Molecular identification of species, individuals and sex

Microsatellites

Pavel Munclinger
Species identification

- DNA barcoding
  - taxon identification using a standardized DNA region

Genebank (NCBI)

BOLDSYSTEM
http://v3.boldsystems.org/

BLAST or special programs
DNA barcoding
Hebert et al. 2003

The use of limited (approximately 600 bp) mitochondrial DNA sequence data as an inexpensive, easy way to “scan” and identify all of life.
... enabling the rapid and inexpensive identification of the estimated 10 million species on Earth. (Savolainen et al. 2005)

Database needed!

**Lepidoptera:** The All Leps campaign is assembling barcodes for 25,000 species of Lepidoptera, focusing on the faunas of Australia, Canada, Costa Rica and the United States. 7512 species barcoded | View

**Fishes:** The FISH-BOL campaign is gathering barcodes for all species of fishes (approximately 30,000) with an emphasis on the 15,000 marine species. 2538 species barcoded | View

**Canadian Fauna:** The Canadian Barcode of Life Network, incorporates an on-going initiative to barcode all Canadian species. 2133 species barcoded | View

**Birds:** The All-Birds Barcoding Initiative (ABBI) is assembling DNA barcodes for all 10,000 bird species, and aims to complete this task within 5 years. 1233 species barcoded | View
Figure 2. The predicted transmembrane structure for cytochrome c oxidase subunit 1 (COI). The area highlighted in yellow that includes five of the twelve transmembrane regions is coded for by the sequence designated by CBOL as the DNA barcode region.

Figure 1. A diagrammatic representation of a mitochondrial genome based on complete mitochondrial DNA sequences from a variety of bird species.
A. Catalogued species (Ntot=2 million)

B. Environmental OTUs (1430 18S V4 rDNA 97%)

- Metazoa
- Fungi
- Alveolata
- Archaeplastida
- Rhizaria
- 'Hacrobia'
- Amoebozoa
- Stramenopila
- Opisthokonta
- Excavata
- Streptophyta
Hesperiidae Costa Rica (Skippers)
The 28 last instar caterpillars
Lowenstein et al. 2009

Sushi

albacore
*Thunnus alalunga*

bluefin tuna
*Thunnus maccoyii*

critically endangered

escolar
*Lepidocybium flavobrunneum*

health concerns
DNA barcoding – parasites

Avian malaria
*Haemoproteus, Plasmodium, Leucocytozoon*

DNA from bird blood, parasite-specific primers → parasite lineages
Alcaide et al. 2009
Arthropod bloodmeal

Culex pipiens
Leopard cat in Pakistan
*Prionailurus bengalensis*

- Shezhad et al. 2012
- 12S rRNA
- Illumina
Metagenomics

- Metagenomics: DNA sequencing of environmental samples
- Water (sea, pool), clouds…
- Reconstruction of ancient vegetation (permafrost)
- Earthworm extracellular DNA from soil…
- Gut microbiome
intraspecific divergence << Interspecific divergence

Figure 1. Comparison of Nucleotide Sequence Differences in COI among 260 Species of North American Birds

Figure 3. Nucleotide divergence in a 617 bp segment of the COI gene in five lepidopteron families at species, genus and family level. Data from Hebert PDN, Cywinska A, Ball SL, deWaar JR. 2003a Proc R Soc Lond B 270:313–322
Barcoding gap
Cypraeidae
Meyer & Paulay 2005
When barcoding fails: Genome chimerization (admixing) and reticulation obscure phylogenetic and taxonomic relationships

*Metschnikowia*

(a) D1/D2 domain sequences. (b) ITS1–5.8S–ITS2 sequences
Hares in Spain and Portugal

- *L. timidus* mtDNA in *Lepus granatensis*, *L. castroviejoi*, and *L. europaeus*

- *L. timidus* retreated from this region at the end of the last ice age

- Similar situation in bats, newts, fish...
Introgression from local into the invading species
DNA Sequence from Cretaceous Period Bone Fragments

Scott R. Woodward,* Nathan J. Weyand, Mark Bunnell

DNA was extracted from 80-million-year-old bone fragments found in strata of the Upper Cretaceous Blackhawk Formation in the roof of an underground coal mine in eastern Utah. This DNA was used as the template in a polymerase chain reaction that amplified and sequenced a portion of the gene encoding mitochondrial cytochrome b. These sequences differ from all other cytochrome b sequences investigated, including those in the GenBank and European Molecular Biology Laboratory databases. DNA isolated from these bone fragments and the resulting gene sequences demonstrate that small fragments of DNA may survive in bone for millions of years.

Consensus

15,627

15,627

Consensus

<table>
<thead>
<tr>
<th>Consensus</th>
<th>15,627</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human numt</td>
<td>→</td>
</tr>
</tbody>
</table>
numts and DNA barcoding
Song et al. 2008
Retension of ancestral polymorphism

Lineage sorting
→ reciprocal monophyly
... more and more problems

<table>
<thead>
<tr>
<th>Problem</th>
<th>Consequence</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Failure to test clear hypotheses</td>
<td>Choice of inappropriate or suboptimal analytical method due to confusion as to the objectives of the study</td>
<td>Explicitly state each hypothesis, and for each distinct aspect of the study present separate headings in methods and results sections</td>
</tr>
<tr>
<td>Inadequate a priori identification of specimens</td>
<td>Conflicting identifications made by different labs can compromise the effectiveness of reference libraries that are ultimately used as a resource for scientific or regulatory purposes</td>
<td>Present a bibliography of references, as well as the distinguishing morphological characters used in the identification process. Follow recommendations outlined by Steinke &amp; Hanner (2011)</td>
</tr>
<tr>
<td>The use of the term 'species identification'</td>
<td>Confusion between identification of individuals, and delimitation/discovery of species</td>
<td>To clarify objectives, use the term 'specimen identification' or 'species discovery' where appropriate</td>
</tr>
<tr>
<td>Inappropriate use of neighbour-joining trees</td>
<td>(a) Relying on strict monophyly for identification can reduce the apparent effectiveness of DNA barcoding as an identification tool. This can be due to either mtDNA paraphyly or misidentification of specimens. (b) For biodiversity assessment and species discovery, NJ trees cannot estimate the number of species independently with respect to the taxonomic names</td>
<td>(a) Alternative criteria such as 'best close match' are readily available, and have higher rates of identification success. This method can be implemented using the free software packages TaxonDNA (Meier et al. 2006) or Spider (Brown et al. 2012). (b) Estimate species richness using ABDG (Puillandre et al. 2012), GMYC (Monaghan et al. 2009) or BOLD’s BIN system (<a href="http://v3.boldsystems.org">http://v3.boldsystems.org</a>)</td>
</tr>
<tr>
<td>Inappropriate use of bootstrap resampling</td>
<td>For specimen identification purposes, bootstrap resampling can further reduce the already low identification success rates associated with NJ trees</td>
<td>Only use bootstrapping where appropriate: e.g., as part of a species delimitation process on preestimated groups</td>
</tr>
<tr>
<td>Inappropriate use of fixed distance thresholds</td>
<td>For specimen identification purposes, a generic threshold which is set too low or high can reduce or bias identification error rates</td>
<td>Thresholds can now be optimized for specific data sets using the method of Virgilio et al. (2012), or with software such as ABDG (Puillandre et al. 2012) and Spider (Brown et al. 2012)</td>
</tr>
<tr>
<td>Incorrectly interpreting the barcoding gap</td>
<td>Overlapping distributions of intra-/interspecific distances do not necessarily mean that barcodes perform poorly for identification</td>
<td>For specimen identification studies, dotplots of intra-/interspecific distances are a better way to illustrate the barcoding gap (e.g. Robinson et al. 2009)</td>
</tr>
</tbody>
</table>
Species delimitation

• **Distances**
  – ABGD
  – (Automatic Barcode Gap Discovery)

• **Phylogenies (trees)**
  – „minimal phylogenetic units“ (OTUs)
  – BPP (Bayesian Phylogenetics and Phylogeography)
  – GMYC, PTP…
**Figure 6.** Summarized results from all species delimitation analyses, represented on a *BEAST* tree resulting from analysis of all sampling locales. Colors correspond to recovered groupings, with each partition represented by a unique color. Colored bars above phylogeny represent hypothesized species groupings based on multiple analyses. Insert image of an adult male *Allatopus starretti*, sp. nov. (Kern Co., Poso Flat Road).
Molecular identification of individuals

• microsatellites
Repeats

- more than half of our genome
- Interspersed (transposony)
- Tandem (mini- a mikrosatellites)
- Minisatellites (longer motifs: 10 - 100 nucleotides)
- Mikrosatellites (1 - 6 nucleotides motifs)
- Human - 700 000 msats, 3 % of genome
• Mutation rate $10^{-3(2)} \cdot 10^{-7}$
• Male germline > female germline
• Slippage

**Figure 1.** Dependence of microsatellite mutability on repeat number (A) and length (B). Mutability is per locus per generation. The bands around the curves indicate the 2.5th and 97.5th percentiles of empirical distributions obtained through a resampling procedure (see Methods). Only points with at least 30 microsatellites are plotted.
• Simple Mendelian inheritance

• Highly variable

• Paternity, population structure…
Fluorescent labels

primer

primer

primer

primer
Multiplex

More loci in one reaction
Primers for new loci: traditional approaches
genome library, hybridization with probe, sequencing

1. Extract total DNA and cut with restriction enzyme
2. Gel-isolate 300-600 bp fragments
3. Ligate fragments into phagemid vector, clone into E. coli, probe with labeled repeat sequence (e.g., GT₆), select positive clones
4. Sequence flanking regions, make PCR primers
5. PCR-amplify each microsatellite locus
6. Score Mendelian polymorphism

Figure 3.18 General protocol for microsatellite assays.

Figure 3.10. General protocol for DNA cloning and genomic library construction.
Alternative – NGS

pyrosequencing

Illumina
Cross-species amplifications

Some msats are surprisingly conserved, but…
Pyrenean brown bears

*Taberlet et al. 1997*

- Faeces, hairs
- 24 msats
- 4 males a 1 female
- Otters, faeces
- →Molecular scatology

Fig. 3 Home range of two Pyrenean brown bears obtained by noninvasive genetic sampling and genotyping.
allelic dropout

allelic dropout and false allele.

allelic dropout

allelic dropout

allelic dropout

→ multiple tube approach
• Clonality

• Genetic chimeras
- Rotifera – Bdelloidea
- Ostracoda (*Darwinula*)
- Partenogenetic clones

*Darwinula stevensoni*
Gynogenesis

Ambystoma – genome of 4 species
Meiotic apomixis

suppression of first division

suppression of second division

no crossing over

1 crossing over

Meiotic apomixis with suppression of the first or second division, with or without recombination

Parthenogenesis

Gynogenesis

Automixis

no crossing over

1 crossing over

cleavage

sister

non-sister

Automixis with fusion of cleavage nuclei, sister nuclei, or nonsister nuclei, with or without recombination
Premeiotic chromosome doubling
Genetic chimeras
Deep sea fish
• marmosets and tamarins
• *Callithrix jacchus* (also *Saguinus*)
• Dizygotic twins
• Hematopoietic chimeras
Canine transmissible venereal sarcoma (CTVS)

Sarcophilus harrisii
Microchimerism
Chang and Eng
Siam twins
born 1811

Fusion of embryos
(heteropaternal superfecundation 2.4%)

→ genotypes of ovaries and somatic tissues may differ

Lydia Fairchild

Nikita from Irkutska

Blaschko’s lines
Lydia Fairchild
Molecular sexing – birds

- CHD1W a CHD1Z (Griffith et al. 1998)
- ATP5A1Z a ATP5A1W (Bantock et al. 2008)
- Genes on sex chromosomes
- Primers amplify introns of both genes
- Introns may differ in length
Molecular sexing - mammals

- Y (Sry) (duplex PCR with an autosomal or X gene)
- Sry DNA-binding motif (HMG box)

- *Microtus cabrerae*
  Sry on Chr X

- *Ellobius, Tokudaia*
  Sry is missing

- *Dicrostonyx torquatus, Mus minutoides*
  feminizing X* → X*Y females

Bryja a Konečný 2003

D. torquatus
M. cabrerae
Tokudaia osimensis
Molecular sexing applicable in 4000 species of lizards and snakes? From dream to real possibility

Michail Rovatsos and Lukáš Kratochvíl

Department of Ecology, Faculty of Science, Charles University in Prague, Vídeňská 1, 12844 Prague, Czech Republic

![qPCR Graph](image-url)
Gynandromorphs

Double fertilization of binucleate eggs

Loss of the W

ZW / ZZ