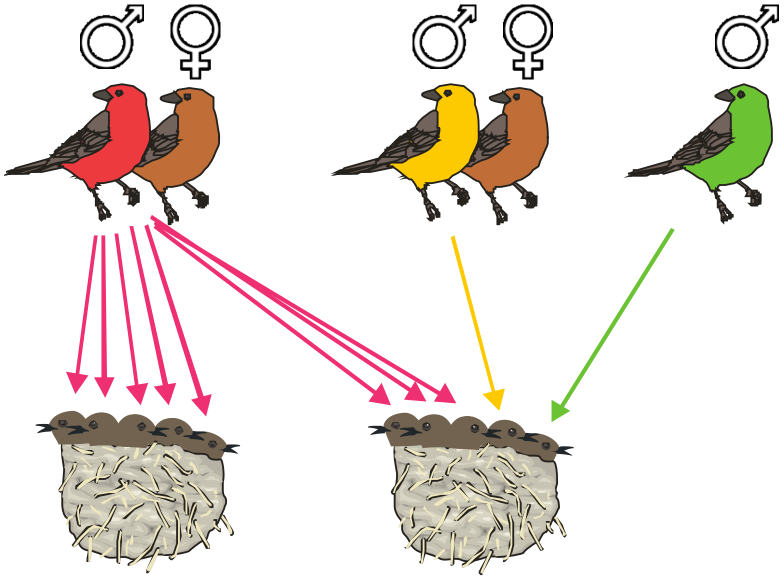
Molecular identification

Species, individual, sex

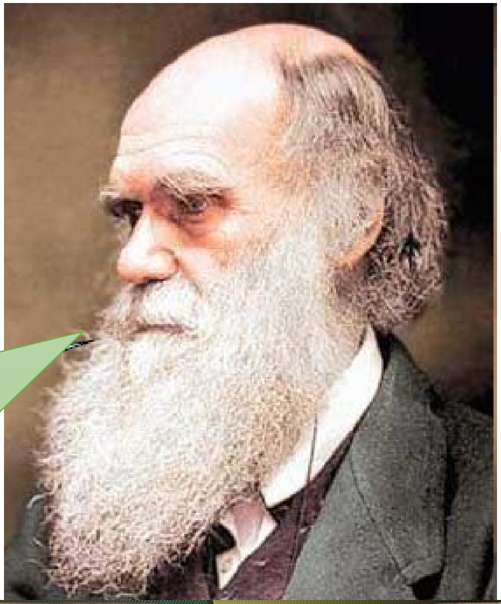


Identification of species

DNA barcoding

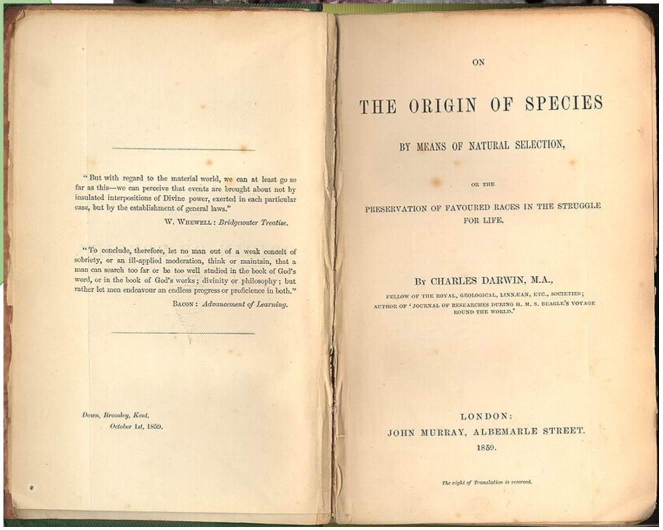
WHAT THE SPECIES IS AND DO WE NEED THEM?

Is it possible to define a species?



It is really laughable to see what different ideas are prominent in various naturalists minds, when they speak of „species“; ... It all comes, I believe, from trying to define the indefinable.

C. Darwin, 24 Dec 1856
(Letter to J.D. Hooker)



"But with regard to the material world, we can at least go so far as this—we can perceive that events are brought about not by insulated interpositions of Divine power, exerted in each particular case, but by the establishment of general laws."

W. Whewell: *Bridgewater Treatise.*

"To conclude, therefore, let no man out of a weak conceit of sobriety, or an ill-applied moderation, think or maintain, that a man can search too far or be too well studied in the book of God's word, or in the book of God's works; divinity or philosophy; but rather let men endeavour an endless progress or proficience in both."

Bacon: *Advancement of Learning.*

Down, Bromley, Kent,
October 1st, 1859.

ON

THE ORIGIN OF SPECIES

BY MEANS OF NATURAL SELECTION,

OR THE

PRESERVATION OF FAVOURED RACES IN THE STRUGGLE
FOR LIFE.

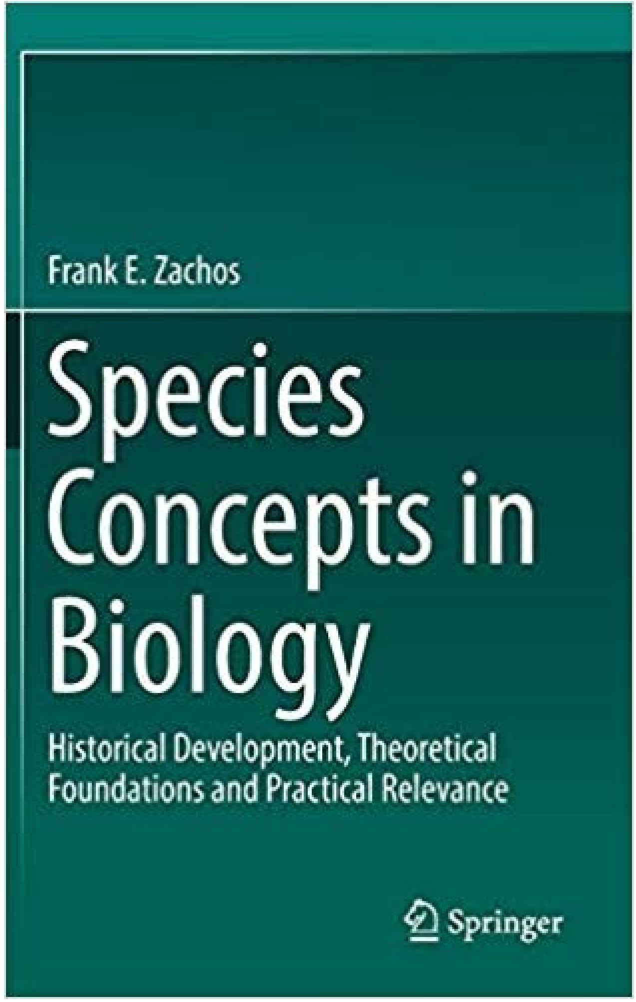
By CHARLES DARWIN, M.A.,
FELLOW OF THE ROYAL, GEOLOGICAL, LINNÆAN, ETC., SOCIETIES;
AUTHOR OF 'JOURNAL OF RESEARCHES DURING H. M. S. BEAGLE'S VOYAGE
ROUND THE WORLD.'

LONDON:
JOHN MURRAY, ALBEMARLE STREET.
1859.

The right of Translation is reserved.

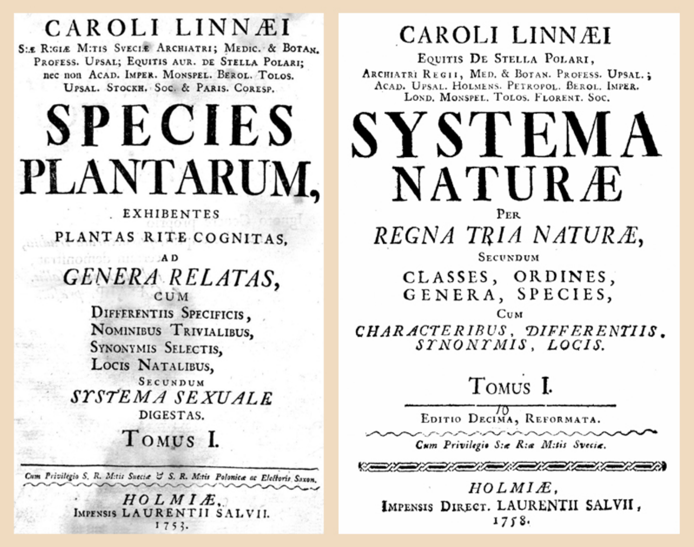
Species concepts in biology

- Agamospecies
- Biological
- Biosimilarity
- Cladistic
- Cohesion
- Compilospecies
- Differential Fitness
- Ecological
- Evolutionarily Significant Unit
- Evolutionary
- Genealogical
- Genealogical Concordance
- General Lineage
- Genetic
- Genic
- Genoypic cluster
- Hennigian
- Internodal
- Least Inclusive Taxonomic Unit
- Morphological
- Non-dimensional
- Nothospecies
- Phenetic
- Phylogenetic (Diagnosability Version)
- Phylogenetic (Monophyly Version)
- Phylo-Phenetic
- Pragmatic
- Recognition
- Reproductive Competition
- Successional
- Taxonomic
- Unified



Morphological Species Concept

- = the smallest groups that are consistently and **persistently distinct, and distinguishable by ordinary means**
- Aristoteles → Linnaeus → rules of zoological nomenclature
- does not take evolution into account



CAROLI LINNÆI
S:æ R:GIÆ M:TIS SVECIÆ ARCHIATRI; MEDIC. & BOTAN.
PROFESS. UPSAL; EQUITIS AUR. DE STELLA POLARI;
nec non ACAD. IMPER. MONSPEL. BEROL. TOLOS.
UPSAL. STOCKH. SOC. & PARIS. CORESP.
SPECIES
PLANTARUM,
EXHIBENTES
PLANTAS RITE COGNITAS,
AD
GENERA RELATAS,
CUM
DIFFERENTIIS SPECIFICIS,
NOMINIBUS TRIVIALIBUS,
SYNONYMIS SELECTIS,
LOCIS NATALIBUS,
SECUNDUM
SYSTEMA SEXUALE
DIGESTAS.
TOMUS I.

Cum Privilegio S. R. M:tis Sueciæ & S. R. M:tis Polonicæ ac Electoris Saxon.

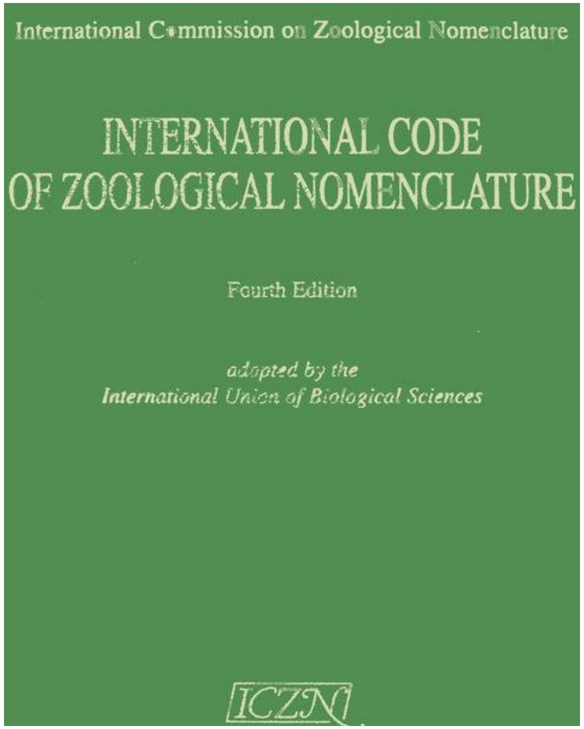
HOLMIÆ,
IMPENSIS LAURENTII SALVII.
1753.

CAROLI LINNÆI
EQUITIS DE STELLA POLARI,
ARCHIATRI REGII, MED. & BOTAN. PROFESS. UPSAL.;
ACAD. UPSAL. HOLMENS. PETROPOL. BEROL. IMPER.
LOND. MONSPEL. TOLOS. FLORENT. SOC.
SYSTEMA
NATURÆ
PER
REGNA TRIA NATURÆ,
SECUNDUM
CLASSES, ORDINES,
GENERA, SPECIES,
CUM
CHARACTERIBUS, DIFFERENTIIS.
SYNONYMIS, LOCIS.
TOMUS I.
10
EDITIO DECIMA, REFORMATA.

Cum Privilegio S:æ R:æ M:tis Sveciæ.

HOLMIÆ,
IMPENSIS DIRECT. LAURENTII SALVII,
1758.

Linnaeus, Carolus (1707 - 1778)



International Commission on Zoological Nomenclature
INTERNATIONAL CODE
OF ZOOLOGICAL NOMENCLATURE
Fourth Edition
adopted by the
International Union of Biological Sciences
ICZN

Biological Species Concept

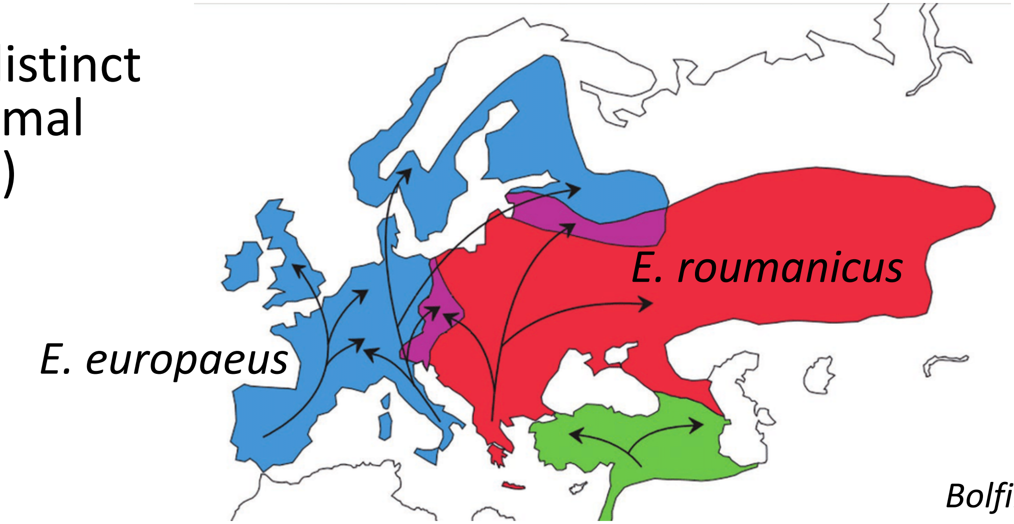
- = **interbreeding natural populations reproductively isolated from other such group**
- reproductive isolation mechanisms (RIM) = post- or prezygotic barriers of reproduction
- **most popular** – it is intuitive and it was promoted most successfully (e.g. by influential evolutionary biologists of the 20th century as main concept of Modern Synthesis)

- problems: allopatric and allochronic populations/species, ...

Mayr, 1942

Complications: Parapatric contact zones

- *Erinaceus* – distinct species (minimal hybridization)



Bolfiková and Hulva 2012

- house mice – distinct subspecies (substantial hybridization)

Macholán et al. 2008

Complications: Ring specie

Complications: Physical constraints

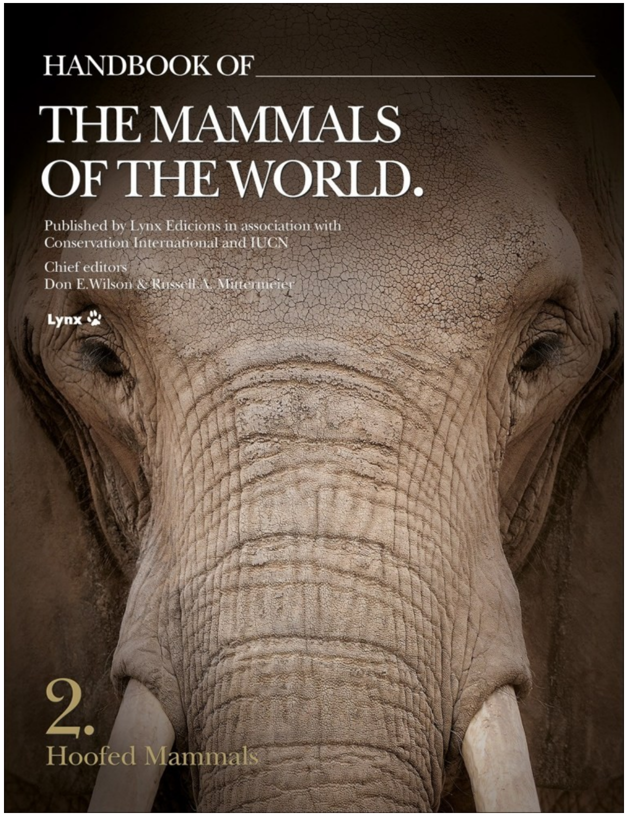
Phylogenetic species concept (Diagnosability Version)

- = the smallest population or group of populations, within which there is a **parental pattern of ancestry and descent**
- two populations are considered species if they are **100% diagnostic** (e.g. discriminant analysis of morphometric data or allele frequency data)
- recent **paradigmatic shift from the Biological Species Concept to Phylogenetic Species Concept**
- extreme cases: descendants of a single mother with a mutation at mtDNA can be 100% diagnosed (i.e. should be considered species)

Cracraft, 1983, Cracraft 1997

Example: Taxonomy of ungulates

- increase from 143 (Grubb 2005) to 279 (Groves and Leslie 2011 - Handbook of the Mammals of the World) species of bovid ungulates



HANDBOOK OF

THE MAMMALS OF THE WORLD.

Published by Lynx Edicions in association with
Conservation International and IUCN

Chief editors
Don E.Wilson & Russell A. Mittermeier

Lynx

2.
Hoofed Mammals

klipspringers
Oreotragus –
from one to 11
species

Consequences intensively debated

- „pros“papers

Horizon Scanning Article

“Taxonomic inflation” in the historical context of mammalogy and conservation

Spartaco Gippoliti[a,*], Colin P. Groves[b]

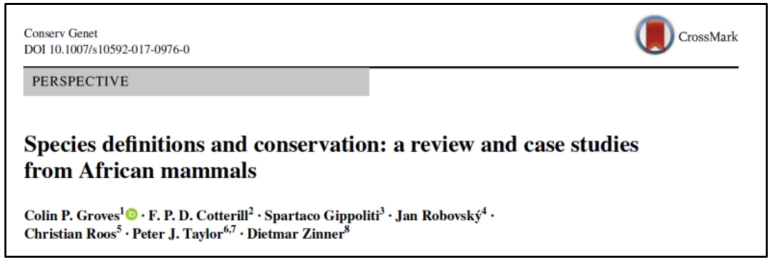
[a]*Viale Liegi 48, 00198 Roma, Italy*
[b]*School of Archaeology & Anthropology, Australian National University, Canberra, Australia*

BIOLOGICAL REVIEWS
Cambridge Philosophical Society

Biol. Rev. (2018), **93**, pp. 115–130. 115
doi: 10.1111/brv.12335

Impacts of taxonomic inertia for the conservation of African ungulate diversity: an overview

Spartaco Gippoliti[1], Fenton P. D. Cotterill[2], Dietmar Zinner[3,*] and Colin P. Groves[4]



Conserv Genet
DOI 10.1007/s10592-017-0976-0
CrossMark

PERSPECTIVE

Species definitions and conservation: a review and case studies from African mammals

Colin P. Groves[1] · F. P. D. Cotterill[2] · Spartaco Gippoliti[3] · Jan Robovský[4] · Christian Roos[5] · Peter J. Taylor[6,7] · Dietmar Zinner[8]

Syst. Biol. 63(5):819–832, 2014
©The Author(s) 2014. Published by Oxford University Press, on behalf of the Society of Systematic Biologists. All rights reserved.
For Permissions, please email: journals.permissions@oup.com
DOI:10.1093/sysbio/syu003
Advance Access publication January 11, 2014

Why One Century of Phenetics is Enough: Response to “Are There Really Twice As Many Bovid Species As We Thought?”

Fenton P. D. Cotterill[1,*], Peter J. Taylor[2], Spartaco Gippoliti[3], Jacqueline M. Bishop[4], and Colin P. Groves[5]

Commentary

Mammal taxonomy without taxonomists: a reply to Zachos and Lovari

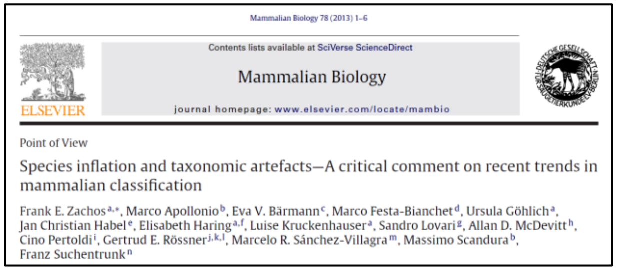
Spartaco Gippoliti[a,*], Fenton P.D. Cotterill[b], Colin P. Groves[c]

- „cons“ papers

Opinion TRENDS in Ecology and Evolution Vol.19 No.9 September 2004
ELSEVIER
Full text provided by www.sciencedirect.com
SCIENCE DIRECT

Taxonomic inflation: its influence on macroecology and conservation

Nick J.B. Isaac[1], James Mallet[2] and Georgina M. Mace[1]



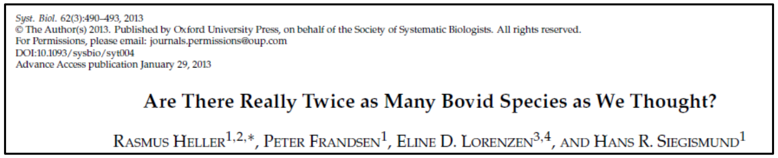
Mammalian Biology 78 (2013) 1–6

Contents lists available at SciVerse ScienceDirect
Mammalian Biology
journal homepage: www.elsevier.com/locate/mambio
ELSEVIER

Point of View

Species inflation and taxonomic artefacts—A critical comment on recent trends in mammalian classification

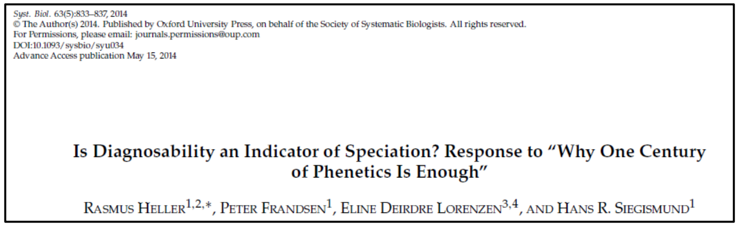
Frank E. Zachos[a,*], Marco Apollonio[b], Eva V. Bärmann[c], Marco Festa-Bianchet[d], Ursula Göhlich[a], Jan Christian Habel[e], Elisabeth Haring[a,f], Luise Kruckenhauser[a], Sandro Lovari[g], Allan D. McDevitt[h], Cino Pertoldi[i], Gertrud E. Rössner[j,k,l], Marcelo R. Sánchez-Villagra[m], Massimo Scandura[b], Franz Suchentrunk[n]



Syst. Biol. 62(3):490–493, 2013
© The Author(s) 2013. Published by Oxford University Press, on behalf of the Society of Systematic Biologists. All rights reserved.
For Permissions, please email: journals.permissions@oup.com
DOI:10.1093/sysbio/syt004
Advance Access publication January 29, 2013

Are There Really Twice as Many Bovid Species as We Thought?

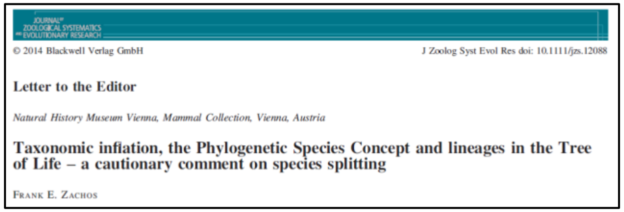
Rasmus Heller[1,2,*], Peter Frandsen[1], Eline D. Lorenzen[3,4], and Hans R. Siegismund[1]



Syst. Biol. 63(5):833–837, 2014
© The Author(s) 2014. Published by Oxford University Press, on behalf of the Society of Systematic Biologists. All rights reserved.
For Permissions, please email: journals.permissions@oup.com
DOI:10.1093/sysbio/syu034
Advance Access publication May 15, 2014

Is Diagnosability an Indicator of Speciation? Response to “Why One Century of Phenetics Is Enough”

Rasmus Heller[1,2,*], Peter Frandsen[1], Eline Deirdre Lorenzen[3,4], and Hans R. Siegismund[1]



JOURNAL of ZOOLOGICAL SYSTEMATICS and EVOLUTIONARY RESEARCH

© 2014 Blackwell Verlag GmbH J Zoolog Syst Evol Res doi: 10.1111/jzs.12088

Letter to the Editor

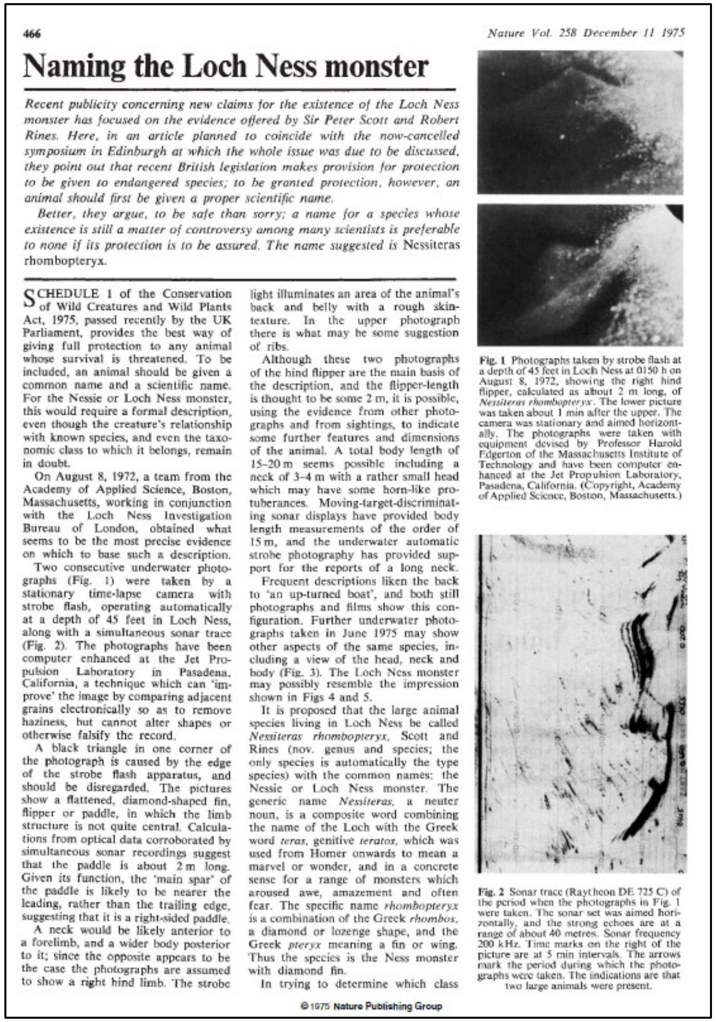
Natural History Museum Vienna, Mammal Collection, Vienna, Austria

Taxonomic inflation, the Phylogenetic Species Concept and lineages in the Tree of Life – a cautionary comment on species splitting

Frank E. Zachos

Why does it matter? The power of names!

- description of *Nessiteras rhombopteryx* in Nature
- „the Loch Ness monster“ – if indeed it does exist, it exists in small numbers and deserves protection
- to be protected **it must have a taxonomic name**



466 Nature Vol. 258 December 11 1975

Naming the Loch Ness monster

Recent publicity concerning new claims for the existence of the Loch Ness monster has focused on the evidence offered by Sir Peter Scott and Robert Rines. Here, in an article planned to coincide with the now-cancelled symposium in Edinburgh at which the whole issue was due to be discussed, they point out that recent British legislation makes provision for protection to be given to endangered species; to be granted protection, however, an animal should first be given a proper scientific name.

Better, they argue, to be safe than sorry; a name for a species whose existence is still a matter of controversy among many scientists is preferable to none if its protection is to be assured. The name suggested is Nessiteras rhombopteryx.

SCHEDULE 1 of the Conservation of Wild Creatures and Wild Plants Act, 1975, passed recently by the UK Parliament, provides the best way of giving full protection to any animal whose survival is threatened. To be included, an animal should be given a common name and a scientific name. For the Nessie or Loch Ness monster, this would require a formal description, even though the creature's relationship with known species, and even the taxonomic class to which it belongs, remain in doubt.

On August 8, 1972, a team from the Academy of Applied Science, Boston, Massachusetts, working in conjunction with the Loch Ness Investigation Bureau of London, obtained what seems to be the most precise evidence on which to base such a description.

Two consecutive underwater photographs (Fig. 1) were taken by a stationary time-lapse camera with strobe flash, operating automatically at a depth of 45 feet in Loch Ness, along with a simultaneous sonar trace (Fig. 2). The photographs have been computer enhanced at the Jet Propulsion Laboratory in Pasadena, California, a technique which can 'improve' the image by comparing adjacent grains electronically so as to remove haziness, but cannot alter shapes or otherwise falsify the record.

A black triangle in one corner of the photograph is caused by the edge of the strobe flash apparatus, and should be disregarded. The pictures show a flattened, diamond-shaped fin, flipper or paddle, in which the limb structure is not quite central. Calculations from optical data corroborated by simultaneous sonar recordings suggest that the paddle is about 2 m long. Given its function, the 'main spar' of the paddle is likely to be nearer the leading, rather than the trailing edge, suggesting that it is a right-sided paddle.

A neck would be likely anterior to a forelimb, and a wider body posterior to it; since the opposite appears to be the case the photographs are assumed to show a right hind limb. The strobe light illuminates an area of the animal's back and belly with a rough skin-texture. In the upper photograph there is what may be some suggestion of ribs.

Although these two photographs of the hind flipper are the main basis of the description, and the flipper-length is thought to be some 2 m, it is possible, using the evidence from other photographs and from sightings, to indicate some further features and dimensions of the animal. A total body length of 15–20 m seems possible including a neck of 3–4 m with a rather small head which may have some horn-like protuberances. Moving-target-discriminating sonar displays have provided body length measurements of the order of 15 m, and the underwater automatic strobe photography has provided support for the reports of a long neck.

Frequent descriptions liken the back to 'an up-turned boat', and both still photographs and films show this configuration. Further underwater photographs taken in June 1975 may show other aspects of the same species, including a view of the head, neck and body (Fig. 3). The Loch Ness monster may possibly resemble the impression shown in Figs 4 and 5.

It is proposed that the large animal species living in Loch Ness be called *Nessiteras rhombopteryx*, Scott and Rines (nov. genus and species; the only species is automatically the type species) with the common names: the Nessie or Loch Ness monster. The generic name *Nessiteras*, a neuter noun, is a composite word combining the name of the Loch with the Greek word *teras*, genitive *teratos*, which was used from Homer onwards to mean a marvel or wonder, and in a concrete sense for a range of monsters which aroused awe, amazement and often fear. The specific name *rhombopteryx* is a combination of the Greek *rhombos*, a diamond or lozenge shape, and the Greek *pteryx* meaning a fin or wing. Thus the species is the Ness monster with diamond fin.

In trying to determine which class

Fig. 1 Photographs taken by strobe flash at a depth of 45 feet in Loch Ness at 0150 h on August 8, 1972, showing the right hind flipper, calculated as about 2 m long, of *Nessiteras rhombopteryx*. The lower picture was taken about 1 min after the upper. The camera was stationary and aimed horizontally. The photographs were taken with equipment devised by Professor Harold Edgerton of the Massachusetts Institute of Technology and have been computer enhanced at the Jet Propulsion Laboratory, Pasadena, California. (Copyright, Academy of Applied Science, Boston, Massachusetts.)

Fig. 2 Sonar trace (Raytheon DE 725 C) of the period when the photographs in Fig. 1 were taken. The sonar set was aimed horizontally, and the strong echoes are at a range of about 40 metres. Sonar frequency 200 kHz. Time marks on the right of the picture are at 5 min intervals. The arrows mark the period during which the photographs were taken. The indications are that two large animals were present.

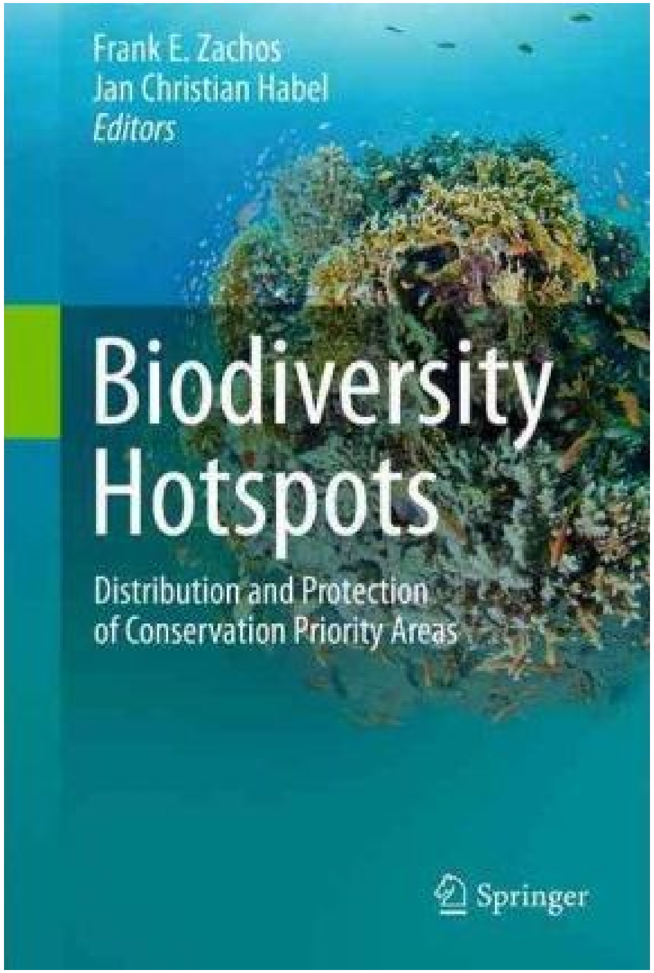
Rinnes and Scott 1975, Nature

Why does it matter? Taxonomic inflation.

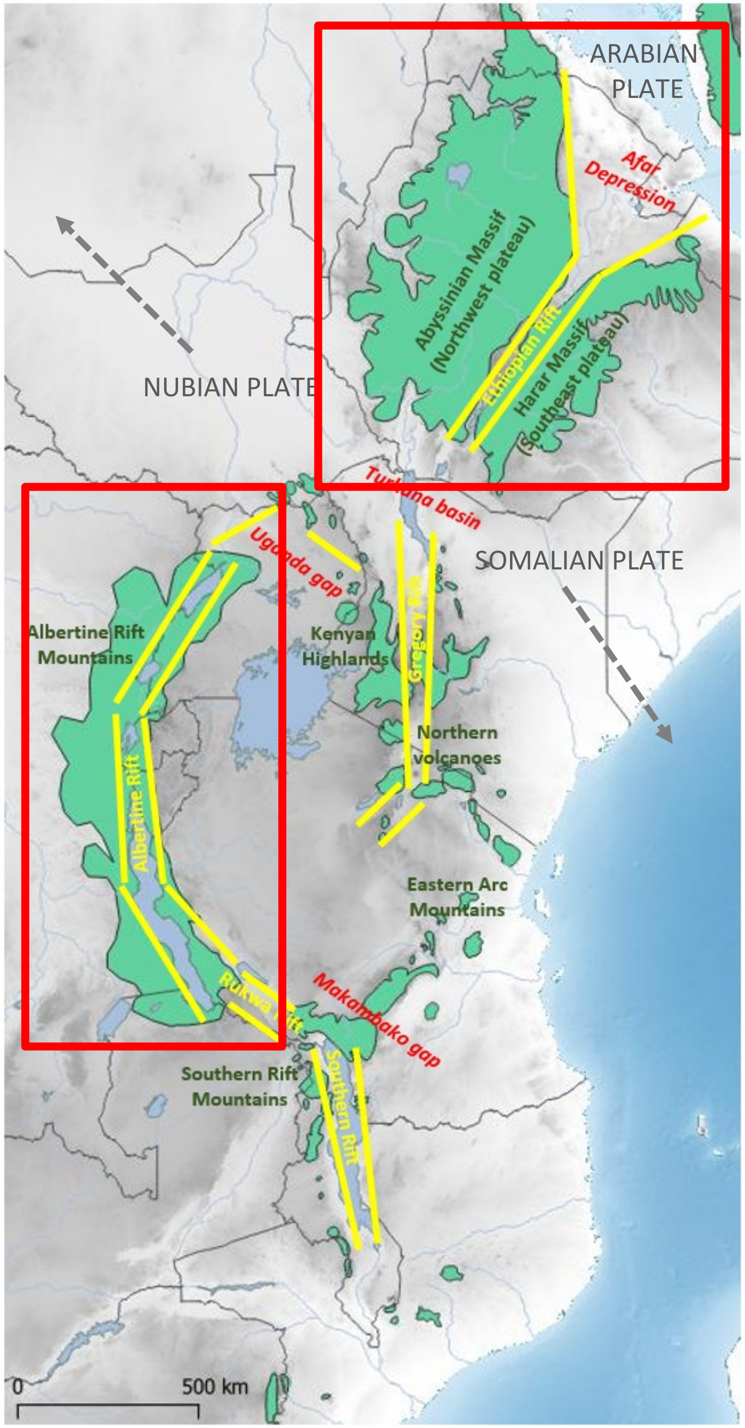
- **inflation leads to devaluation**
- tigers (*Panthera tigris*) have been split into two species based on 3 diagnostic bp in the mitochondrial cytochrome *b* (*P. tigris* and *P. sumatrae*) (Cracraft et al. 1998)
- genetic drift of some other populations in India has already led to the fixation of unique haplotypes
- *„The fact that tigers are dwindling towards extinction will thus cause a multitude of new tiger „species“ – before they all vanish“* (Zachos 2016)
- PSC → increase of threatened species – many species will have **low population sizes and distribution ranges (IUCN RED List criteria)** – e.g. US Endangered Species Act – increase from US$4.6 billion to US$7.6 billion for full recovery of all species

Why does it matter? Biodiversity research.

- **species richness is a function of the underlying species concept**
- often comparing „apples and oranges“
- 36 biodiversity hotspots (at least 1500 endemic vascular plants species and 70% of primary vegetation has been destroyed)
- more than US$1 billion for conservation in biodiversity hotspots



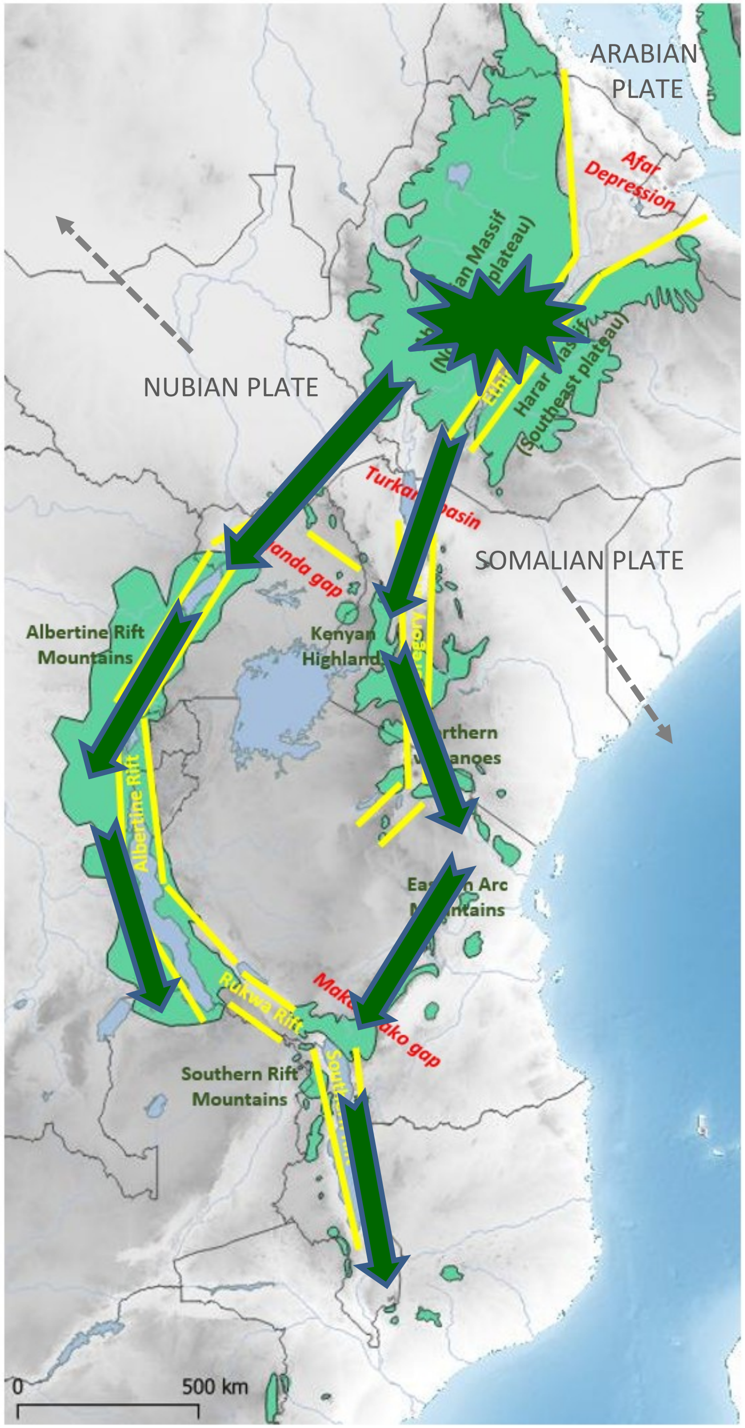
Mittermeier et al. 2011



Eastern Afromontane Biodiversity Hotspot (EABH)

- Albertine Rift – considered as the most diverse part of EABH
- Ethiopian Highlands – the most neglected part of EABH despite the large area and geomorphological diversity
- examples from rodents

„Ethiopian craddle“



Mus (*Nannomys*)

Otomys typus group

Lophuromys flavopunctatus group

Tachyoryctes

e.g. Bryja et al. 2019, Folia Zoologica

„out of Ethiopia“

root rats (*Tachyoryctes*)

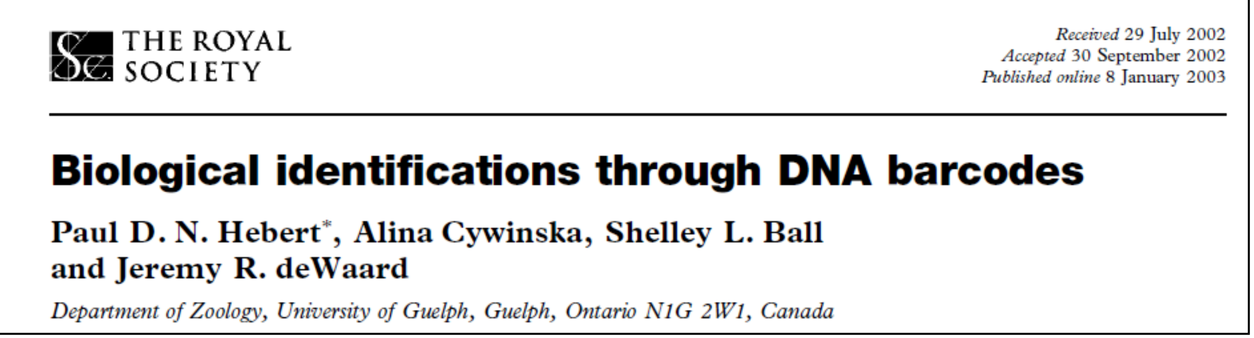
T. macrocephalus (ETH)

T. „splendens“ (1 sp. in Ethiopia, 12 spp. in Kenya and Albertine rift)

- the highest evolutionary diversity in Ethiopia (5 species using PSC)
- a single colonization of Kenyan Highlands and Albertine Rift Mts.

Šumbera, ..., Bryja 2018, Mol Phyl Evol

HOW CAN GENETICS BE HELPFUL IN SPECIES IDENTIFICATION?


THE ROYAL SOCIETY

Received 29 July 2002
Accepted 30 September 2002
Published online 8 January 2003

Biological identifications through DNA barcodes

Paul D. N. Hebert*, Alina Cywinska, Shelley L. Ball and Jeremy R. deWaard

Department of Zoology, University of Guelph, Guelph, Ontario N1G 2W1, Canada

first idea in 2003

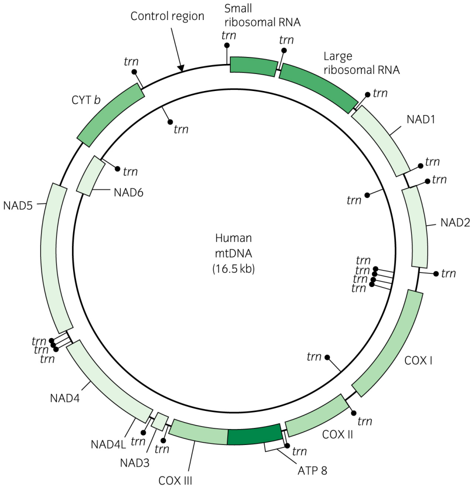
CONSORTIUM FOR THE BARCODE OF LIFE

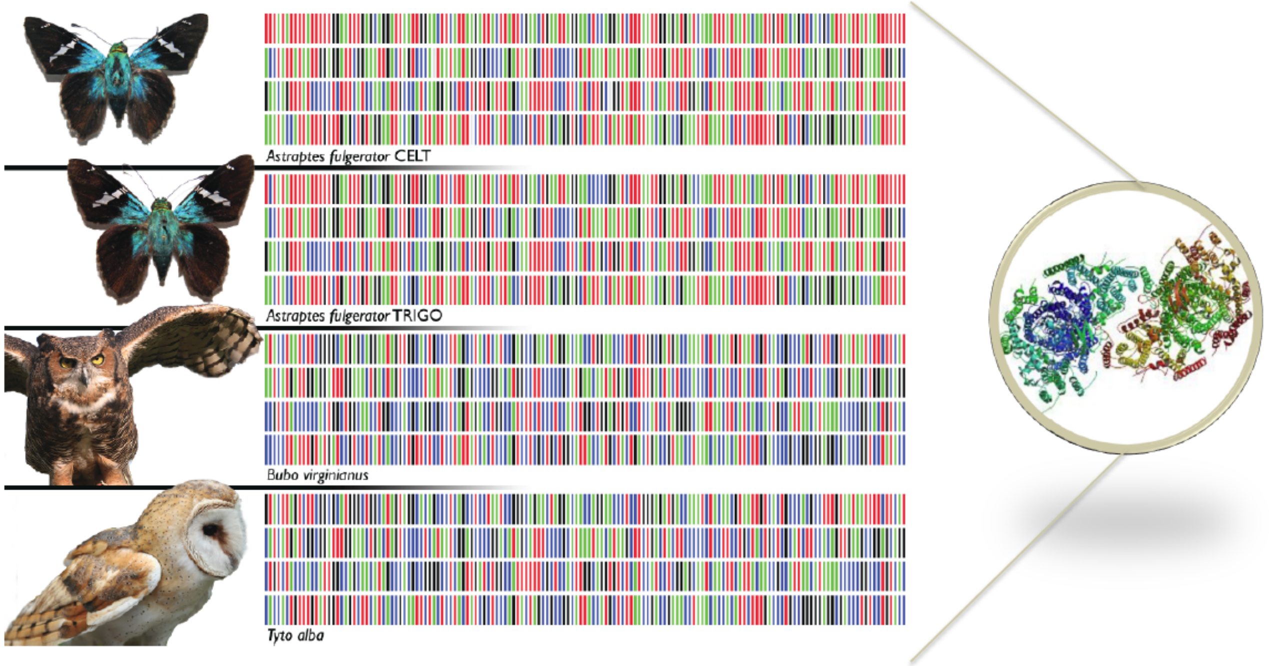
CBOL in 2005

international BARCODE OF LIFE

iBOL 2010-2015

500 000 species barcoded in 2015



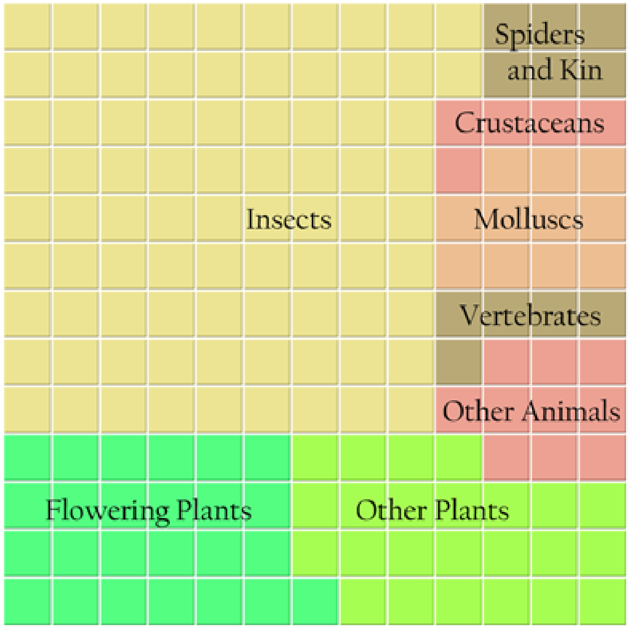


„DNA barcode“ – short fragment of mitochondrial DNA

Why barcode animal and plant species?

Known Biodiversity

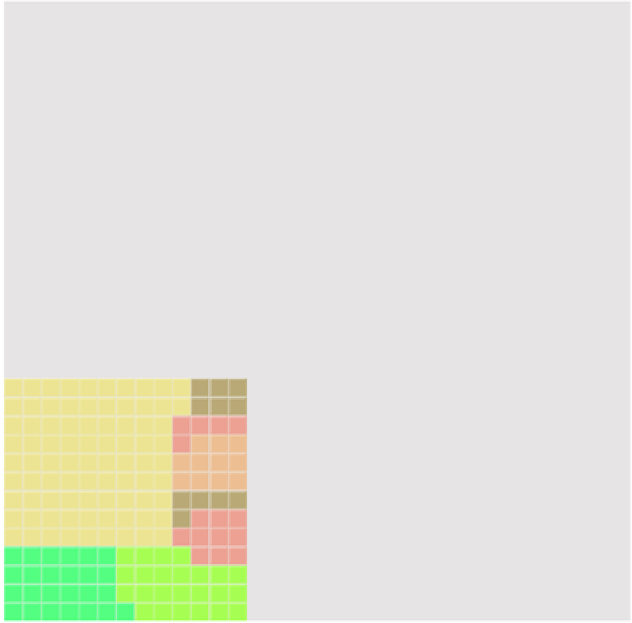
Approximately 1.7 million named species of plants and animals.



1 square = 10,000 species

Estimated Biodiversity

10 million species



Crisis of biodiversity and classical taxonomy

DNA barcoding is important part of „**integrative taxonomy**“

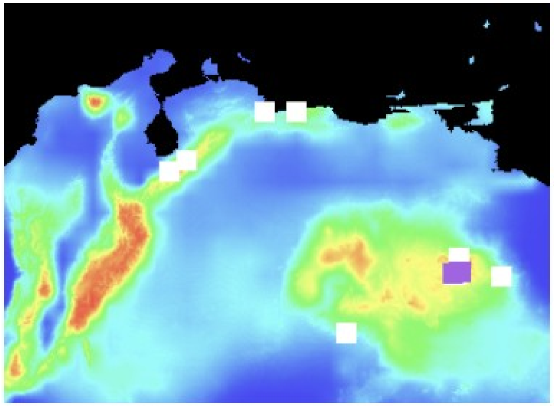
Integrative taxonomy

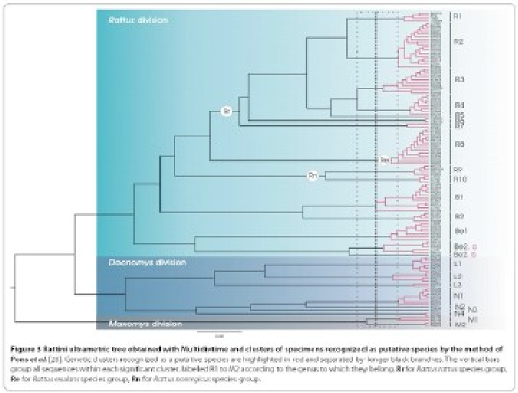
Integr. Taxonomy

Ecology

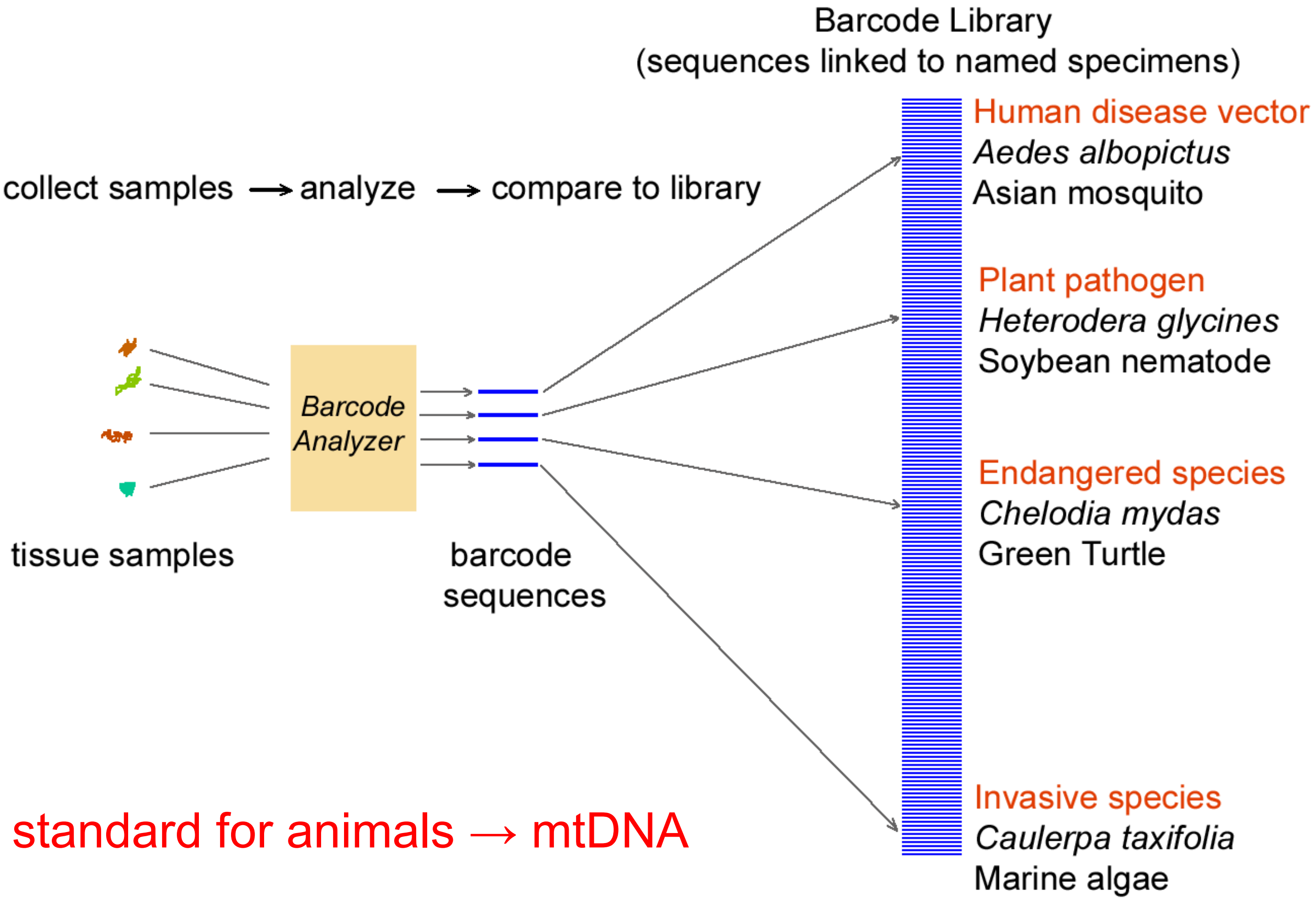
Genetics

Behavioural Biol.





What are the benefits of standardization?



Barcode Library
(sequences linked to named specimens)

collect samples → analyze → compare to library

tissue samples

barcode sequences

Human disease vector
Aedes albopictus
Asian mosquito

Plant pathogen
Heterodera glycines
Soybean nematode

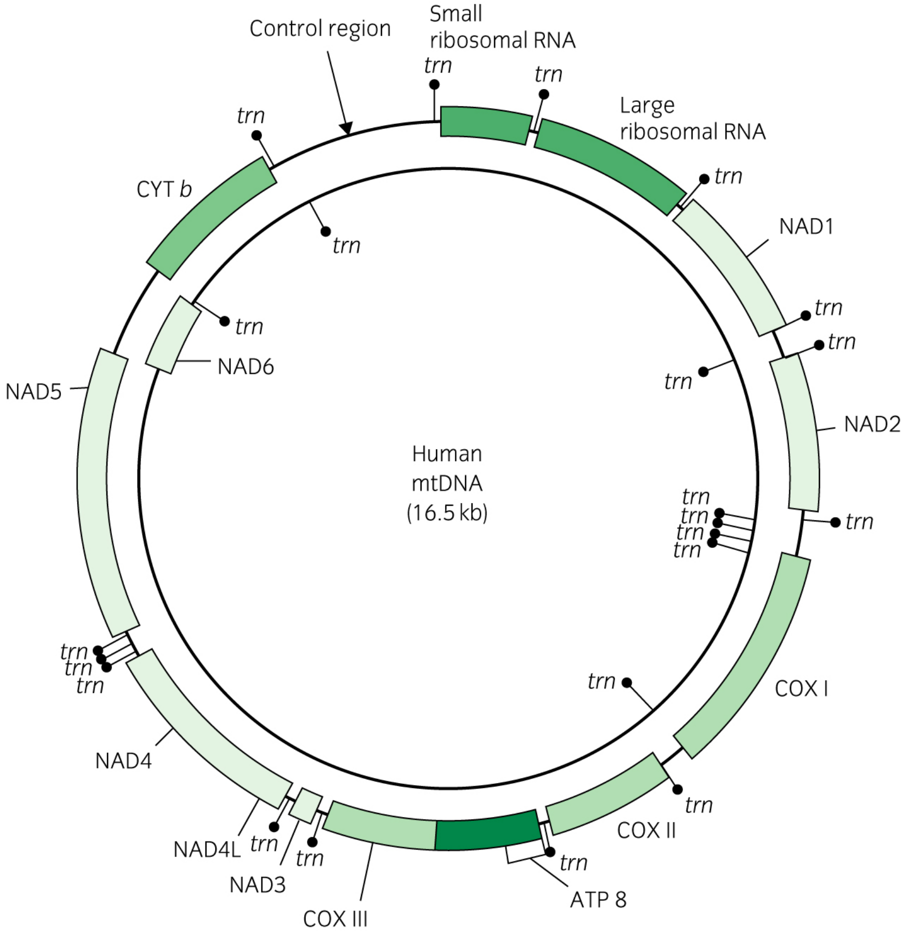
Endangered species
Chelodia mydas
Green Turtle

Invasive species
Caulerpa taxifolia
Marine algae

Suitable standard for animals → mtDNA

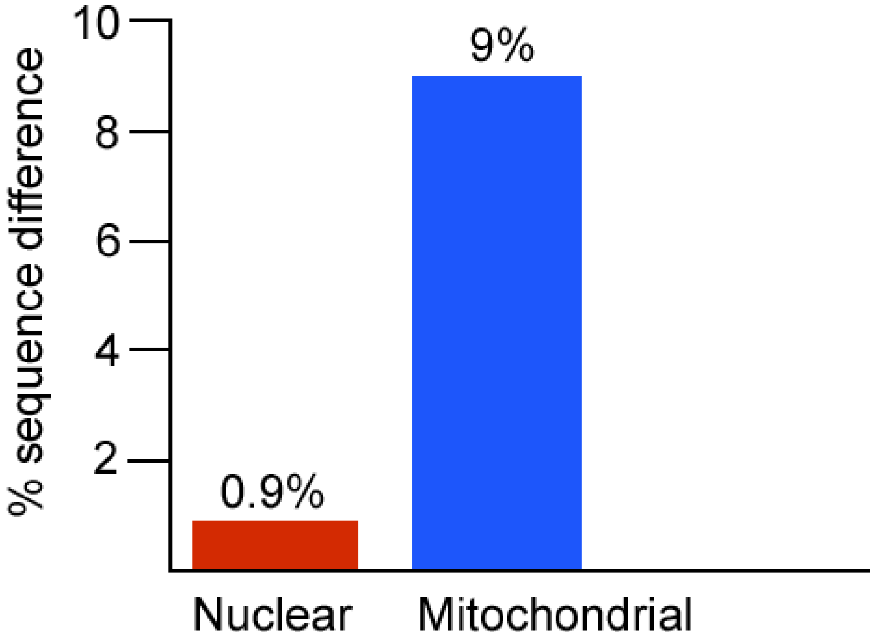
Why barcode animals with mitochondrial DNA?

Four properties make mitochondrial genomes especially suitable for identifying species



1. Greater differences among species, on average 5- to 10-fold higher in mitochondrial than in nuclear genes (lower N_e for mtDNA). Thus shorter segments distinguish among species, and because shorter, less expensively.

Average sequence differences in nuclear and mitochondrial DNA between human and chimp

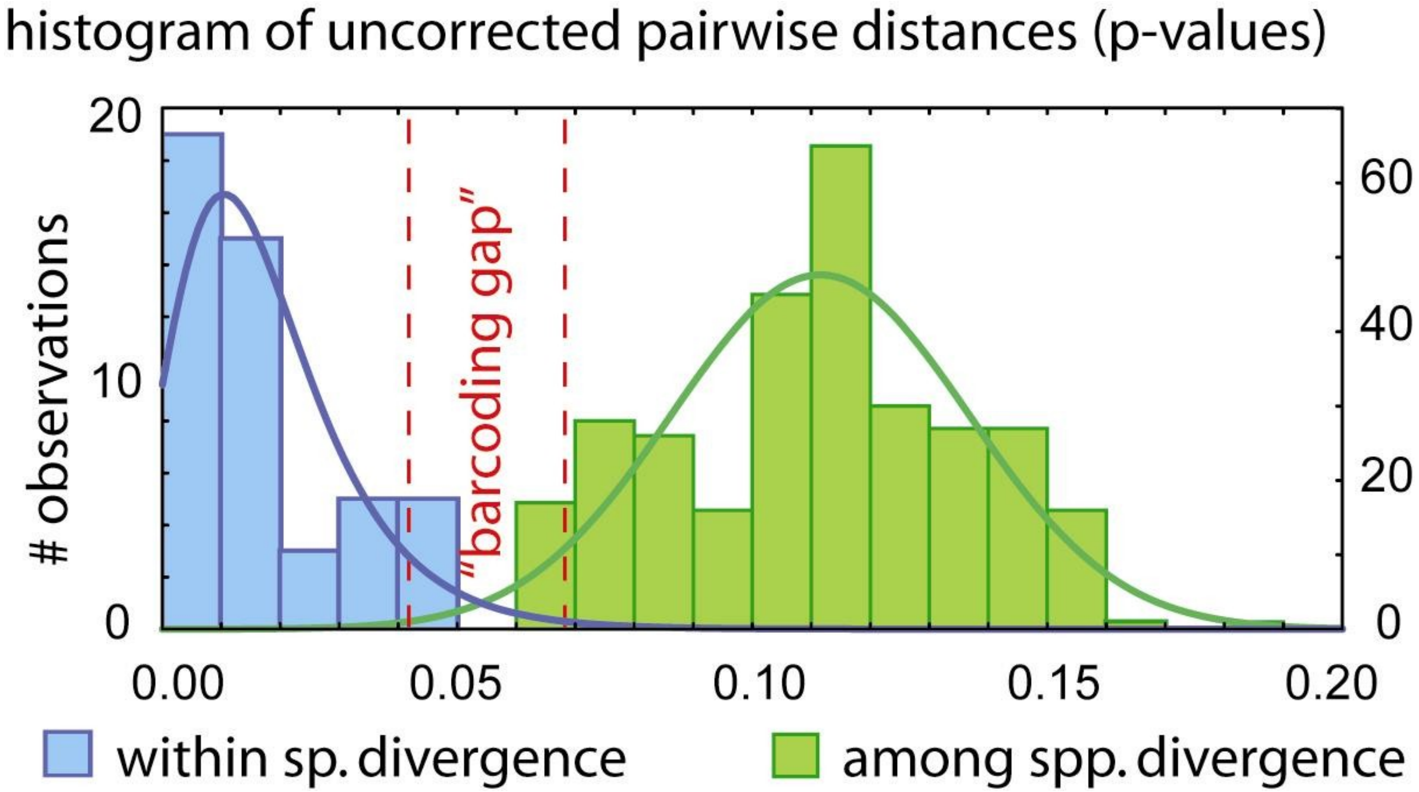


2. Relatively few differences within species in most cases. Small intraspecific and large interspecific differences signal distinct genetic boundaries between most species, enabling precise identification with a barcode.

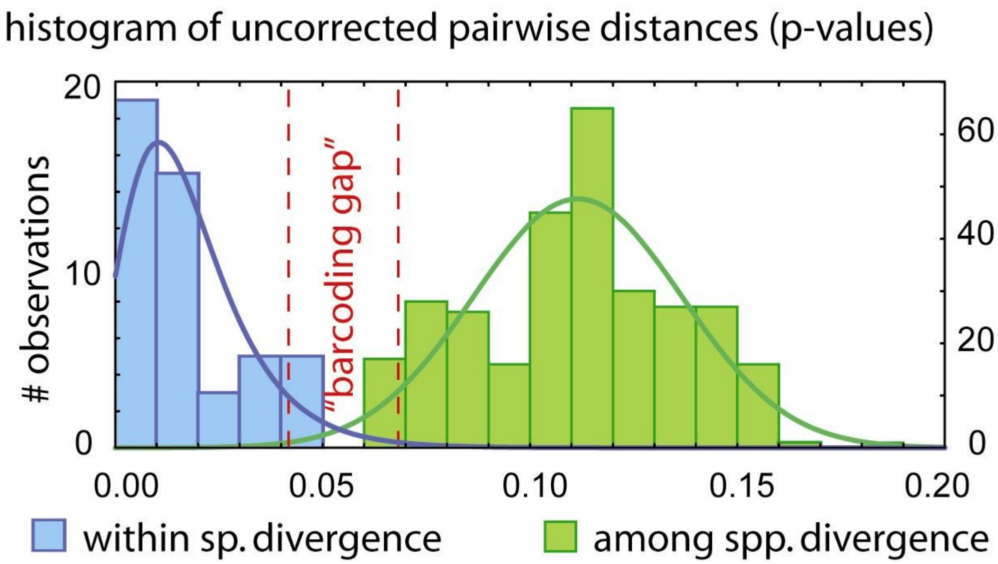
3. Copy number There are 100-10,000 more copies of mitochondrial than nuclear DNA per cell, making recovery, especially from small or partially degraded samples, easier and cheaper.

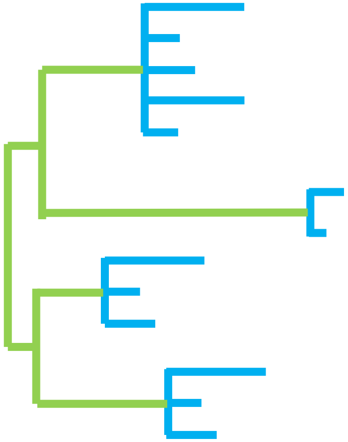
4. Introns, which are non-coding regions interspersed between coding regions of a gene, are absent from mitochondrial DNA of most animal species, making amplification straightforward. Nuclear genes are often interrupted by introns, making amplification difficult or unpredictable.

Barcoding principle



Barcoding principle



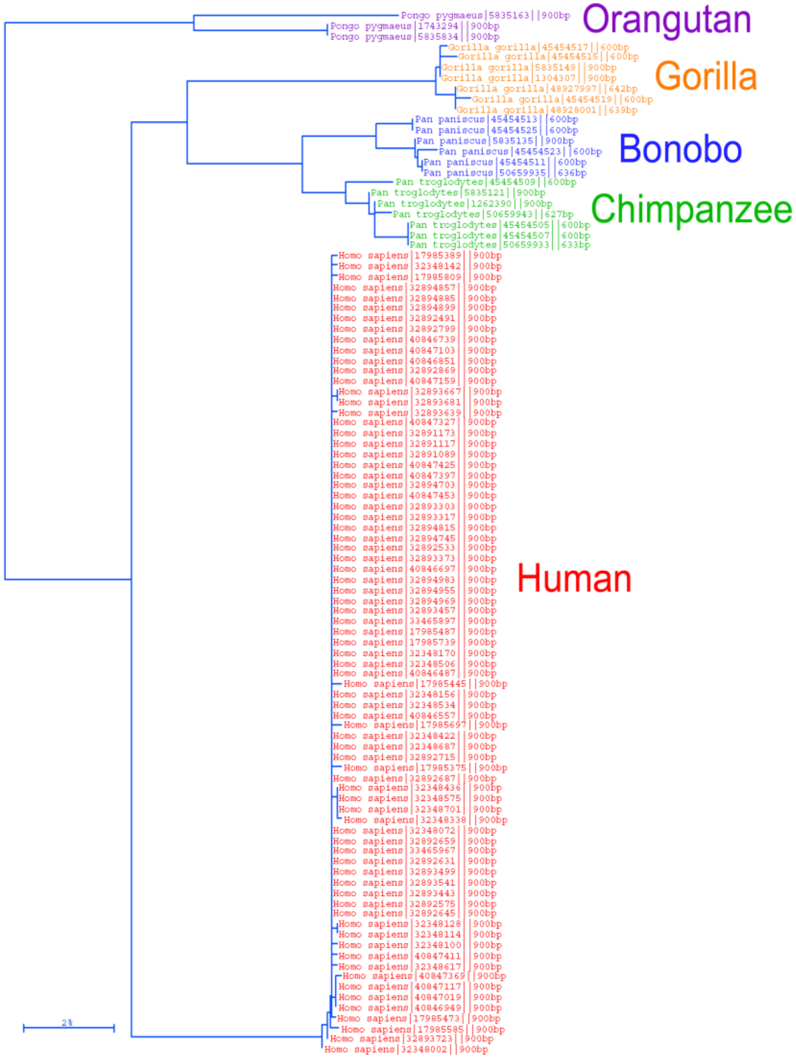


For animals, a 658 base-pair fragment of the mitochondrial gene, **cytochrome oxidase subunit I** (mtCOI) – consensus for iBOL consortium

- for particular taxonomic groups, also other barcodes are widely used - e.g. cytochrome *b* for mammals

Barcodes affirm the unity of the species *Homo sapiens*

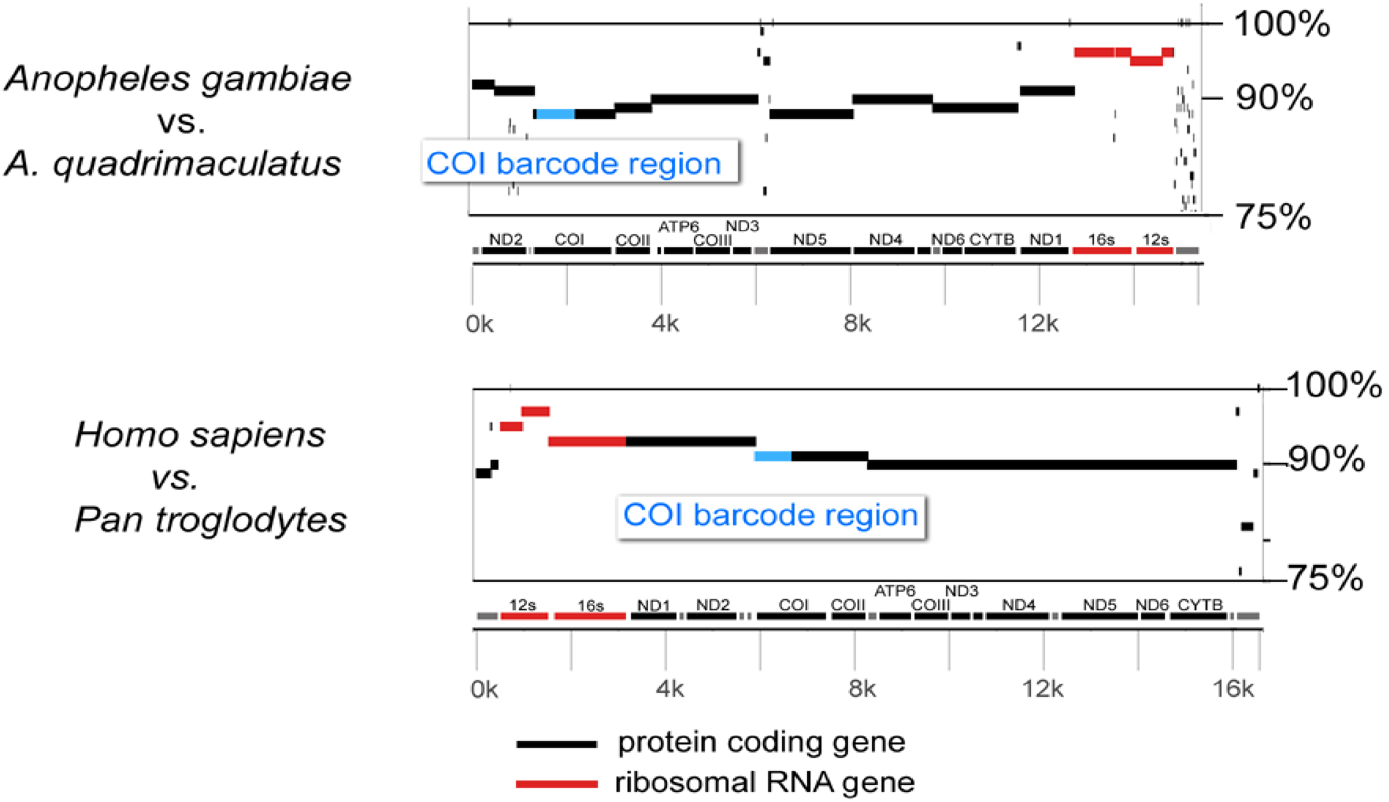
Comparisons show we differ from one another by only 1 or 2 nucleotides out of 648, while we differ from chimpanzees at 60 locations and gorillas at 70 locations.



Cytochrome *c* oxidase I (COI or CoxI) contains differences representative of those in other protein-coding genes

Possible gains in accuracy or cost using a different protein-coding gene would likely be small.

Percent identity plot (PIP) analysis of complete mitochondrial genomes



Focus to date

- For animals, a 658 base-pair fragment of the mitochondrial gene, **cytochrome oxidase subunit I** (mtCOI) – consensus for iBOL consortium

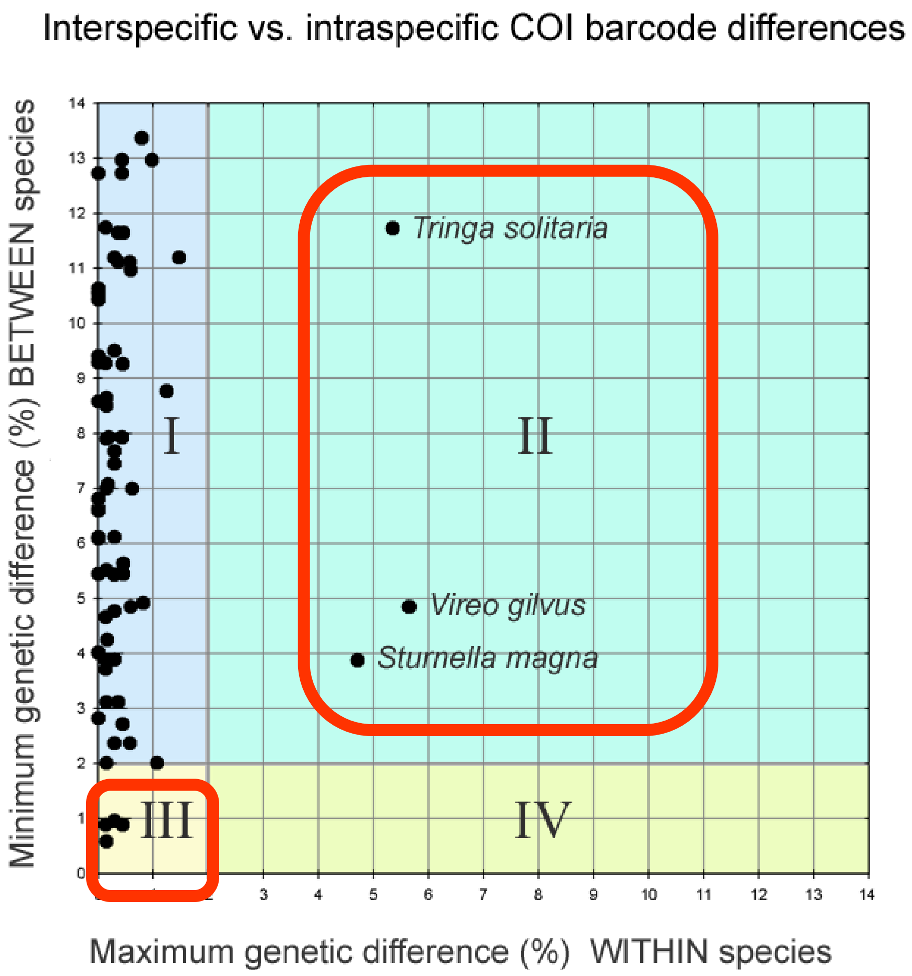
- For plants, mitochondrial genes do not differ sufficiently to distinguish among closely related species. Promising markers are genes on cpDNA: matK and rbcL

- For bacteria, a 16S-rDNA emerges as very useful marker (especially when using next-generation sequencing)

What do barcode differences among and within animal species studied so far suggest?

- barcodes identify most animal species unambiguously
- approximately 2-5% of recognized species have shared barcodes with closely-related species - many of them hybridize regularly
- in all groups studied so far, distinct barcode clusters with biological co-variation suggest cryptic species

Barcoding North American birds



Interspecific vs. intraspecific COI barcode differences

Results for 73 species of North American are shown. Quadrants represent different categories of species:

- I. consistent with current taxonomy
- II. probable lumped species (candidate for taxonomic split)
- III. recent divergence, hybridization, or synonomy.
- IV. probable taxonomic misidentification

A barcoder?

Mark Stoeckle The Rockefeller University
Paul Waggoner Connecticut Agricultural Experiment Station
Jesse Ausubel Alfred P. Sloan Foundation

Next generation sequencing of amplicons

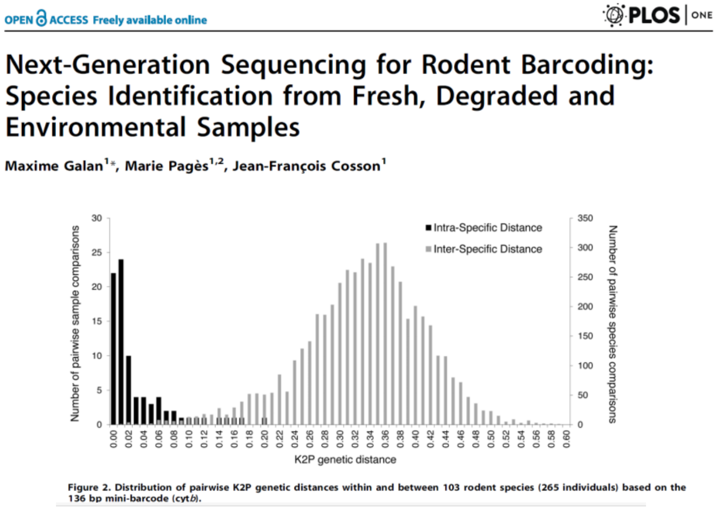
1. Metagenomics

- community of microorganisms
- PCR of 16S (18S) rRNA
- it is also possible to quantify (to some extent)

2. Diet composition

- COI barcoding (carnivores)
- chloroplast (cp)DNA (herbivores)

3. Analysis of contaminated samples



OPEN ACCESS Freely available online

PLOS | ONE

Next-Generation Sequencing for Rodent Barcoding: Species Identification from Fresh, Degraded and Environmental Samples

Maxime Galan[1]*, Marie Pagès[1,2], Jean-François Cosson[1]

Figure 2. Distribution of pairwise K2P genetic distances within and between 103 rodent species (265 individuals) based on the 136 bp mini-barcode (cyt*b*).

|SE|S|AM|E| BARCODE

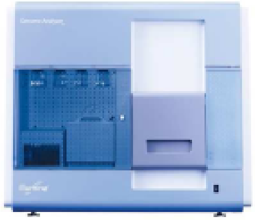
SEquences Seeker & AMplicons Explorer for Barcoding

Genotyping and barcoding based on
high-throughput multiplex amplicon sequencing

Illumina sequencers

Illumina MiSeq
4 millions reads/run
150 bp/read

Illumina GAIIx
300 millions reads/run
150 bp/read

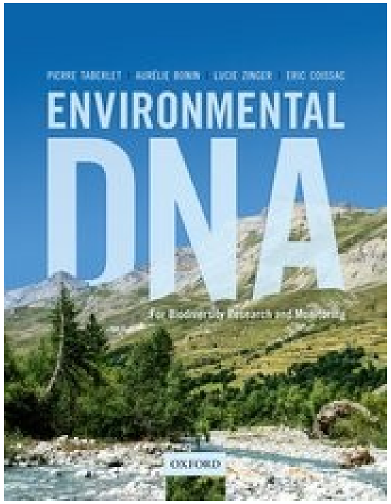


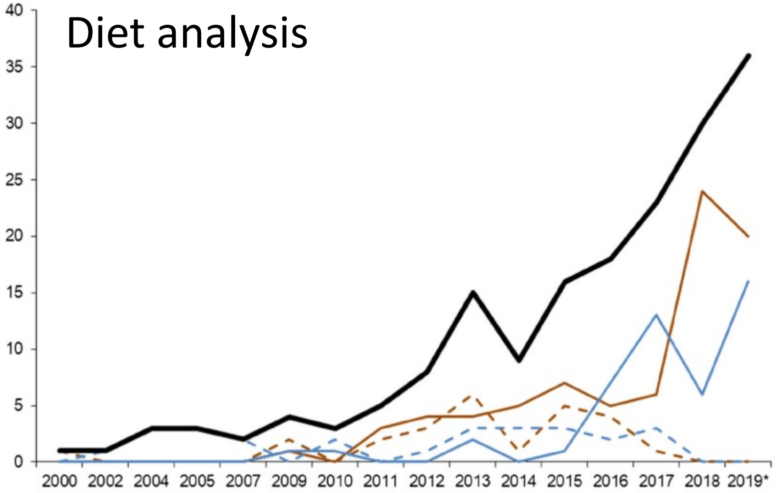
Illumina HighSeq
1500 – 3000 millions reads/run
100 bp/read

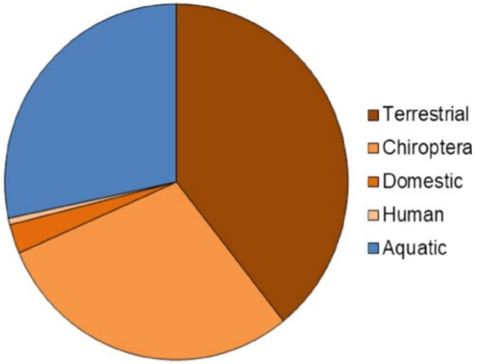
A barcoder? ... COMING SOON

Mark Stoeckle	The Rockefeller University
Paul Waggoner	Connecticut Agricultural Experiment Station
Jesse Ausubel	Alfred P. Sloan Foundation

Metabarcoding/eDNA



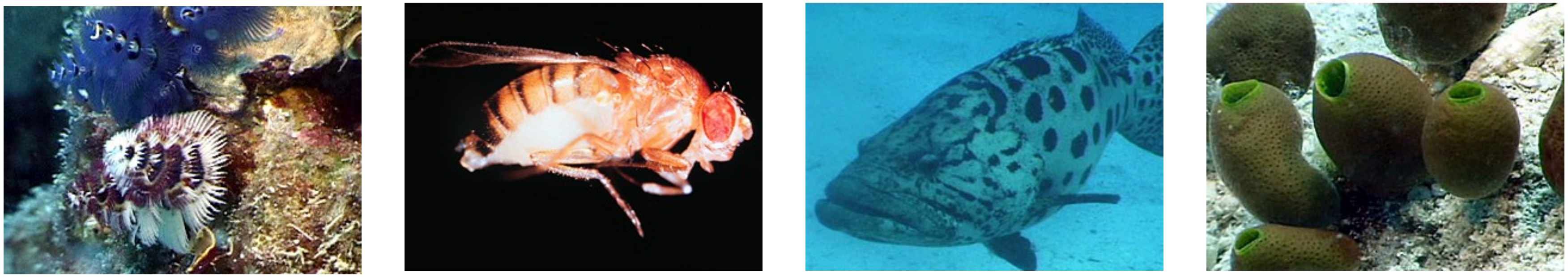

Diet analysis



Bold black line = Total number of studies per year; *thin straight lines*: high-throughput next-generation sequencing techniques; *thin dashed lines*: Sanger sequencing method or other traditional molecular technique, e.g., RFLP)

- identification of **MOTUs** („molecular operational taxonomic units“)

What *isn’t* DNA Barcoding?

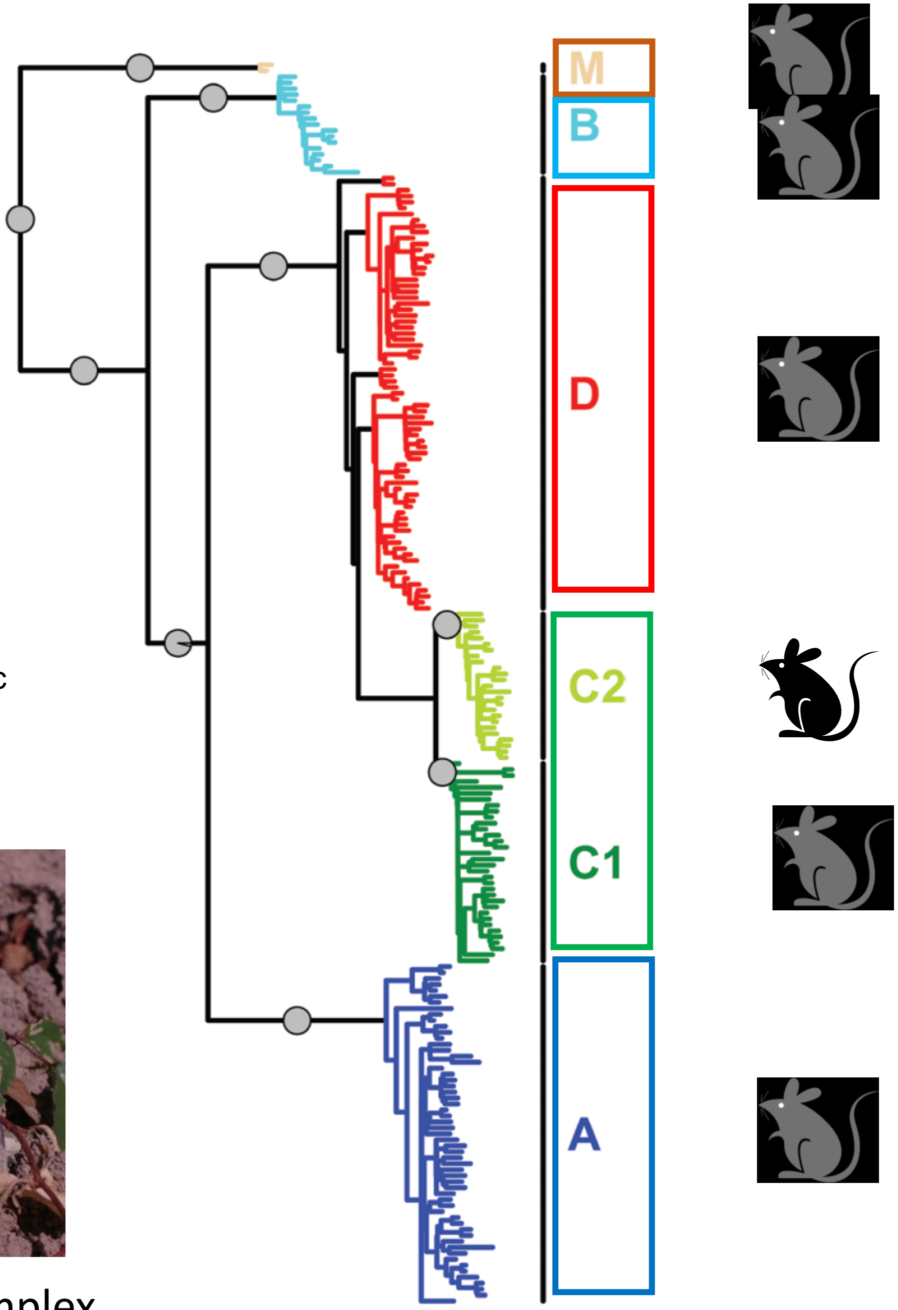


- it is not intended to, in any way, supplant or invalidate existing taxonomic practice
- it is not DNA taxonomy; it does not equate species identity, formally or informally, with a particular DNA sequence
- it is not intended to duplicate or compete with efforts to resolve deep phylogeny (e.g., Assembling the Tree of Life, ATOL)

DO WE REALLY BARCODE SPECIES?

What is a species?

mtDNA tree
(=DNA barcoding)



Five species based on mtDNA barcoding – min. 7% divergence (cyt b)

Two species based on phenotype?

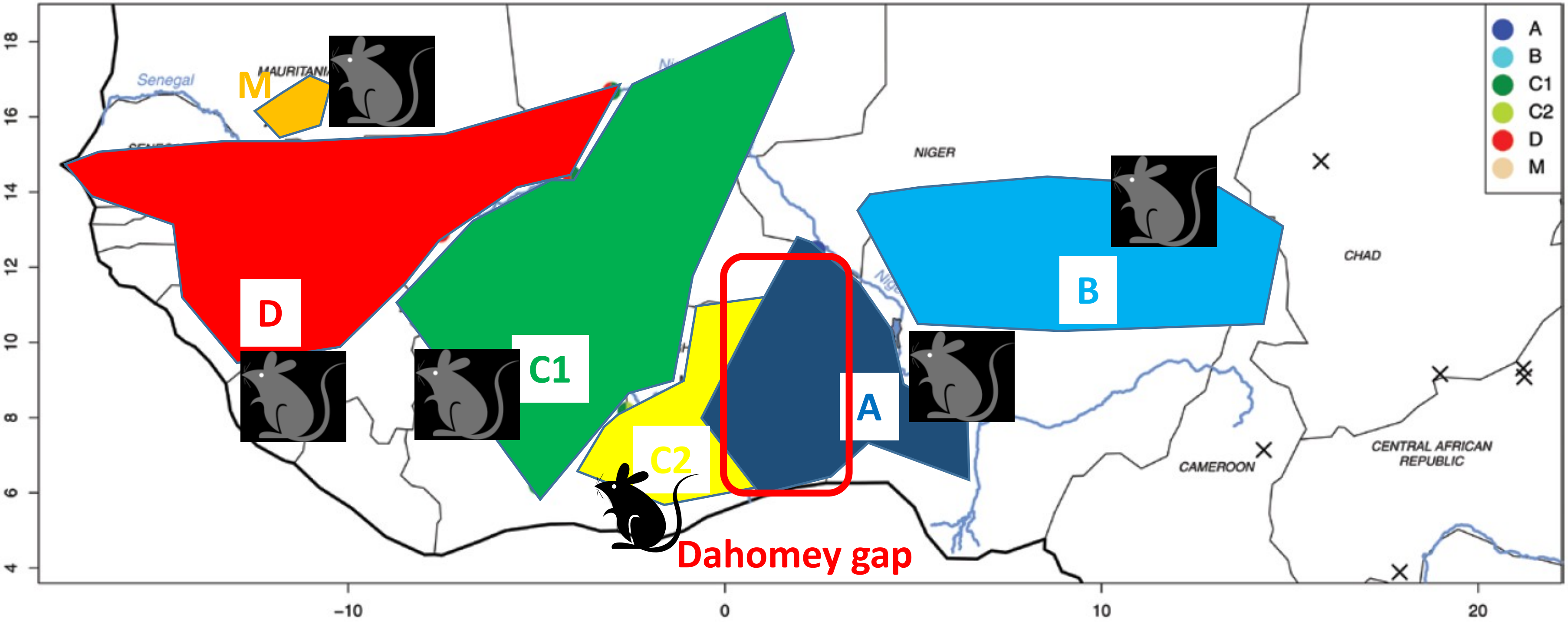
daltoni - paraphyletic

derooi - commensal

Praomys daltoni complex

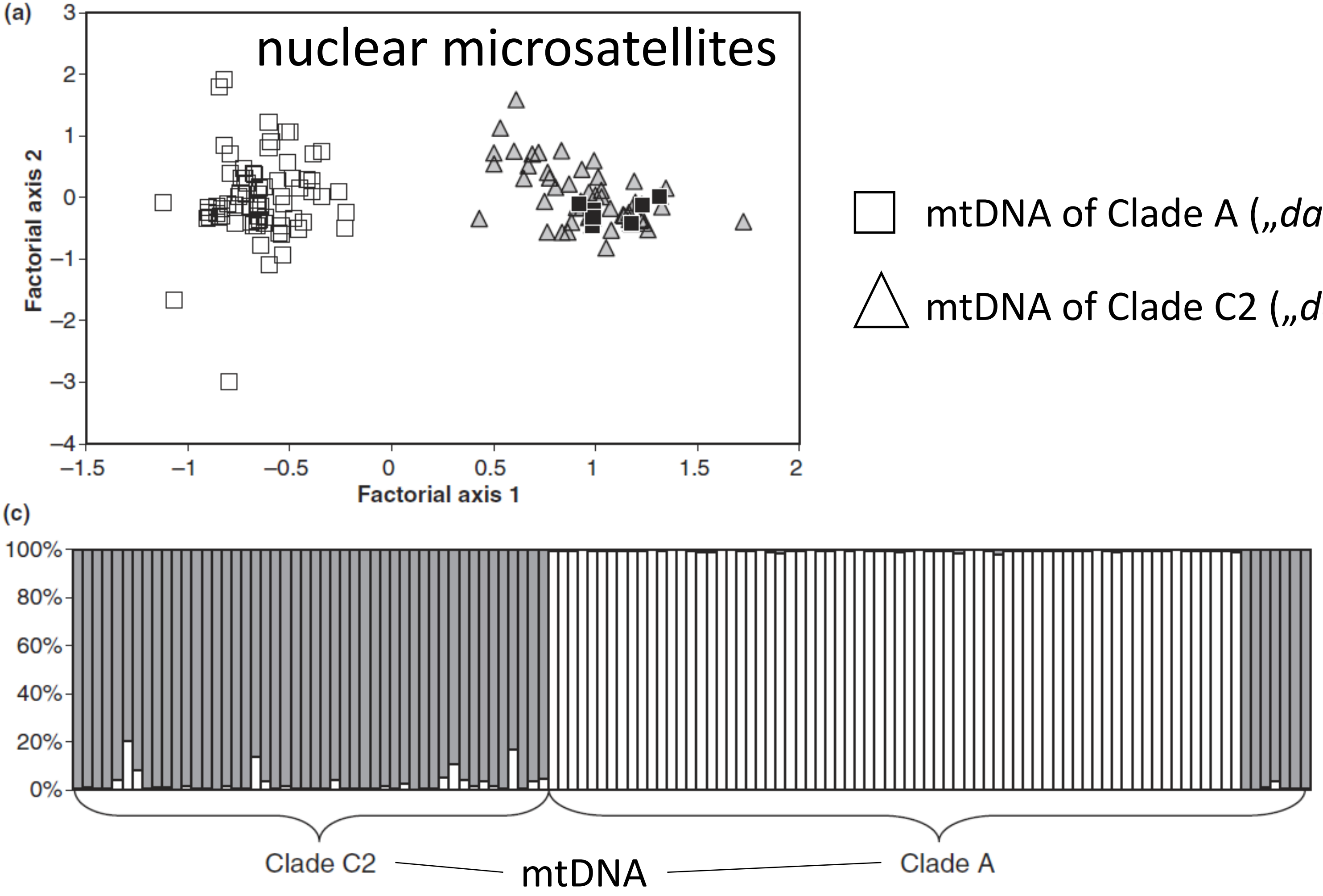
Bryja et al. 2010 Mol Ecol; Mikula, ..., Bryja 2020 Biol J Linn Soc

Phylogeographic structure at mtDNA



Bryja et al. 2010 Mol Ecol; Mikula, ..., Bryja 2020 Biol J Linn Soc

Partial mtDNA introgression in Dahomey gap



nuclear microsatellites

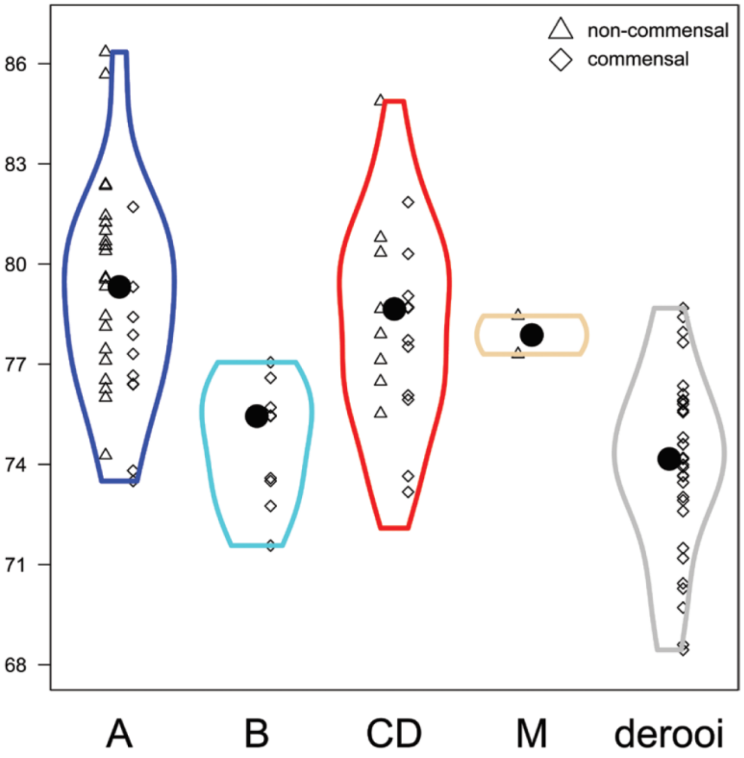
mtDNA

□ mtDNA of Clade A („*daltoni*“)

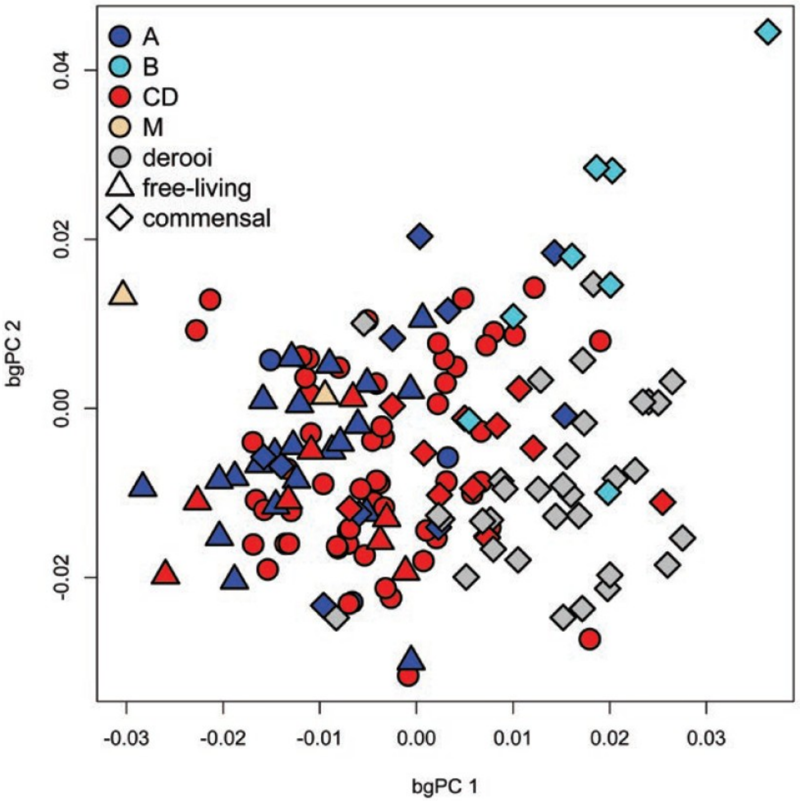
△ mtDNA of Clade C2 („*derooi*“)

Bryja et al. 2010 Mol Ecol

Morphological differentiation



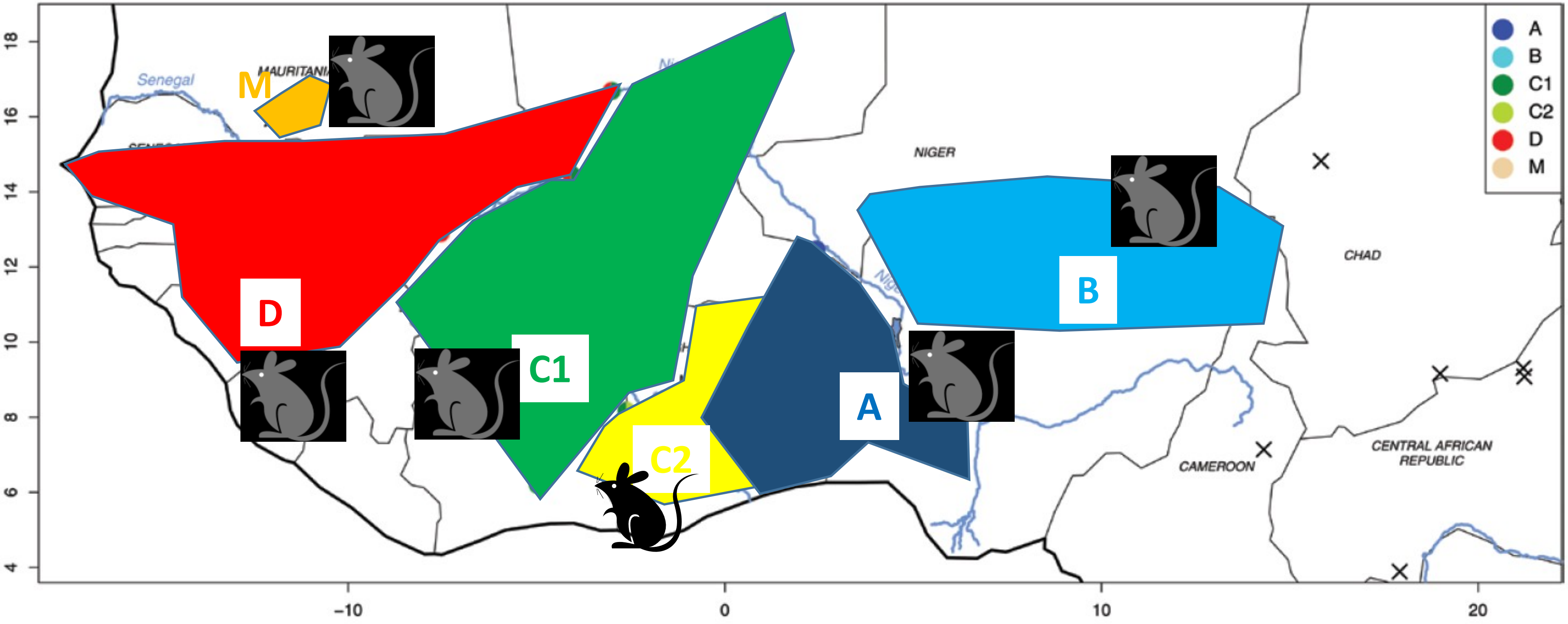
Skull size



Skull shape

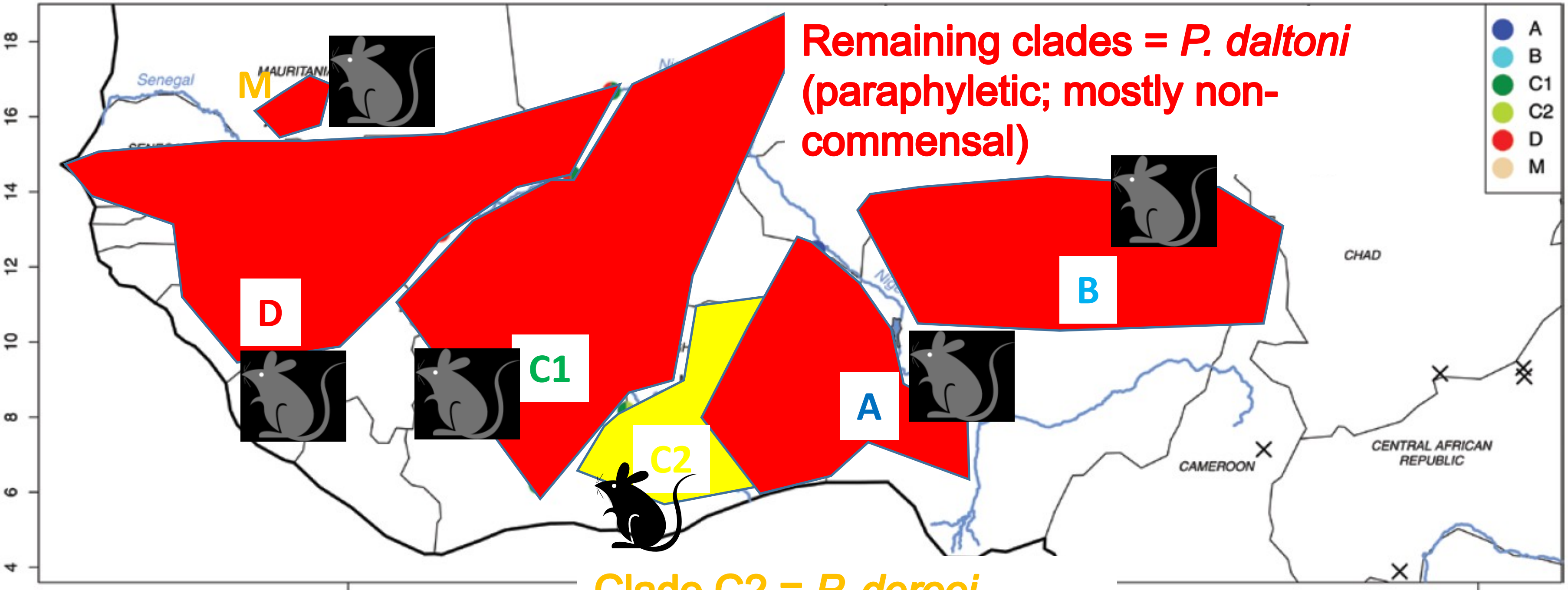
Mikula, ..., Bryja 2020 Biol J Linn Soc

What is a species?



Bryja et al. 2010 Mol Ecol; Mikula, ..., Bryja 2020 Biol J Linn Soc

Morphology and ecology



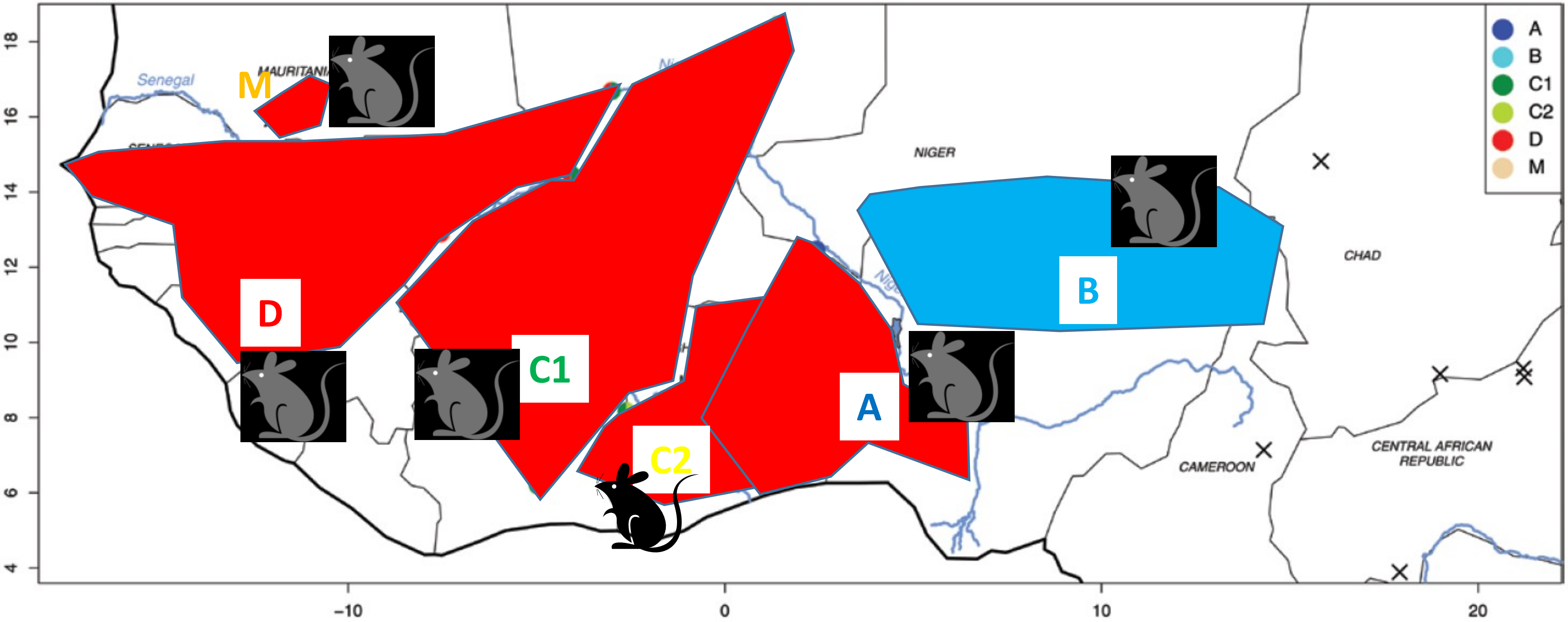
Remaining clades = *P. daltoni* (paraphyletic; mostly non-commensal)

Clade C2 = *P. derooi* (exclusively commensal)

2 species

Bryja et al. 2010 Mol Ecol; Mikula, ..., Bryja 2020 Biol J Linn Soc

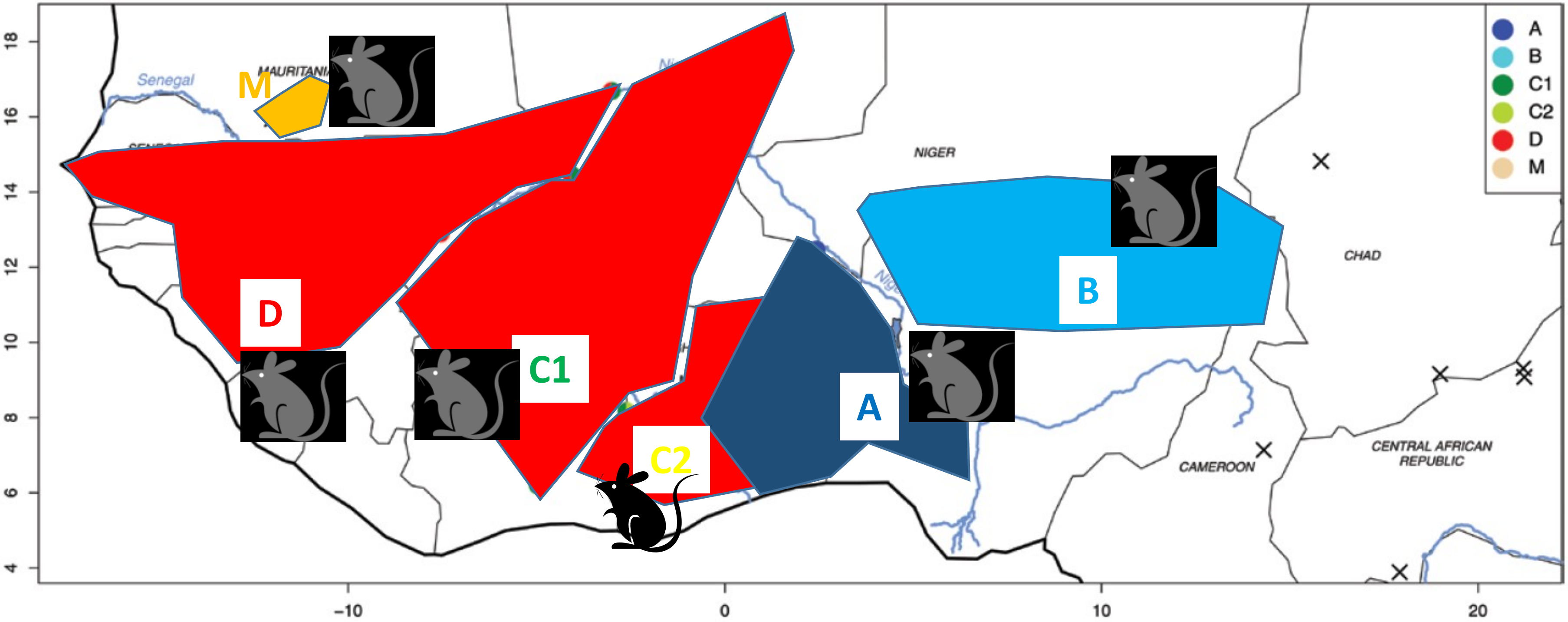
Karyotypes



2 species

Bryja et al. 2010 Mol Ecol; Mikula, ..., Bryja 2020 Biol J Linn Soc

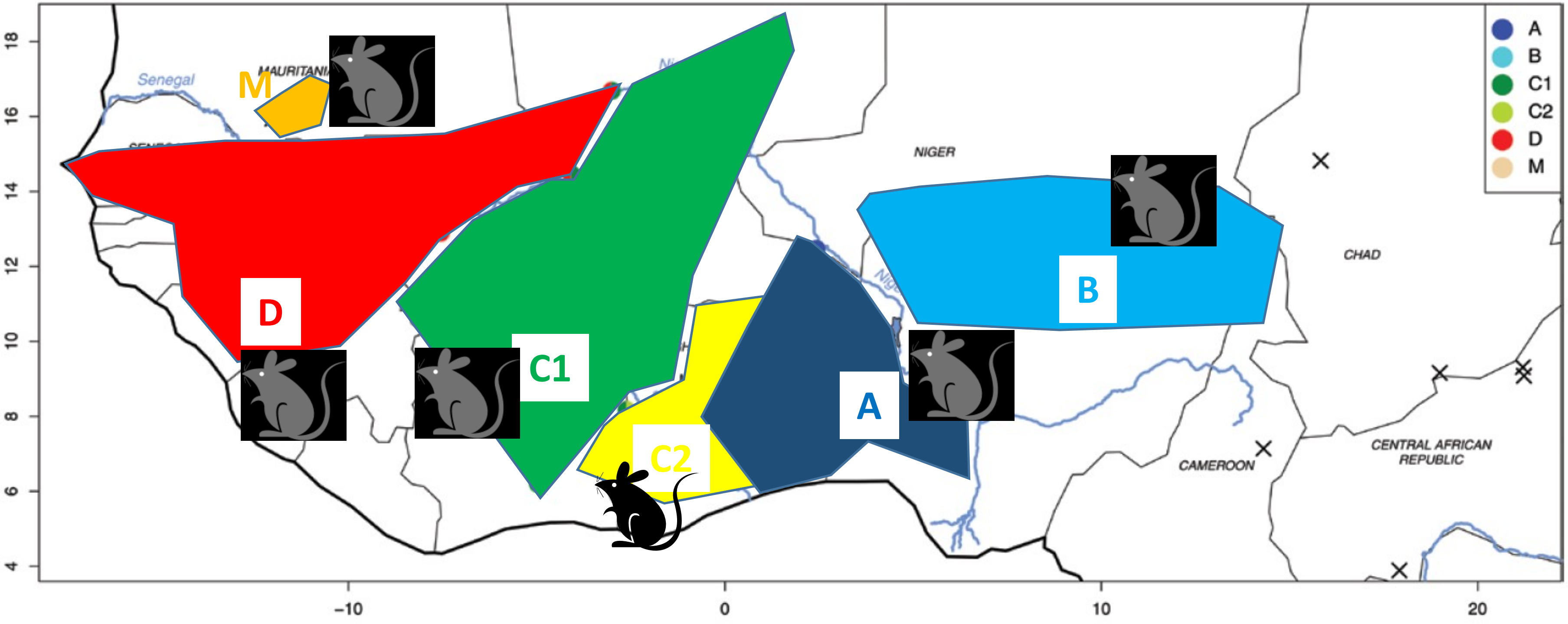
MtDNA phylogeny + microsatellites + karyotypes



4 species

Bryja et al. 2010 Mol Ecol; Mikula, ..., Bryja 2020 Biol J Linn Soc

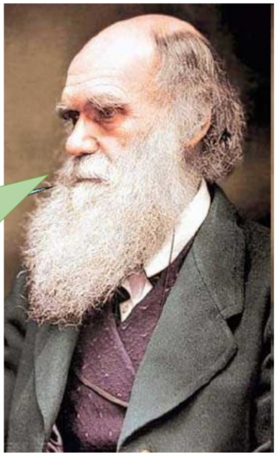
Phylogenetic species concept – splitting approach



6 species

Bryja et al. 2010 Mol Ecol; Mikula, ..., Bryja 2020 Biol J Linn Soc

What is a species



It is really laughable to see what different ideas are prominent in various naturalists minds, when they speak of „species“; ... It all comes, I believe, from trying to define the indefinable.

Darwin 1856

„A fundamental difficulty facing biologists interested in genetic delimitation of species is that in order to delimit species ***they must first be defined****. Species definitions intermingle with species concepts and the* ***lack of consensus in this field poses a serious dilemma*** *for the „delimiters“.* ***If systematists cannot agree on what defines a species, how can geneticists possibly develop objective methods to identify one?****“*

Bruce Rannala, Current Zoology 2015

What are the main limits to barcoding encountered so far?

What are the main limits to barcoding encountered so far?

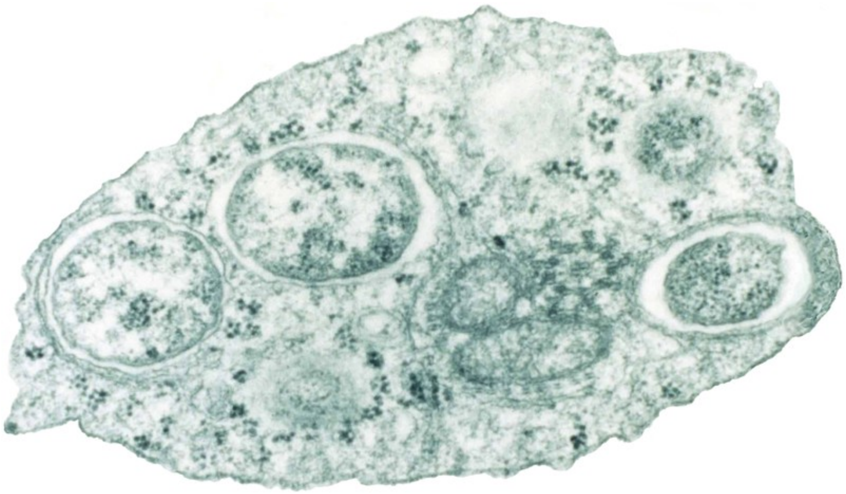
1) horizontal gene transfer
2) nuclear pseudogenes
3) heteroplasmy (paternal leakage)
4) gene tree vs. species tree
5) hybrids – mtDNA introgression

1. Horizontal gene transfer

DNA barcoding cannot reliably identify species of the blowfly genus *Protocalliphora* (Diptera: Calliphoridae)

T.L Whitworth, R.D Dawson, H Magalon and E Baudry

Proc. R. Soc. B 2007 **274**, doi: 10.1098/rspb.2007.0062, published 22 July 2007



Wolbachia within an insect cell
(25-70% species of insects)

Results of nuclear and mitochondrial DNA do not match

AFLP

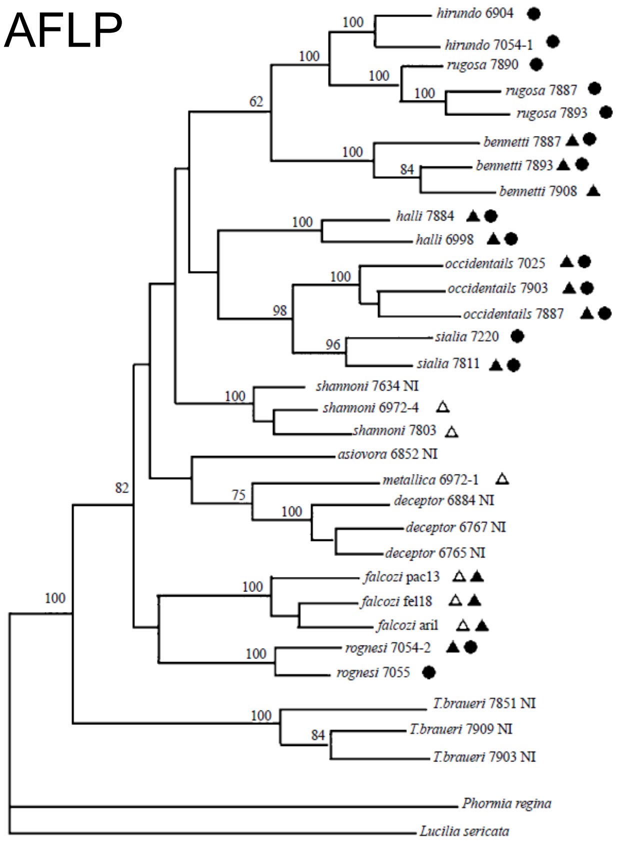


Figure 1. Phylogram of the *Protocalliphora* genus based upon AFLP data. The tree was generated by parsimony analysis using a heuristic search with tree bisection–reconnection. Bootstrap values are shown as percentage of 1000 replicates at each node only if they are 50% or greater. The *Wolbachia* infection status of each individual is shown on the tree. Individuals infected with *w*A1, *w*A2 or *w*B *Wolbachia* strains are respectively represented by an open triangle, a solid triangle and a circle. Non-infected individuals are symbolized by NI.

Symbols correspond to the type of *Wolbachia* infection

mtDNA

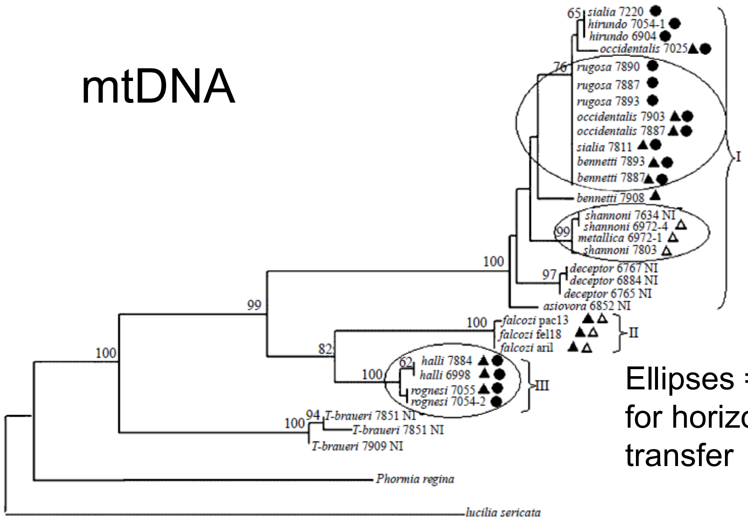
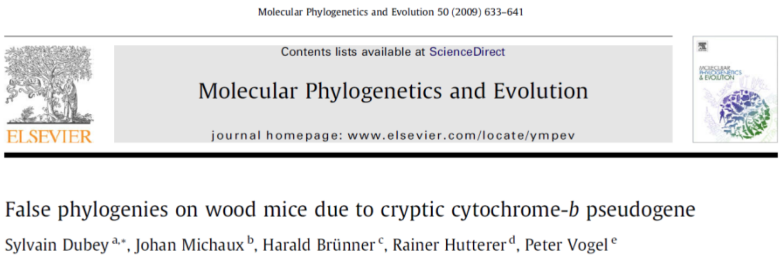


Figure 2. Phylogram of the *Protocalliphora* genus based upon COI and COII data (total of 953 bp). The tree was generated by maximum likelihood analysis using a heuristic search with tree bisection–reconnection. Bootstrap values are shown as percentage of 1000 replicates at each node only if they are 50% or greater. The *Wolbachia* infection status of each individual is shown on the tree. Individuals infected with *w*A1, *w*A2 or *w*B *Wolbachia* strains are respectively represented by an open triangle, a solid triangle and a circle. Non-infected individuals are symbolized by NI. Three clusters defined using 3 or 1.8% divergence as threshold values (§3) are shown on the figure. The three ellipses indicate cases where horizontal transfer of *Wolbachia* between species seems probable (§3).

Ellipses = candidates for horizontal gene transfer

Horizontal transfer of mtDNA through *Wolbachia*
(among closely related species, at the level of genera the barcoding is OK)

2. Pseudogenes



Molecular Phylogenetics and Evolution 50 (2009) 633–641

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

False phylogenies on wood mice due to cryptic cytochrome-*b* pseudogene

Sylvain Dubey[a,*], Johan Michaux[b], Harald Brünner[c], Rainer Hutterer[d], Peter Vogel[e]

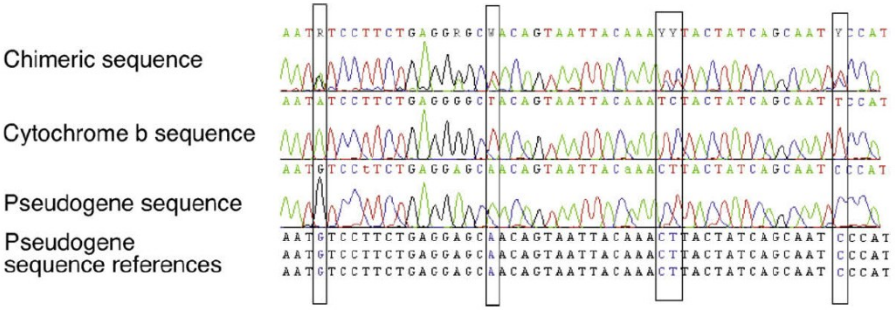
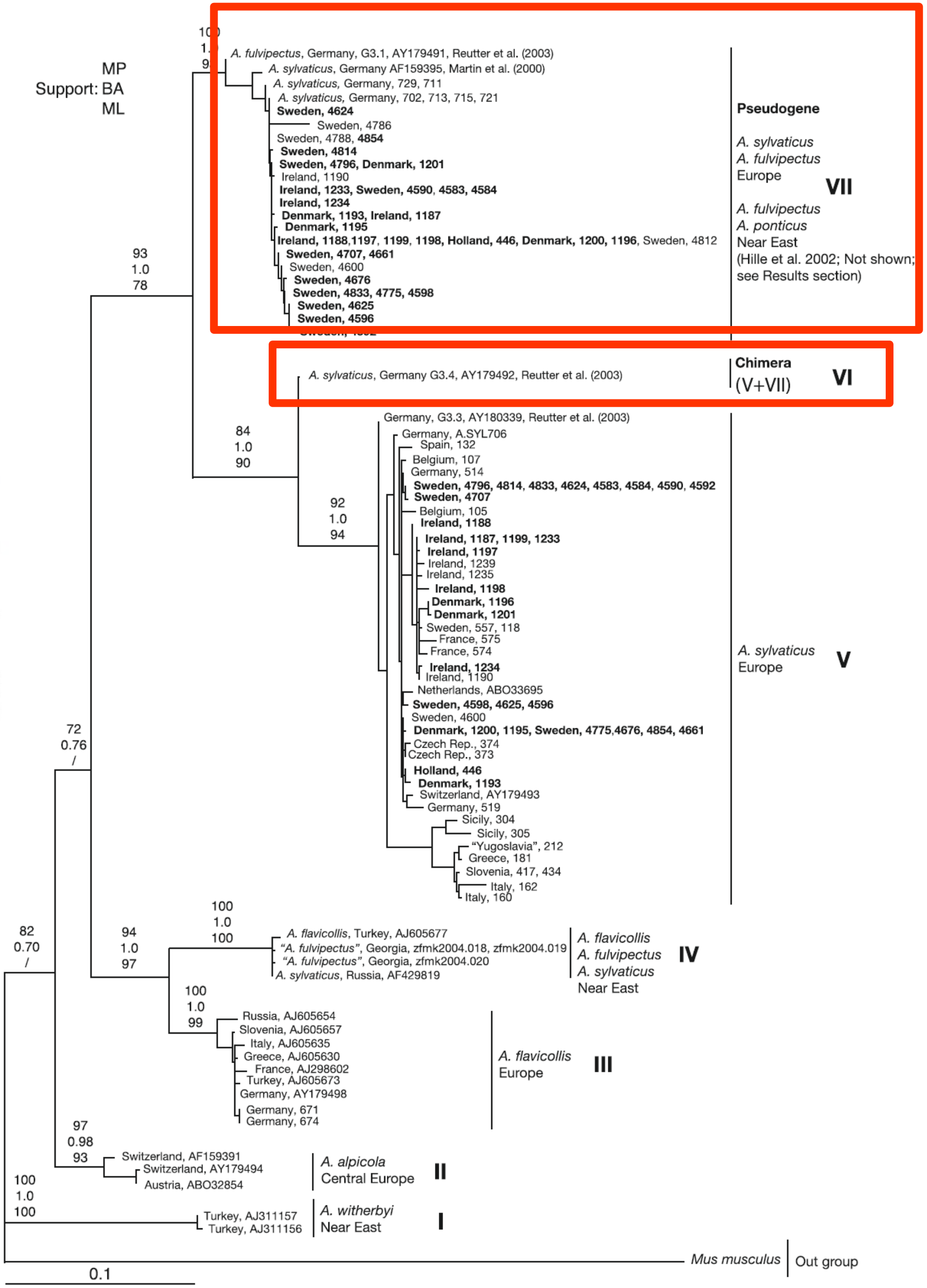


Fig. 2. Examples of electropherograms of a chimeric *cyt-b* sequence of lineage V/VII (VI) and of pure lineages V and VII.

Heterozygotes in mtDNA → be careful!

NUMTS = „**nu**clear copy of **mt**DNA **s**equences

Relatively often for cytochrome *b*



How to recognize numt?

- ultracentrifugation (fresh samples required)

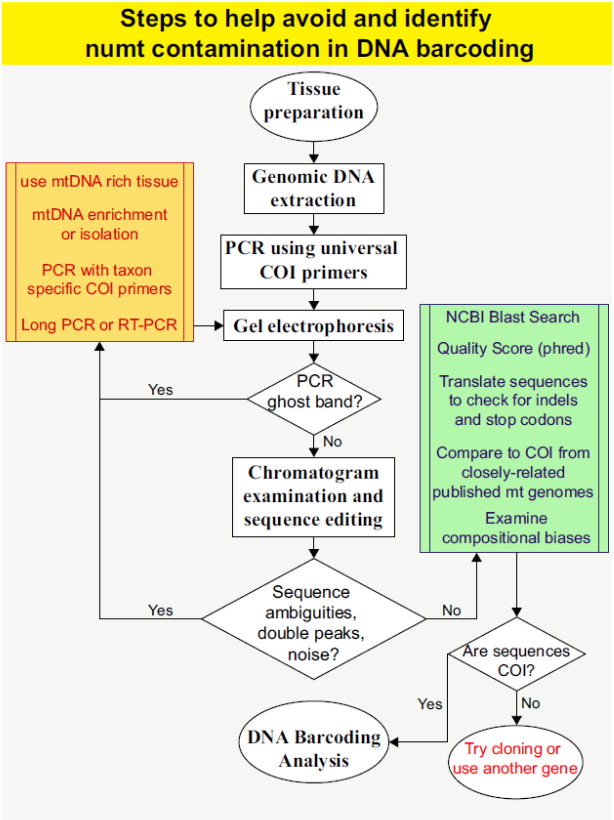
- the use of tissues with high proportion of mitochondria (e.g. muscles)

- *long-range PCR* (or sequence complete mtDNA)

- RT-PCR (pseudogenes are not transcribed)

- indels, stop codons

- cloning



Steps to help avoid and identify numt contamination in DNA barcoding

Tissue preparation

Genomic DNA extraction

PCR using universal COI primers

Gel electrophoresis

PCR ghost band?

Yes

No

Chromatogram examination and sequence editing

Sequence ambiguities, double peaks, noise?

Yes

No

Are sequences COI?

Yes

No

DNA Barcoding Analysis

Try cloning or use another gene

use mtDNA rich tissue

mtDNA enrichment or isolation

PCR with taxon specific COI primers

Long PCR or RT-PCR

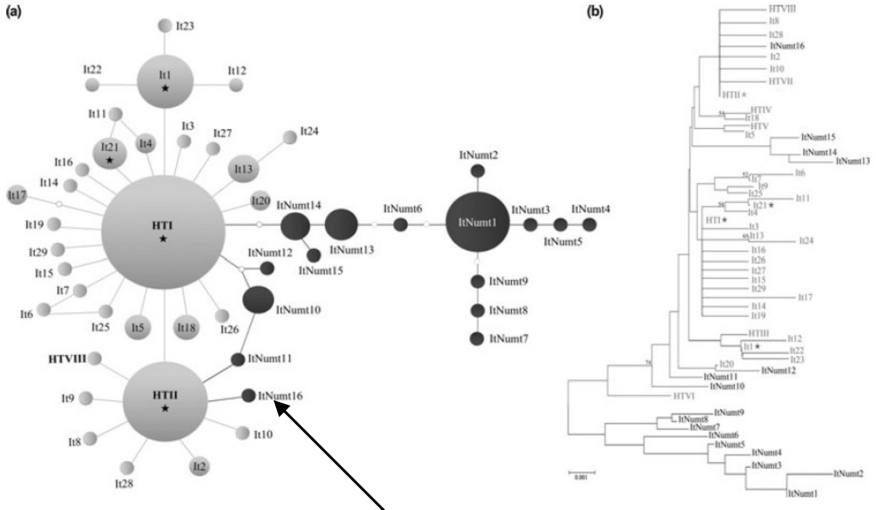
NCBI Blast Search

Quality Score (phred)

Translate sequences to check for indels and stop codons

Compare to COI from closely-related published mt genomes

Examine compositional biases



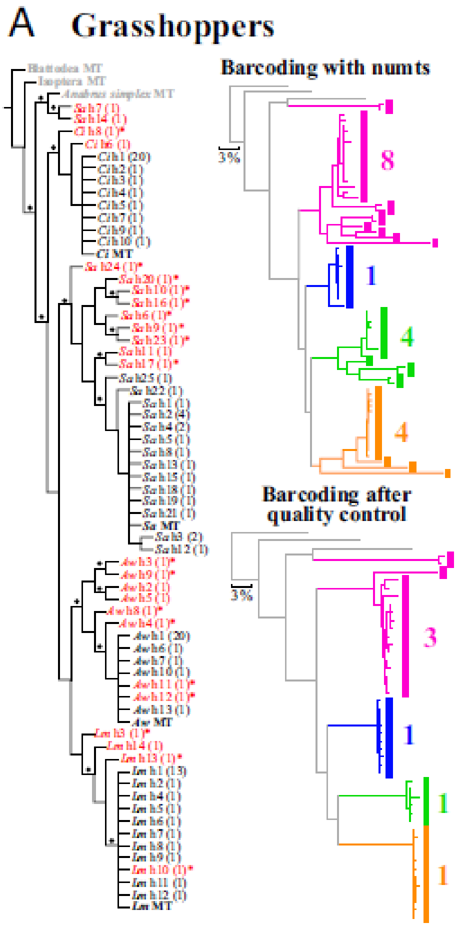
cryptic numts

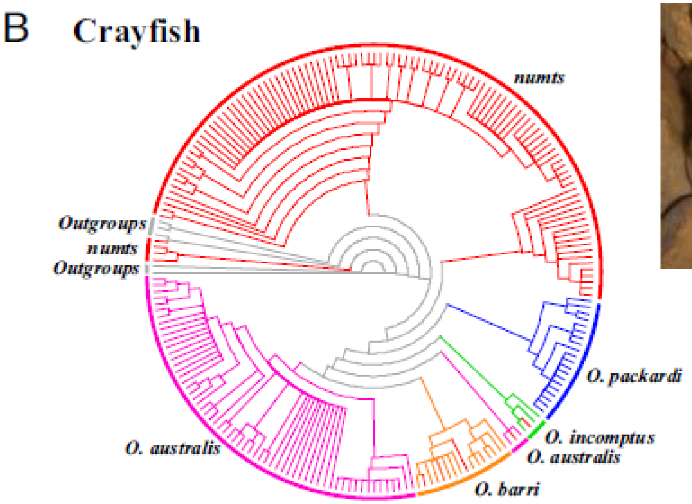
Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified

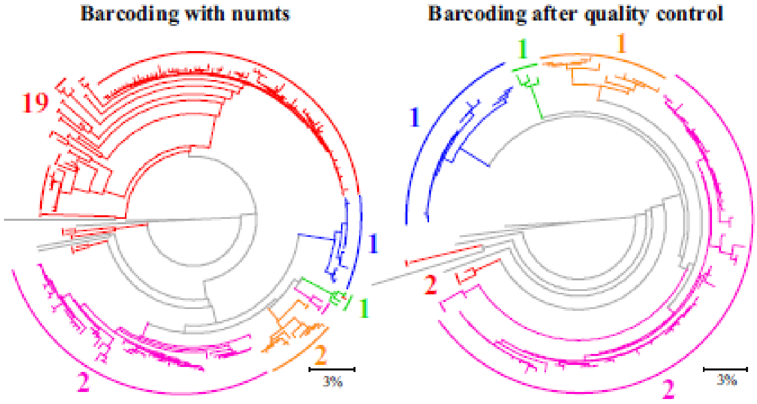
Hojun Song*†, Jennifer E. Buhay*‡, Michael F. Whiting*, and Keith A. Crandall*

*Department of Biology, Brigham Young University, Provo, UT 84602; and ‡Belle W. Baruch Institute for Marine Sciences, University of South Carolina, Columbia, SC 29208

Edited by W. Ford Doolittle, Dalhousie University, Halifax, NS, Canada, and approved July 14, 2008 (received for review March 28, 2008)

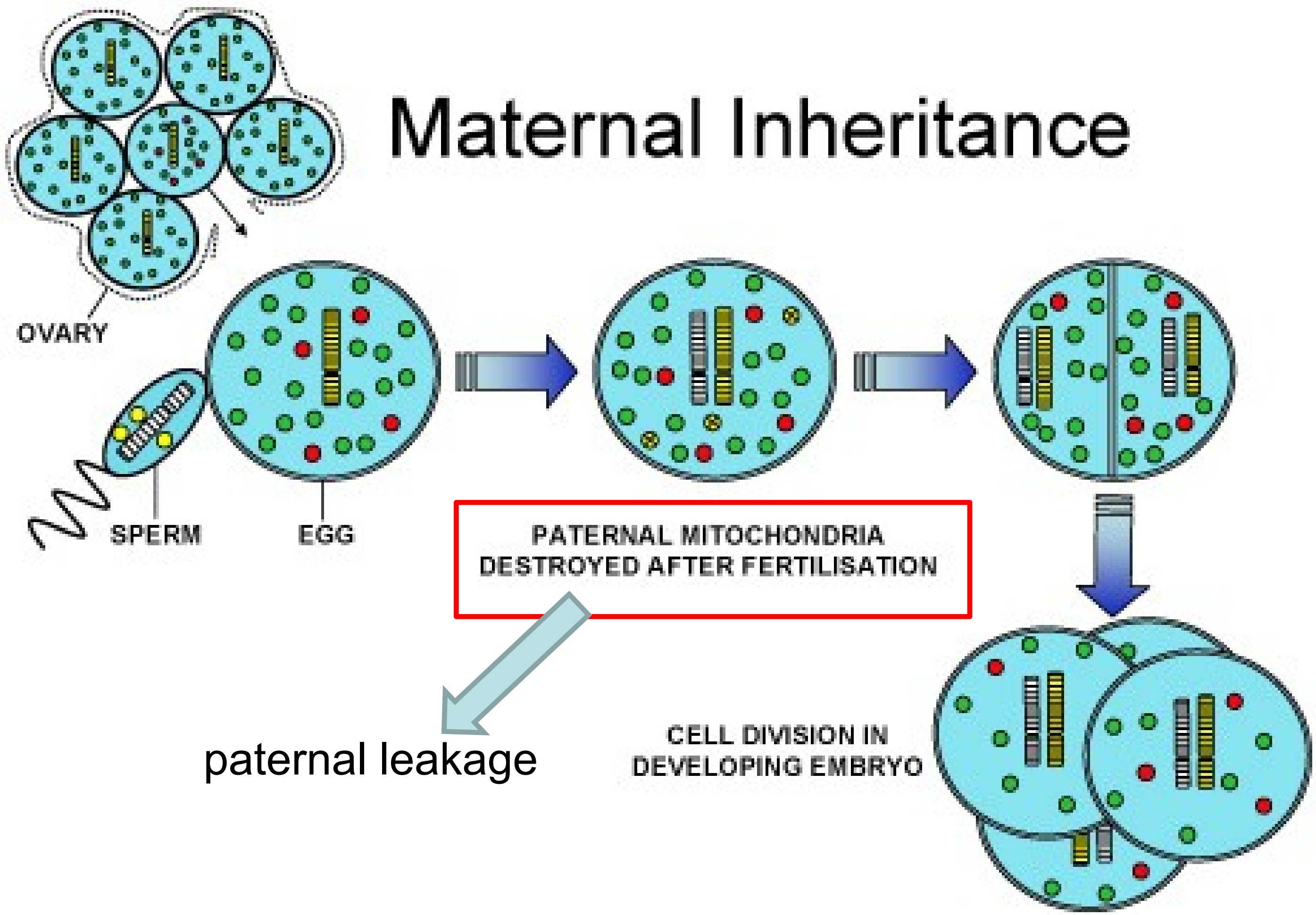






number of suggested barcoded taxa based on 3% divergence on COI with/without numts (identified by stop-codons and indels)

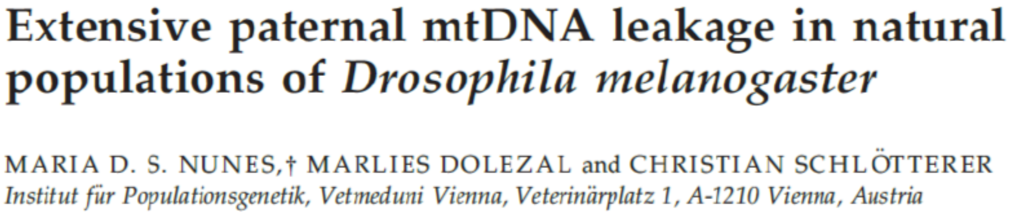
3. Heteroplasmy



paternal leakage

- well studied mitochondrial disorders in human
- low N_e of mtDNA → usually fast fixation of new mutations – mitochondrial bottleneck

Paternal leakage



Extensive paternal mtDNA leakage in natural populations of *Drosophila melanogaster*

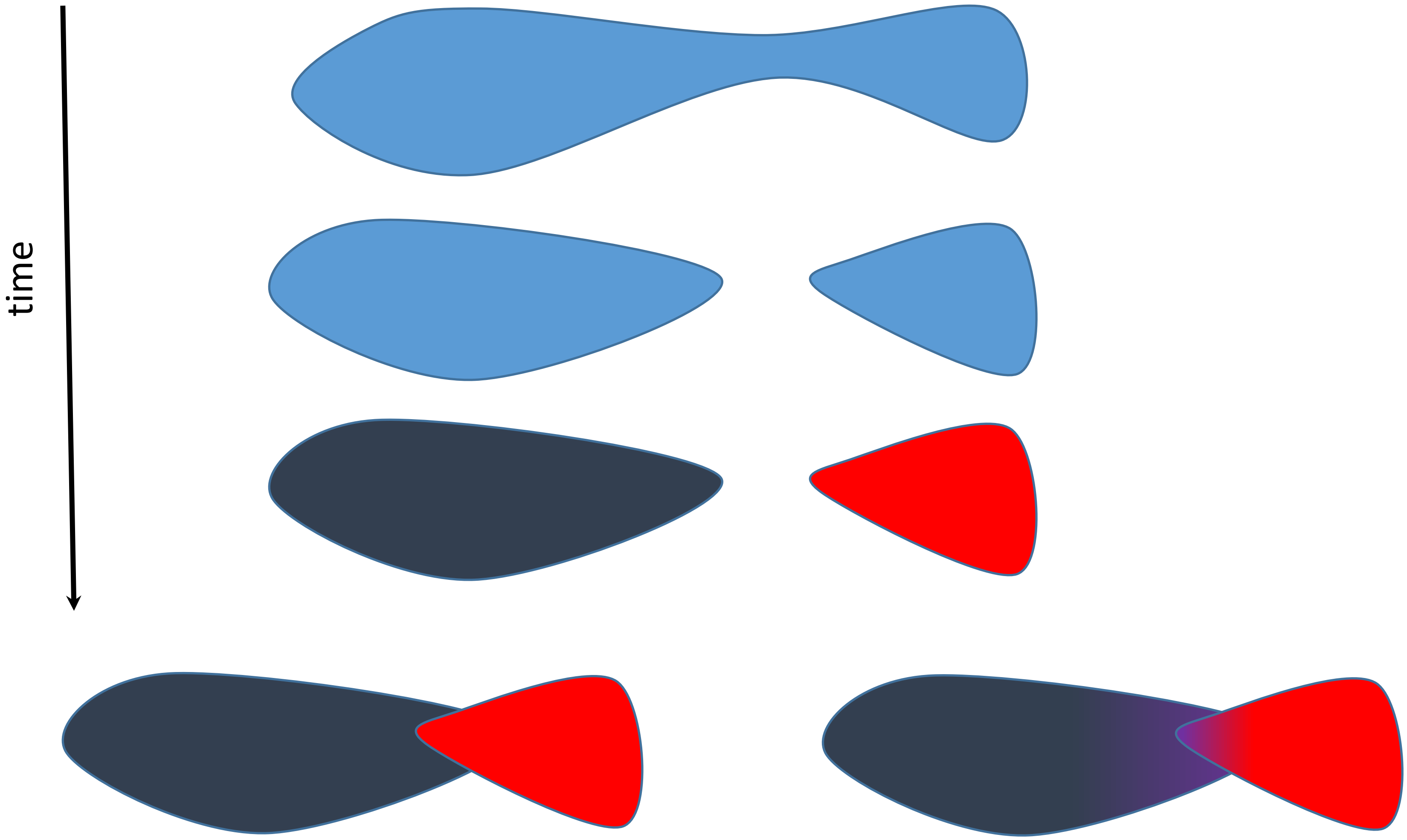
MARIA D. S. NUNES,† MARLIES DOLEZAL and CHRISTIAN SCHLÖTTERER
Institut für Populationsgenetik, Vetmeduni Vienna, Veterinärplatz 1, A-1210 Vienna, Austria

- allele-specific real-time quantitative PCR (RT-qPCR) → heteroplasmie je asi častý jev
- 14 % jedinců, ale velmi nízká frekvence druhého haplotypu
- paternal leakage 6 %

4. GENE TREES VS. SPECIES TREES

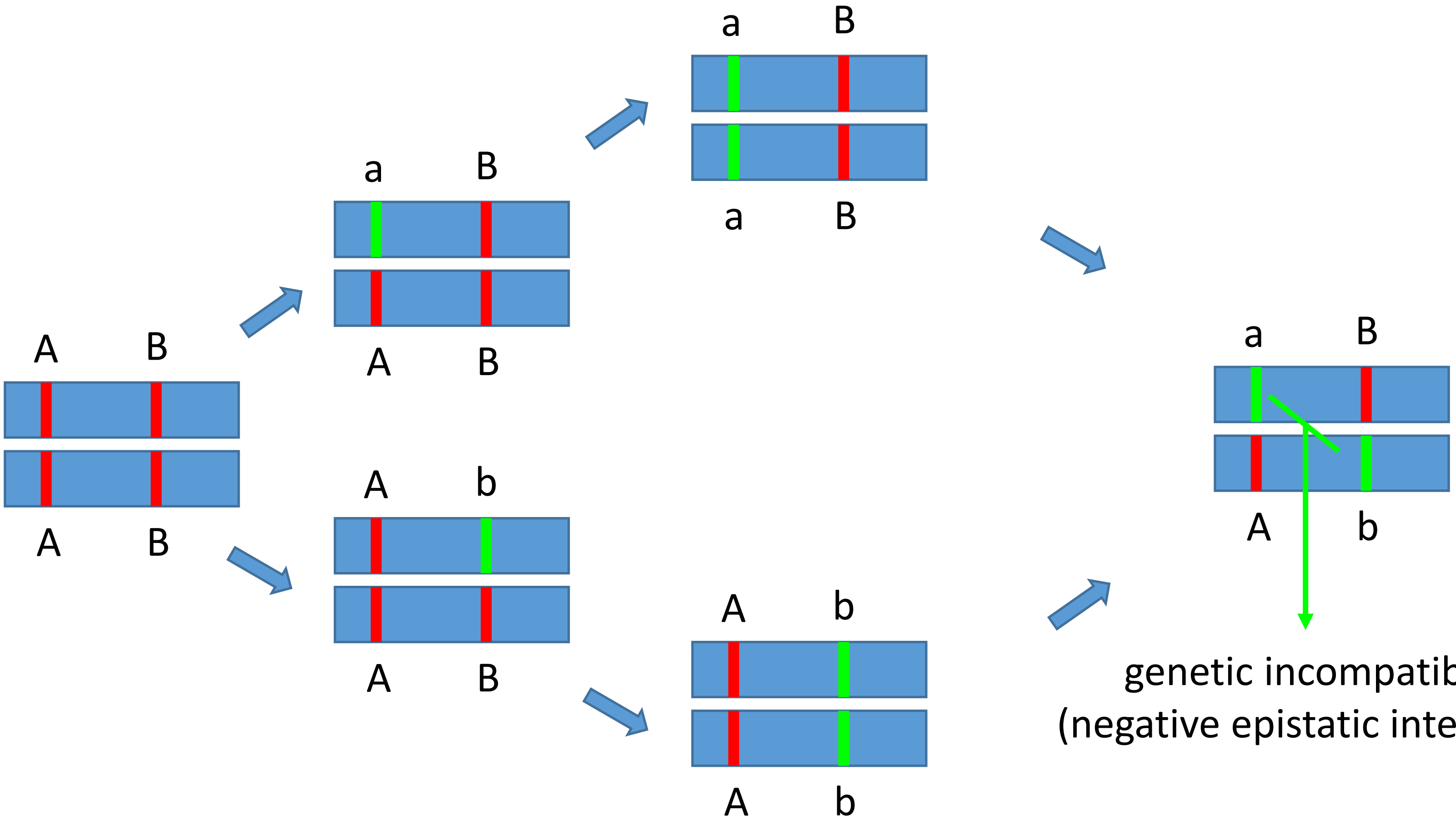
STATISTICAL MULTI-LOCUS SPECIES DELIMITATION

Allopatric speciation model



Mayr 1942

Dobzhansky-Muller incompatibility



genetic incompatibility
(negative epistatic interaction)

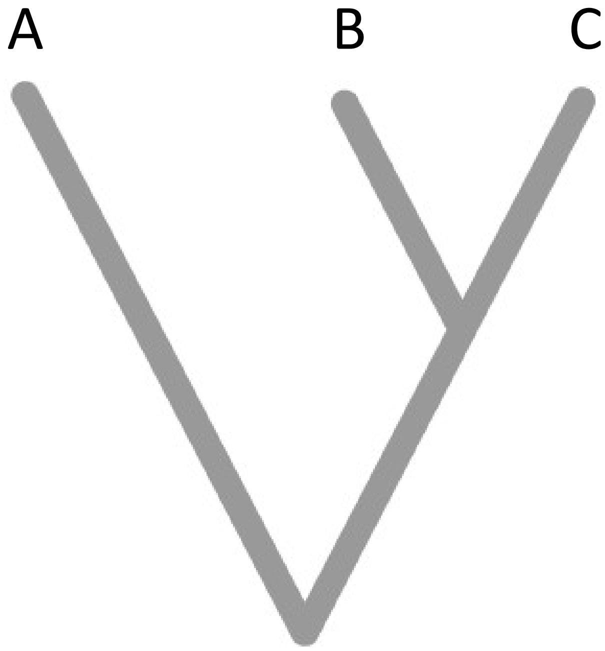
Dobzhansky 1936, Muller 1942

Phylogeny at the level of populations and species

Species are metapopulation lineages > new methods for DNA-based species delimitation

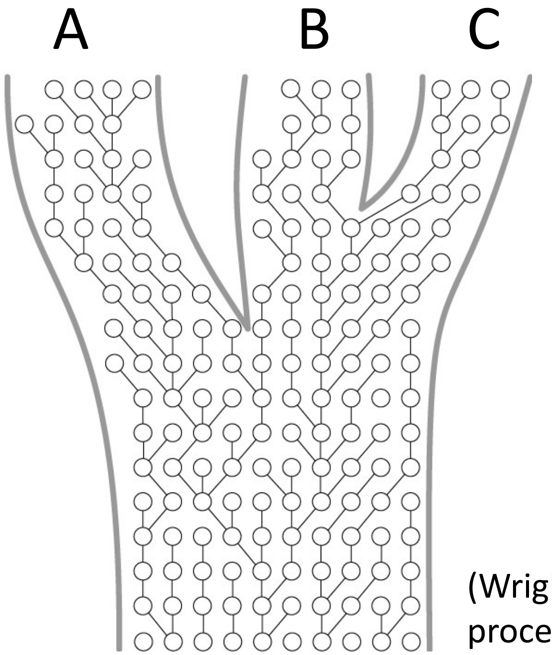
Gene trees

- Vital to understanding the process of speciation
- Span **intraspecific** and **interspecific** evolution



Species phylogeny

\>>

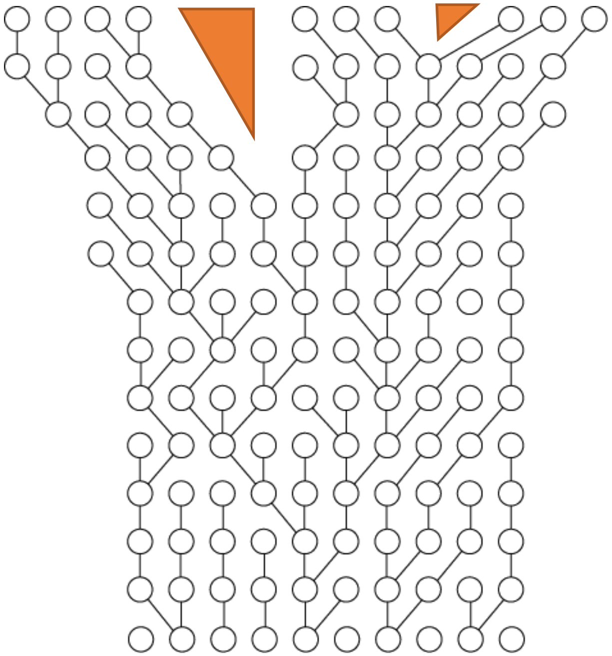


(Wright-Fisher process)

Population genetics:
coalescence process

Phylogeny at the level of populations and species

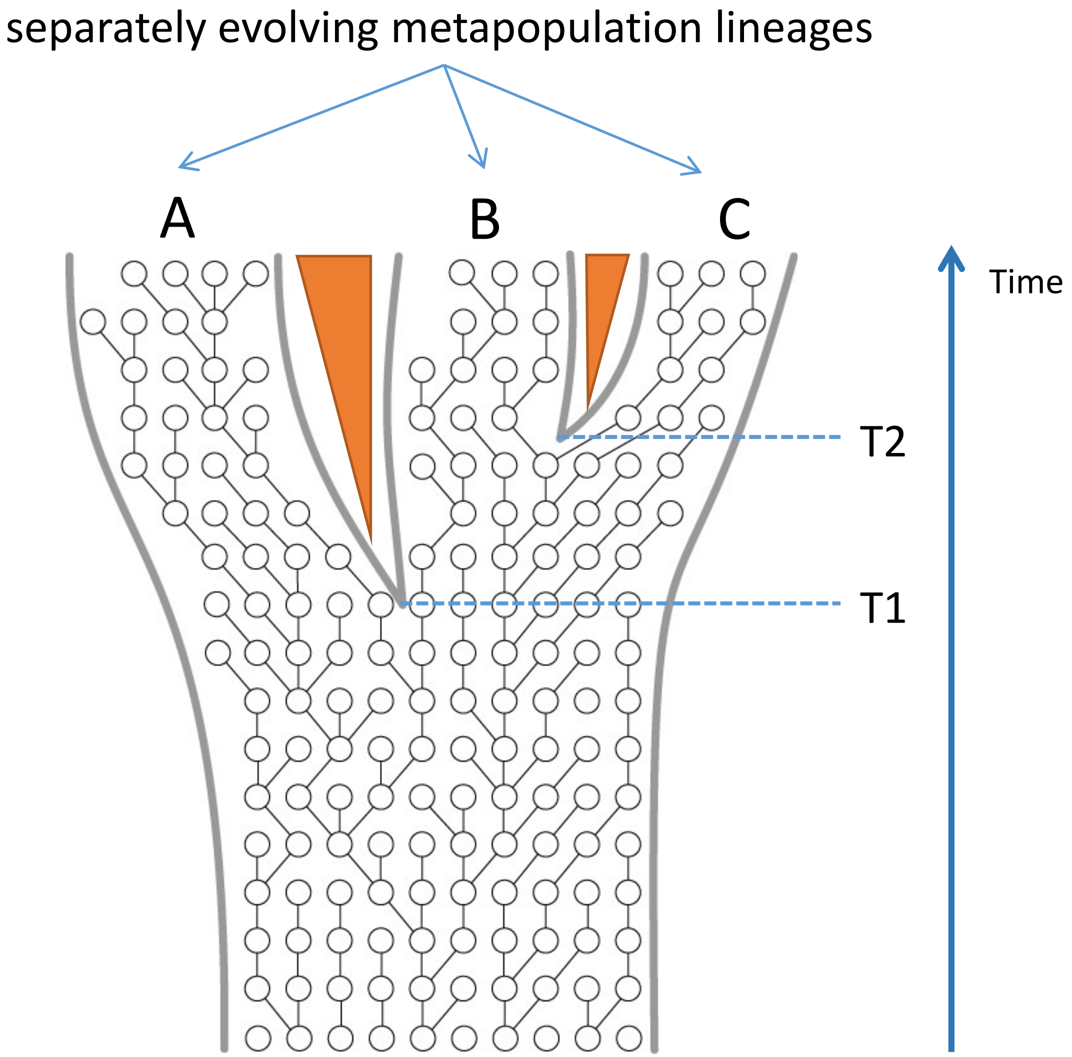
Barriers to gene flow



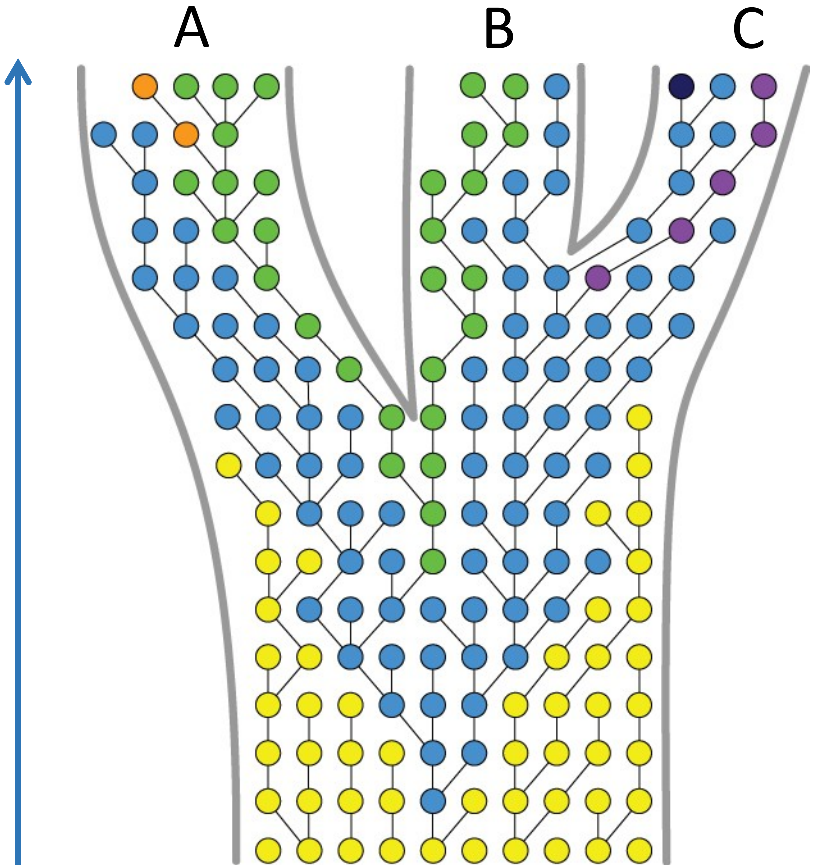
coalescence process

○: individual organisms / allele copy

Phylogeny at the level of populations and species



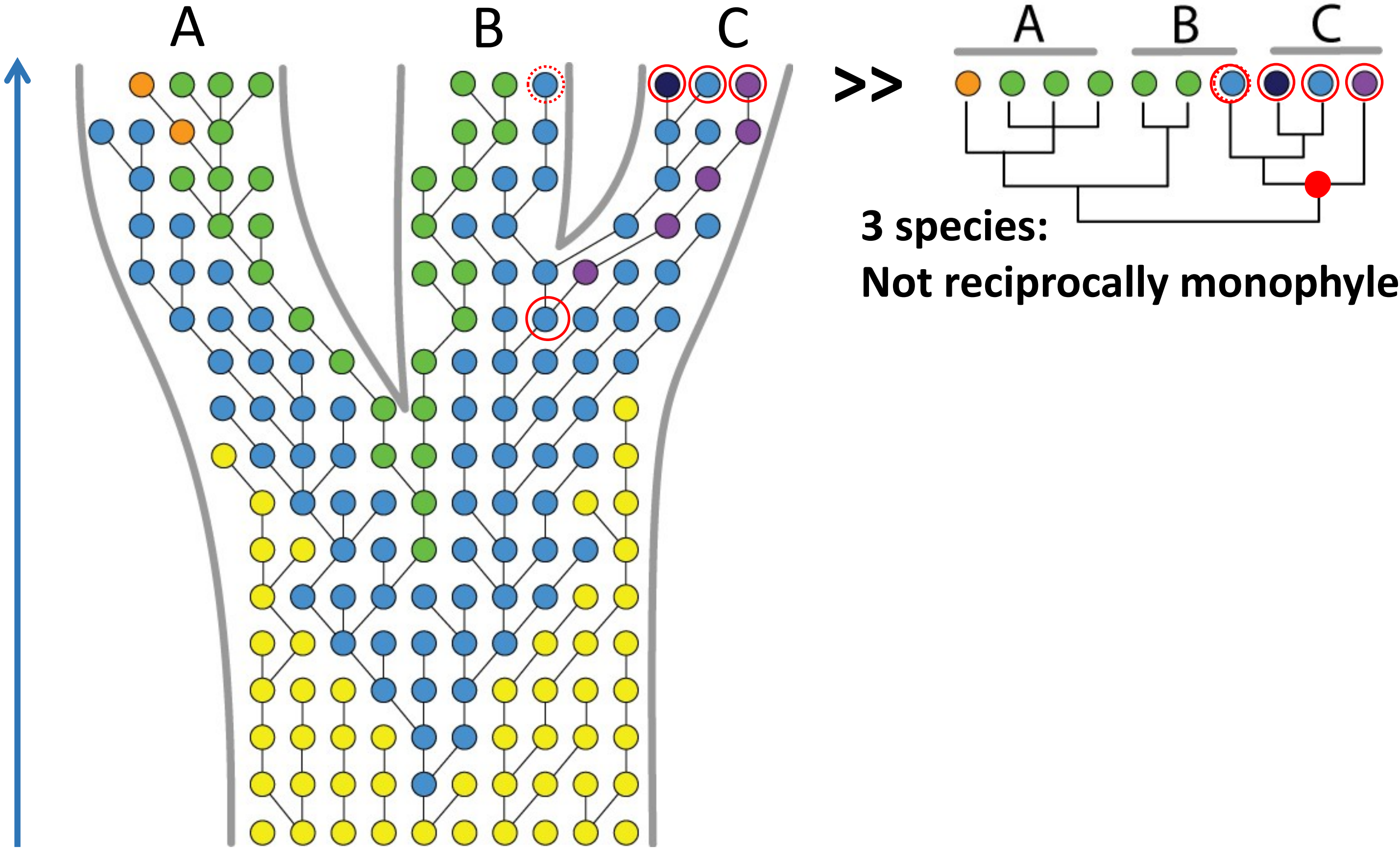
Phylogeny at the level of populations and species



Trans-species polymorphism

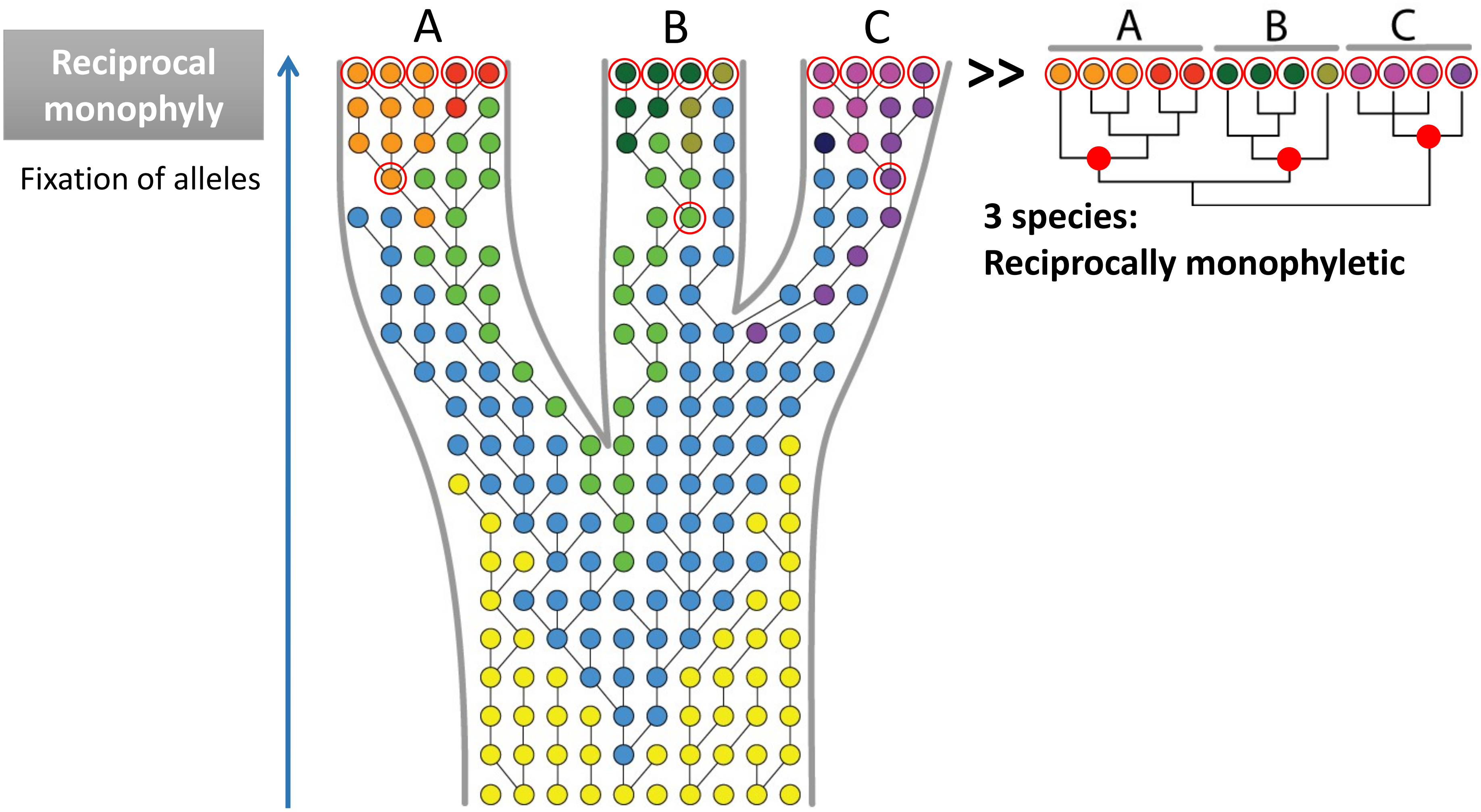
Phylogeny at the level of populations and species

Trans-species polymorphism



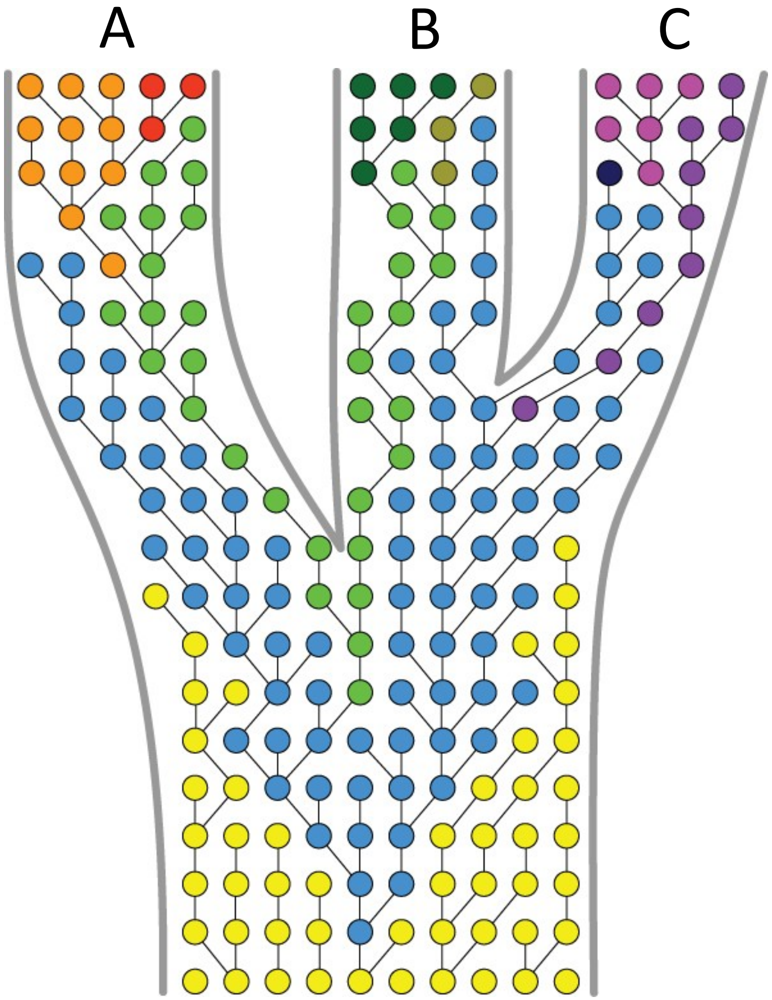
3 species:
Not reciprocally monophyletic

Phylogeny at the level of populations and species



3 species:
Reciprocally monophyletic

Phylogeny at the level of populations and species



- Gene genealogies below and above the species level are different in nature

species delimitation

- population genetics
- phylogenetics

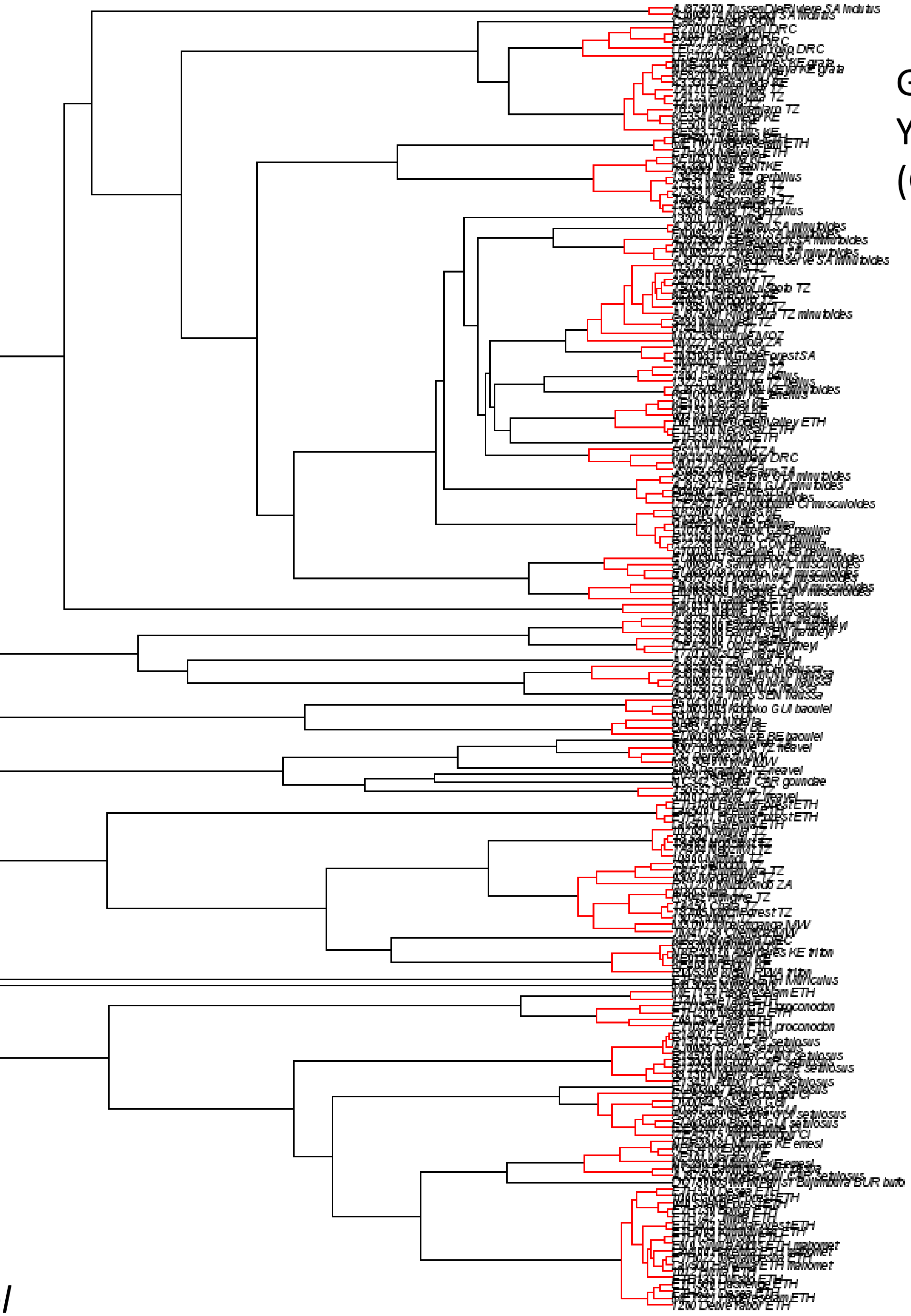
Single-locus delimitation methods

- gene tree lineages that are found in different species cannot coalesce to a common ancestor („no interspecific gene flow“)
- **general mixed Yule coalescent model** (**GMYC**) – model the transition point between cladogenesis and allele coalescence
- prone to over-delimitation

- similar appraches: **mPTP**, **ABGD** („automated barcoding gap detection“)

- should be combined with other approaches

Mus (*Nannomys*) *mattheyi*
Senegal, Dar Salam, 2006
Photo by J. Červený



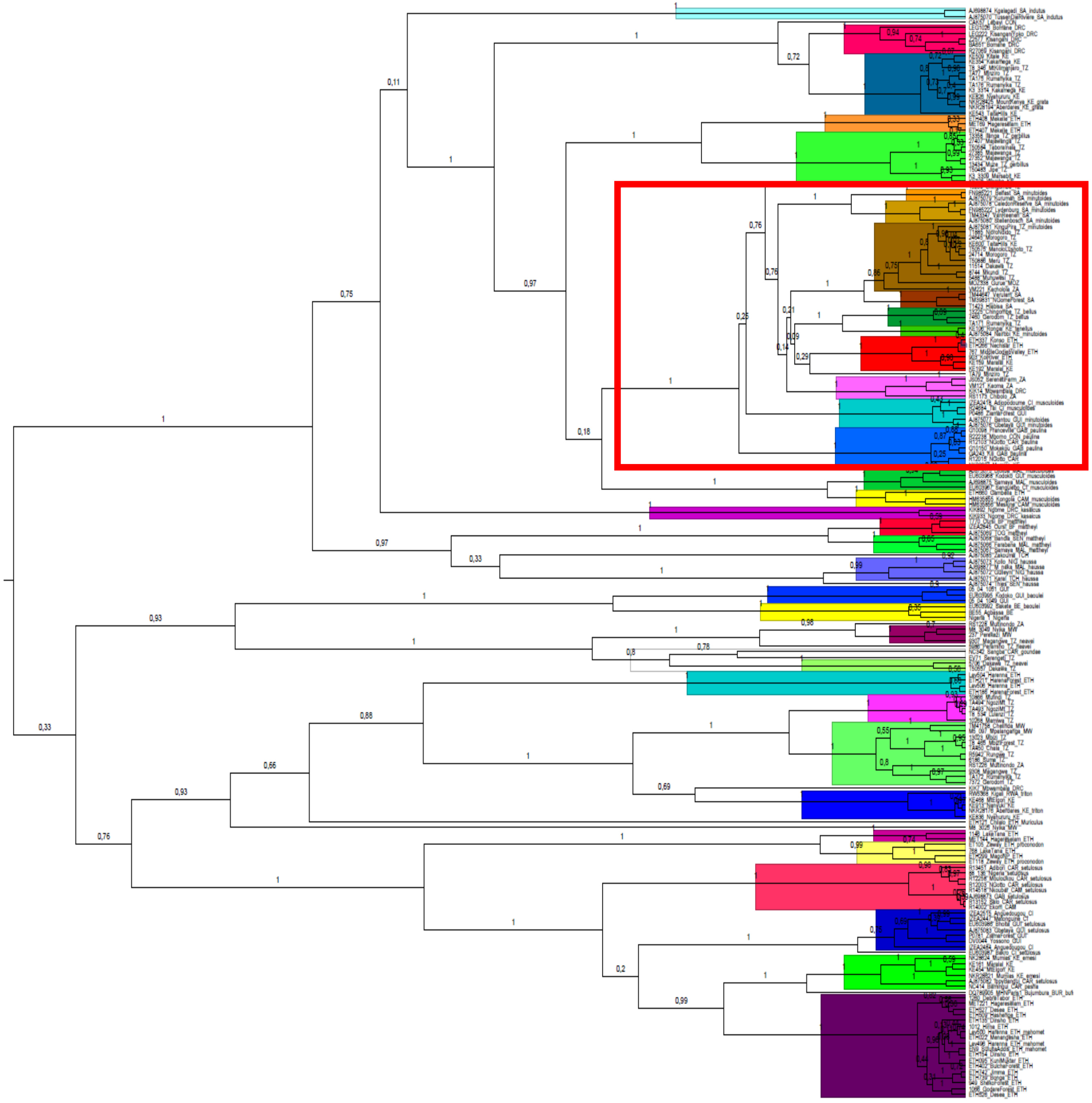
Generalized mixed Yule-coalescent model (GMYC; Pons et al. 2006)

red = intraspecific branching (coalescent process)

black = branching between species (**Yule model including speciation and extinction rates**)

→ splitting of species („taxonomic inflation“)

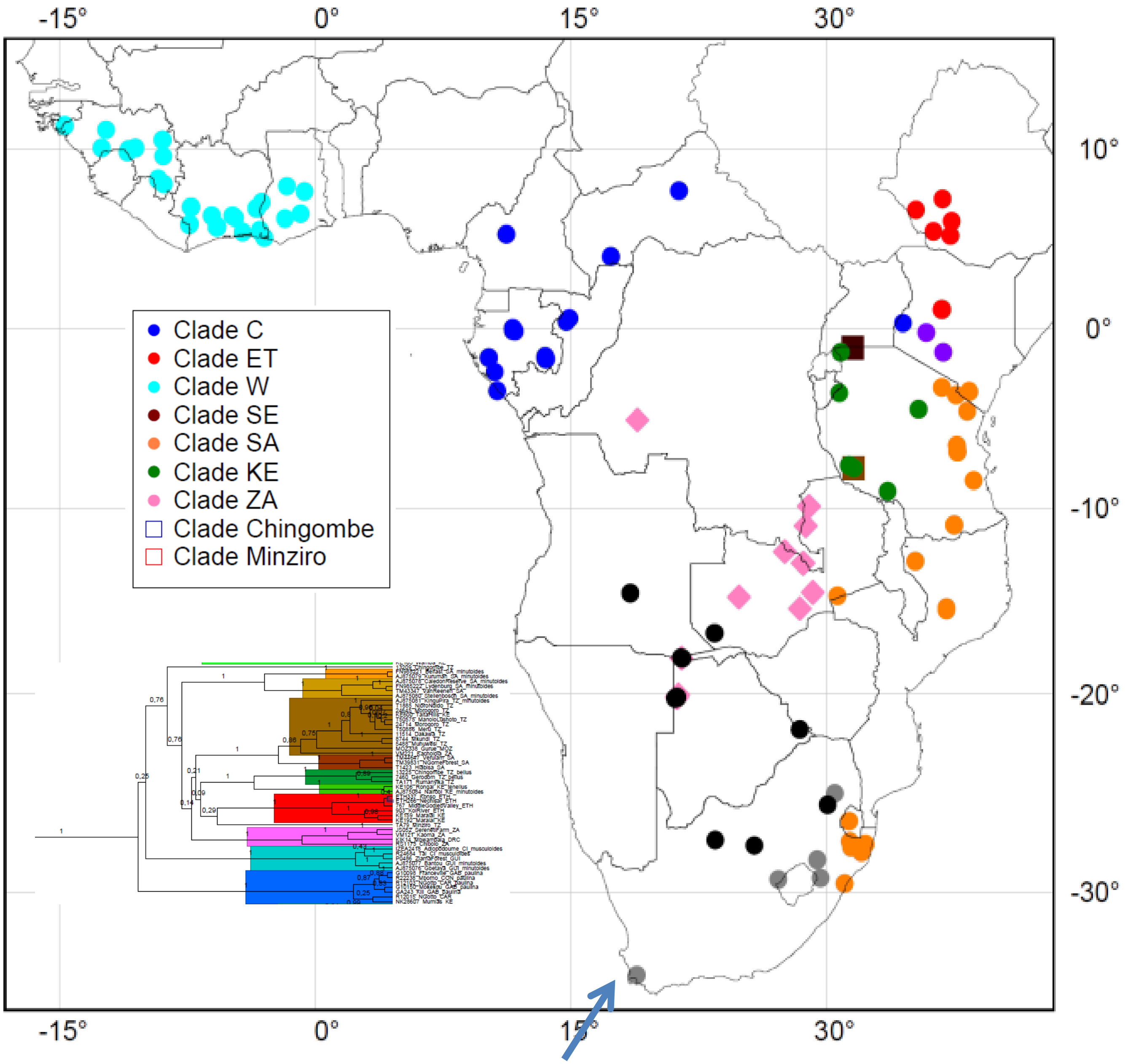
Bryja et al. 2014, BMC Evol Biol



Mus minutoides

Bryja et al. 2014 BMC Evol Biol

Mus minutoides



Bryja et al. 2014 BMC Evol Biol

type locality

- monophyletic clade *M. minutoides*
- **12 „species“** identified by GMYC approach (**3.27-6.96% K2P-distance** among lineages)
- mostly parapatric distribution
- spatial genetic structure is similar to other widely distributed savanna species = **intraspecific phylogeographic structure**

→ Correction of number of possible taxa

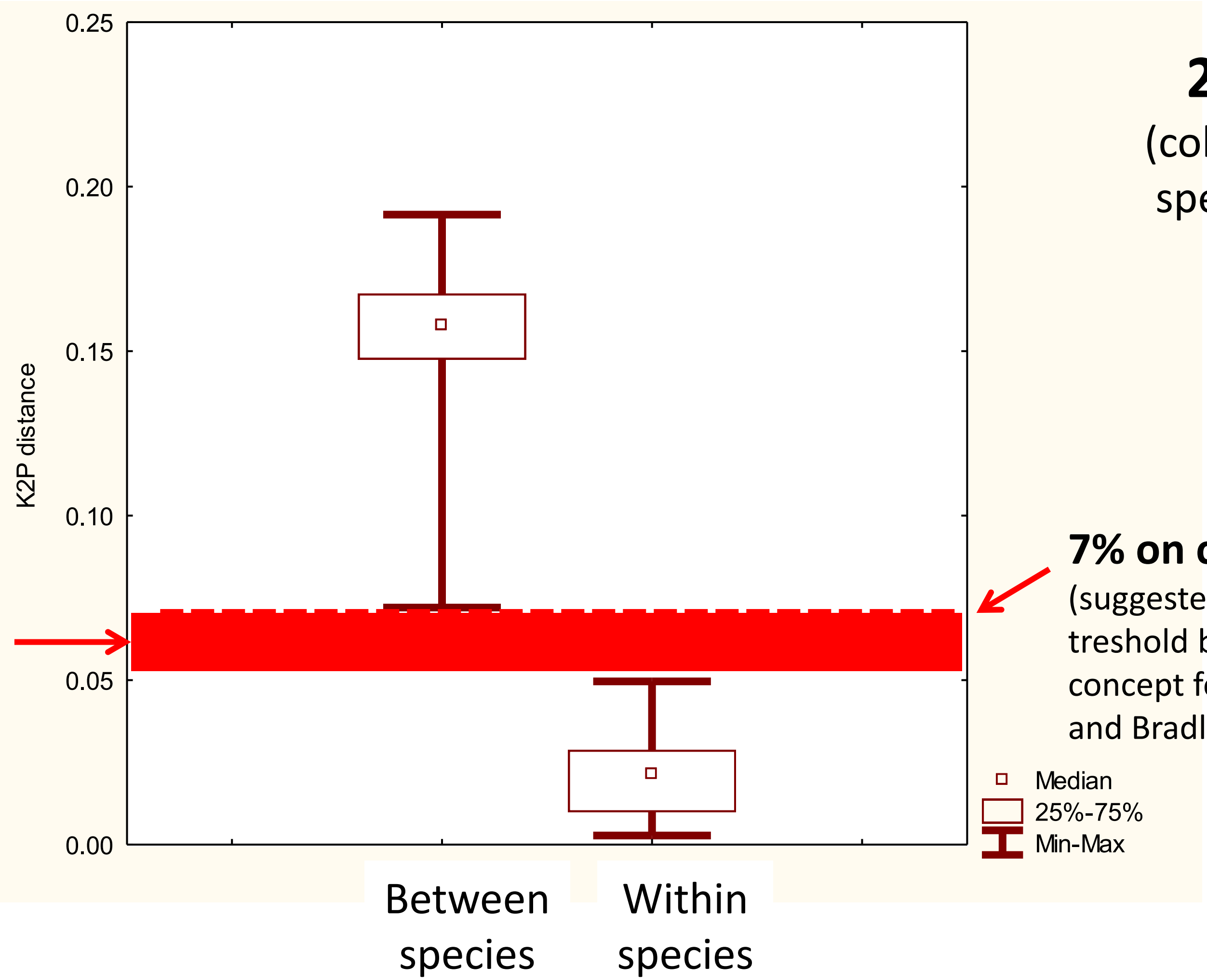
- „GMYC identified species“ collapsed to **MOTUs** (= putative species) if:

(1) parapatric distribution of sister lineages

AND

(2) genetic distances among sister lineages lower than 7% (treshold based on *M. minutoides*)

MOTUs vs. genetic distances

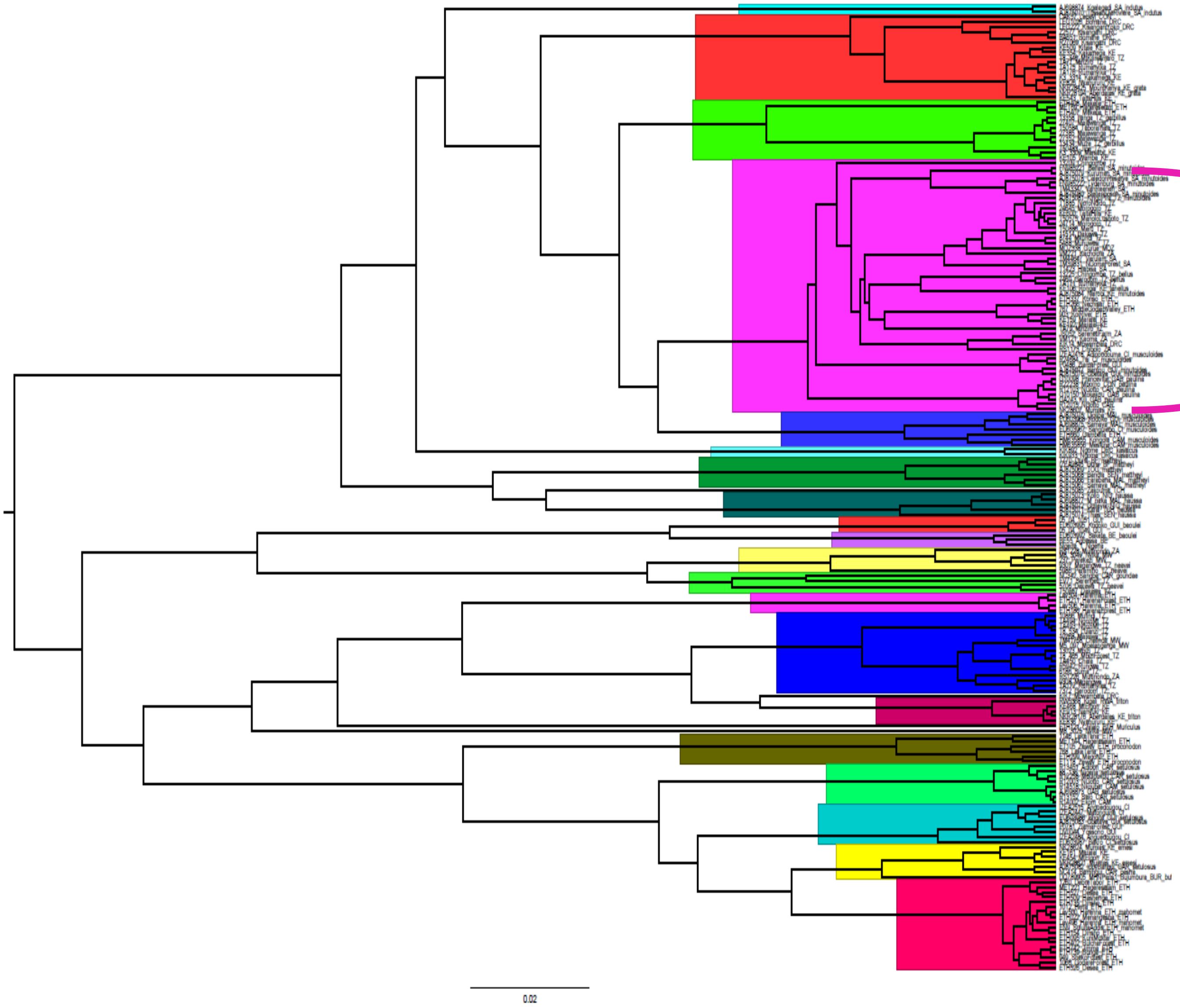


27 MOTUs
(collapsed from 49 species identified by GMYC)

7% on cytochrome *b*
(suggested as the arbitrary treshold by the genetic species concept for mammals; Baker and Bradley, 2006)

„barcoding gap“

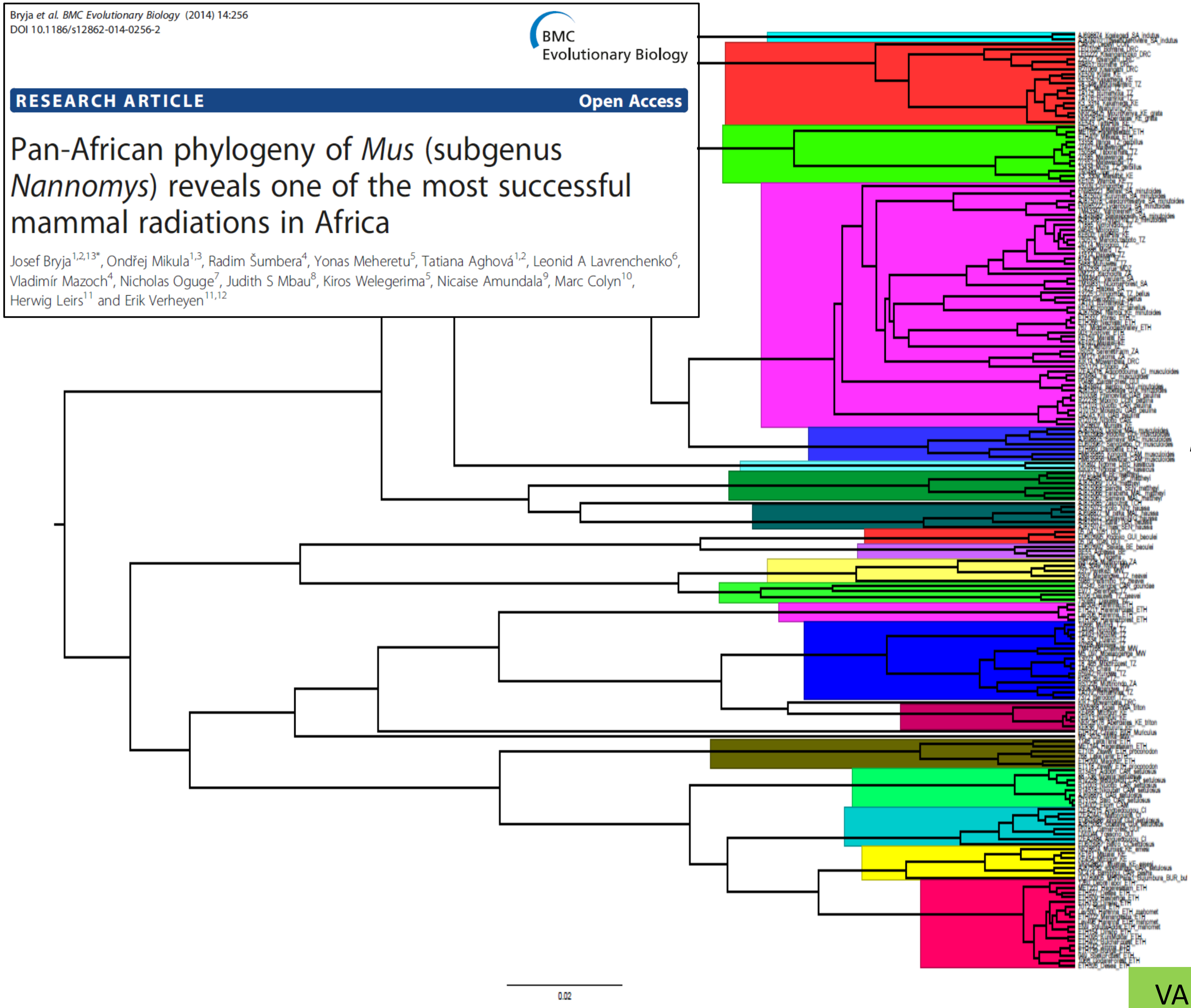
27 MOTUs after correction of GMYC delimitations



Mus minutoides

- 27 MOTUs (**7.2-19.2% distance**)

How many species of *Nannomys*?



Bryja *et al. BMC Evolutionary Biology* (2014) 14:256
DOI 10.1186/s12862-014-0256-2

BMC Evolutionary Biology

RESEARCH ARTICLE **Open Access**

Pan-African phylogeny of *Mus* (subgenus *Nannomys*) reveals one of the most successful mammal radiations in Africa

Josef Bryja[1,2,13*], Ondřej Mikula[1,3], Radim Šumbera[4], Yonas Meheretu[5], Tatiana Aghová[1,2], Leonid A Lavrenchenko[6], Vladimír Mazoch[4], Nicholas Oguge[7], Judith S Mbau[8], Kiros Welegerima[5], Nicaise Amundala[9], Marc Colyn[10], Herwig Leirs[11] and Erik Verheyen[11,12]

indutus
tenellus?
minutoides
musculoides
mattheyi
haussa
baoulei
neavei
sorella
triton
imberbis
setulosus
bufo
mahomet

Sp. 1 *(gratus?)*
Sp. 2 *(gerbillus?)*
Sp. 3 *(kasaicus?)*
Sp. 4 (Tchad)
Sp. 5 *(cf. baoulei?)*
Sp. 6 (Dakawa)
Sp. 7 (Harena)
Sp. 8 (Kikwit)
Sp. 9 (Kenya)
Sp. 10 *(proconodon?)*
Sp. 11 *(cf. setulosus?)*

VALID NAMES
(14-15)

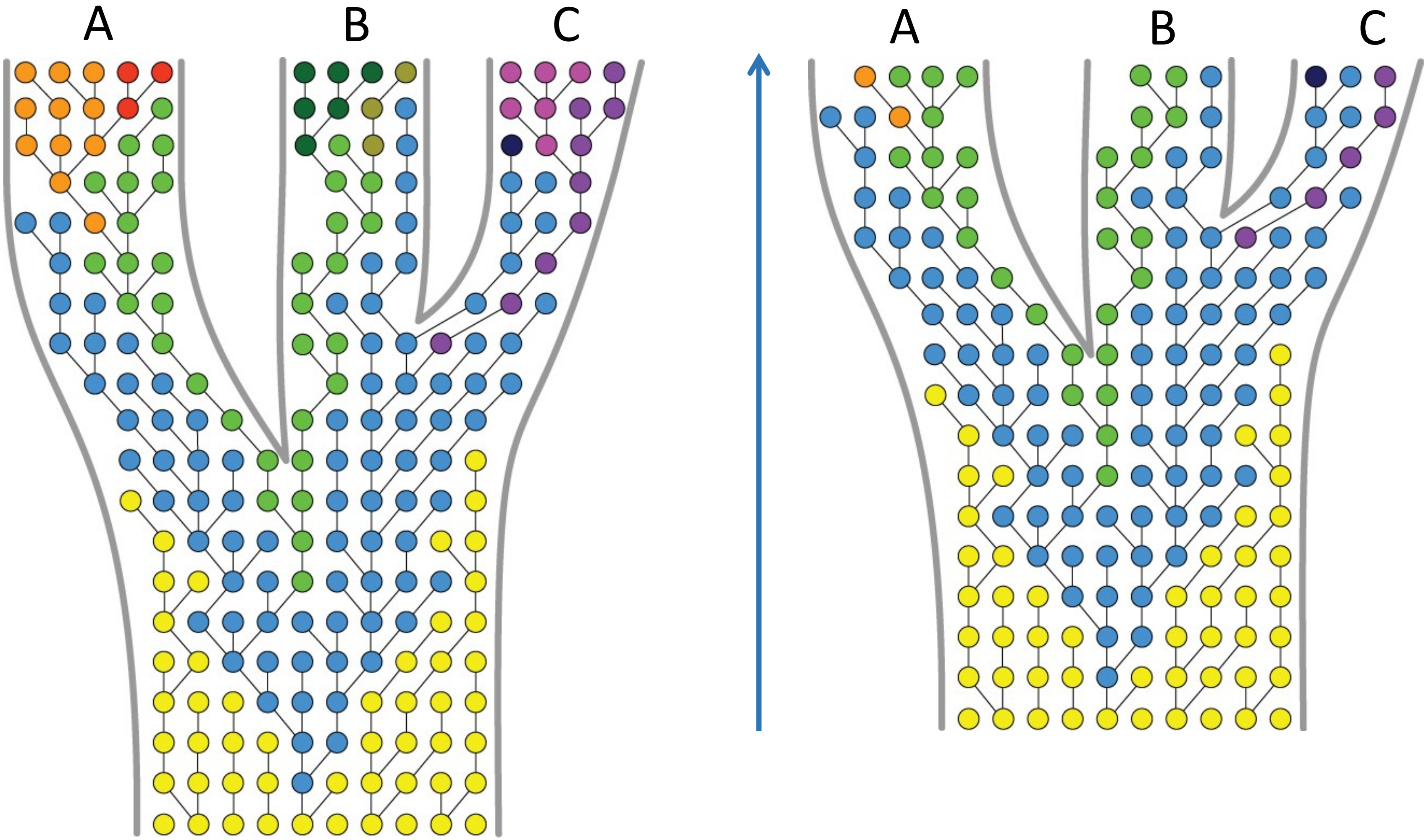
NEW SPECIES
(11-12)

Bryja et al. 2014 BMC Evol Biol

Multi-species coalescence for species delimitation

- **multi-locus data**
- **BPP** (joint estimates of species delimitation and species tree)
- spedeSTEM
- STACEY
- and others

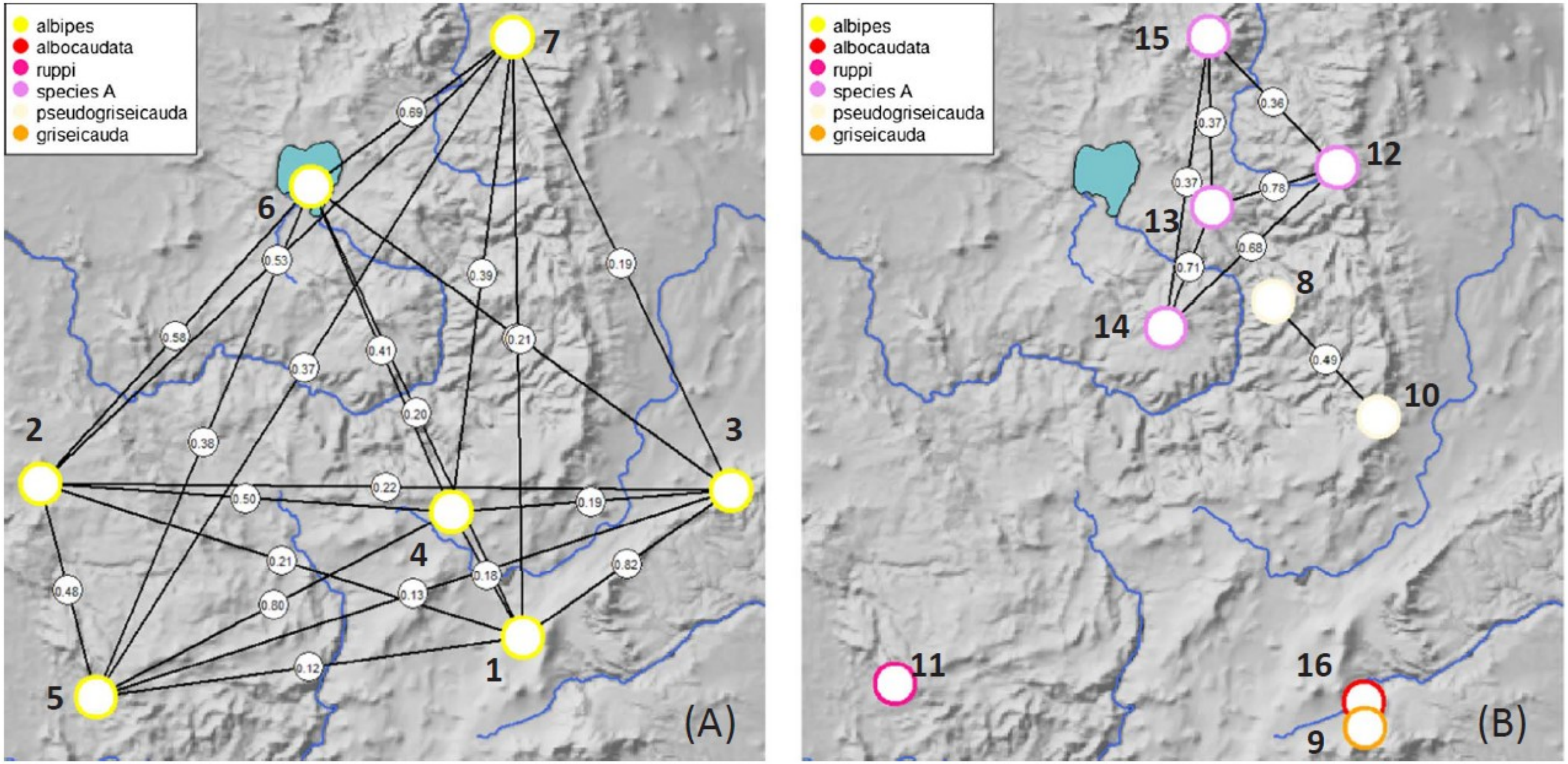
Multilocus data, multi-species coalescence



locus 1

locus 2

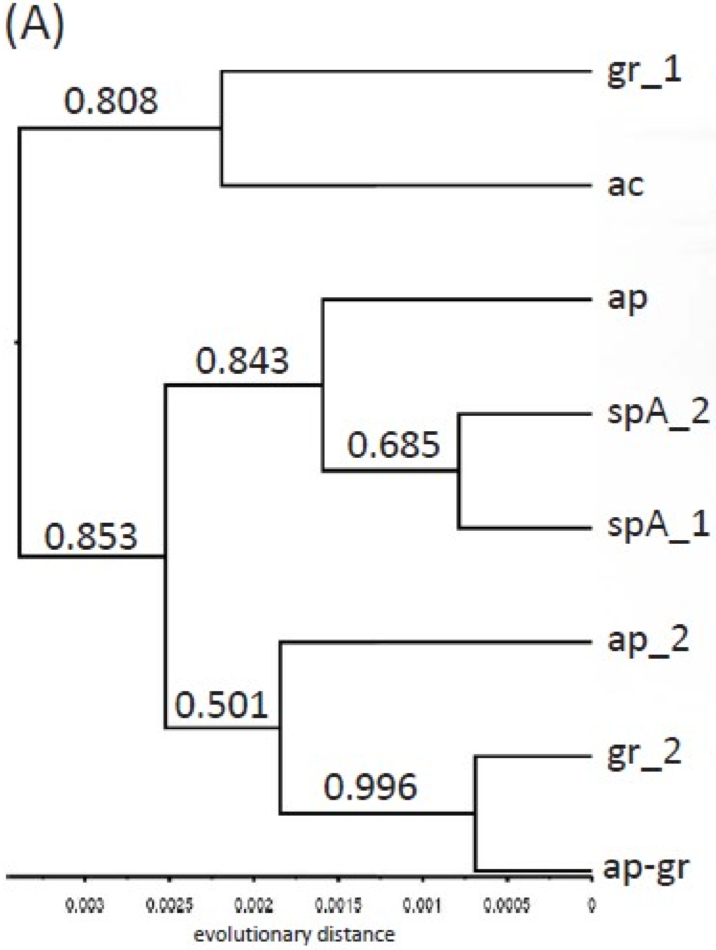
Example: 6 nucler loci in *Stenocephalemys*

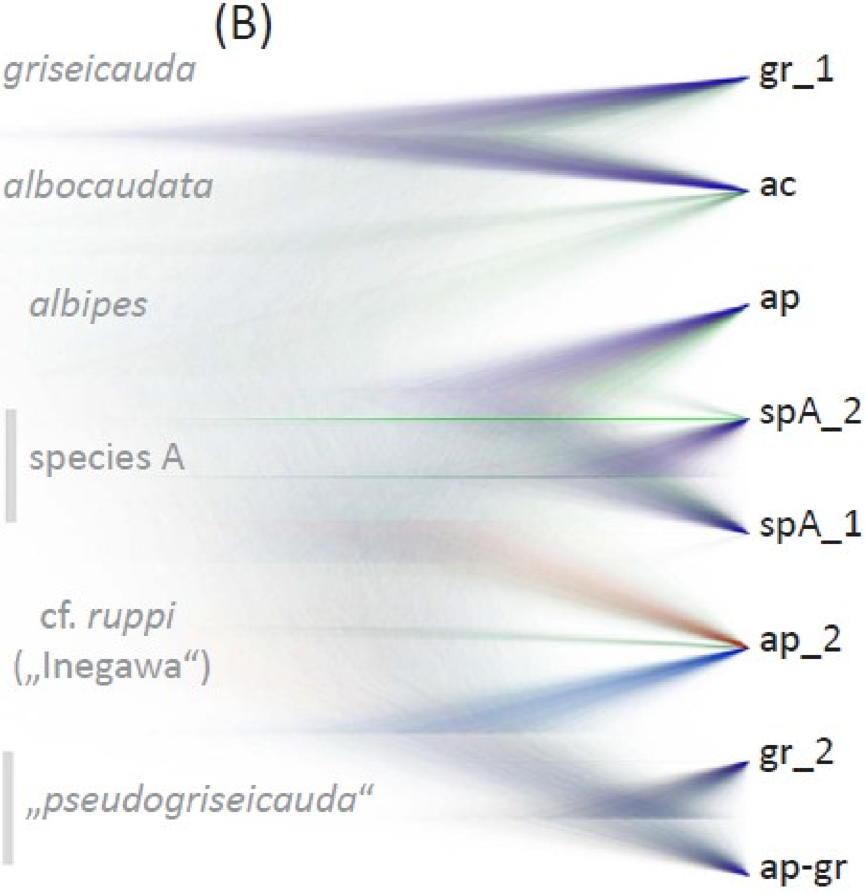


STACEY - species delimitation

Bryja et al. 2018, Mol Phyl Evol

Example: 6 nucler loci in *Stenocephalemys*

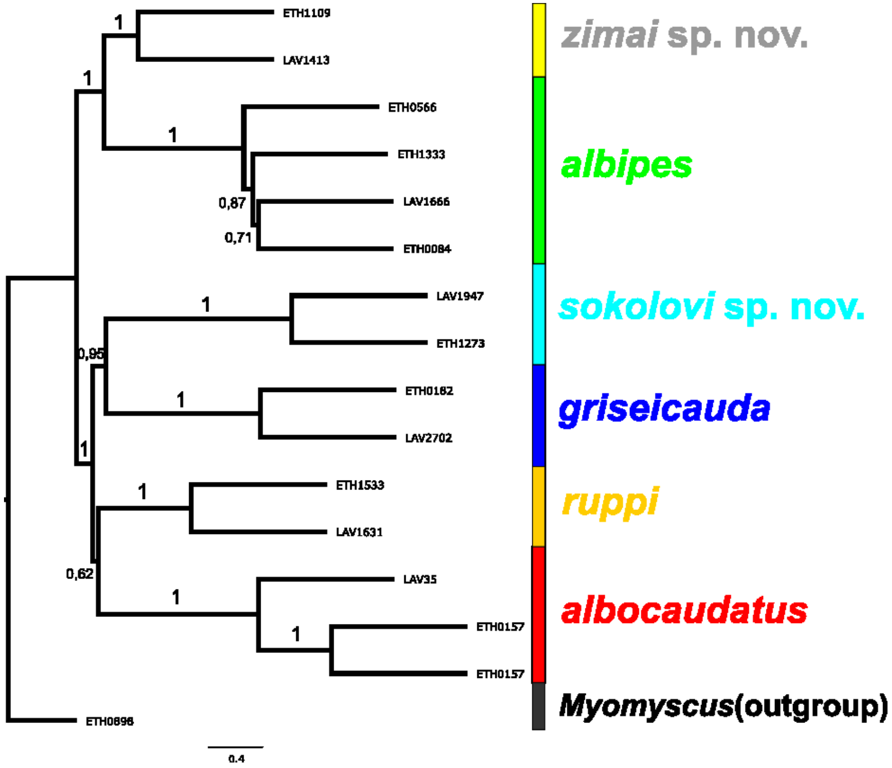




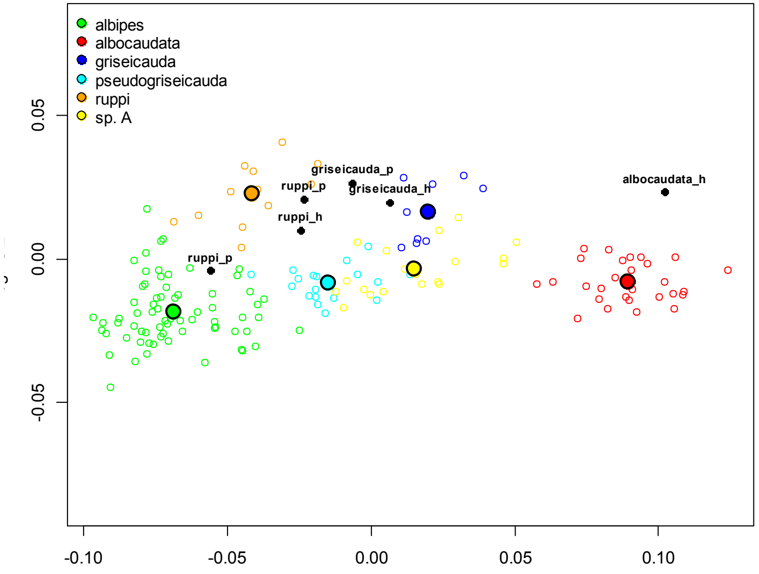
StarBEAST – species tree

Bryja et al. 2018, Mol Phyl Evol

Integrative taxonomy of *Stenocephalemys*



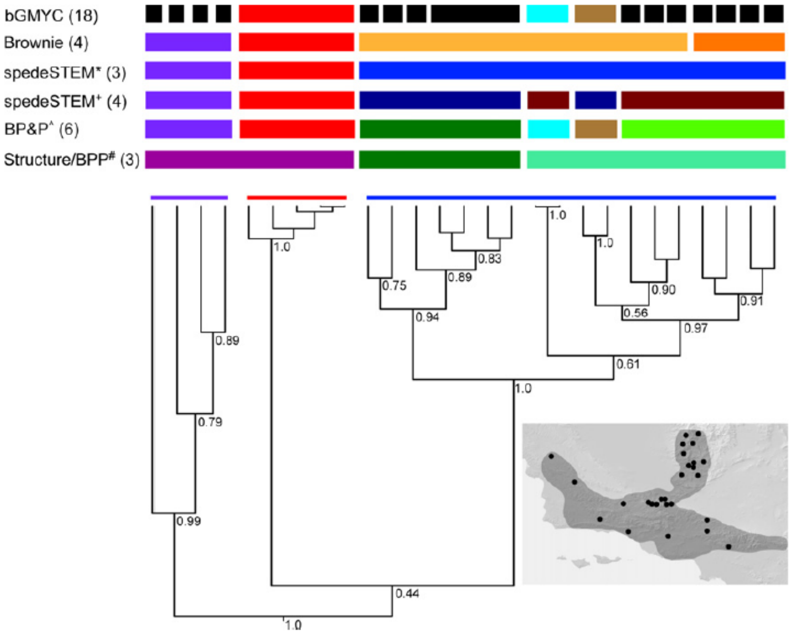
anchored phylogenomics
(388 nuclear loci)

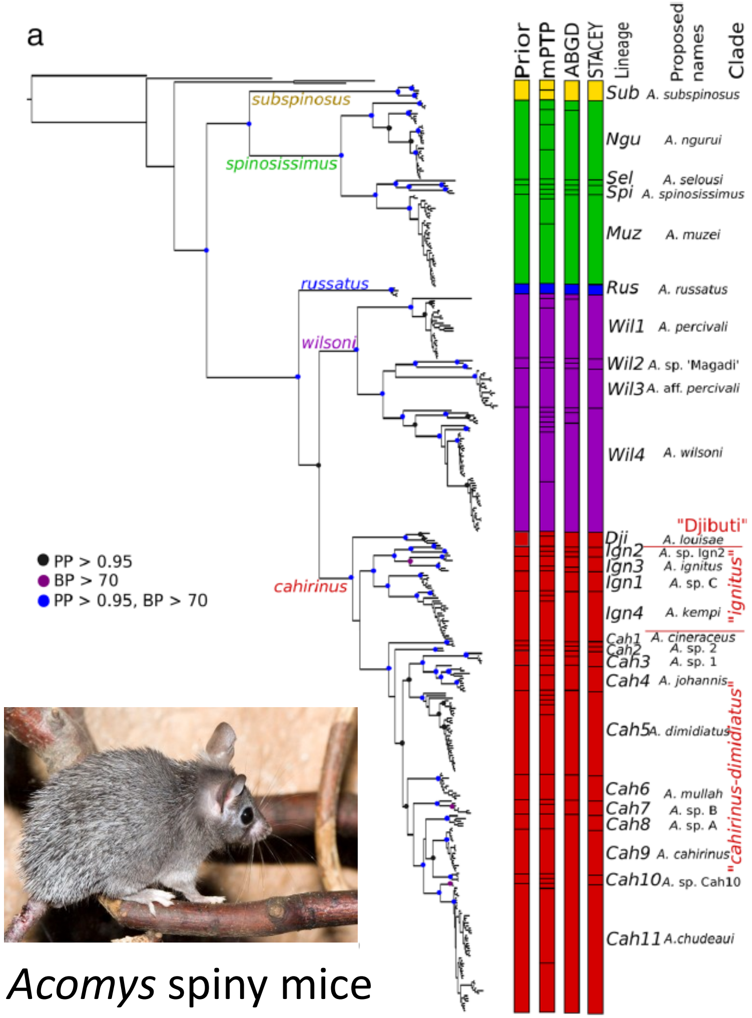


Mizerovská et al. 2020

Incongruent results from empirical systems

Aliatypus spiders



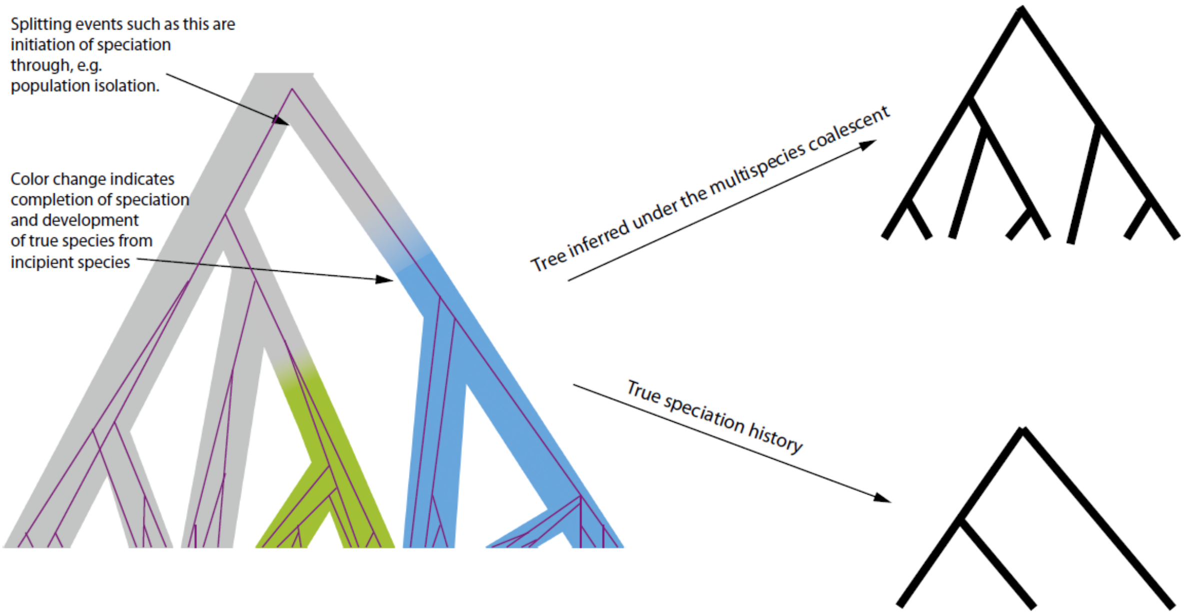


Acomys spiny mice

Carstens et al. 2013

Aghová, ..., Bryja 2019, BMC Evol Biol

MSC identifies population structure, not species



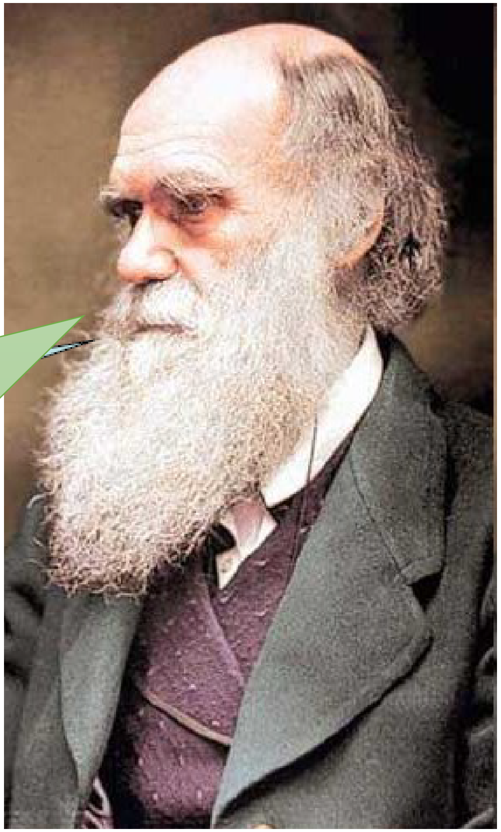
Splitting events such as this are initiation of speciation through, e.g. population isolation.

Color change indicates completion of speciation and development of true species from incipient species

- *„until we develop genomic-based species delimitation approaches that are able to discriminate between population- and species-level structuring, it is important not just to recognize, but to treat and report* ***the units delimited under MSC at best as tentative hypotheses of species****“*

Sukumaran and Knowles, 2017, PNAS

Is it possible to define a species?



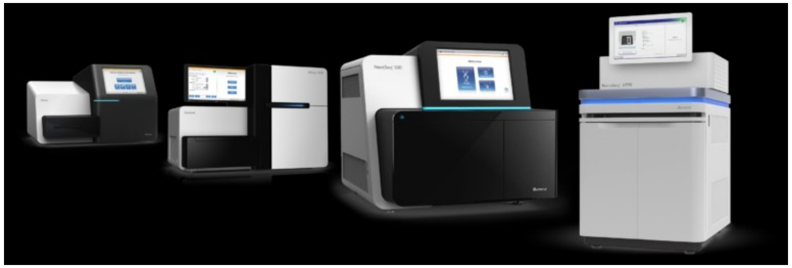
It is really laughable to see what different ideas are prominent in various naturalists minds, when they speak of „species“; ... It all comes, I believe, from trying to define the indefinable.

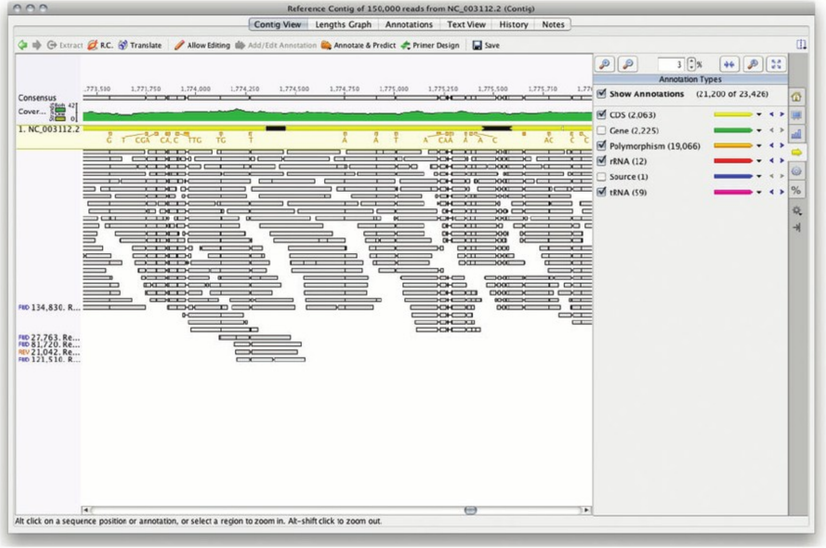
Darwin 1856

THE RISE OF GENOMICS – CAN IT HELP?

High-throughput sequencing (HTS)

- unprecedented increase of genetic data
- „several-loci“ approaches for species delimitation applicable with difficulties – computationally demanding → **new approaches for HTS data are highly required** (and are intensively developed)

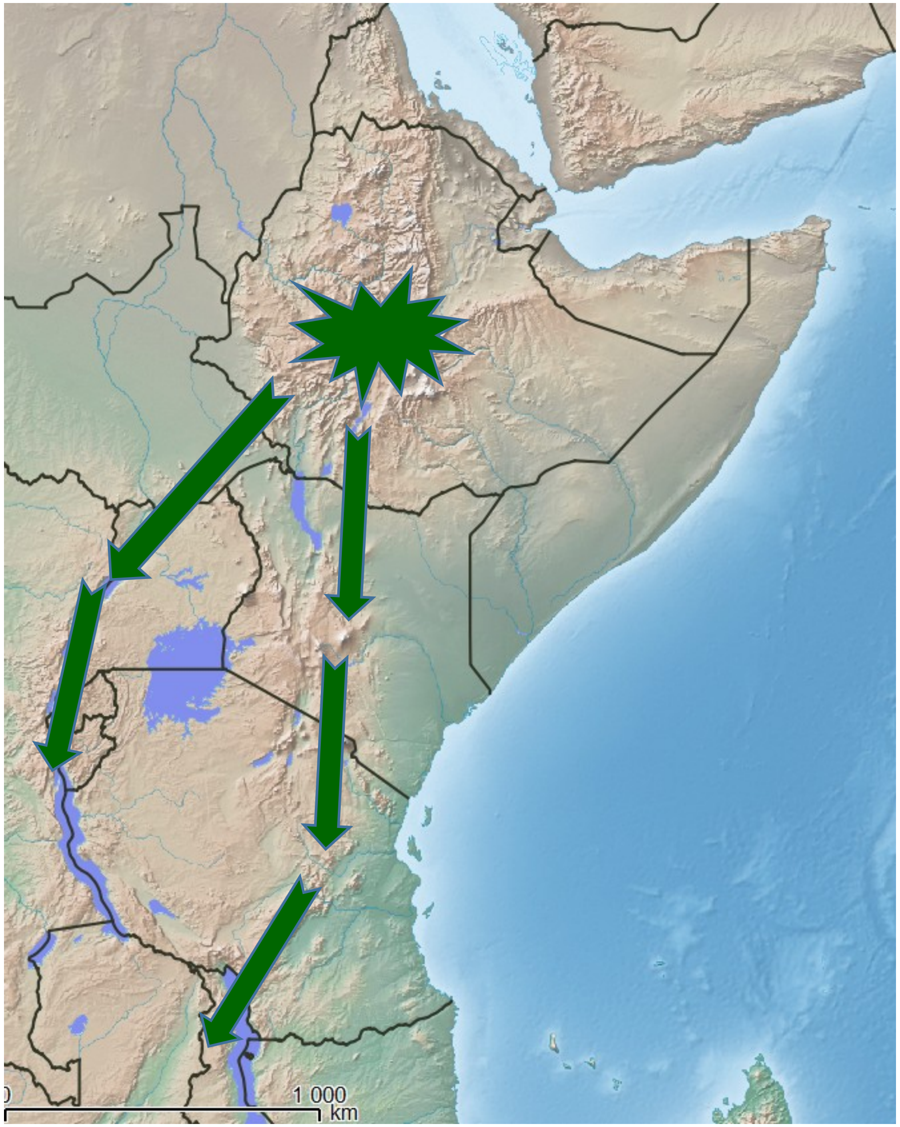




☑ Show Annotations (21,200 of 23,426)
☑ CDS (2,063)
☐ Gene (2,225)
☑ Polymorphism (19,066)
☑ rRNA (12)
☐ Source (1)
☑ tRNA (59)

Alt click on a sequence position or annotation, or select a region to zoom in. Alt-shift click to zoom out.

Alternative approaches for species delimitation from genomic data



- ancestral lineage in Ethiopian highlands, where diversified and sourced the colonization of other mountains (**mostly in Pleistocene**)
- ***Lophuromys flavopunctatus* complex**

V. Komarova

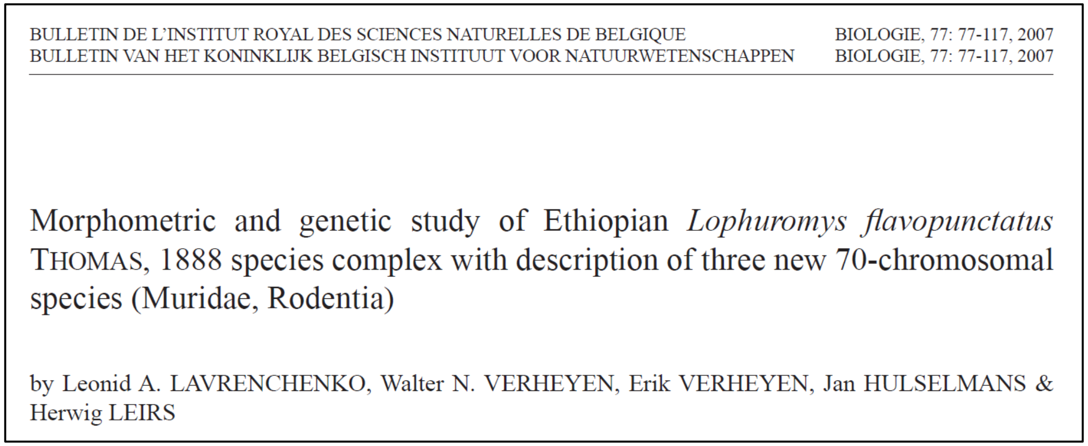
D. Kostin

O. Mikula

A. Bryjová

D. Čížková

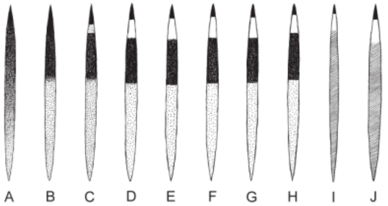
Nine endemic species in Ethiopia

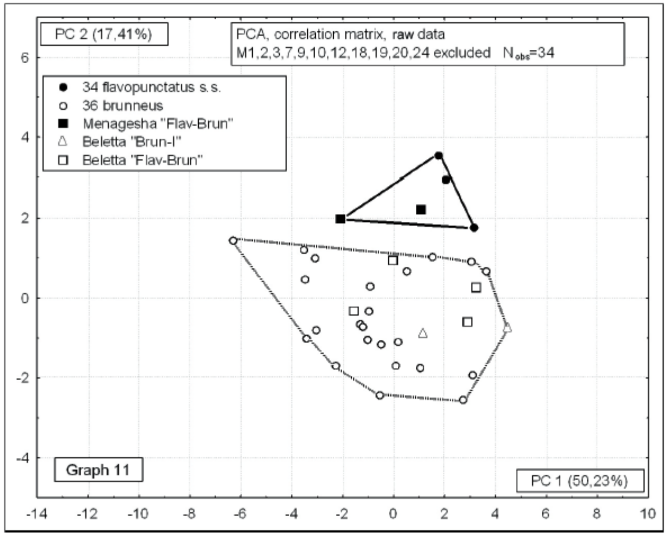


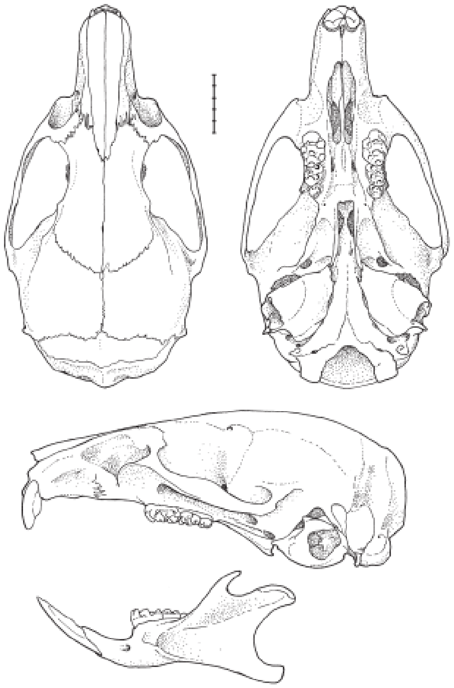
BULLETIN DE L'INSTITUT ROYAL DES SCIENCES NATURELLES DE BELGIQUE — BIOLOGIE, 77: 77-117, 2007
BULLETIN VAN HET KONINKLIJK BELGISCH INSTITUUT VOOR NATUURWETENSCHAPPEN — BIOLOGIE, 77: 77-117, 2007

Morphometric and genetic study of Ethiopian *Lophuromys flavopunctatus* THOMAS, 1888 species complex with description of three new 70-chromosomal species (Muridae, Rodentia)

by Leonid A. LAVRENCHENKO, Walter N. VERHEYEN, Erik VERHEYEN, Jan HULSELMANS & Herwig LEIRS





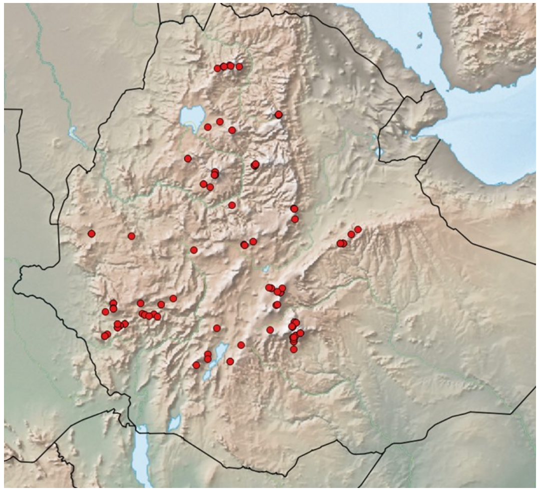


3.2. Views of skull and mandible of *Lophuromys menageshae* n.sp. (ZMMU S-165969, holotype). Scale bar = 5 mm.

Lophuromys flavopunctatus group in Ethiopia

- Are there really nine well-delimited species?
- What is their distribution, co-occurrence patterns, ecological requirements? -> IUCN assessment, etc.

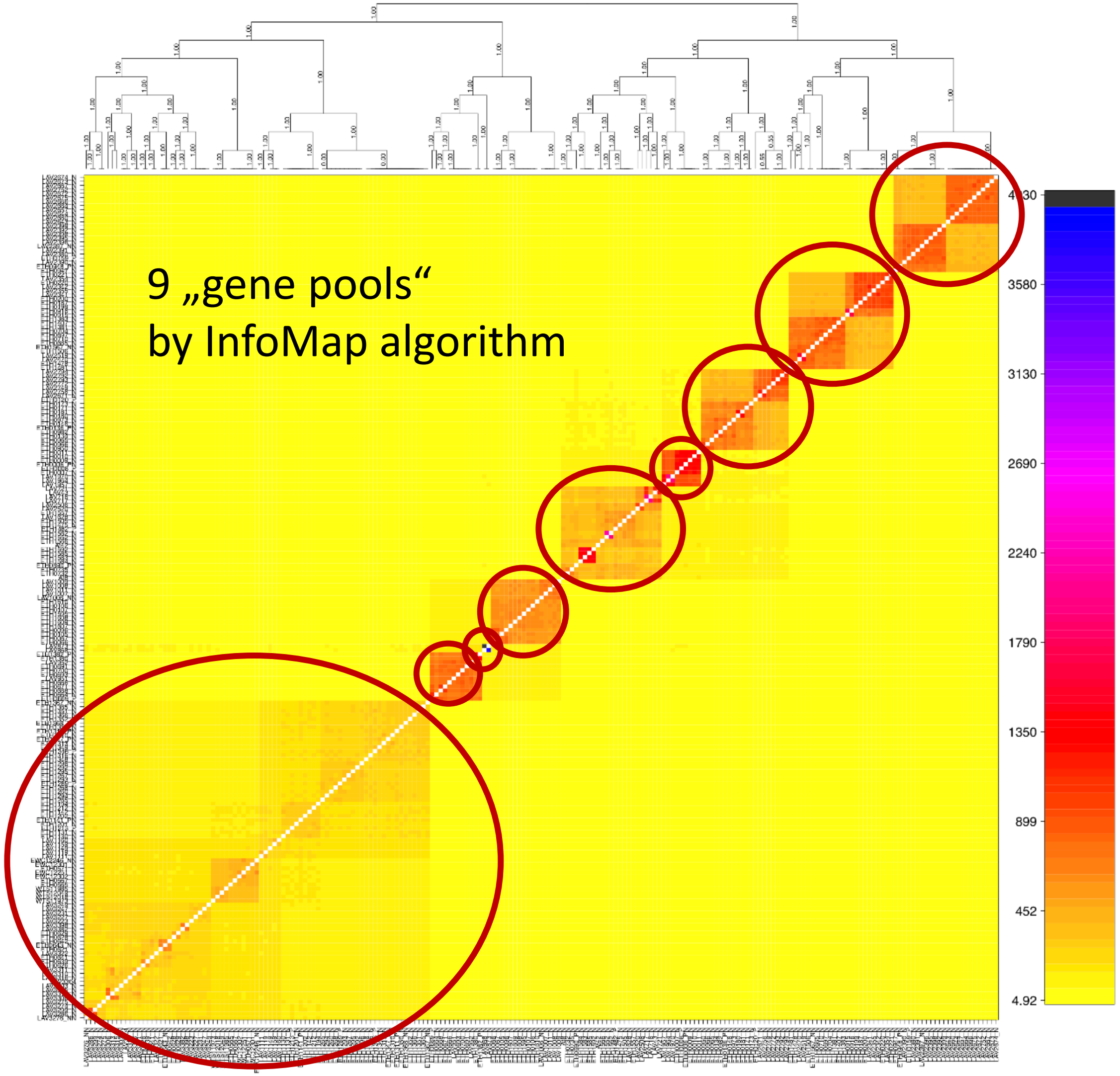
- cca 500 specimens from all major mountain ranges barcoded at mtDNA
- 4 nuclear markers (two introns + two exons)
- **genomic approach (ddRAD sequencing) → thousands of SNPs across the genome**



ddRADseq: co-ancestry matrix

209
individuals

15 623
informative
loci

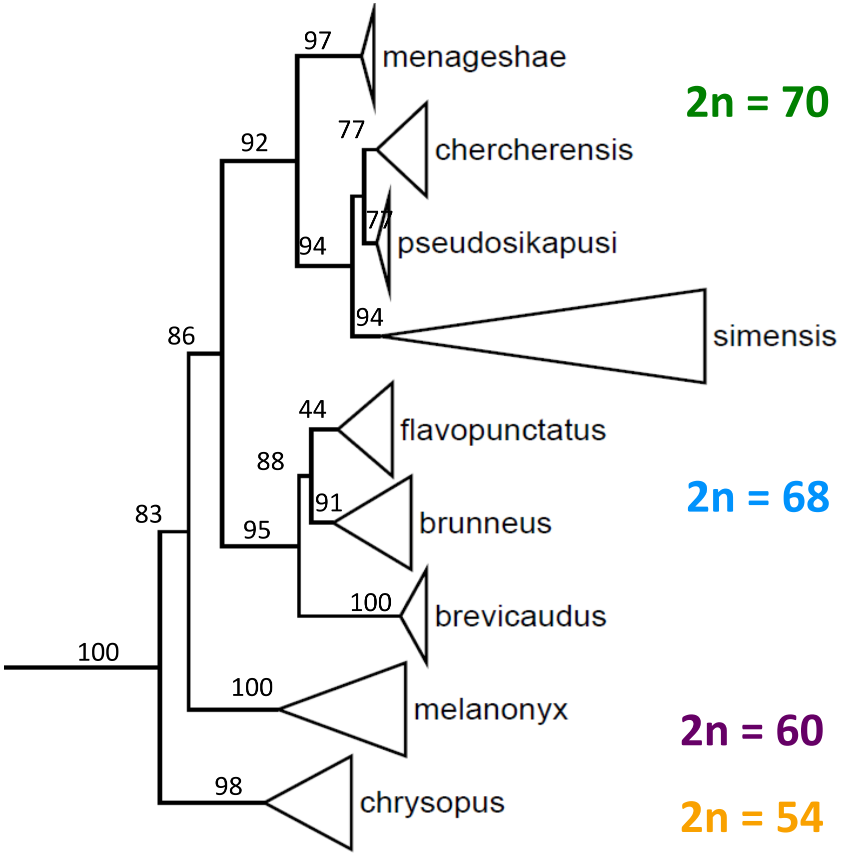


9 „gene pools“
by InfoMap algorithm

Mikula et al., in prep.

Maximum likelihood analysis of concatenated nuclear dataset

Sanger sequencing



4 nuclear markers
(2 604 bp concatenated dataset)

2n = 70

menageshae

chercherensis

pseudosikapusi

simensis

2n = 68

flavopunctatus

brunneus

brevicaudus

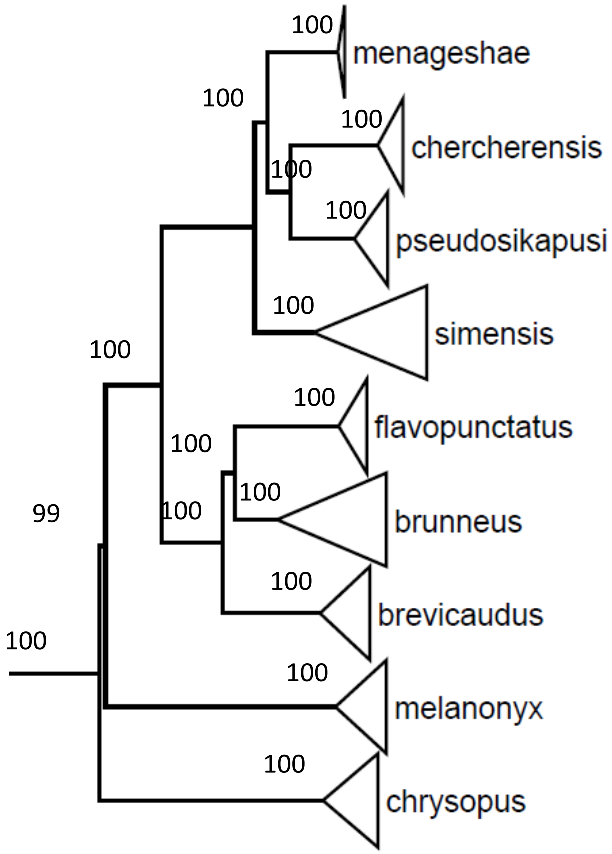
2n = 60

melanonyx

2n = 54

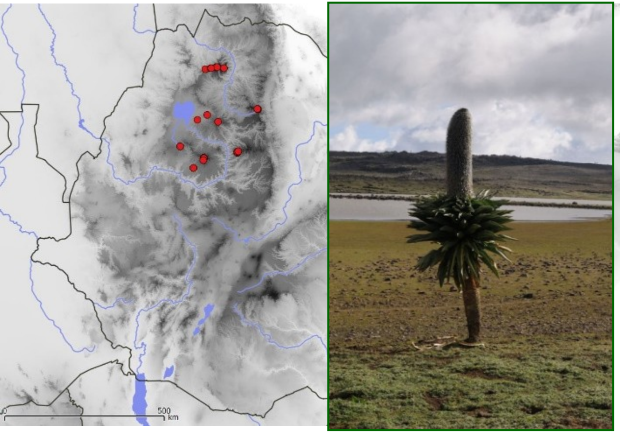
chrysopus

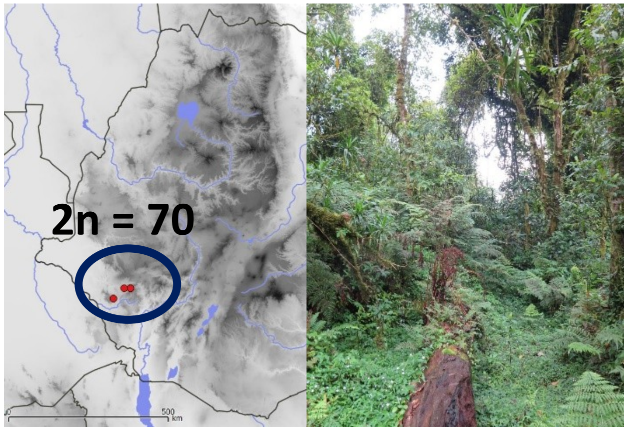
ddRADseq

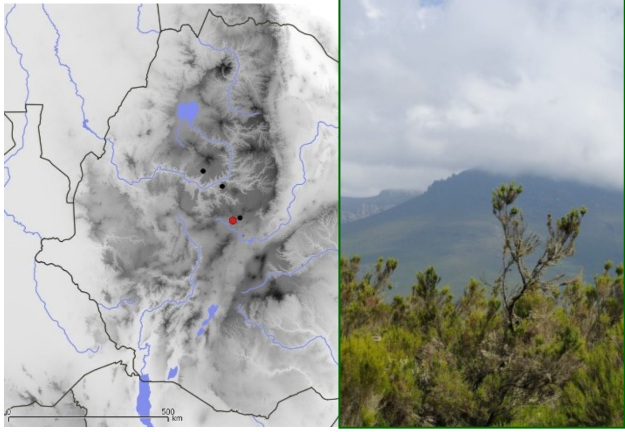


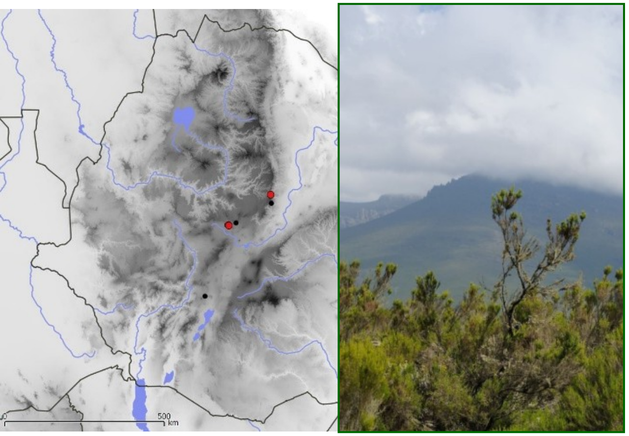
15 623 informative loci

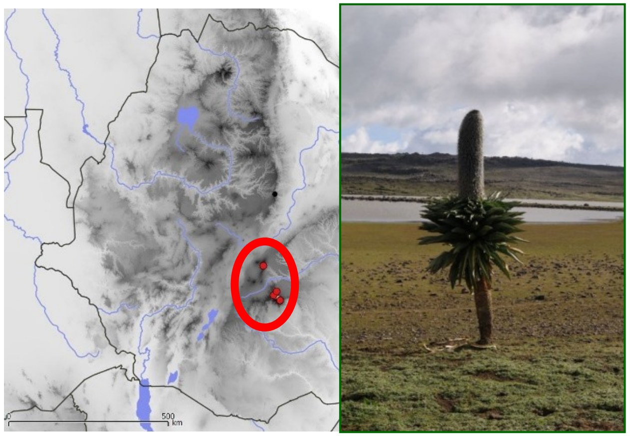
Komarova, Kostin, Bryja et al., Mol. Ecol., 2020

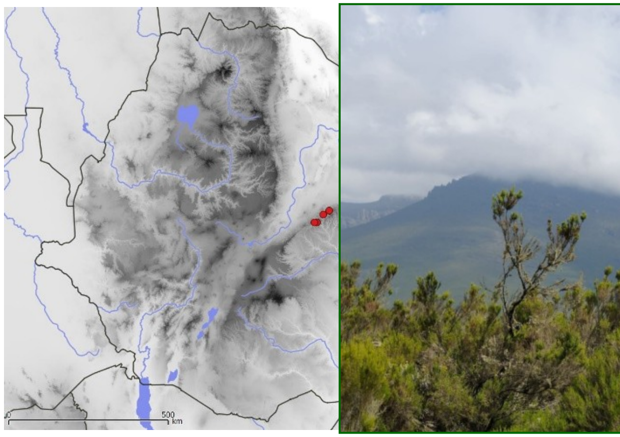

500 km

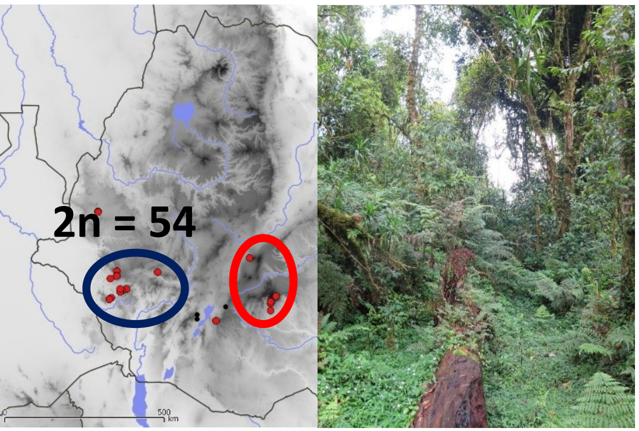

2n = 70
500 km

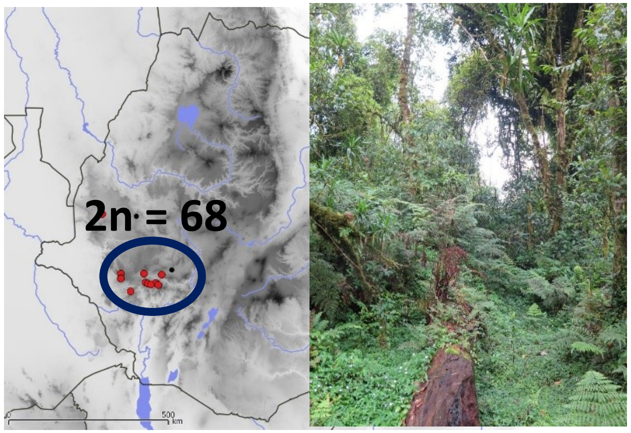

500 km

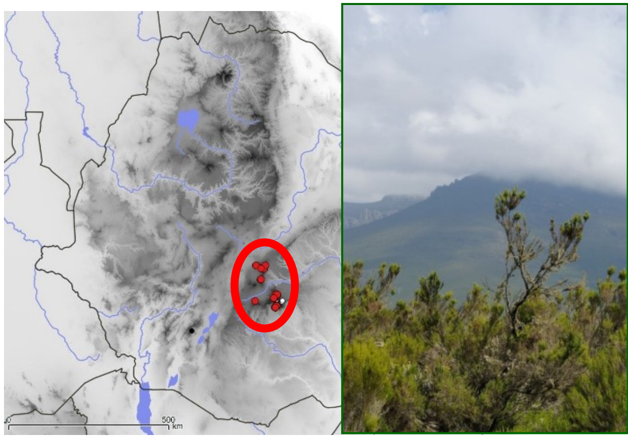

500 km


500 km


500 km

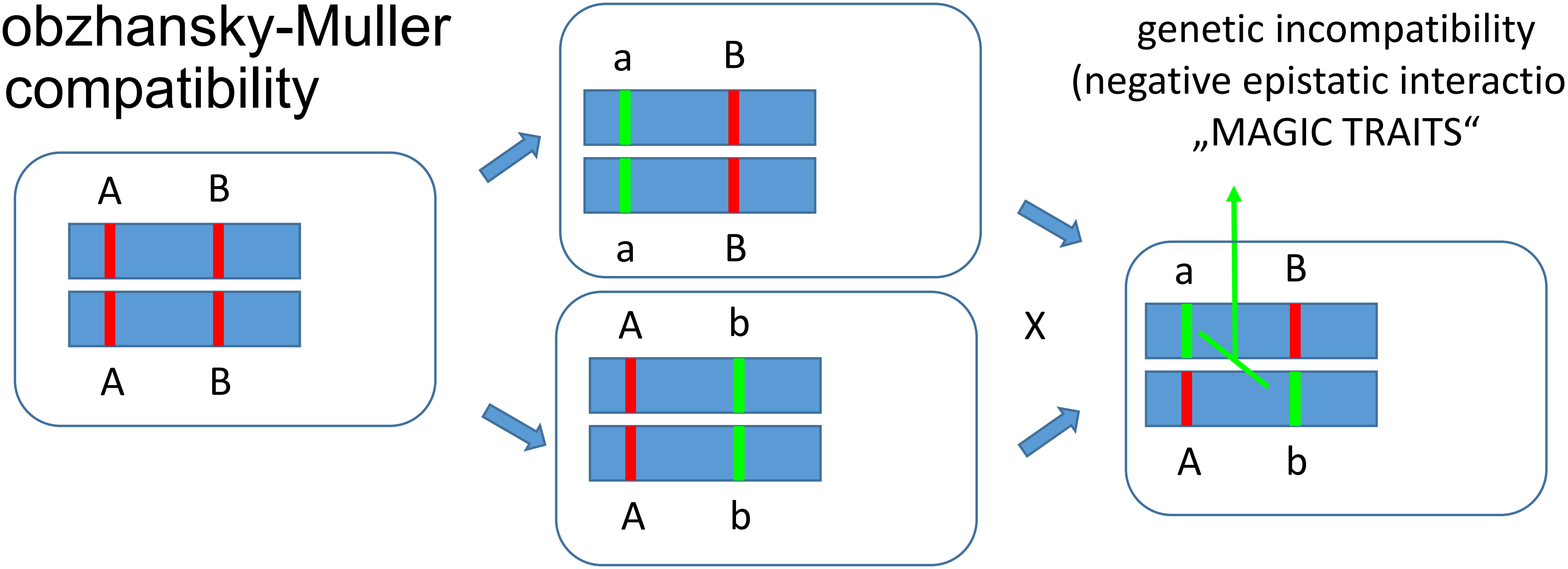

2n = 54
500 km


2n = 68
500 km

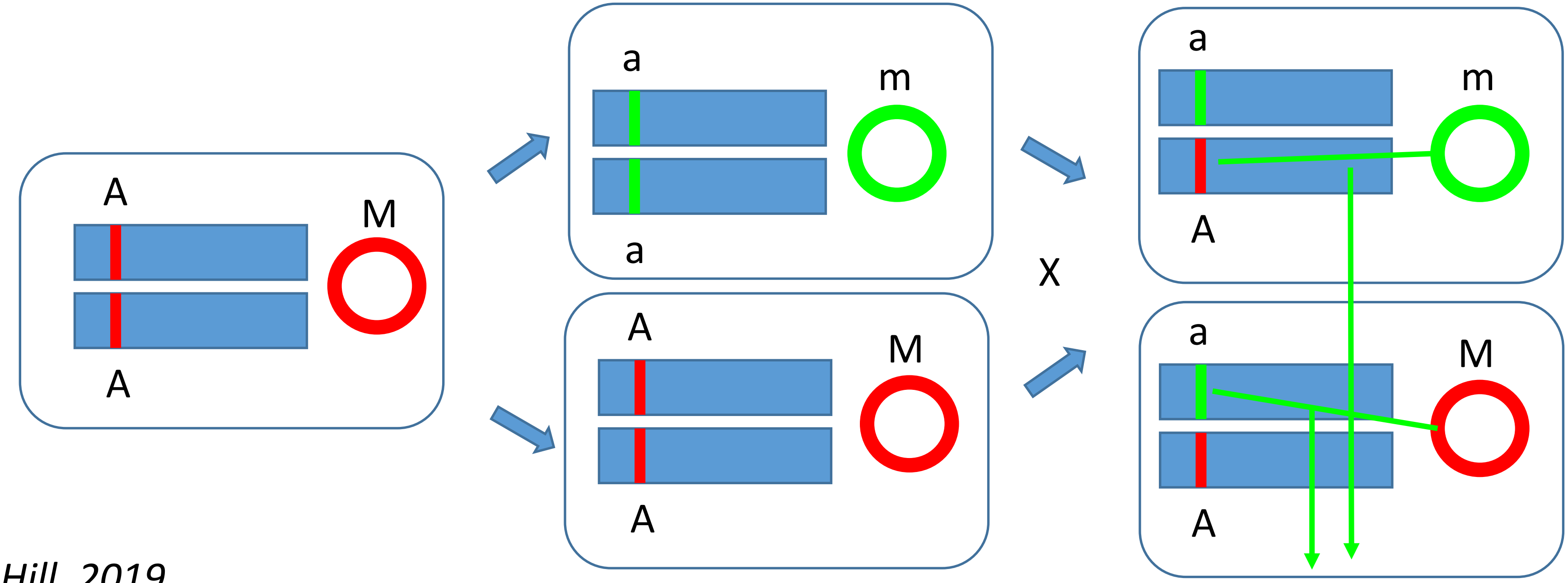

500 km

5. MITOCHONDRIAL INTROGRESSION AND A UNIFYING SPECIES CONCEPT (?)

Dobzhansky-Muller incompatibility



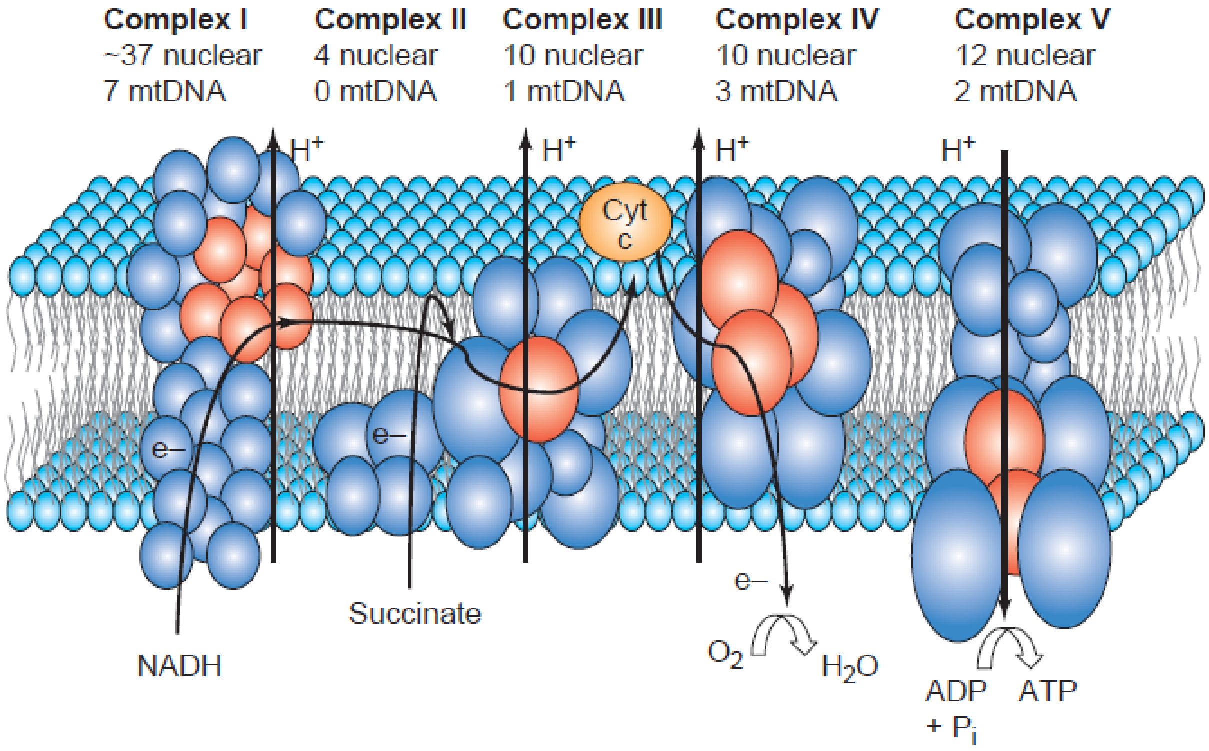
genetic incompatibility
(negative epistatic interaction)
„MAGIC TRAITS“



mitonuclear incompatibility

Hill, 2019

Oxidative phosphorylation (OXPHOS) by electron transport system (ETS)

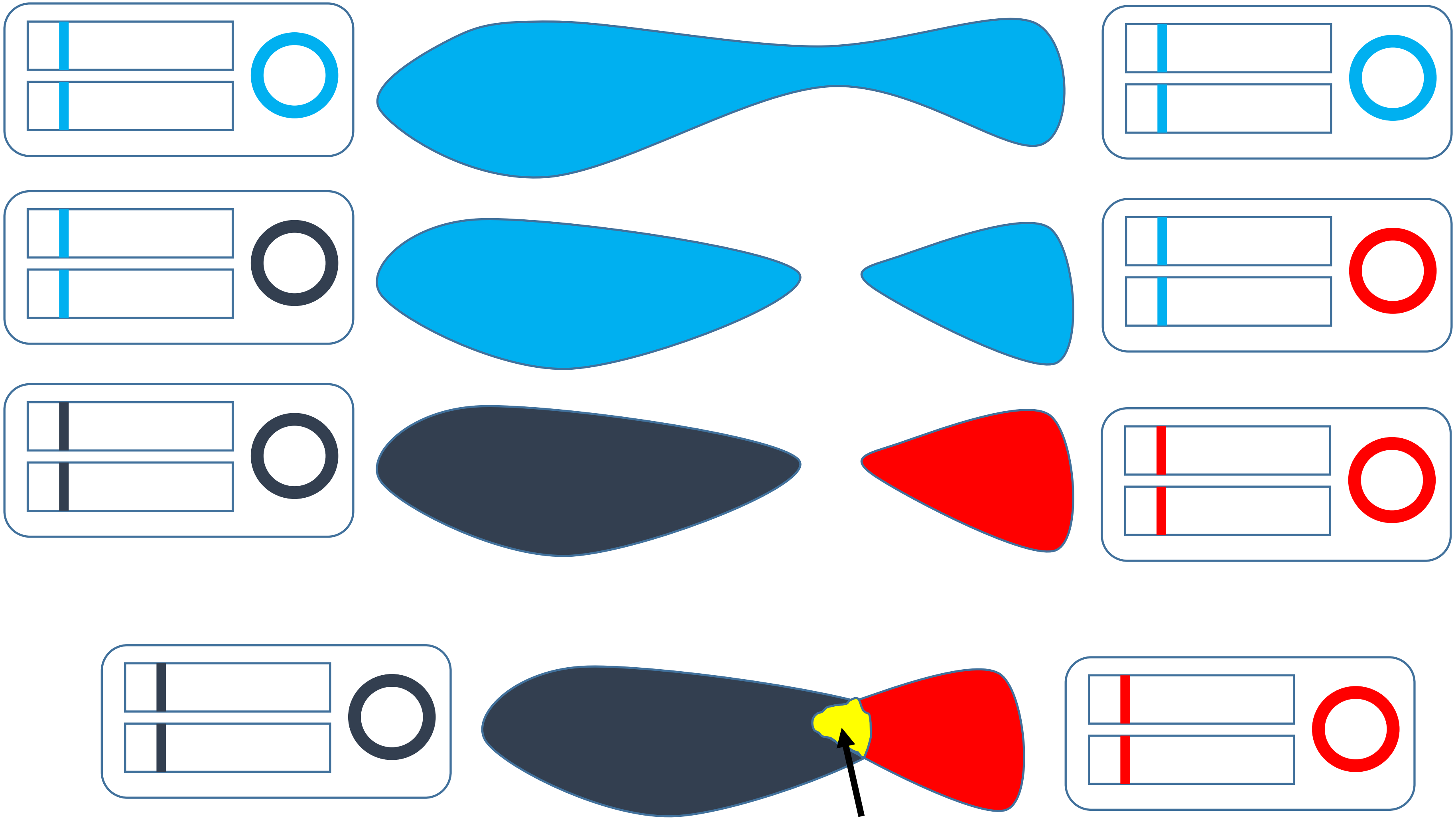


red = mt genes
(mtDNA)

blue = N-mt genes
(nDNA)

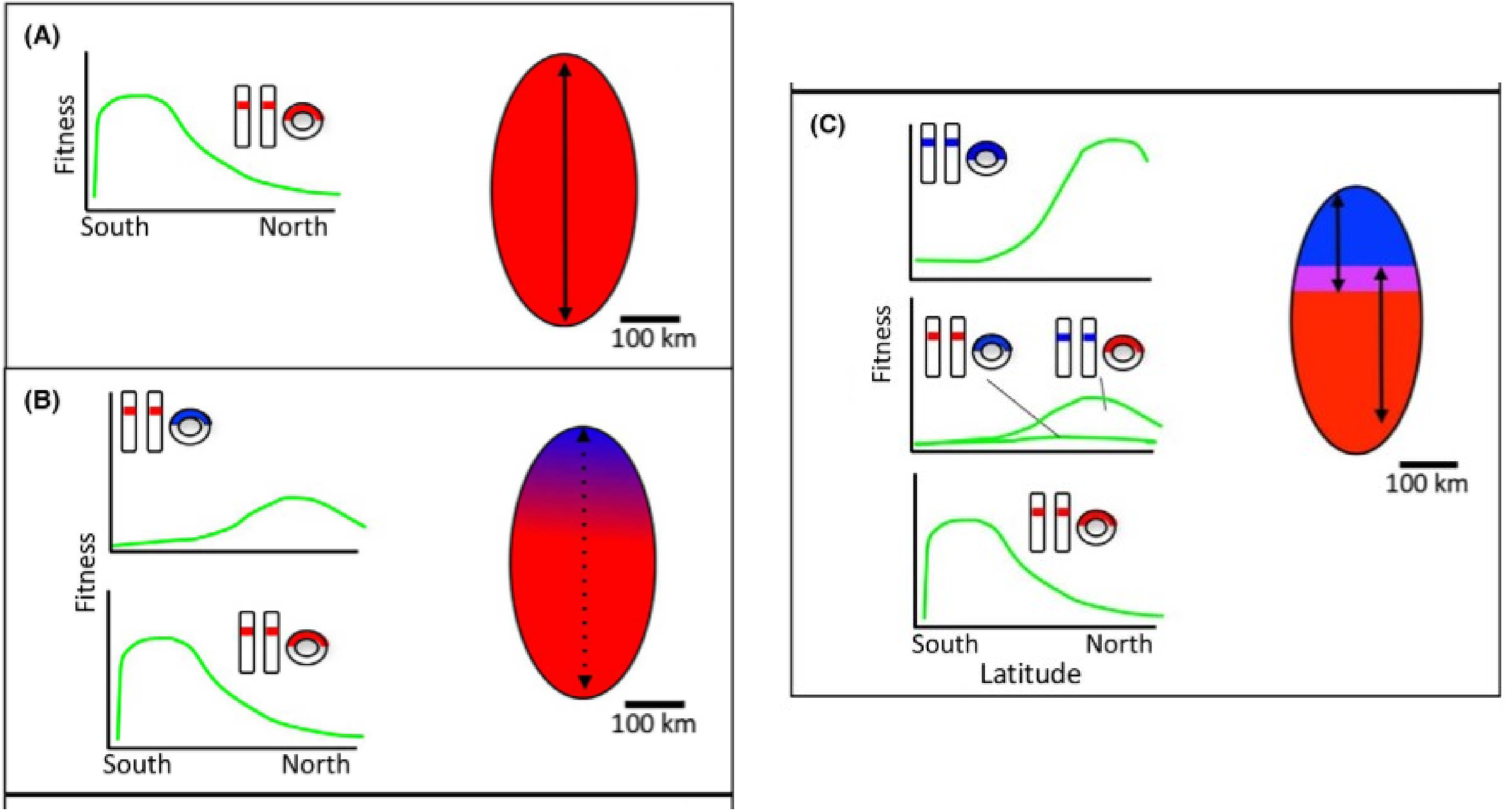
Rand et al. 2004

Co-adaptation of mt and N-mt genes



Hill 2019

Mitonuclear coevolution as the genesis of speciation

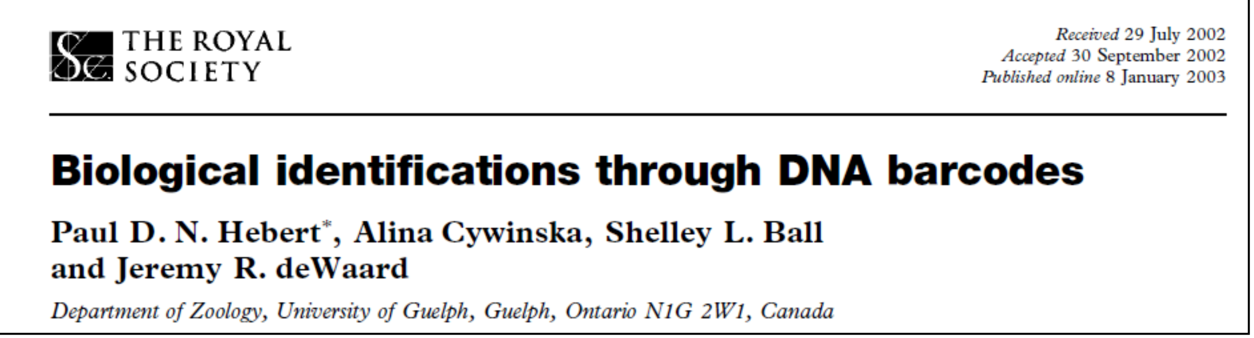


Hill 2016, 2017

Mitonuclear compatibility species complex

- = **a species is a population that is genetically isolated from other populations by incompatibilities in uniquely coadapted mt and N-mt genes**
- the need of mitonuclear coadaptation is **universal** among eukaryotes
- species boundaries become objective and defensible
- determining species boundaries would no longer be an esoteric intellectual exercise
- **„species“ is exclusively a eukaryotic concept**

Hill 2016, 2017, 2018, 2019



THE ROYAL SOCIETY

Received 29 July 2002
Accepted 30 September 2002
Published online 8 January 2003

Biological identifications through DNA barcodes

Paul D. N. Hebert*, Alina Cywinska, Shelley L. Ball and Jeremy R. deWaard

Department of Zoology, University of Guelph, Guelph, Ontario N1G 2W1, Canada

first idea in 2003

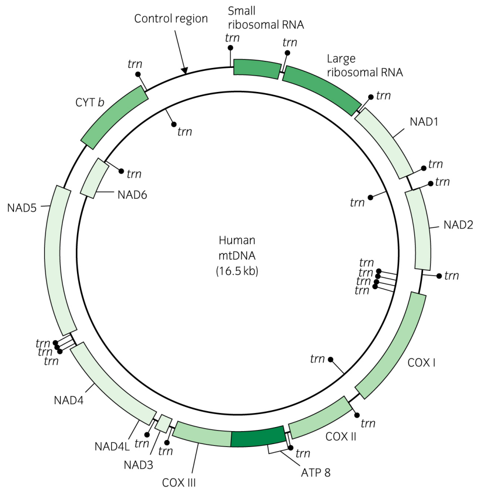
CONSORTIUM FOR THE BARCODE OF LIFE

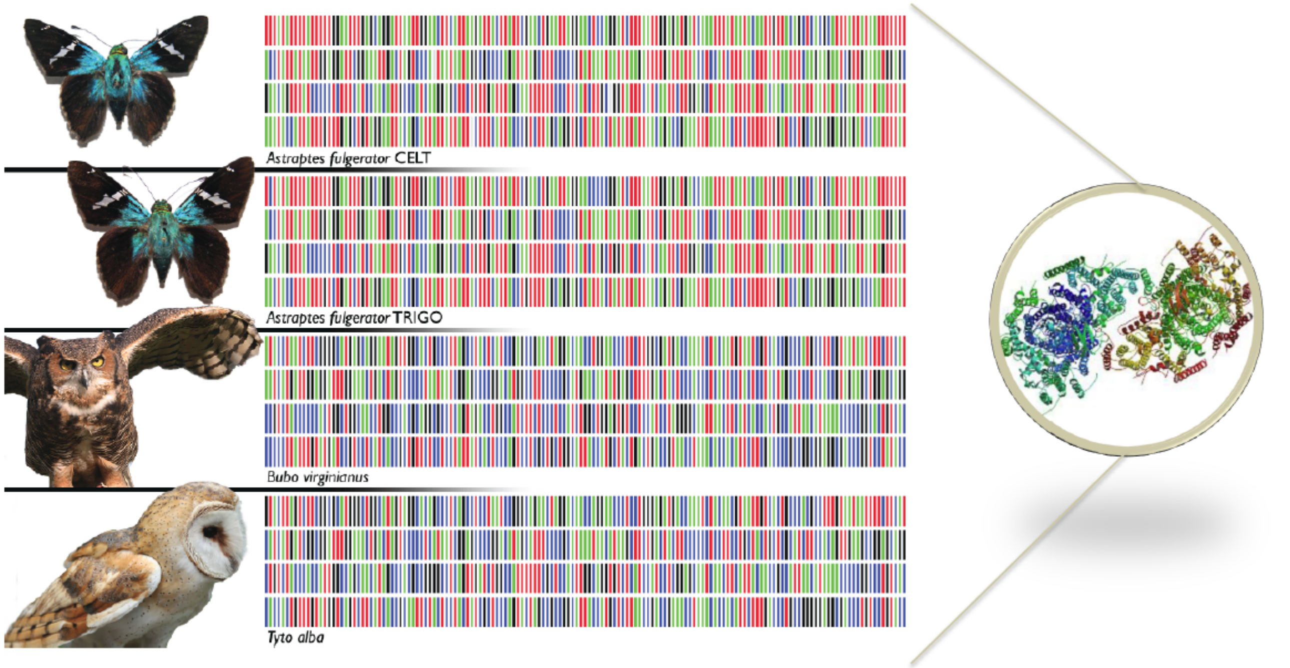
CBOL in 2005

international BARCODE OF LIFE

iBOL 2010-2015

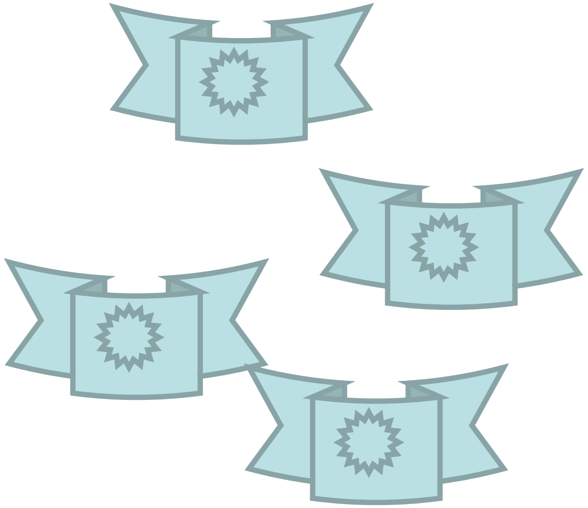
500 000 species barcoded in 2015

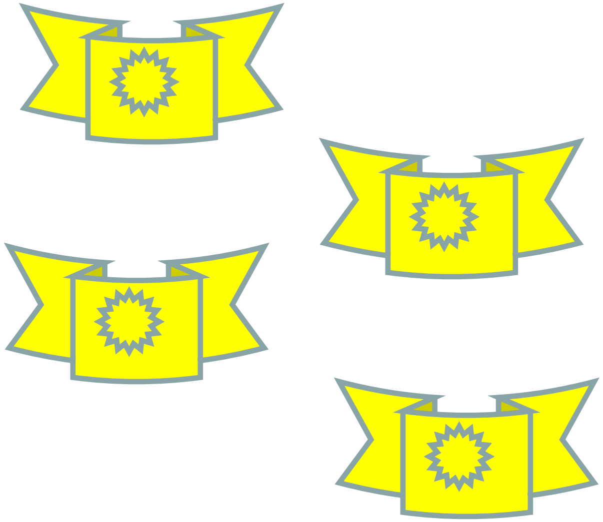




„DNA barcode“ – short fragment of mitochondrial DNA

Introgression/replacement of mtDNA in *Myotis*



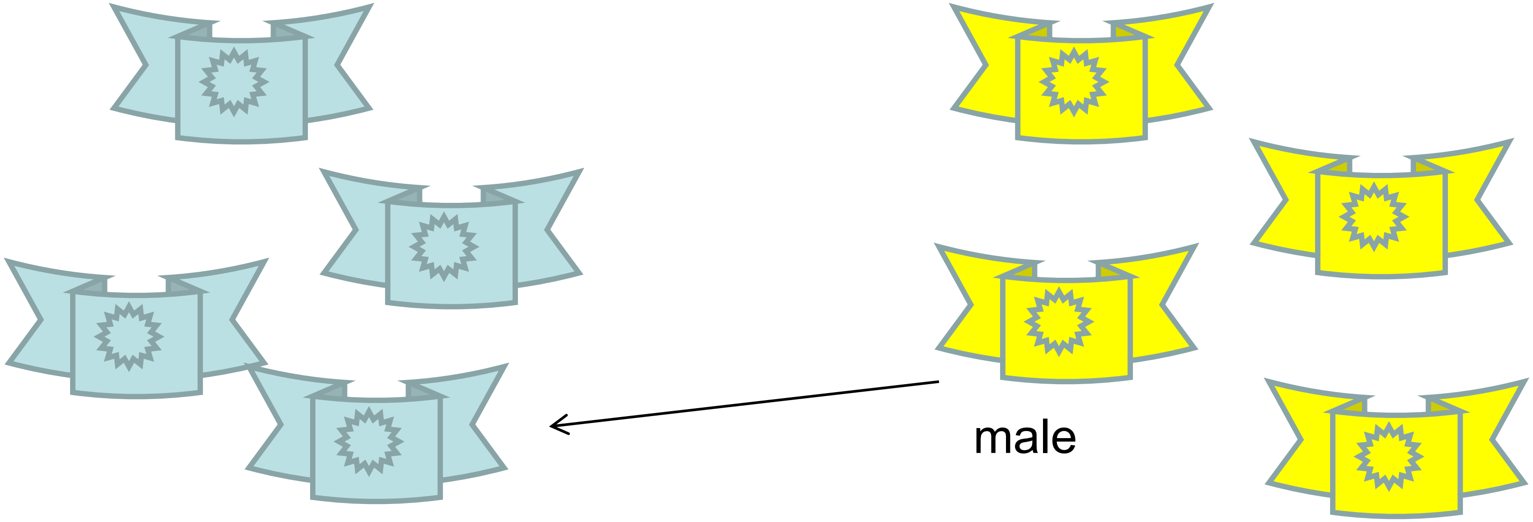


Myotis myotis - Europe

Myotis blythii - Asia

Berthier et al. 2006

Myotis blythii vs. *Myotis myotis* - mtDNA replacement

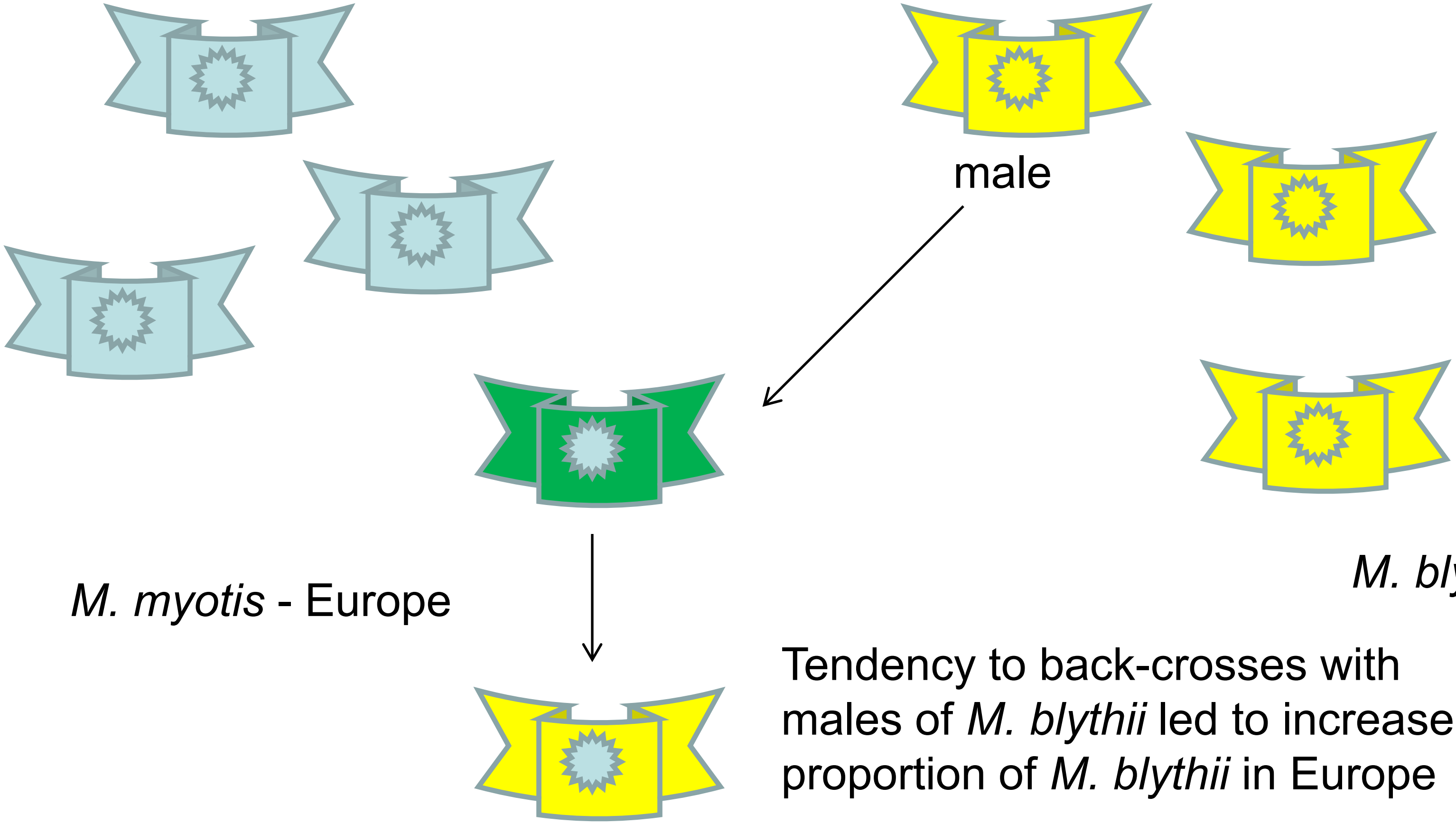


male

M. myotis - Europe

M. blythii - Asia

Myotis blythii vs. *Myotis myotis* - mtDNA replacement



male

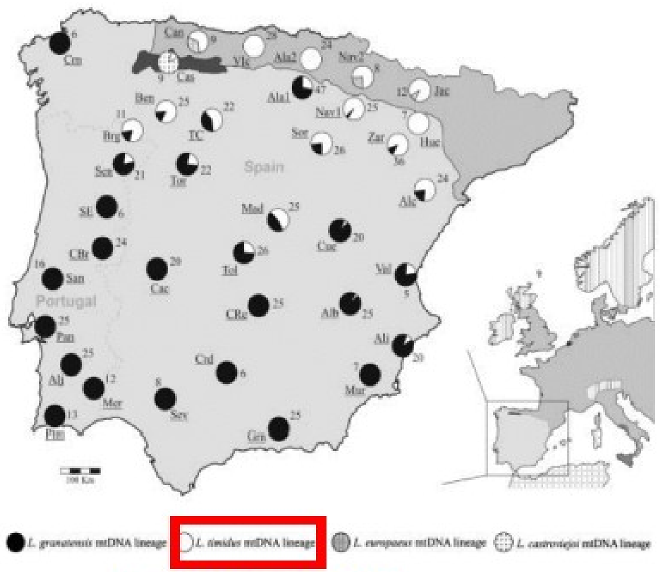
M. myotis - Europe

M. blythii - Asia

Tendency to back-crosses with males of *M. blythii* led to increase of proportion of *M. blythii* in Europe

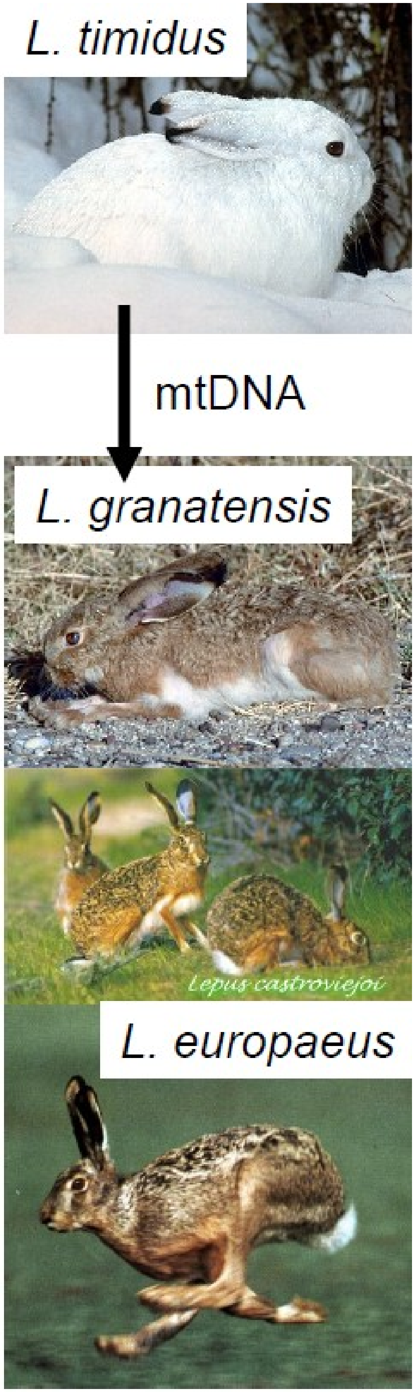
Colonizing (invasive) species often adopt mtDNA of original speices *(Currat et al. 2008)*

Hares in Spain and Portugal



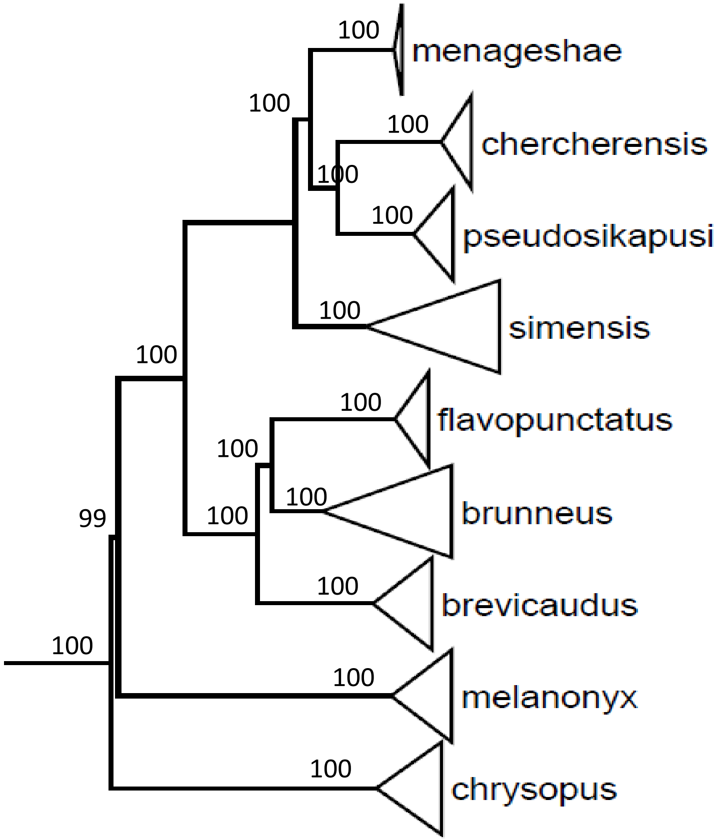
L. *timidus* mtDNA

- three species of *Lepus* in Iberia have often mtDNA of *L. timidus*
- but *L. timidus* dissappeared from Iberia at the end of last glacial period
- neutral process – consequence of spatial expansion



Interspecific mtDNA introgressions in *Lophuromys*

ddRADseq



menangeshae

chercherensis

pseudosikapusi

simensis

2n = 70

flavopunctatus

brunneus

brevicaudus

2n = 68

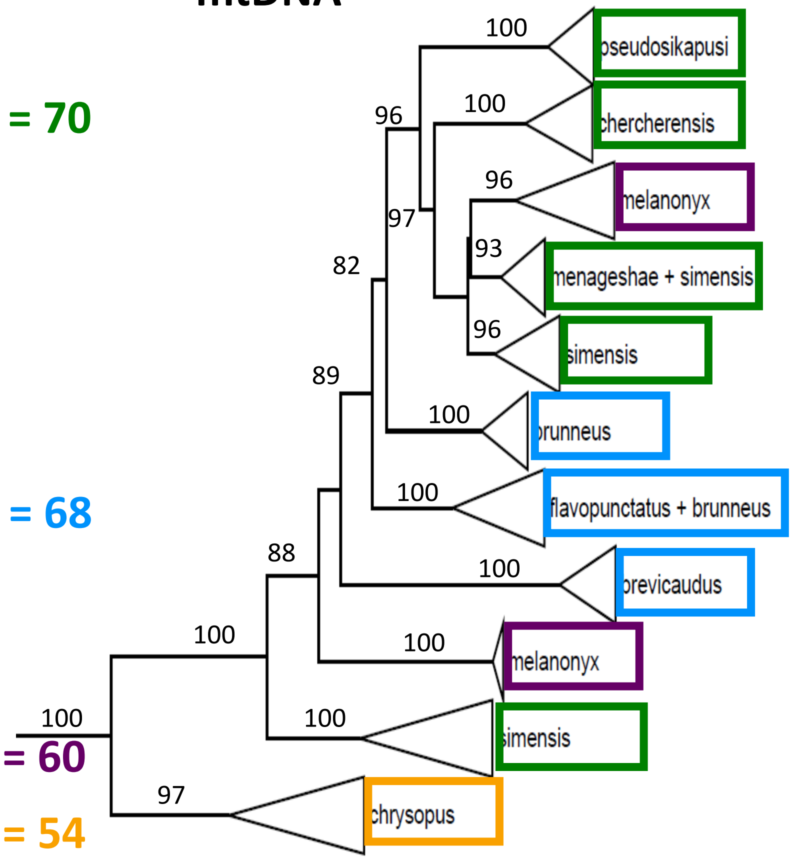
melanonyx

2n = 60

chrysopus

2n = 54

mtDNA



15 623 informative loci

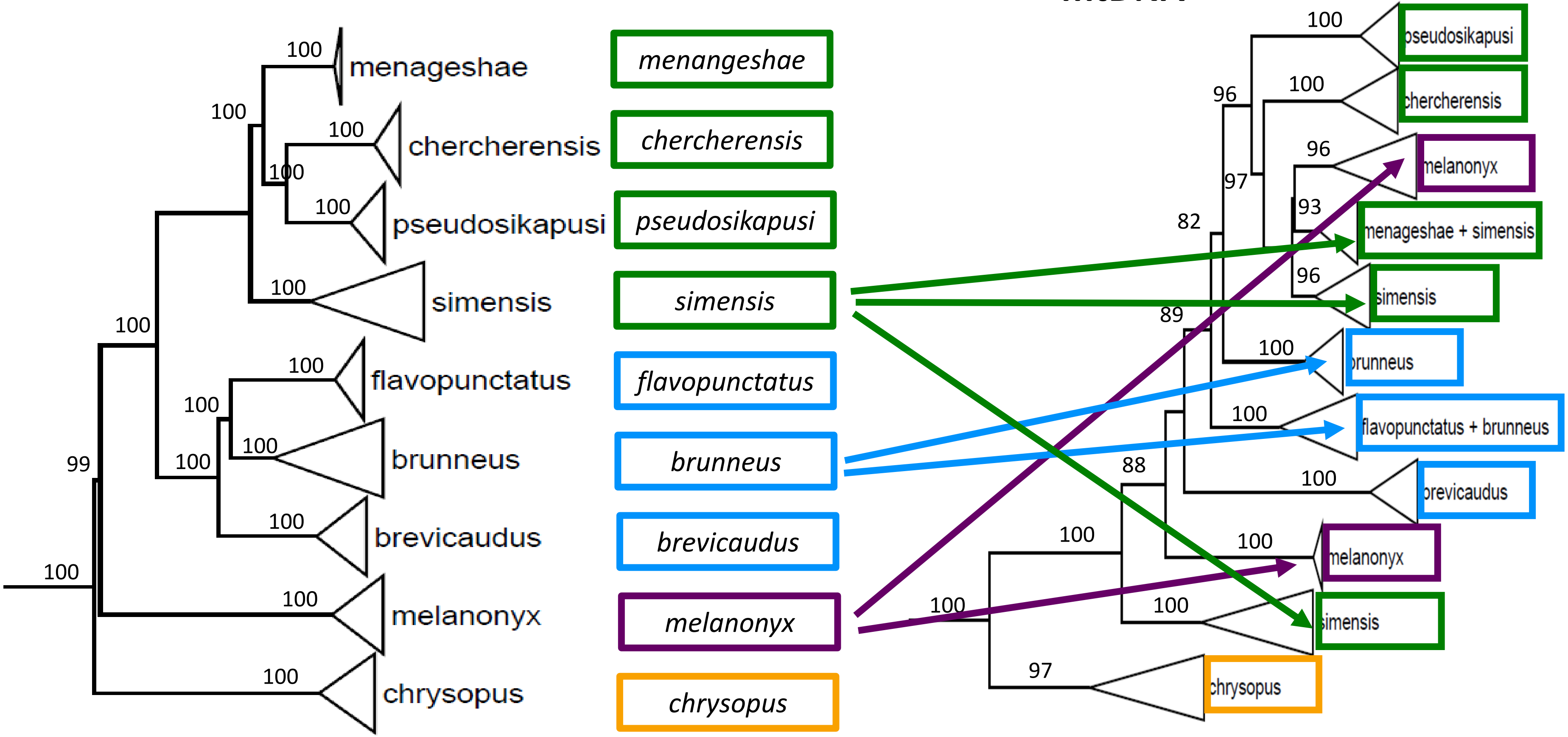
cytochrome *b* (1140 bp)

Komarova, Kostin, Bryja et al., Mol. Ecol., 2020

Interspecific mtDNA introgressions in *Lophuromys*

ddRADseq

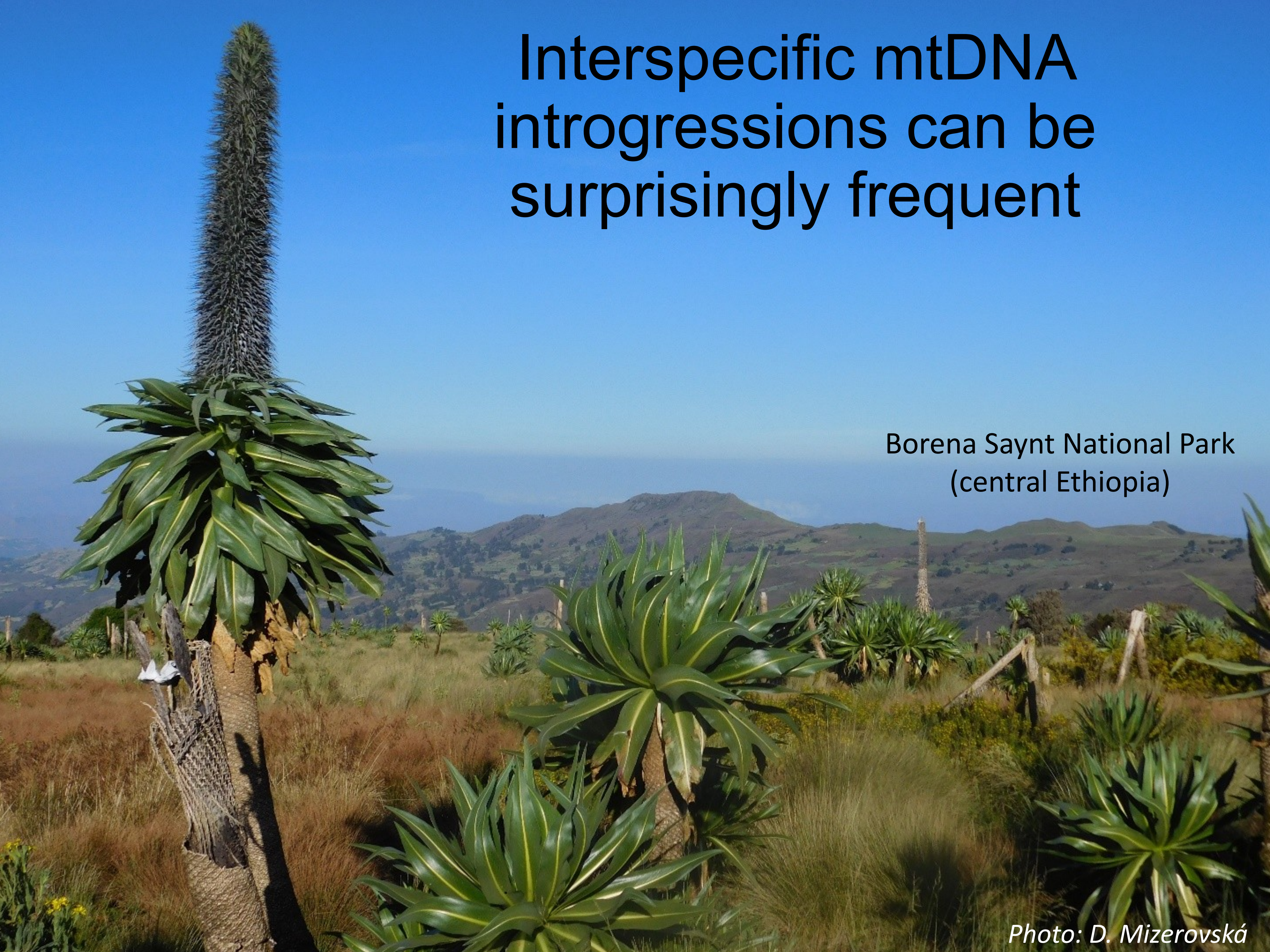
mtDNA



15 623 informative loci

cytochrome *b* (1140 bp)

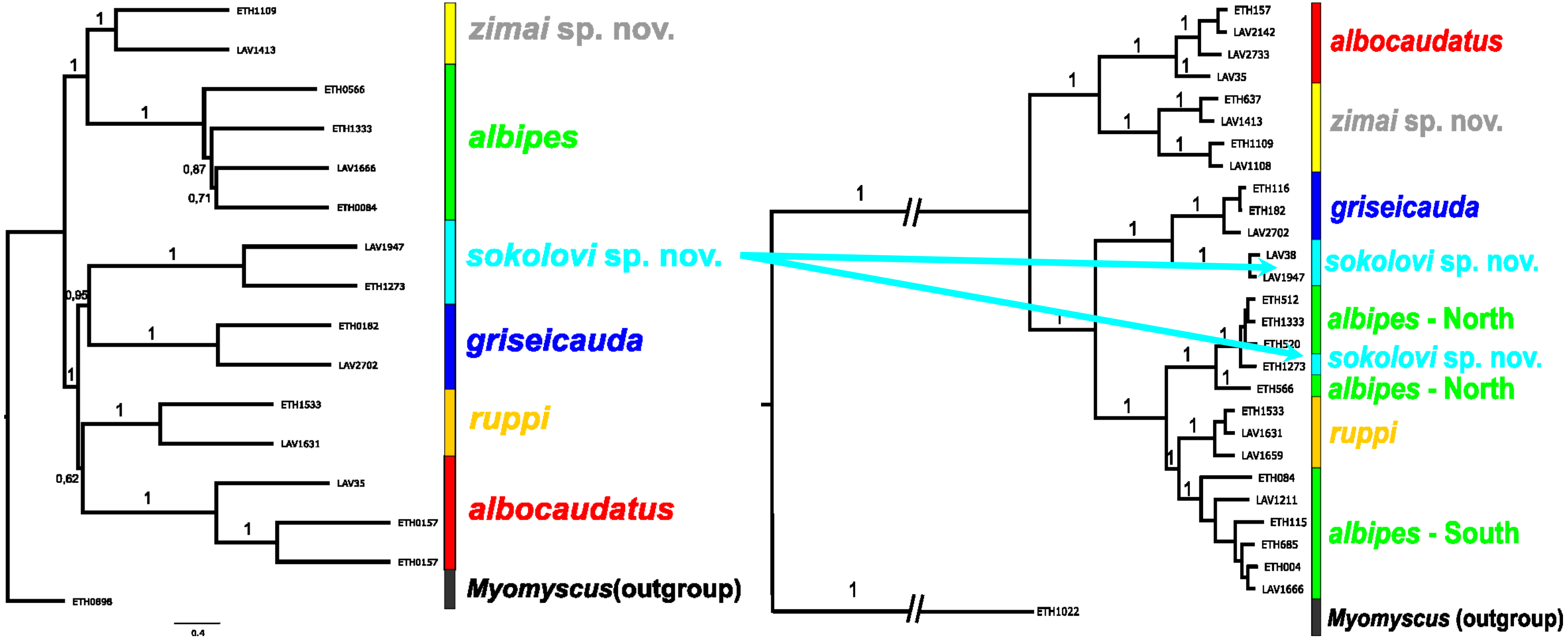
Interspecific mtDNA introgressions can be surprisingly frequent



Borena Saynt National Park
(central Ethiopia)

Photo: D. Mizerovská

Stenocephalemys (Ethiopian endemics)



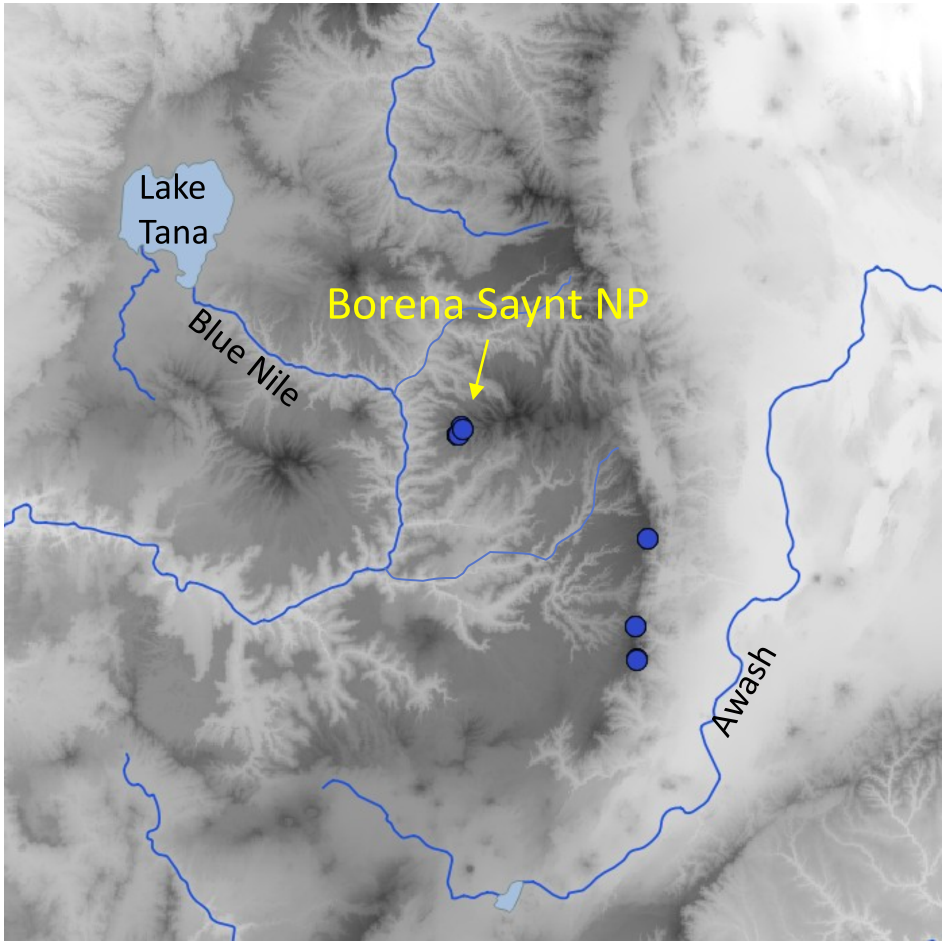
(A) anchored phylogenomics
(388 nuclear loci)

(B) complete mitogenomes

Mizerovská et al. 2020, J. Vert. Biol.

Borena Saynt NP

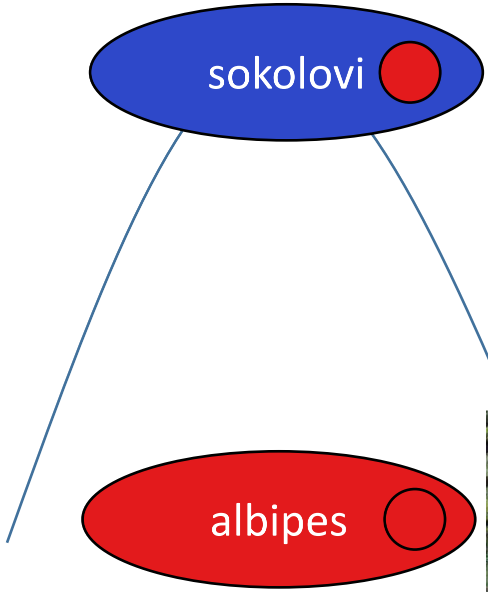
Stenocephalemys



● Distribution of *S. sokolovi* sp. nov.

Borena Saynt NP

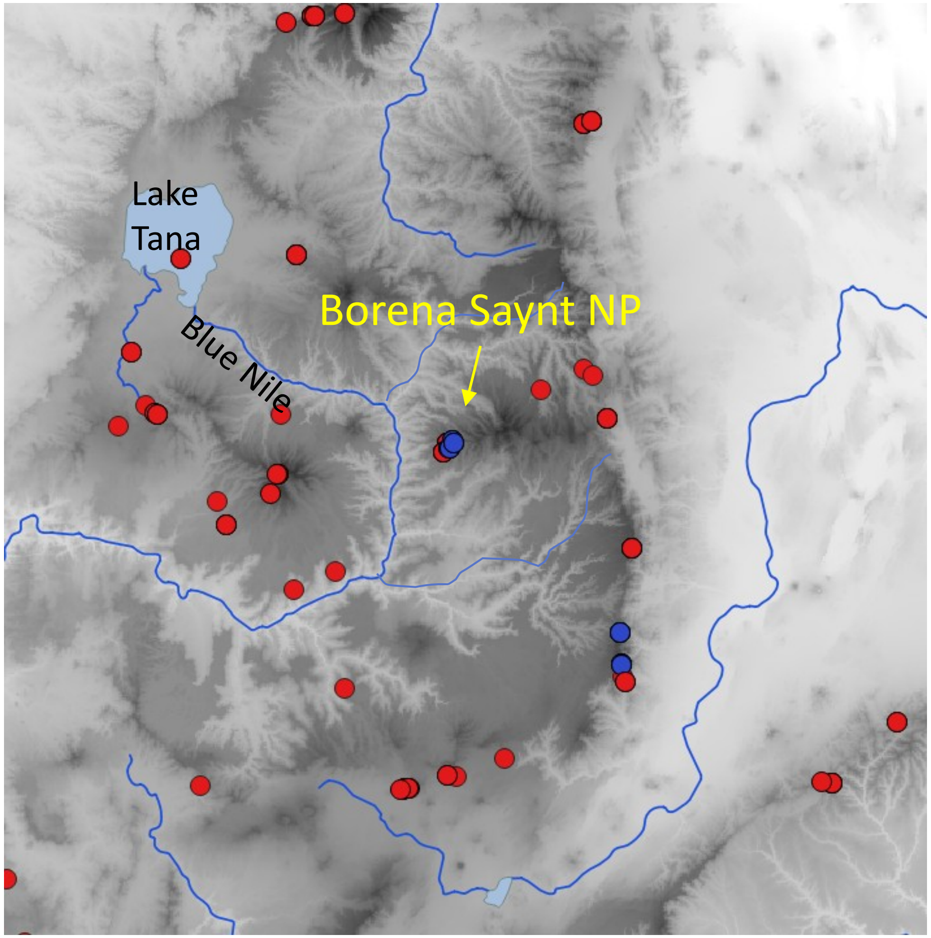
3500 m a.s.l.


sokolovi

albipes

2800 m a.s.l.

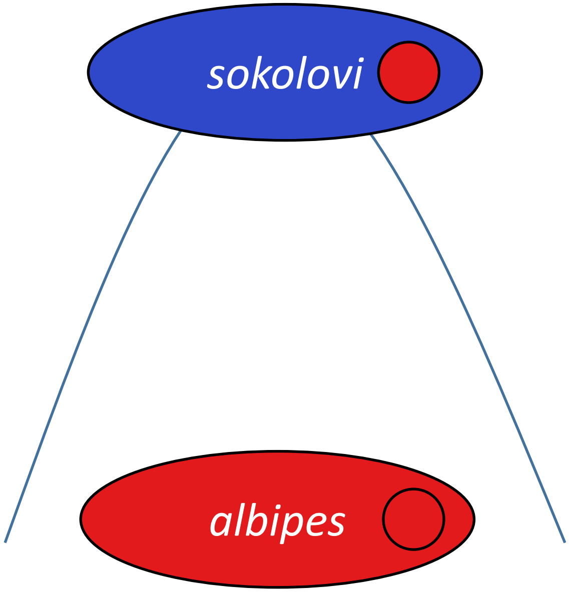
Stenocephalemys



- 🔵 Distribution of *S. sokolovi* sp. nov.
- 🔴 Distribution of *S. albipes*

Borena Saynt NP

anchored phylogenomic

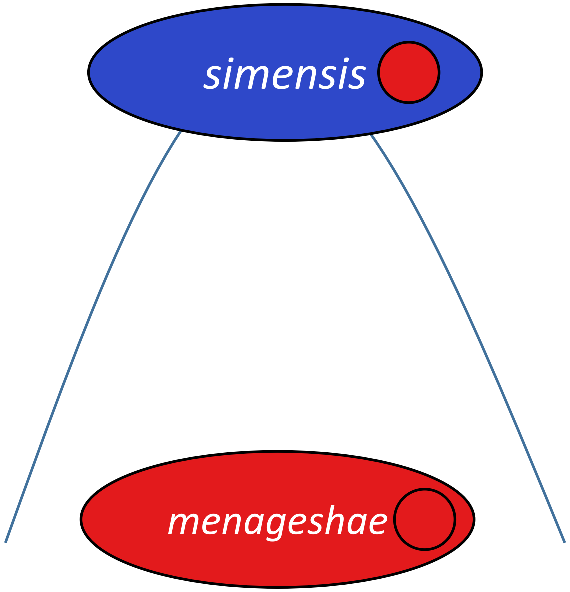


sokolovi

albipes

Stenocephalemys
(Mizerovská et al. 2020)

ddRADseq

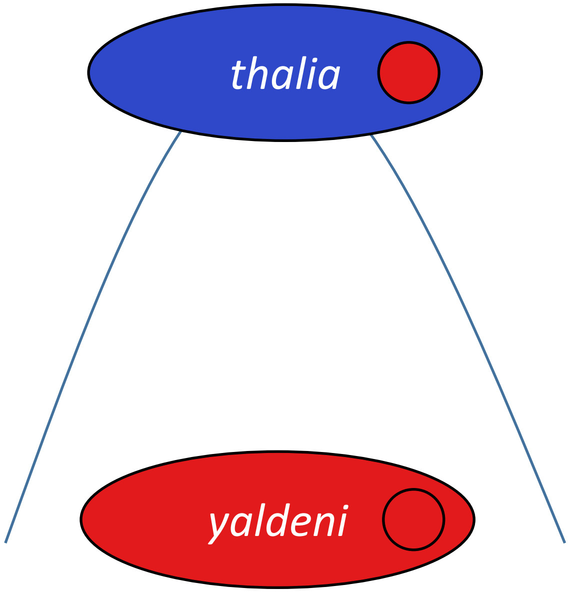


simensis

menageshae

Lophuromys
(Komarova et al. 2021)

morphology



thalia

yaldeni

Crocidura
(Konečný et al. 2020)

Borena Saynt NP, 3700 m a.s.l.

Is there somebody with „own“ mtDNA?

Borena Saynt NP, 3700 m a.s.l.

Otomys simiensis
Otomys typus

Arvicanthis abyssinicus

Dendromus sp. n. 1

Crocidura makeda n.sp.

Borena Saynt NP, 3700 m a.s.l.

Otomys simiensis

Otomys typus
(ddRADseq)

Arvicanthis abyssinicus
(ddRADseq)

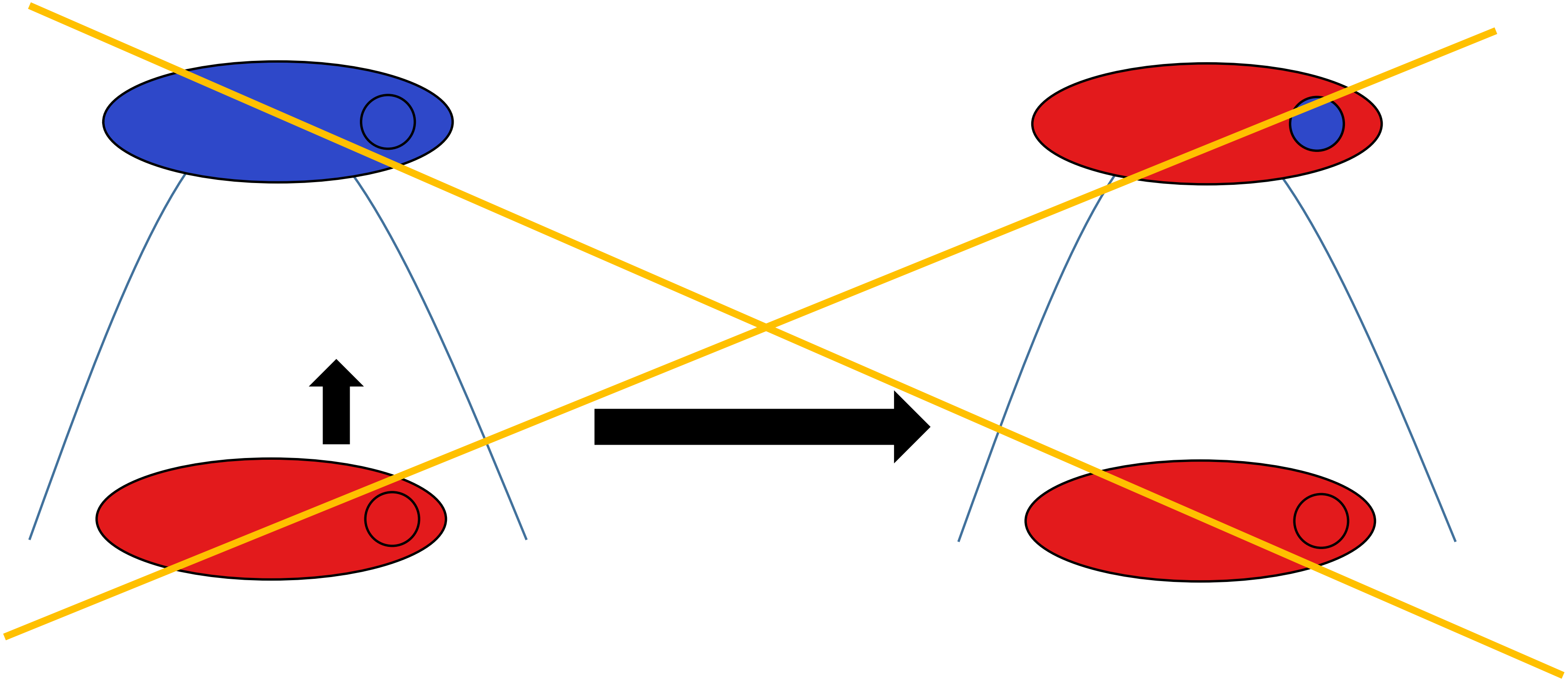
Dendromus
sp. n. 1

Crocidura
makeda n.sp.

Possible evolutionary explanations

Non-adaptive explanation

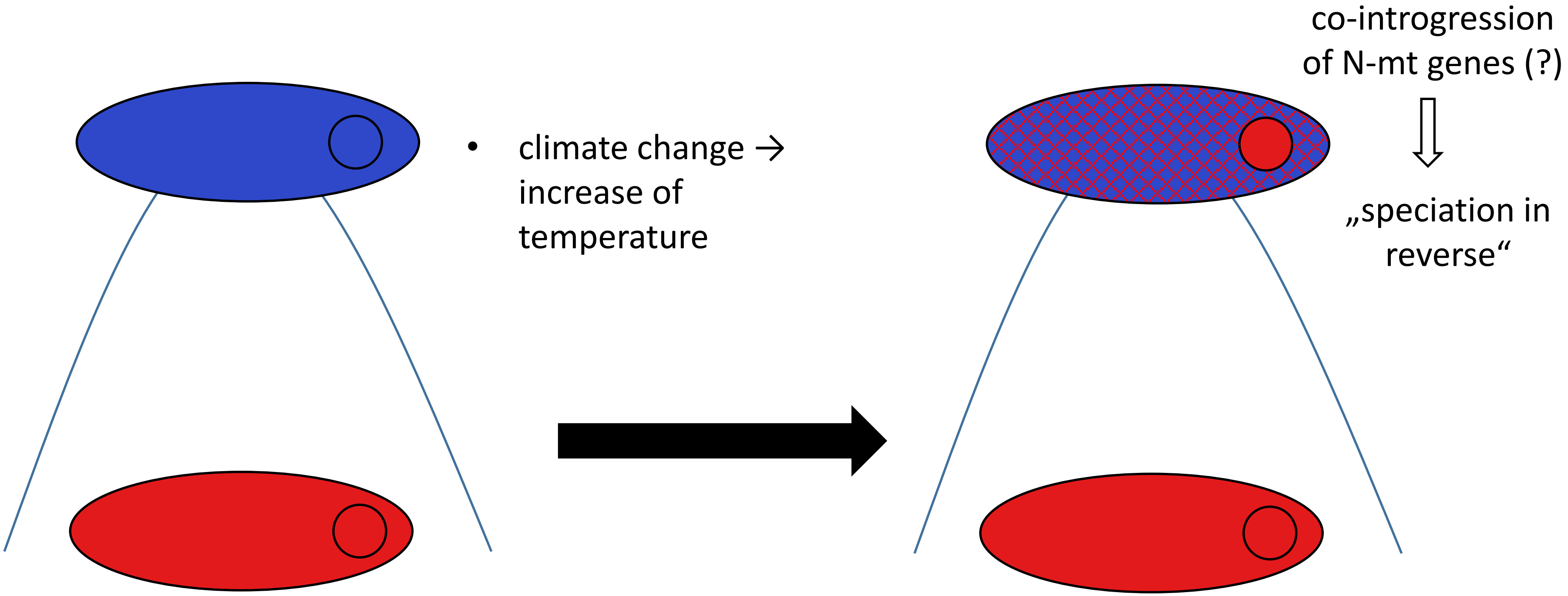
- rodents have male-biased dispersal – expanding males captures mtDNA of local species (see example with hares)
- expansion due to climate change → low-elevation species should move up and capture high-elevation mtDNA



Possible evolutionary explanations

Adaptive explanation no. 1

- advantageous OXPHOS genes of low-elevation taxon
- → studies of energetic metabolism and co-introgression are required



- climate change → increase of temperature

co-introgression of N-mt genes (?)

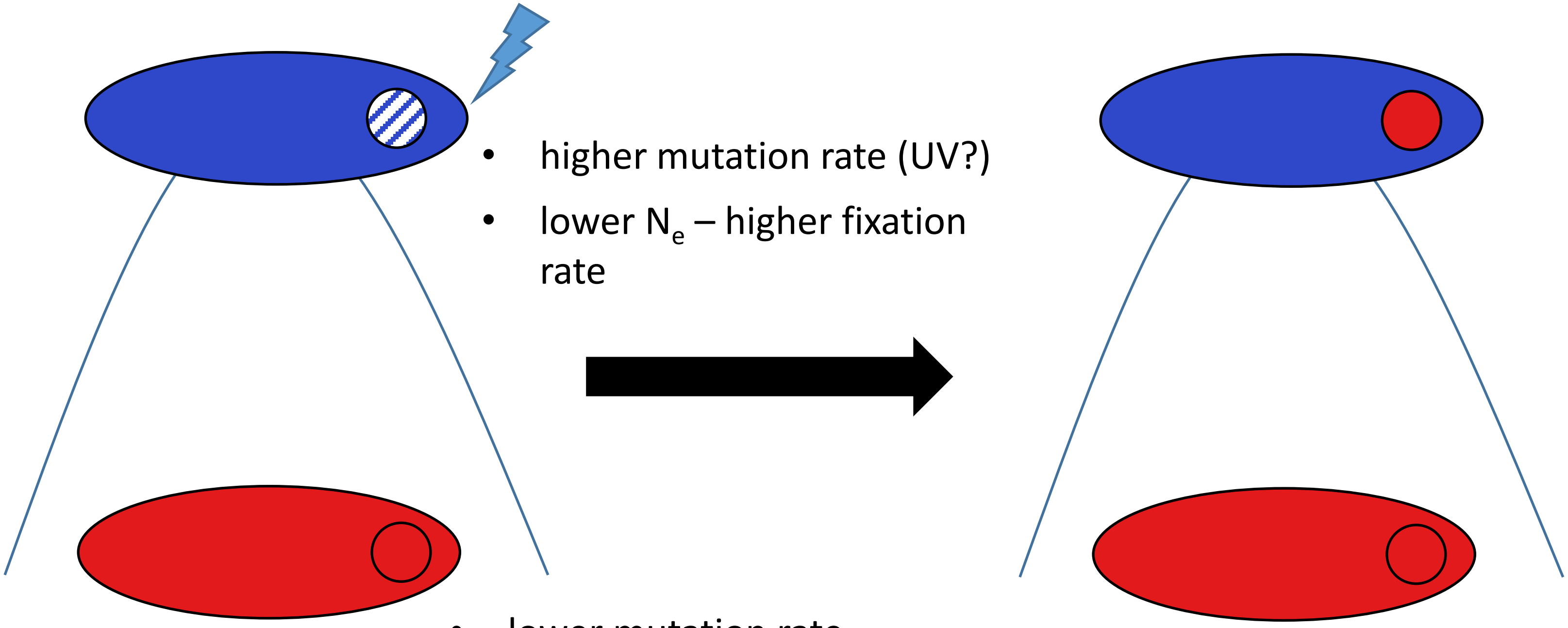
⇩

„speciation in reverse“

Possible evolutionary explanations

Adaptive explanation no. 2

- „mutational erosion“ - replacement of non-functional high-elevation mtDNA (accumulation of mutations in small populations in fragmented Afroalpine habitats) – „best of bad options“



- higher mutation rate (UV?)
- lower N_e – higher fixation rate

- lower mutation rate
- higher N_e

CONCLUSIONS

Conclusions

- species concepts and species delimitations are **crucial parts of biodiversity studies**
- enormous value in applied **conservation biology**
- **genomics can be useful** tool for species delimitation (analytical approaches still in development) – once they will be able to discriminate between population- and species-level structuring
- mitonuclear compatitibility species concept is a good candidate for **a unifying species concept**