



# Metodologie molekulární fylogeneze a taxonomie hmyzu

## Bi7770

Andrea Tóthová



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

# Dva pohledy genomiky

- Vertikální: kompletné genomy, hluboké, ale omezené informace
  - několik druhů a jedinců
  - úspěšně ukončeno u člověka, Drosophily a některých plodin
- Horizontální: krátké cílené sekvence, plytké, ale široké informace
  - mnoho druhů a jedinců
  - např. DNA barcoding

# Historické výzvy

- Problémy s konceptem druhu a jeho aplikacemi
- Problémy s druhovou identifikací
- Systém znaků – morfologie, genetika, atd.
- Přístup k existujícím informacím
- Snižování odbornosti
- Snižování dostupných služeb
- Genomika a Internet nabízí nové možnosti

# Asociace vývojových stádií, zpracovaných částí a dimorfických pohlaví

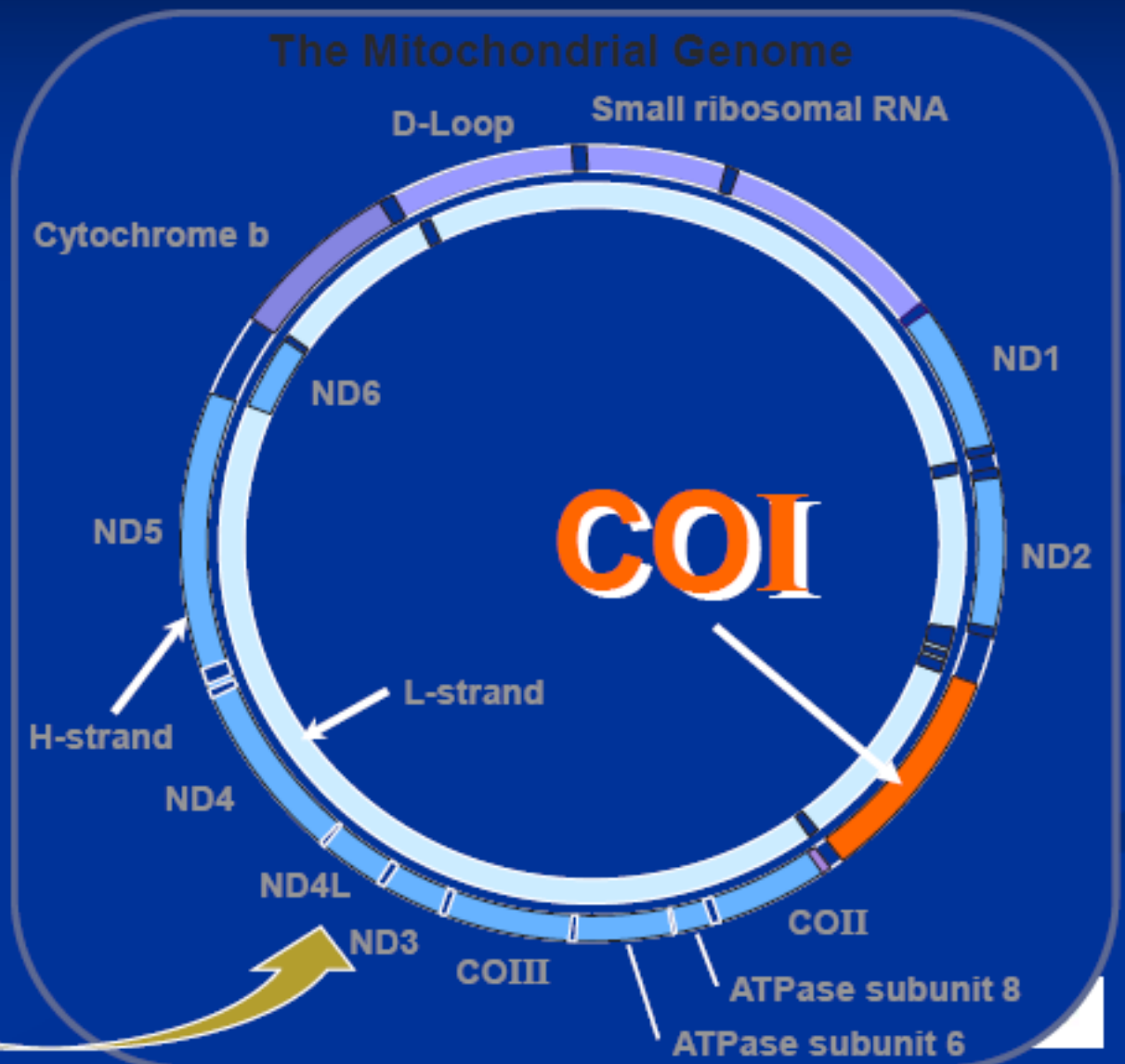


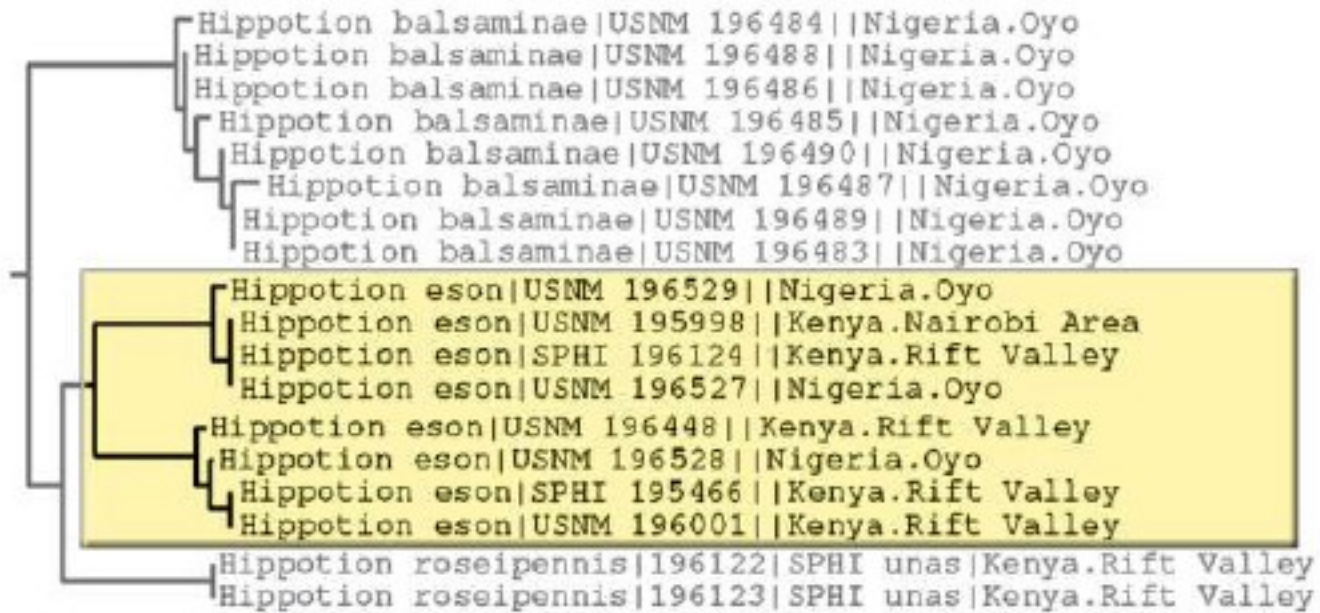
# System čárového kódu

- **DNA barcode je krátká genová sekvence standardizované části genomu použitá k druhové identifikaci**



# Interné ID pro všechny organizmy na Zemi





*Hippotion eson*

←1cm→



USNM slide 125245



USNM slide 125247

# Příklad: Afriční motýli

Tvoří species komplexy

Jejich parazitoidi (Tachinidae) také  
(Dittrich et al 2006)

*Leguminivora ptychora* na luštěninách je  
také species komplex

Ale někteří škůdci jsou široko rozšíření,  
např. *Spoladea recurvalis* a *Maruca vitrata*



# Rychlé a efektivní

- Čeleď Sphingidae – vzorky 49 druhů za 6 měsíců (téměř kompletní lokální fauna)
- DNA barcoding rozlišil druhy jak v lokálním, tak v globálním měřítku
- Místní knihovny mohou být rychle srovnány a přispět ke globálním knihovnám

# Jak Barcoding funguje

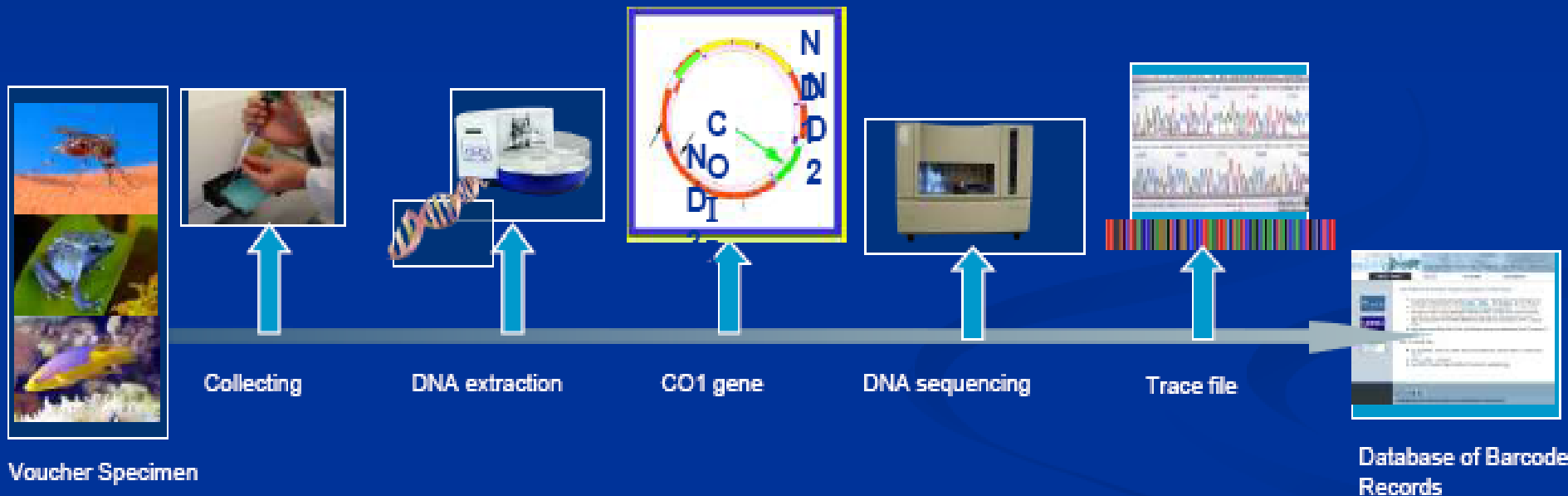
Tvorba referenční knihovny:

- Správně určený jedinec (vouchers)
- Vzorek tkáně
- DNA extrakce, PCR amplifikace
- DNA sekvenování
- Odeslání dat do GenBanku

Použití referenční knihovny :

- Neurčené druhy
- Tkáň, DNA, sekvenování
- Srovnání s referenčními sekvencemi

# Jak se to vše děje od jedince přes sekvenci po druh?



# Produkce dat v r. 2007



ABI 3100 sekvenátor

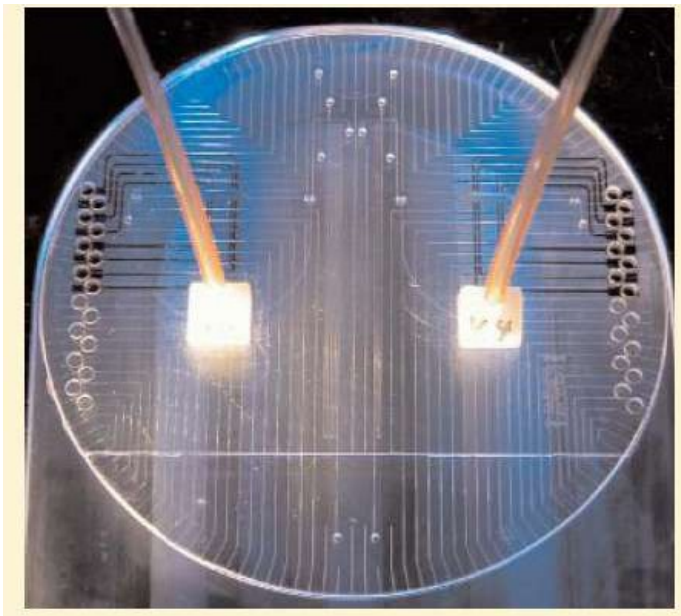
Stovky vzorků denně, cena od několika centů po dolary

PCR amplifikační jednotka



# Produkce dat v r. 2008

Rychlejší a přenosnější systém – stovky vzorků za hodinu



Integrované DNA mikročipy



Stolní mikrofluidné systémy

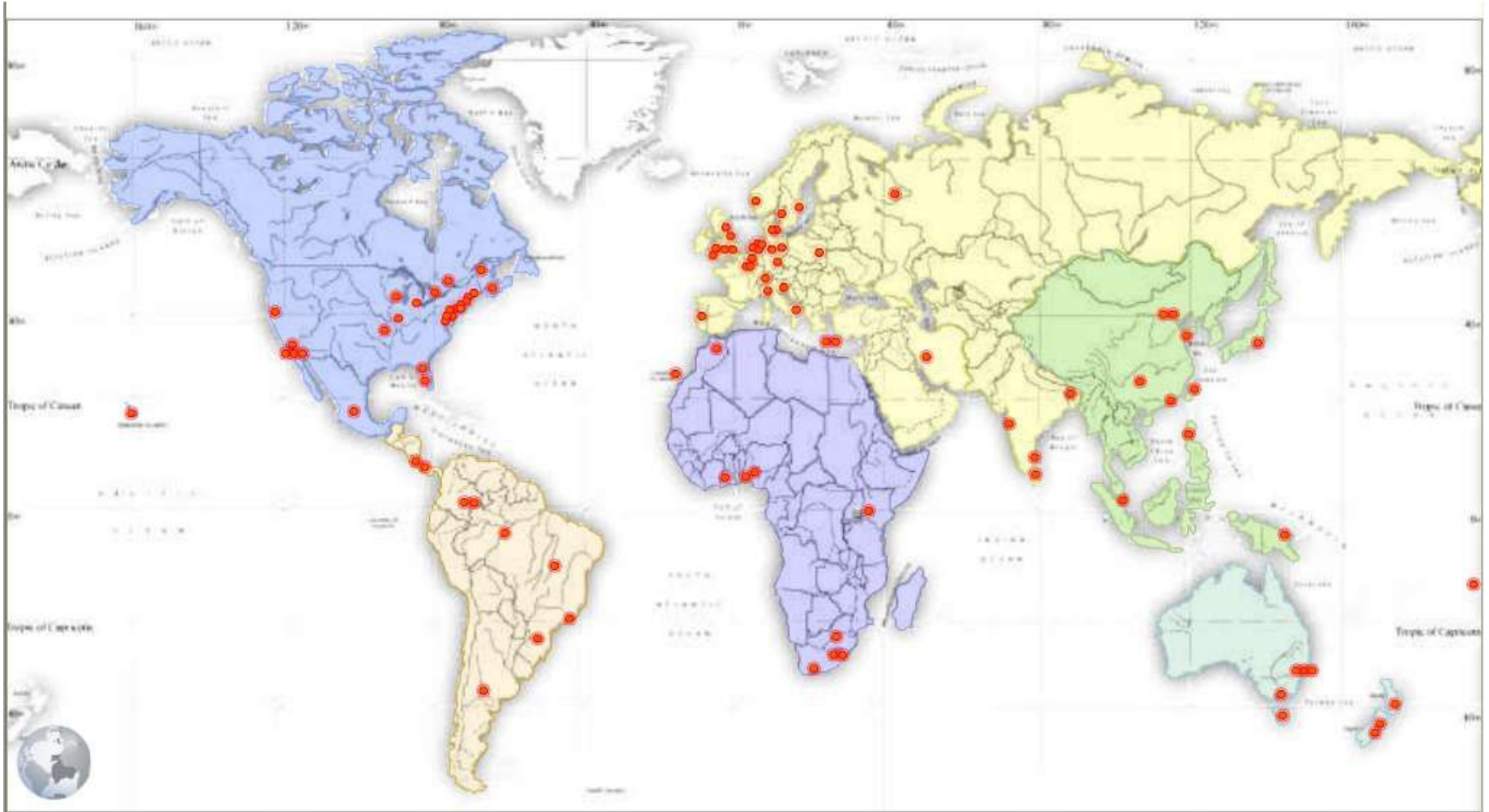
# Produkce dat v budoucnosti?



- Získání dat kdekoli, hned
- Cena několik haléřů
- Link do referenční databáze
- Taxonomická GPS
- Použitelné nespecialisty

# CBOL – organizace členů od r. 2008

Více než 170 organizací z více než 50 zemí (z toho 54 organizací z 20 rozvojových zemí)



# Mise CBO Lu: uvést DNA Barcoding jako globální standard


1. Vyvinout a zvednout standardy komunity
2. Barcode projekty plnit databáze
3. Globální participace a koordinace
4. Přijetí taxonomickou komunitou
5. Koordinace s jinými oblastmi vědy
6. Přijetí regulačními agenturami
7. Vyvíjení produktů soukromými společnostmi



# Propojení GenBanku s vouchery

## Registry of Biological Repositories

Institutional Acronyms and Collections Codes



Home Institutional Repositories Non-Institutional Repositories FAQ Contact Us

### Institution

Search by

or Institution Name      Acronym      or Location      search

Please find your institution and edit the associated data. The institution will be contacted and the new data will be confirmed before it is posted.

Click on the column header to sort institutions by Acronym, Name or Country

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

1 2 3 Next > Last >

Acronym	Institution Name	Address	City	Country	Status	Action
A	Arnold Arboretum, Harvard University	22 Divinity Avenue, Cambridge, Massachusetts, 02138		U.S.A.	Unconfirmed	Edit
AA	Ministry of Science, Academy of Sciences	44 Temirajzev Street, Alma-Ata 480070		Kazakhstan	Unconfirmed	Edit
AAAG	Alan and Anita Gilgoly				Unconfirmed	Edit
AAH	Arnold Arboretum, Harvard University	The Arborway, Jamaica Plain, Massachusetts, 02130		U.S.A.	Unconfirmed	Edit
AAPL	Plant Industry Laboratory	Alberta, Edmonton, Alberta Agriculture		Canada	Unconfirmed	Edit

#### On This Page

- Instructions for Users
- Searching datafields
- Sorting columns
- Alphabet index
- Status column

#### Progress

- 6,636 Total Institutional Records
- 3 Confirmed Institutional Records
- 0 Confirmed Non-Institutional Records
- 2 Confirmed Collections

#### Sponsors

CONSORTIUM FOR THE BARCODE OF LIFE



# BOLD Specimen Webpage

**BOLDSYSTEMS** | Management & Analysis

Hesperidae of the AGG 1 [CSOR]

**Specimen Identifiers**

Sample ID:	02-0704-16276	Museum ID:	02-0704-16276
Isolate / Field Name:		Collection Code:	
Donated By:		Deposited In:	Smithsonian Institution

**Taxonomy**

Identifier:	19
phylum:	Arthropoda
class:	Insecta
order:	Lepidoptera
family:	Hesperidae
subfamily:	Pyrginae
genus:	Anastreus
species:	Anastreus okeanus

**Specimen Details**

Voucher Type:	
Issue Type:	
Extra Info:	Pyrginae
Sex:	m
Reproduction:	s
Life Stage:	

**Collection Data**

Collector:	Proter Moraga
Date Collected:	12-Jul-2002
Country:	Costa Rica
State/Province:	Guanacaste
Region/Country:	Area de Conservacion 276
Sector:	Ciel Oro
Exact Site:	Urcaria
Latitude:	11.8291
Longitude:	-85.4732
Coord. Source:	
Deviation Depth:	300



**Photographs**

Dorsal View

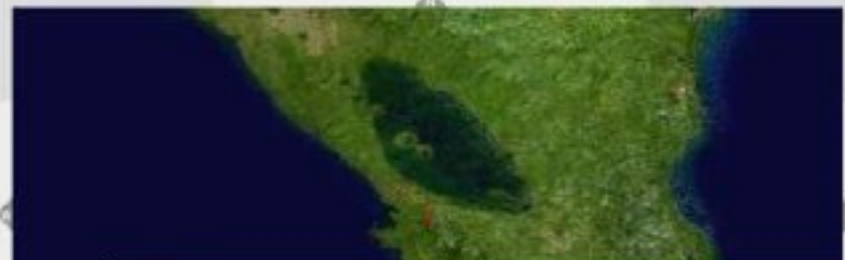


Ventral View



**BOLDSYSTEMS** | Management & Analysis

Hesperidae of the AGG 1 [CSOR]



**Specimen Identifiers**

Sample ID:	02-0704-16276	Catalog Number:	02-0704-16276
Isolate / Field Name:		Collection Code:	
Donated By:		Vouchered At:	3

**BOLDSYSTEM** | Management & Analysis


Hesperidae of the AGG 1 [CSOR]

**Specimen Identifiers**

Sample ID:	02-0704-16276	Catalog Number:	02-0704-16276
Isolate / Field Name:		Collection Code:	
Donated By:		Vouchered At:	3

**Photographs**

Dorsal View



# BOLD Sequence Webpage

BOLDSYSTEM

Management & Analysis

PROFIT



*Hesperidae of the ACG 1 [CSCR]*

## Barcode Identifiers

Barcode ID:	C8CR010-04	Sample ID:	
Gene:	COXI	GenBank Accession:	
Last Updated:		Translation Matrix:	



Model 377  
Version 3.3  
LR-377  
Version 3.3.1b2

19-TZBNA 238-03  
BF1  
TZBNA 238-03  
Lane 12

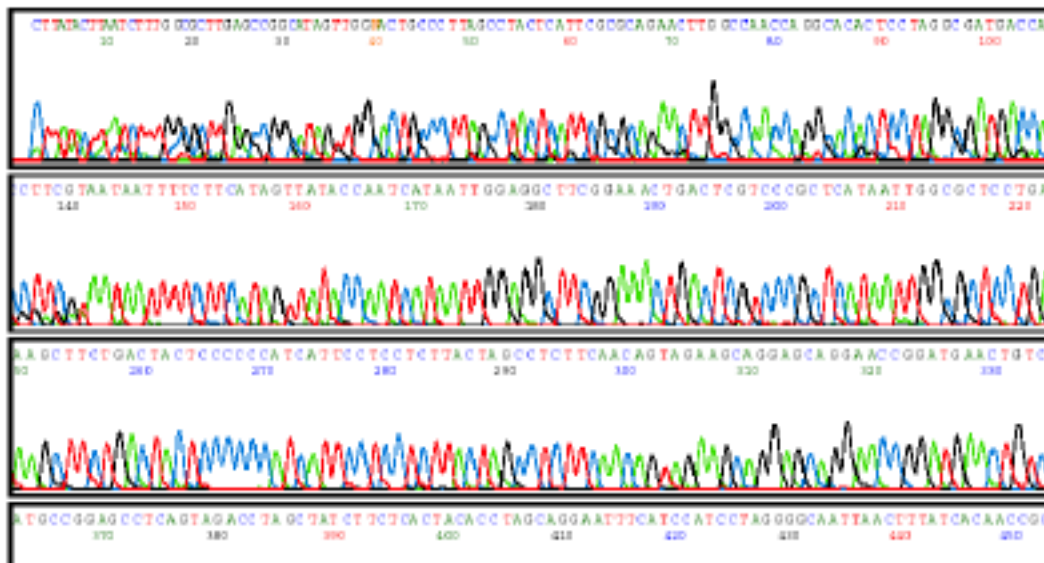
Signal G:117 A:154 T:91 C:178  
DT377\_BDv3\_v2.r10b  
214BDv3  
Points 1390 to 15200 Pk 1 Loc:

## Sequencing Runs

Run Date	Run Site	Direction	Trace File
----------	----------	-----------	------------

## Nucleotide Sequence

Length:	617	MMACCTTTATATTTTATTTTGGAAITTGAGCAGGAATAGT
Comp. A:	203	TTAGGTAACCCAGGATCTTTAATTTGGAGATGATCAAATTTA
Comp. G:	85	ATTTTTTTTATAGTAATACCAATTATAAATTGGAGGATTTGG
Comp. C:	92	GATATAGCATTTCACCAATAAATAATATAAGATTTTGACT
Comp. T:	237	AGAAITTTGAGAAATGGAGCAGGAACAGGATGAACTGTTTA
Updated:	2005-09-09	TCCTCTGTAGACTTAGCTATTTTTCATTACATTTAGCAGG
		ACAACAATTATTAATATACGAATTAGAAATTTATCATTGAT
		ACCGCACTTCTTTACTTTTATCTTTTACCTGTTTGTAGCTGG
		AATACATCATTCTTGGATCMMMMMMMMMMMMMMMMMMMM



## Amino Acid Sequence

Length:	220	MTLYFIFGIMAGHYGTSLSLIRTELGGPGLIGDDOITYTT
		DMAPFRMMNMSFQLLPFSINLLISSIVENGAGTGTUYVYPLSANIAHQSSVDLAIPLHLGAGISSILGAINYI
		TTIIMRISINLSFDQMLPVMVGVITALLLLSLPVLGAIYTNLITDRDLETSPDX-----

## Illustrative Barcode



# Ceratopogonidae – 105 barcoding sekvenci

**BOLDSYSTEMS v2.5** | Taxonomy Browser

Home | Taxonomy Browser | Request an Account | Identify Specimen | FAQs | Documentation

**Ceratopogonidae**

Sub-taxa

- Ceratopogonidae
- Dasyhelea
- Forcipomyia
- Leptoconops

Genera

- Bezzia
- Bezzia

**BOLD Systems - Sequence Download - Mozilla Firefox**

www.boldsystems.org/views/sequencedownload.php?&

**BOLDSYSTEMS v2.5** | Management & Analysis

**Sequence Download [Published Sequences]**

**Sequence Download**

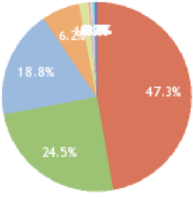
Count: 335

Fasta File:

**HEVLASY.CZ** | Mapy Google | Poslat SMS



da, no-see-ums, midgies, sand flies, punkies, and others), are a family of small flies (1-lack flies), and Thaumaleidae. They are found in almost any aquatic or semiaquatic kind of host animal. *Culicoides*, *Forcipomyia* (*Lasiohelea*), and *Leptoconops* suck *lasiohelea* feed exclusively on nectar. Species in other genera are predatory on other... full

**Sample Sources**



Source	Count	Percentage
Elizabeth Macarthur Agricultural Institute, New South Wales	1273	47.3%
Wagga Wagga Agricultural Institute	978	24.5%
Biodiversity Institute of Ontario	978	18.8%
Mined from GenBank, NCBI	323	6.2%
York University	95	2.8%
Stroud Water Research Center	15	0.4%
Canadian National Collection of Insects, Arachnids and Nematodes	9	0.3%
Mahidol University, Department of Medical Entomology	9	0.3%
Australian Quarantine and Inspection Service, Northern Australia	8	0.2%
Research Collection of Graeme V. Cocks	8	0.2%
Others	15	0.4%

images representing subtaxa of Ceratopogonidae



# Globální projekty CBOL

- Fish Barcode of Life (FISH-BOL) - 30 000 mořských/sladkovodných druhů do r. 2010
- All Birds Barcoding Initiative (ABBI) - 10 000 druhů do r. 2010
- Tephritidae – 2 000 škůdců/prospěšných druhů do r. 2008
- Komáry - 3 300 druhů do r. 2008
- Ohrožené druhy
- Trees of the world

# Staré a nové techniky

- CBOL staví na současných taxonomických poznatcích
- Sequence knihovny založeny na voucher jedincích, co dělá vědu opakovatelnou a testovatelnou
- Voucher jedince propojují historické, současné a budoucí výzkum

Příklad: CSIRO studie na bzučivkách mapující rezistence na insekticidy a zjišťování historie pomocí DNA z muzejních jedinců (PNAS 103: 8757)

# DNA Taxonomie – pomoc při řešení problému nebo vnášení chaosu?

První záznamy COI do GenBanku – 1996, od té doby ca. 1000 sekvencí dvoukřídlých

V současnosti – ca. 153 000 popsanych druhů dipter – méně než 1% je zařazeno do „Barcoding procesu“

COI nevhodný pro odlišení blízkých druhů

Stanovení hranic druhu – podobnost sekvencí (pairwise distances) - PROBLÉM

Fylogenetická rekonstrukce příbuzenských vztahů – možné řešení – multigenový přístup

Světové sbírky hmyzu – nemožnost použít materiál pro analýzy - PROBLÉM

Taxonomie založená výlučně/převážně na DNA analýze – zkreslený pohled

**Potřeba propojit s ostatními přístupy – INTEGRATIVNÍ TAXONOMIE**

# GenBank

- <http://www.ncbi.nlm.nih.gov/genbank/>
- Několik databází – Nucleotide, Protein, PubMed, CoreNucleotide, Structure, Genome, etc.
- Věrohodnost sekvencí vyšší než v databázích CBOLu
- Součástí je BLAST - „multialign tool“



# Po zadání hesla – Insect...



HOME SEARCH SITE MAP

PubMed

All Databases

Human Genome

GenBank

Map Viewer

BLAST

Search across databases

insect

GO

Clear

Help

- Result counts displayed in gray indicate one or more terms not found

210933



PubMed: biomedical literature citations and abstracts



737



Books: online books



52759



PubMed Central: free, full text journal articles



204



OMIM: online Mendelian Inheritance in Man



1



Site Search: NCBI web and FTP sites



236769



Nucleotide: Core subset of nucleotide sequence records



5



dbGaP: genotype and phenotype



564429



EST: Expressed Sequence Tag records



2130



UniGene: gene-oriented clusters of transcript sequences



1



GSS: Genome Survey Sequence records



107



CDD: conserved protein domain database



85726



Protein: sequence database



3



UniSTS: markers and mapping data



172



Genome: whole genome sequences



530



PopSet: population study data sets



919



Structure: three-dimensional macromolecular structures



60427



GEO Profiles: expression and molecular abundance profiles



none



Taxonomy: organisms in GenBank



1194



GEO DataSets: experimental sets of GEO data



1



SNP: single nucleotide polymorphism



none



Epigenomics: Epigenetic maps and data sets



# Musca domestica cytochrome oxidase subunit II (COII) gene, partial cds; mitochondrial

GenBank: DQ133110.1

[FASTA](#) [Graphics](#) [PopSet](#)

[Go to:](#)

LOCUS DQ133110 663 bp DNA linear INV 13-DEC-2005  
DEFINITION Musca domestica cytochrome oxidase subunit II (COII) gene, partial  
cds; mitochondrial.

ACCESSION DQ133110  
VERSION DQ133110.1 GI:72398995

KEYWORDS .  
SOURCE mitochondrion Musca domestica (house fly)

ORGANISM [Musca domestica](#)  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Muscoidea; Muscidae; Musca.

REFERENCE 1 (bases 1 to 663)  
AUTHORS Dittmar,K., Porter,M.L., Murray,S. and Whiting,M.F.  
TITLE Molecular phylogenetic analysis of nycteribiid and streblid bat  
flies (Diptera: Brachycera, Calyptratae): implications for host  
associations and phylogeographic origins

JOURNAL Mol. Phylogenet. Evol. 38 (1), 155-170 (2006)  
PUBMED [16087354](#)

REFERENCE 2 (bases 1 to 663)  
AUTHORS Dittmar,K., Porter,M.L., Murray,S. and Whiting,M.F.  
TITLE Direct Submission

JOURNAL Submitted (18-JUL-2005) Integrative Biology, Brigham Young  
University, 401 WIDB, Provo, UT 84602, USA

FEATURES  
source Location/Qualifiers  
1..663  
/organism="Musca domestica"  
/organelle="mitochondrion"  
/mol\_type="genomic DNA"  
/specimen\_voucher="Di180"  
/db\_xref="taxon:7370"  
[gene](#) <1..>663  
/gene="COII"  
[CDS](#) <1..>663  
/gene="COII"  
/codon\_start=1  
/transl\_table=5  
/product="cytochrome oxidase subunit II"  
/protein\_id="AAZ72903.1"  
/db\_xref="GI:72398996"  
/translation="MSTWANLGLQDSSSPLMELIFFHDHALMILVMITVLVGYLMFT  
LFFNKYVNRLLHGQTEIIEIWTILPAIILLFIAFPSRLRLLYLLDEINEPSVTLKAIGH  
QWWSYSEYSDFNVEFDSYMIPTNELPVDGFRLLDNDNRVVLPMNSQIRILLVTAADVI  
HSWTVPALGVKVDGTPGRLNQTNFLINRPLGYQCSEICGANHNSFMPIVIESIPVNY  
FIK"

ORIGIN  
1 atgtcaacat gagcaaattt aggtttaca gatagttctt ctccattaat agaacaatta  
61 attttttttc atgatcatgc attaataaatt ttagtaataa ttacagtatt agtcggatat  
121 ttaatgttta cattattttt taataaatat gttaatcgtt atttattaca tggacaaaca  
181 attgaaatta tttgaactat tttacctgca attattttat tattcattgc tttcccttct  
241 ttacgattat tatacttatt agatgaattt aatgaacct cagtaacttt aaaggctatt  
301 ggatcatcaat gatattgaag gatattgaat tcagatttta ataatgttga atttgattct

Customize view

Analyze this sequence

Run BLAST

Pick Primers

Find in this Sequence

Related information

Related Sequences

PopSet

Protein

PubMed

Taxonomy

Recent activity

[Turn Off](#) [Clear](#)

Musca domestica cytochrome oxidase subunit II (COII) gene, partial cds; mitochon... Nucleotide

Musca domestica mitochondrial (108) Nucleotide

Musca domestica (1272) Nucleotide

insect (236769) Nucleotide

Nucleotide Sequence (1495 letters) BLAST

[See more...](#)

Nucleotide

Nucleotide

Search

Limits Advanced

Help

Display Settings: FASTA

Send:

Change region shown

Customize view

## Musca domestica cytochrome oxidase subunit II (COII) gene, partial cds; mitochondrial

GenBank: DQ133110.1

[GenBank](#) [Graphics](#) [PopSet](#)

```
>gi|72398995|gb|DQ133110.1| Musca domestica cytochrome oxidase subunit II (COII)
gene, partial cds; mitochondrial
ATGTCACATGAGCAAATTTAGGTTTACAAGATAGTCTTCTCCATTAATAGAACAATTAATTTTTTTC
ATGATCATGCATTAATAATTTTAGTAATAATTACAGTATTAGTCGGATATTTAATGTTTACATTATTTT
TAATAAATATGTTAATCGTTATTTATTACATGGACAACAATTGAAATTAATTTGAACTATTTTACCTGCA
ATTATTTTATTATTCAATGCTTCCCTTCTTTACGATTATTATACTTATTAGATGAAATTAATGAACCAT
CAGTAACITTAAGGCTATTGGTCATCAATGATATTGAAGTTATGAATATTAGATTTAATAATGTTGA
ATTTGATTCCTATATAATTCCTACAAATGAATTACCAGTAGACGGATTTCGTTTATTAGATGTAGATAAT
CGAGTAGTTTTACCAATAAATTCCAAATTCGAATTTAGTAACTGCTGCTGATGTAATTCATTCAATGAA
CTGTTCTGCTTTAGGTGTAAGGTTGATGGTACTCCTGGTCGCTAAATCAAATAATTTCTTAATTA
TCGACCAGGTTTATTCTATGGACAATGTT CAGAAAATTTGTGGAGCTAATCATAGTTTATACCAATTGTA
ATTGAAAGTATTCTGTAAATTTTATTAAAG
```

### Analyze this sequence

- Run BLAST
- Pick Primers
- Find in this Sequence

### Related information

- Related Sequences
- PopSet
- Protein
- PubMed
- Taxonomy


### Recent activity

Turn Off Clear

- Musca domestica cytochrome oxidase subunit II (COII) gene, partial cds; mitochon... Nucleotide
- Musca domestica mitochondrial (188) Nucleotide
- Musca domestica (1272)

# BLAST

<http://blast.ncbi.nlm.nih.gov/Blast.cgi>

 **BLAST**<sup>®</sup> *Basic Local Alignment Search Tool*

[Home](#) [Recent Results](#) [Saved Strategies](#) [Help](#)

My NCBI [Sign In](#) [Register](#)

► [NCBI/BLAST Home](#)

BLAST finds regions of similarity between biological sequences. [more...](#)

**New** Aligning Multiple Protein Sequences? Try the [COBALT Multiple Alignment Tool](#). [Go](#)

### BLAST Assembled RefSeq Genomes

Choose a species genome to search, or [list all genomic BLAST databases](#).

- [Human](#)
- [Mouse](#)
- [Rat](#)
- [Arabidopsis thaliana](#)
- [Oryza sativa](#)
- [Bos taurus](#)
- [Danio rerio](#)
- [Drosophila melanogaster](#)
- [Gallus gallus](#)
- [Pan troglodytes](#)
- [Microbes](#)
- [Apis mellifera](#)

### Basic BLAST

Choose a BLAST program to run.

<a href="#">nucleotide blast</a>	Search a <b>nucleotide</b> database using a <b>nucleotide</b> query <i>Algorithms: blastn, megablast, discontinuous megablast</i>
<a href="#">protein blast</a>	Search <b>protein</b> database using a <b>protein</b> query <i>Algorithms: blastp, psi-blast, phi-blast</i>
<a href="#">blastx</a>	Search <b>protein</b> database using a <b>translated nucleotide</b> query
<a href="#">tblastn</a>	Search <b>translated nucleotide</b> database using a <b>protein</b> query
<a href="#">tblastx</a>	Search <b>translated nucleotide</b> database using a <b>translated nucleotide</b> query

#### Your Recent Results **New!**

[Nucleotide Sequence \(1495 let...](#)

#### News

[SOAP BLAST](#)

A SOAP based BLAST service is available.  
Mon, 18 Jul 2011 08:00:00 EST

[More BLAST news...](#)

#### Tip of the Day

[Use Genomic BLAST to see the genomic context](#)

If you are interested in the evolution of a particular gene or gene family it is often interesting to examine the intro-exon structure even across species.

[More tips...](#)

Other reports: Search Summary [Taxonomy reports] [Distance tree of results]

Graphic Summary

Distribution of 100 Blast Hits on the Query Sequence



Descriptions

Legend for links to other resources: U UniGene E GEO G Gene S Structure M Map Viewer P PubChem BioAssay

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
<a href="#">GQ465784.1</a>	Musca domestica cytochrome oxidase subunit I (COI) gene, partial c	<a href="#">2761</a>	2761	100%	0.0	100%	
<a href="#">AB479529.1</a>	Musca domestica mitochondrial COI gene for cytochrome oxidase sub	<a href="#">2750</a>	2750	100%	0.0	99%	
<a href="#">AB479528.1</a>	Musca domestica mitochondrial COI gene for cytochrome oxidase sub	<a href="#">2750</a>	2750	100%	0.0	99%	
<a href="#">AY526196.1</a>	Musca domestica tRNA-Tyr gene, partial sequence; and cytochrome	<a href="#">2750</a>	2750	100%	0.0	99%	
<a href="#">EU154477.1</a>	Musca domestica mitochondrion, partial genome	<a href="#">2739</a>	2739	100%	0.0	99%	
<a href="#">EU627696.1</a>	Musca conducens cytochrome oxidase subunit I (COI) gene, partial c	<a href="#">2244</a>	2244	99%	0.0	93%	
<a href="#">EU627694.1</a>	Musca asiatica cytochrome oxidase subunit I (COI) gene, partial cds;	<a href="#">2222</a>	2222	99%	0.0	93%	
<a href="#">EU627693.1</a>	Musca sorbens cytochrome oxidase subunit I (COI) gene, partial cds;	<a href="#">2222</a>	2222	99%	0.0	93%	
<a href="#">EU627700.1</a>	Musca larvipara cytochrome oxidase subunit I (COI) gene, partial cds	<a href="#">2206</a>	2206	99%	0.0	93%	
<a href="#">EU627698.1</a>	Musca confisicata cytochrome oxidase subunit I (COI) gene, partial c	<a href="#">2178</a>	2178	99%	0.0	92%	
<a href="#">AB479530.1</a>	Musca crassirostris mitochondrial COI gene for cytochrome oxidase su	<a href="#">2167</a>	2167	99%	0.0	92%	
<a href="#">AB479531.1</a>	Musca crassirostris mitochondrial COI gene for cytochrome oxidase su	<a href="#">2161</a>	2161	99%	0.0	92%	
<a href="#">EU627701.1</a>	Musca convexifrons cytochrome oxidase subunit I (COI) gene, partial	<a href="#">2156</a>	2156	99%	0.0	92%	
<a href="#">EU627702.1</a>	Musca formosana cytochrome oxidase subunit I (COI) gene, partial c	<a href="#">2089</a>	2089	99%	0.0	91%	
<a href="#">AB479533.1</a>	Musca bezzii mitochondrial COI gene for cytochrome oxidase subunit	<a href="#">2061</a>	2061	99%	0.0	91%	
<a href="#">EU627695.1</a>	Musca crassirostris cytochrome oxidase subunit I (COI) gene, partial	<a href="#">2056</a>	2056	99%	0.0	91%	
<a href="#">AB479532.1</a>	Musca bezzii mitochondrial COI gene for cytochrome oxidase subunit	<a href="#">2056</a>	2056	99%	0.0	91%	
<a href="#">EU815009.1</a>	Musca domestica isolate JIA-A-1 cytochrome oxidase subunit I (COI)	<a href="#">2049</a>	2049	74%	0.0	100%	
<a href="#">EU627699.1</a>	Musca inferior cytochrome oxidase subunit I (COI) gene, partial cds;	<a href="#">2045</a>	2045	99%	0.0	91%	
<a href="#">EU814999.1</a>	Musca domestica isolate jia21 cytochrome oxidase subunit I (COI) ge	<a href="#">2045</a>	2045	74%	0.0	100%	
<a href="#">EU814993.1</a>	Musca domestica isolate jia14 cytochrome oxidase subunit I (COI) ge	<a href="#">2043</a>	2043	73%	0.0	100%	
<a href="#">EU814992.1</a>	Musca domestica isolate jia13 cytochrome oxidase subunit I (COI) ge	<a href="#">2039</a>	2039	73%	0.0	100%	

# FLY TREE

2004-2008, 30 mil. USD, 649 taxonů, desítky tisíc bp



FLYTREE

## Assembling the Diptera Tree of Life

### FLYTREE

[Introduction](#)  
[About this Grant](#)  
[About & Contact Us](#)  
[Opportunities](#)

### Features

[About Flies](#)  
[Pictures](#)  
[Fly Morphology](#)  
[Fly Nomenclature](#)  
[Species Highlights](#)  
[Phylogeny](#)  
[Publications & Products](#)

### News

[Press Releases](#)  
[Talking About Flies](#)

[Buzz About Flies](#)  
[Additional Buzz](#)

[Diptera.org](#)

### Latest FLYTREE News:

[view all recent posts](#)

October 5, 2011

[Finding The Fly Tree of Life - The Poster!](#)

March 14, 2011

[Map of the Fly Tree of Life Published!](#)

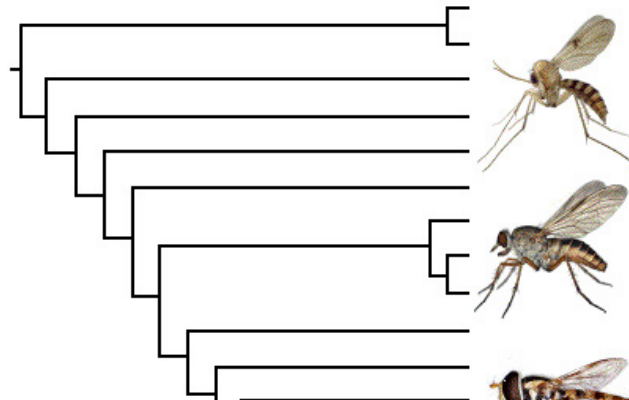
December 8, 2010

[Rediscovering World's Rarest Fly](#)

With over 158,000 described species, flies (the insect order Diptera), are among the most diverse groups of living organisms on the planet.

This diversity transcends simple species numbers and is demonstrated in the great breadth of morphological, ecological, and behavioral variation found in the group.

Flies have a deep evolutionary history that extends back to the Permian Period, over 250 million years ago.





# Builders of the Dipteran Tree...

## FLYTREE

Introduction  
About this Grant  
About & Contact Us  
Opportunities

## Features

About Flies  
Pictures  
Fly Morphology  
Fly Nomenclature  
Species Highlights  
Phylogeny  
Publications & Products

## News

Press Releases  
Talking About Flies

## Buzz About Flies

Additional Buzz  
Diptera.org  
EDIT Diptera  
Tree of Life



<http://www.cals.ncsu.edu/entomology/wiegmann/>



## FLYTREE

# Builders of the Dipteran Tree...

## FLYTREE

Introduction  
About this Grant  
About & Contact Us  
Opportunities

## Features

About Flies  
Pictures  
Fly Morphology  
Fly Nomenclature  
Species Highlights  
Phylogeny  
Publications & Products

## News

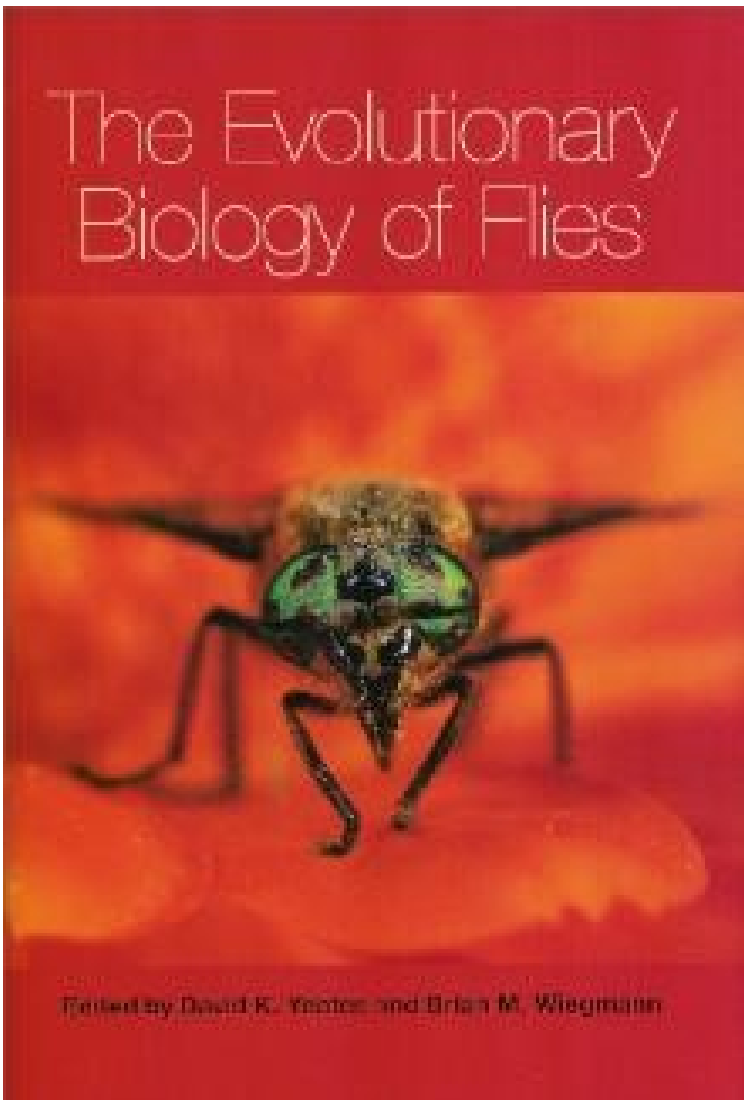
Press Releases  
Talking About Flies

## Buzz About Flies

Additional Buzz  
Diptera.org  
EDIT Diptera



# OUTPUTS





# MEGA v. 5 – úprava sekvencí (alignment)

MEGA v. 5.0.16.0 interface showing a multiple sequence alignment of DNA sequences. The alignment is displayed in a grid format with columns representing positions and rows representing individual sequences. The sequences are labeled on the left, including accession numbers (e.g., 1S, 2S, 3S, 4S, 5S, 7S, 8S, 9S, 11S, 13S, 15S, 18S, 20S, 21S, 22S, 39S, 40S, 41S, 46S, 47S, 48S, 49S, 51S, 44S, 54S, 55S) and identifiers (e.g., 16S, c1c3, c1c2).

The alignment shows conserved regions (indicated by asterisks) and gaps (indicated by dashes). The sequences are color-coded by nucleotide type: A (green), C (blue), G (red), and T (orange). The alignment is displayed in a grid format with columns representing positions and rows representing individual sequences. The sequences are labeled on the left, including accession numbers (e.g., 1S, 2S, 3S, 4S, 5S, 7S, 8S, 9S, 11S, 13S, 15S, 18S, 20S, 21S, 22S, 39S, 40S, 41S, 46S, 47S, 48S, 49S, 51S, 44S, 54S, 55S) and identifiers (e.g., 16S, c1c3, c1c2).

The alignment shows conserved regions (indicated by asterisks) and gaps (indicated by dashes). The sequences are color-coded by nucleotide type: A (green), C (blue), G (red), and T (orange). The alignment is displayed in a grid format with columns representing positions and rows representing individual sequences. The sequences are labeled on the left, including accession numbers (e.g., 1S, 2S, 3S, 4S, 5S, 7S, 8S, 9S, 11S, 13S, 15S, 18S, 20S, 21S, 22S, 39S, 40S, 41S, 46S, 47S, 48S, 49S, 51S, 44S, 54S, 55S) and identifiers (e.g., 16S, c1c3, c1c2).

# Fylogenetické analýzy

Fylogenetický strom – hypotéza, která vznikla co nejlepším odhadem na základě omezeného zdroje informací

## Metody FA – dva přístupy

1. Algoritmus – jde přímo k výsledku, co je jediný strom (odpadá srovnání vzájemně si konkurujících stromů) – metody shlukové analýzy (UPGMA), Neighbour-joining (NJ) – obě využívají data vzdáleností (distance)
2. Kritérium optimálnosti – **dva kroky** – definování kritéria, podle kterého je hodnocen každý strom určitým skóre, které se použije k následnému srovnání všech stromů
  - použití specifického algoritmu pro výpočet funkce (kritérium optimálnosti) a pro získání stromu s nejlepší hodnotou této funkce

# Jaká by měla vybraná metoda být?

Výkonnost – „tempus fugit“ nebo „time is money“

pomoc – heuristické metody hledání v případě vyššího počtu taxonů či znaků

Síla – kolik dat musíme shromáždit, aby byly výsledky správné

Konzistence – s přidáváním dalších znaků spějeme k správnému výsledku

Robustnost – do jaké míry vedou drobné odchýlky od vstupných předpokladů k nesprávným závěrům

Falzifikovatelnost – určení nevhodnosti modelu na základě odchýlky od předpokladu

**IDEÁLNÍ METODA NEEXISTUJE...**

# Metoda maximální parsimonie – úspornosti (MP)

Jedna z nejpoužívanějších metod - rychlá, jednoduchá preferuje jednodušší hypotézy před složitějšími (široká filozofická platnost), tzn. vybere možnost (strom) s minimálním počtem evolučních kroků nutných k vysvětlení vstupních dat

Ne všechny znaky jsou použitelné, parsimony - informative

+

Dobrá pochopitelnost, jednoduchost, rychlost, nízký počet předpokladů (předpokládá, že jakákoli evoluční změna je vzácná, takže MP strom se dá považovat za nejlepší odhad skutečné evoluce)

-

Nekonzistentnost, přitažlivost dlouhých větví (LBA)

# Metoda maximální pravděpodobnosti (Maximum likelihood, ML)

- posuzují se jednotlivé hypotézy o evoluční historii zkoumaných taxonů z hlediska pravděpodobnosti, že jsou v souladu se získanými daty, výsledek – maximálně pravděpodobný odhad

Tři součásti - vstupné data

evoluční model

fylogenetický strom s topologií i délkou větví

+

Nízka náchylnost k chybě, robustnost vůči odchýlkám

-

Vysoká výpočetní náročnost

# Bayesian inference

Výpočet pravděpodobnosti na základě specifikovaného modelu a na základě toho, co jsme o charakteru dat zjistili

Základ – strom s danou topologií a délkami větví, model nukleotidových substitucí a rozložení substitučních frekvencí mezi jednotlivými nukleotidy

Princip přístupu jako u ML

VÝHODY – menší časová náročnost, strom zohledňující fylogenetický signál v datasetu, možnost použít i pro smíšený dataset

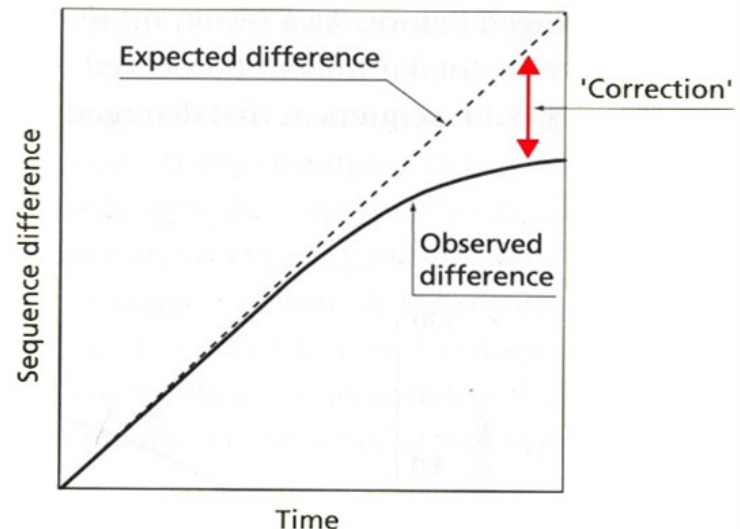
# Distanční metody

Založené na podobnostech (vzdálenostech, rozdílech)

Poznání skutečné evoluční vzdálenosti mezi všemi členy studovaného souboru taxonů umožňuje velmi lehkou rekonstrukci evoluční historie těchto taxonů

Opakované změny  
jednoho znaku –  
korigované distance (jako  
u pravděpodobnosti)

Nekorigovaná vzdálenost  
– p-distance  
Korekce: JC, F81, K2P,  
F84, GTR



# Přepříprava☺

- Úprava sekvencí (Sequencher)
- Vytvoření alignmentu (MEGA) - .fas, .nex
- Analýza MP (Paup)
- Analýza NJ (Paup, MEGA)
- Vytvoření souboru pro MrModeltest (PAUP)
- MrModeltest
- Příprava souboru pro MrBayes
- Ukázka práce s SequenceMatrix pro velké datasey