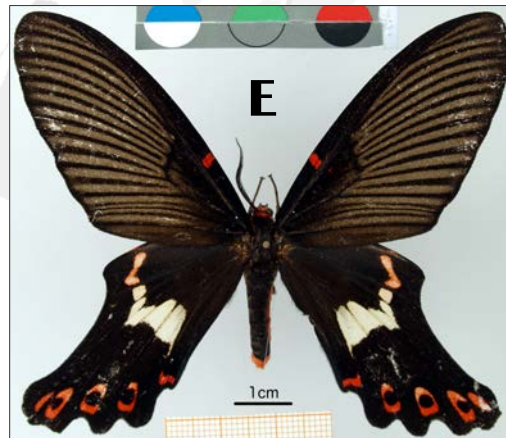
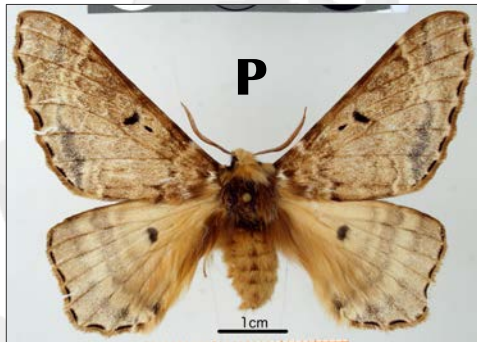
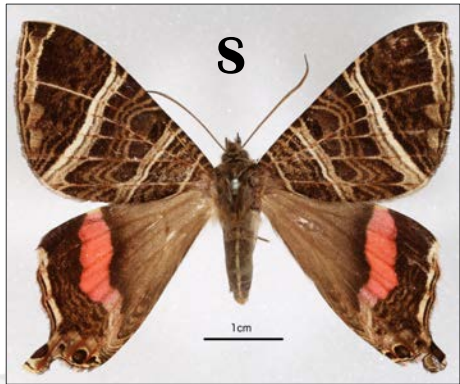
Museomics: Phylogenomics of the Moth Family Epicopeiidae (Lepidoptera) Using Target Enrichment

Elsa Call,^{1,5} Christoph Mayer,² Victoria Twort,^{1,3} Lars Dietz,² Niklas Wahlberg,¹ and Marianne Espeland⁴

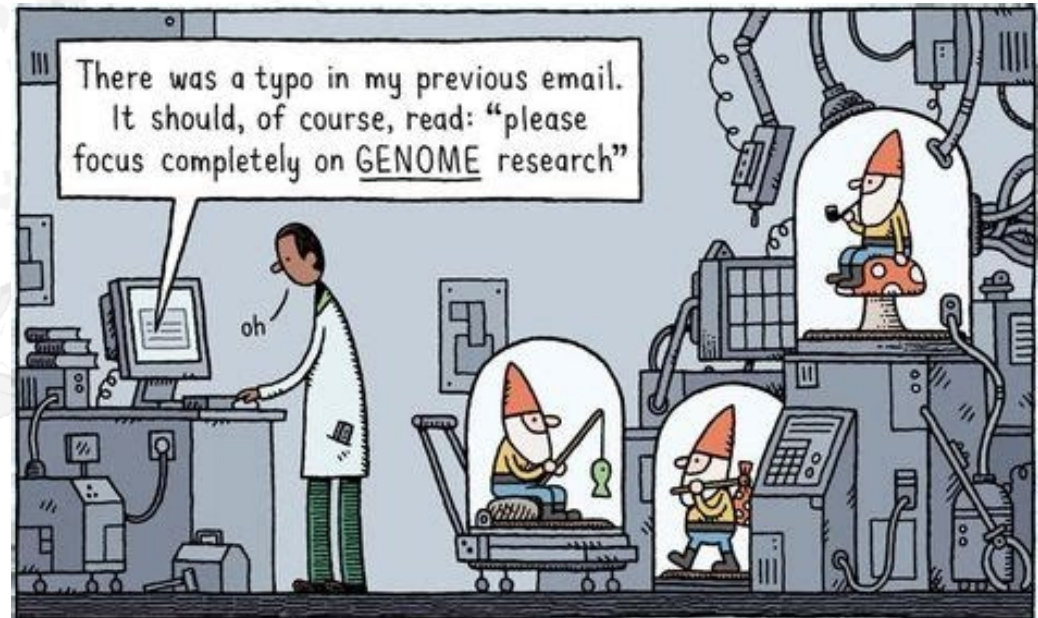
Number of raw loci recovered per year of collection



● Semturiidae sister to Pseudobistonidae+Epicopeiidae

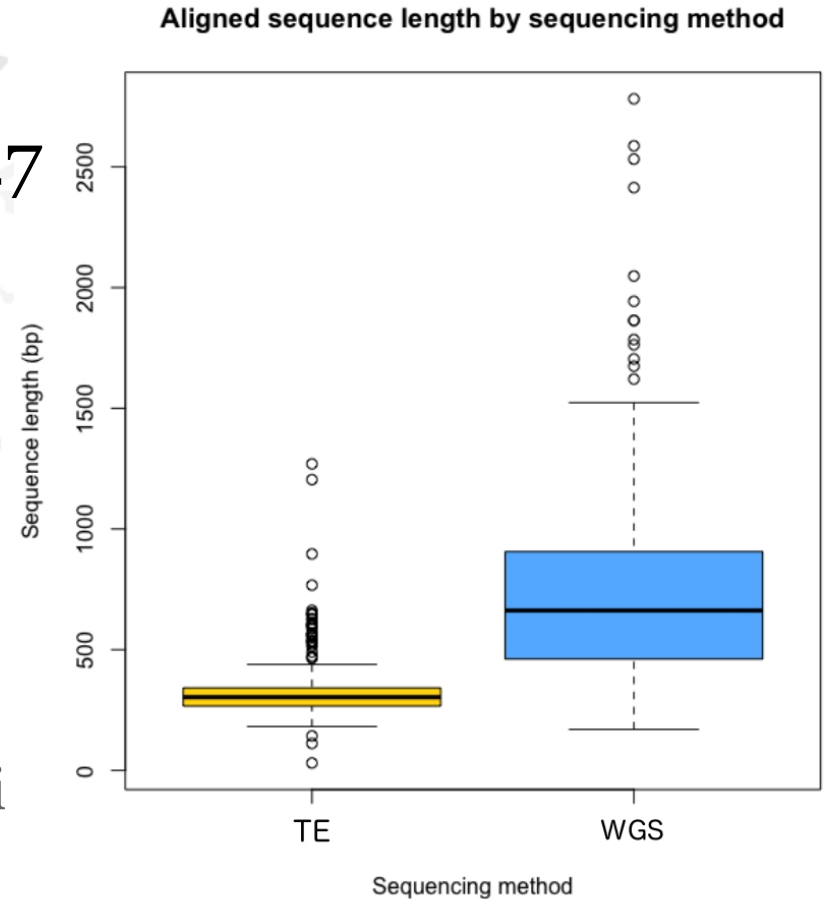


WHOLE-GENOME SEQUENCING (WGS) RESULTS & DISCUSSION

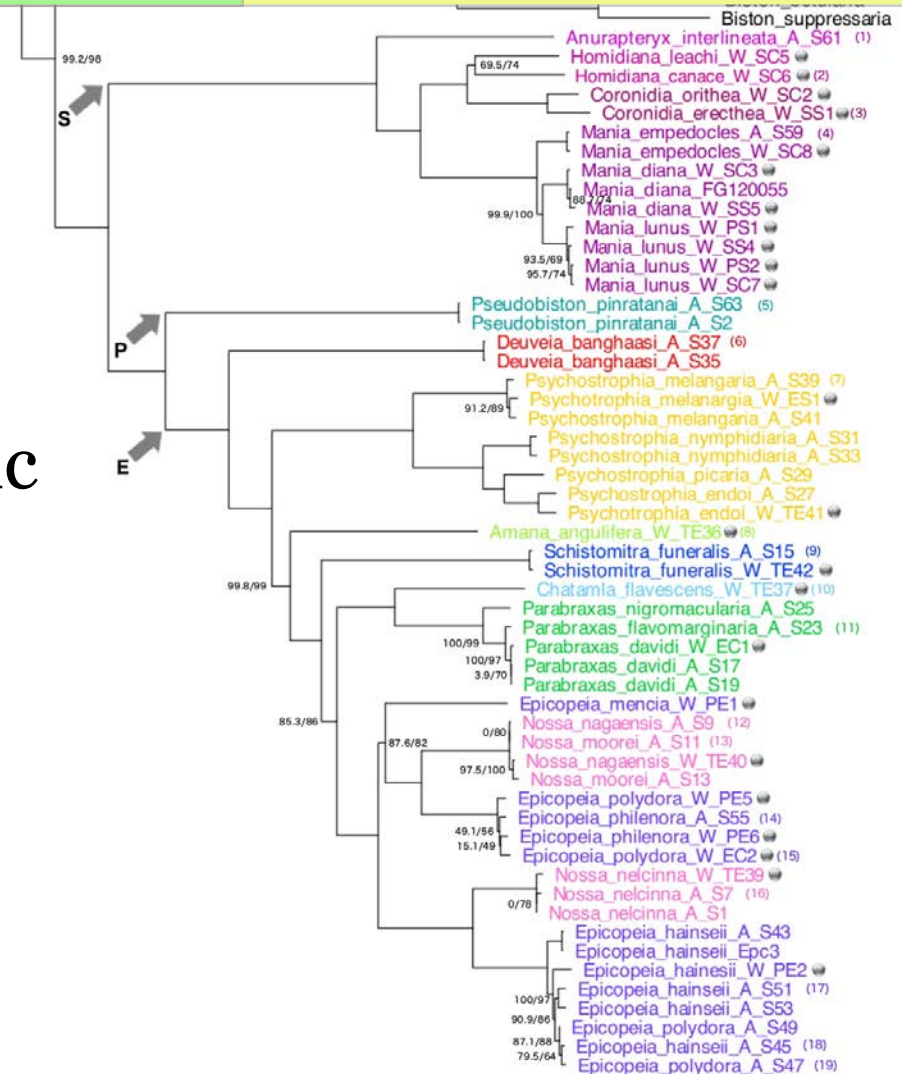


- *De novo* assembly 205 Mbp
- 308 genes recovered from the 327
(Call *et al.*, 2021)

- *De novo* assembly 205 Mbp
- 308 genes recovered from the 327 (Call *et al.*, 2021)
- WGS length longer
 - On average 786 bp
- TE get targeted genes
- Bias for TE
 - Investigate the 1,753 genes & 2,953 loci
 - Other genes sets



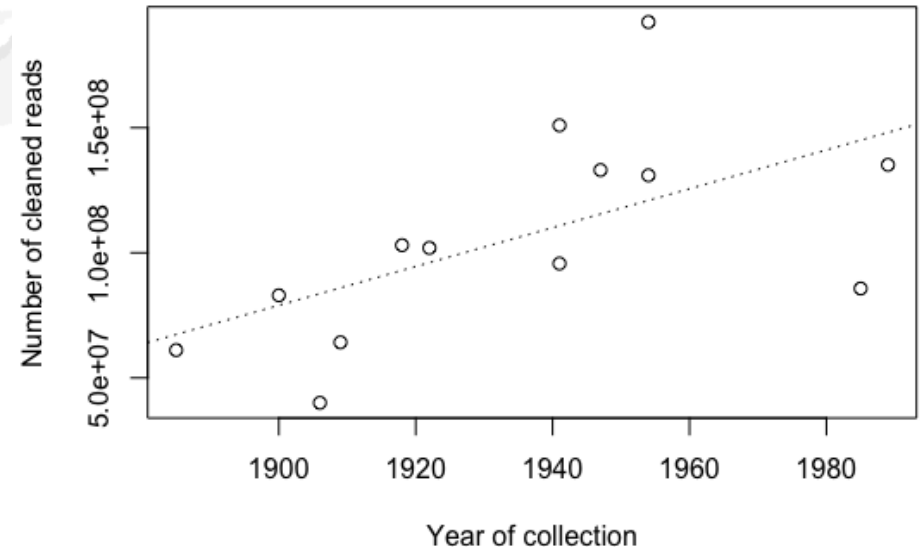
- Hypothesis of Call *et al.*, 2021 confirmed
- Misidentifications
- *Nossa* & *Epicopeia* paraphyletic



- Population genetics(?)
- 81.40% of sequences are from *Pieris napi*
- 100% mitochondrial genome
 - Coverage of 1,963X
- Oldest specimens have fewer loci
 - From 1885 \Rightarrow 80.23% of the nuclear genome

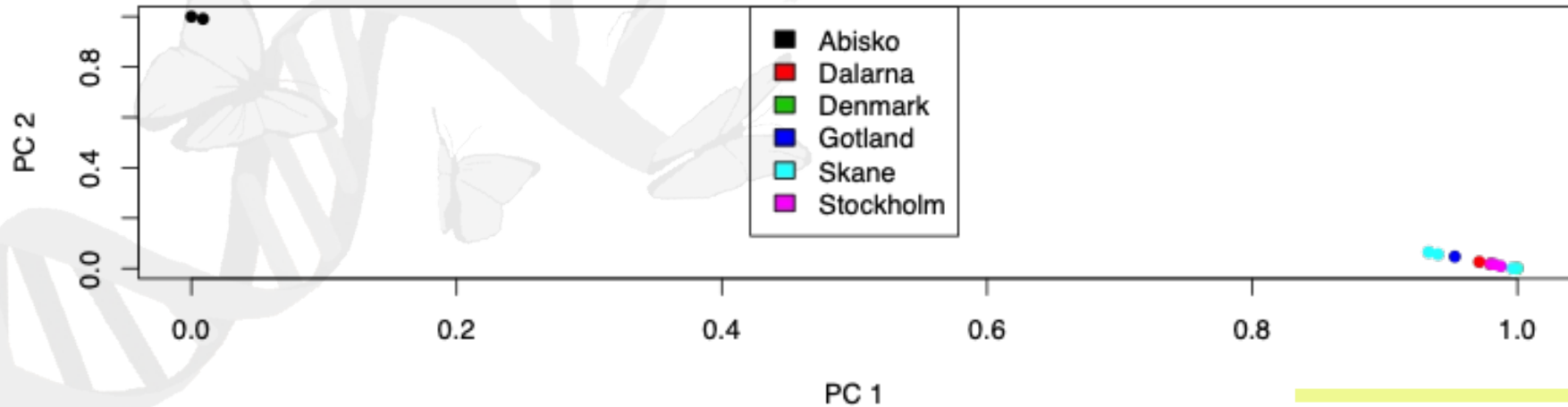


Number of cleaned reads recovered per year of collection



- Population genetics statistics
 - Heterozygosity, F_{ST} , admixture & inbreeding
- Abisko significantly different from other populations
 - *P. n. adalwinda* (Abisko) subspecies of *Pieris napi*

Principal components



- Recover good proportion of insects' genomes
 - Both mitochondrial & nuclear
- Phylogenetic relationships + use for population genetics
- Comparison between TE & WGS



THANK YOU FOR YOUR ATTENTION

SPECIAL THANKS TO:

Niklas WAHLBERG

Victoria TWORT

Marianne ESPELAND

Christoph MAYER

Christopher W. WHEAT

Department of Biology
(Lund University)

BIG4 Members
(supervisors & students)

Museums: Lund, Bonn,
Copenhagen; Leiden,
Stockholm & Tokyo



This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 6422141



LUND
UNIVERSITY

