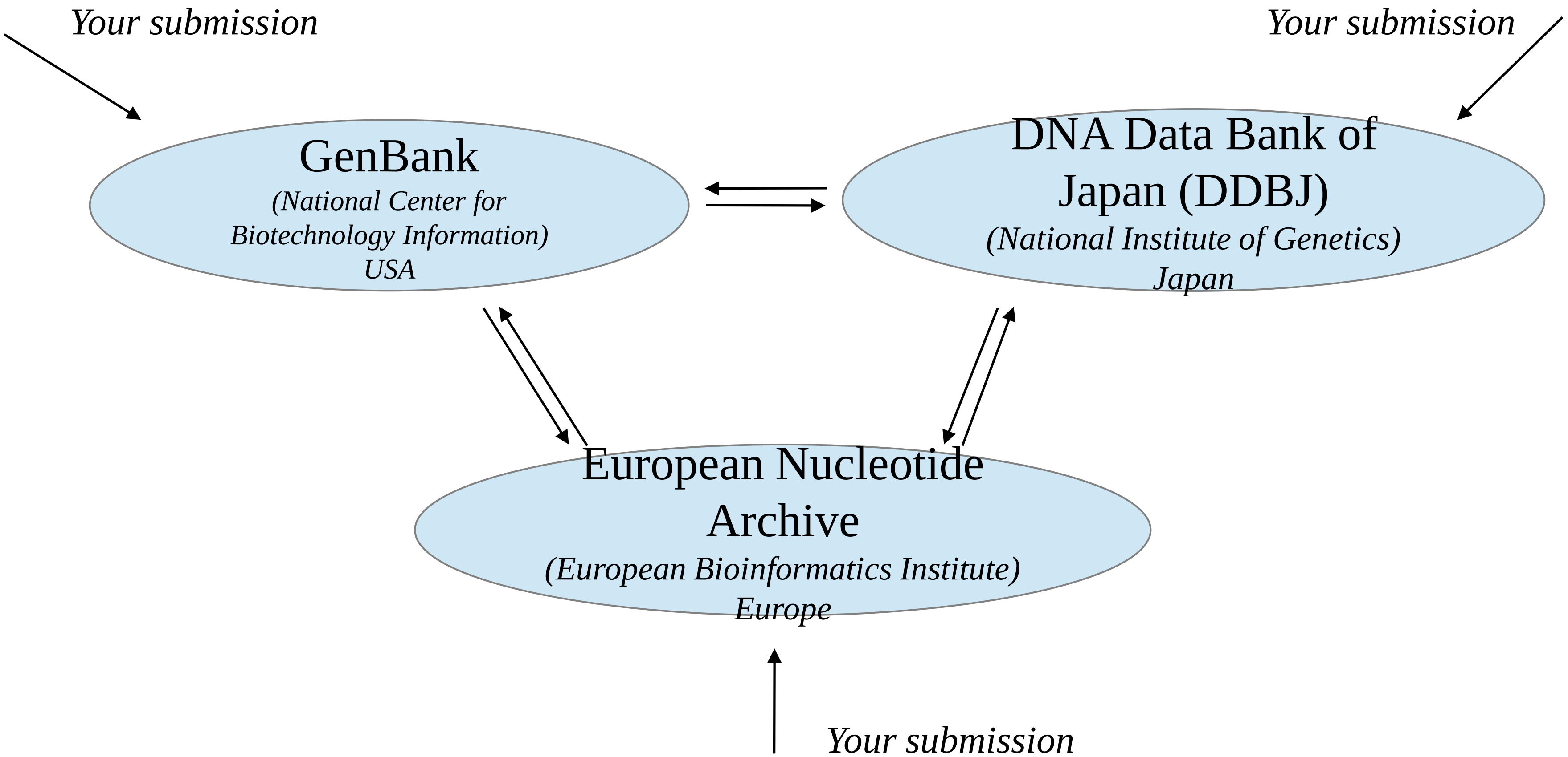
Molecular Applications in Zoology

Sequence data handling - practicals

Zuzana Starostová

Sequences “stored“ in databases

International Nucleotide Sequence Databases (INSD)



Your submission

Your submission

GenBank
(National Center for
Biotechnology Information)
USA

DNA Data Bank of Japan (DDBJ)
(National Institute of Genetics)
Japan

European Nucleotide Archive
(European Bioinformatics Institute)
Europe

Your submission

Sequence data handling and manipulation

sequence format: text file (can be edited in e.g. notepad or some specialized programs for sequence editing – BioEdit, Geneious, …)

FASTA (.fa, .fas, .fasta)
GenBank (.gb)

FASTA

- only basic information about the sequence, input format for programs

```
>gi|148832288|gb|EF443167.1| Rhinopoma hardwickei haplotype 2949 cytochrome b gene, partial cds; mitochondrial
ATGACCCACATCCGAAAATCCCACCCCTTATTCAAAATTATCAACGACTCATTCATCGACCTACC
AGCTCCATCAAACATTTCCTCCTGATGAAATTTTGGGTCCCTACTAGGTATTTGTTTAGCTGTAC
AAATCTTAACAGGACTGTTCCTAGCAATACATTATACATCAGATACCACAACCGCCTTCTACTCT
GTTACCCATATCTGCCGAGACGTAAATTACGGCTGAATCCTACGTTACCTCCATGCCAACGGAGC
ATCCATATTCTTCATCTGCCTATTTATACATGTAGGCCGAGGCATCTATTACGGCTCATACCTAT
TCACAGAAACATGAAACATTGGCATTATCCTTCTATTCGCCGTAATAGCAACAGCATTCATAGGC
TATGTCCTCCCA
>gi|...
ATGA...
```

GenBank

- Detailed information about the sequence

```
•LOCUS       EF443167                 402 bp    DNA     linear   MAM 15-OCT-2007
•DEFINITION  Rhinopoma hardwickei haplotype 2949 cytochrome b gene, partial cds;
•            mitochondrial.
•ACCESSION   EF443167
•VERSION     EF443167.1  GI:148832288
•KEYWORDS    .
•SOURCE      mitochondrion Rhinopoma hardwickii (Lesser mouse-tailed bat)
•  ORGANISM  Rhinopoma hardwickii
•            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
•            Mammalia; Eutheria; Laurasiatheria; Chiroptera; Microchiroptera;
•            Rhinopomatidae; Rhinopoma.
•REFERENCE   1  (bases 1 to 402)
•  AUTHORS   Hulva,P., Horacek,I. and Benda,P.
•  TITLE     Molecules, morphometrics and new fossils provide an integrated view
•            of the evolutionary history of Rhinopomatidae (Mammalia:
•            Chiroptera)
•  JOURNAL   BMC Evol. Biol. 7, 165 (2007)
•   PUBMED   17868440
•  REMARK    Publication Status: Online-Only
•REFERENCE   2  (bases 1 to 402)
•  AUTHORS   Hulva,P.
•  TITLE     Direct Submission
•  JOURNAL   Submitted (18-FEB-2007) Zoology, Charles University, Vinicna 7,
•            Prague, 2 12843, Czech Republic
•FEATURES             Location/Qualifiers
•     source          1..402
•                     /organism="Rhinopoma hardwickii"
•                     /organelle="mitochondrion"
•                     /mol_type="genomic DNA"
•                     /db_xref="taxon:124756"
•                     /haplotype="2949"
•     CDS             1..>402
•                     /codon_start=1
•                     /transl_table=2
•                     /product="cytochrome b"
•                     /protein_id="ABR14084.1"
•                     /db_xref="GI:148832289"
•                     /translation="MTHIRKSHPLFKIINDSFIDLPAPSNISSWWNFGSLLGICLAVQ
•                     ILTGLFLAMHYTSDTTTAFYSVTHICRDVNYGWILRYLHANGASMFFICLFMHVGRGI
•                     YYGSYLFTETWNIGIILLFAVMATAFMGYVLP"
•ORIGIN
•        1 atgacccaca tccgaaaatc ccacccctta ttcaaaatta tcaacgactc attcatcgac
•       61 ctaccagctc catcaaacat ttcctcctga tgaaattttg ggtccctact aggtatttgt
•      121 ttagctgtac aaatcttaac aggactgttc ctagcaatac attatacatc agataccaca
•      181 accgccttct actctgttac ccatatctgc cgagacgtaa attacggctg aatcctacgt
•      241 tacctccatg ccaacggagc atccatattc ttcatctgcc tatttataca tgtaggccga
•      301 ggcatctatt acggctcata cctattcaca gaaacatgaa acattggcat tatccttcta
•      361 ttcgccgtaa tagcaacagc attcataggc tatgtcctcc ca
•//
```

DEFINITION

List of genes in the DNA fragment

```
•LOCUS       EF443167                 402 bp    DNA     linear   MAM 15-OCT-2007
•DEFINITION  Rhinopoma hardwickei haplotype 2949 cytochrome b gene, partial cds;
•            mitochondrial.
•ACCESSION   EF443167
•VERSION     EF443167.1  GI:148832288
•KEYWORDS    .
•SOURCE      mitochondrion Rhinopoma hardwickii (Lesser mouse-tailed bat)
•  ORGANISM  Rhinopoma hardwickii
•            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
•            Mammalia; Eutheria; Laurasiatheria; Chiroptera; Microchiroptera;
•            Rhinopomatidae; Rhinopoma.
•REFERENCE   1  (bases 1 to 402)
•  AUTHORS   Hulva,P., Horacek,I. and Benda,P.
•  TITLE     Molecules, morphometrics and new fossils provide an integrated view
•            of the evolutionary history of Rhinopomatidae (Mammalia:
•            Chiroptera)
•  JOURNAL   BMC Evol. Biol. 7, 165 (2007)
•   PUBMED   17868440
•  REMARK    Publication Status: Online-Only
•REFERENCE   2  (bases 1 to 402)
•  AUTHORS   Hulva,P.
•  TITLE     Direct Submission
•  JOURNAL   Submitted (18-FEB-2007) Zoology, Charles University, Vinicna 7,
•            Prague, 2 12843, Czech Republic
•FEATURES             Location/Qualifiers
•     source          1..402
•                     /organism="Rhinopoma hardwickii"
•                     /organelle="mitochondrion"
•                     /mol_type="genomic DNA"
•                     /db_xref="taxon:124756"
•                     /haplotype="2949"
•     CDS             1..>402
•                     /codon_start=1
•                     /transl_table=2
•                     /product="cytochrome b"
•                     /protein_id="ABR14084.1"
•                     /db_xref="GI:148832289"
•                     /translation="MTHIRKSHPLFKIINDSFIDLPAPSNISSWWNFGSLLGICLAVQ
•                     ILTGLFLAMHYTSDTTTAFYSVTHICRDVNYGWILRYLHANGASMFFICLFMHVGRGI
•                     YYGSYLFTETWNIGIILLFAVMATAFMGYVLP"
•ORIGIN
•        1 atgacccaca tccgaaaatc ccacccctta ttcaaaatta tcaacgactc attcatcgac
•       61 ctaccagctc catcaaacat ttcctcctga tgaaattttg ggtccctact aggtatttgt
•      121 ttagctgtac aaatcttaac aggactgttc ctagcaatac attatacatc agataccaca
•      181 accgccttct actctgttac ccatatctgc cgagacgtaa attacggctg aatcctacgt
•      241 tacctccatg ccaacggagc atccatattc ttcatctgcc tatttataca tgtaggccga
•      301 ggcatctatt acggctcata cctattcaca gaaacatgaa acattggcat tatccttcta
•      361 ttcgccgtaa tagcaacagc attcataggc tatgtcctcc ca
•//
```

ACCESSION

VERSION

```
•LOCUS       EF443167                 402 bp    DNA     linear   MAM 15-OCT-2007
•DEFINITION  Rhinopoma hardwickei haplotype 2949 cytochrome b gene, partial cds;
•            mitochondrial.
•ACCESSION   EF443167
•VERSION     EF443167.1  GI:148832288
•KEYWORDS    .
•SOURCE      mitochondrion Rhinopoma hardwickii (Lesser mouse-tailed bat)
•  ORGANISM  Rhinopoma hardwickii
•            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
•            Mammalia; Eutheria; Laurasiatheria; Chiroptera; Microchiroptera;
•            Rhinopomatidae; Rhinopoma.
•REFERENCE   1  (bases 1 to 402)
•  AUTHORS   Hulva,P., Horacek,I. and Benda,P.
•  TITLE     Molecules, morphometrics and new fossils provide an integrated view
•            of the evolutionary history of Rhinopomatidae (Mammalia:
•            Chiroptera)
•  JOURNAL   BMC Evol. Biol. 7, 165 (2007)
•   PUBMED   17868440
•  REMARK    Publication Status: Online-Only
•REFERENCE   2  (bases 1 to 402)
•  AUTHORS   Hulva,P.
•  TITLE     Direct Submission
•  JOURNAL   Submitted (18-FEB-2007) Zoology, Charles University, Vinicna 7,
•            Prague, 2 12843, Czech Republic
•FEATURES             Location/Qualifiers
•     source          1..402
•                     /organism="Rhinopoma hardwickii"
•                     /organelle="mitochondrion"
•                     /mol_type="genomic DNA"
•                     /db_xref="taxon:124756"
•                     /haplotype="2949"
•     CDS             1..>402
•                     /codon_start=1
•                     /transl_table=2
•                     /product="cytochrome b"
•                     /protein_id="ABR14084.1"
•                     /db_xref="GI:148832289"
•                     /translation="MTHIRKSHPLFKIINDSFIDLPAPSNISSWWNFGSLLGICLAVQ
•                     ILTGLFLAMHYTSDTTTAFYSVTHICRDVNYGWILRYLHANGASMFFICLFMHVGRGI
•                     YYGSYLFTETWNIGIILLFAVMATAFMGYVLP"
•ORIGIN
•        1 atgacccaca tccgaaaatc ccacccctta ttcaaaatta tcaacgactc attcatcgac
•       61 ctaccagctc catcaaacat ttcctcctga tgaaattttg ggtccctact aggtatttgt
•      121 ttagctgtac aaatcttaac aggactgttc ctagcaatac attatacatc agataccaca
•      181 accgccttct actctgttac ccatatctgc cgagacgtaa attacggctg aatcctacgt
•      241 tacctccatg ccaacggagc atccatattc ttcatctgcc tatttataca tgtaggccga
•      301 ggcatctatt acggctcata cctattcaca gaaacatgaa acattggcat tatccttcta
•      361 ttcgccgtaa tagcaacagc attcataggc tatgtcctcc ca
•//
```

KEYWORDS

```
•LOCUS       EF443167                 402 bp    DNA     linear   MAM 15-OCT-2007
•DEFINITION  Rhinopoma hardwickei haplotype 2949 cytochrome b gene, partial cds;
•            mitochondrial.
•ACCESSION   EF443167
•VERSION     EF443167.1  GI:148832288
•KEYWORDS    .
•SOURCE      mitochondrion Rhinopoma hardwickii (Lesser mouse-tailed bat)
•  ORGANISM  Rhinopoma hardwickii
•            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
•            Mammalia; Eutheria; Laurasiatheria; Chiroptera; Microchiroptera;
•            Rhinopomatidae; Rhinopoma.
•REFERENCE   1  (bases 1 to 402)
•  AUTHORS   Hulva,P., Horacek,I. and Benda,P.
•  TITLE     Molecules, morphometrics and new fossils provide an integrated view
•            of the evolutionary history of Rhinopomatidae (Mammalia:
•            Chiroptera)
•  JOURNAL   BMC Evol. Biol. 7, 165 (2007)
•   PUBMED   17868440
•  REMARK    Publication Status: Online-Only
•REFERENCE   2  (bases 1 to 402)
•  AUTHORS   Hulva,P.
•  TITLE     Direct Submission
•  JOURNAL   Submitted (18-FEB-2007) Zoology, Charles University, Vinicna 7,
•            Prague, 2 12843, Czech Republic
•FEATURES             Location/Qualifiers
•     source          1..402
•                     /organism="Rhinopoma hardwickii"
•                     /organelle="mitochondrion"
•                     /mol_type="genomic DNA"
•                     /db_xref="taxon:124756"
•                     /haplotype="2949"
•     CDS             1..>402
•                     /codon_start=1
•                     /transl_table=2
•                     /product="cytochrome b"
•                     /protein_id="ABR14084.1"
•                     /db_xref="GI:148832289"
•                     /translation="MTHIRKSHPLFKIINDSFIDLPAPSNISSWWNFGSLLGICLAVQ
•                     ILTGLFLAMHYTSDTTTAFYSVTHICRDVNYGWILRYLHANGASMFFICLFMHVGRGI
•                     YYGSYLFTETWNIGIILLFAVMATAFMGYVLP"
•ORIGIN
•        1 atgacccaca tccgaaaatc ccaccccta ttcaaaatta tcaacgactc attcatcgac
•       61 ctaccagctc catcaaacat ttcctcctga tgaaattttg ggtccctact aggtatttgt
•      121 ttagctgtac aaatcttaac aggactgttc ctagcaatac attatacatc agataccaca
•      181 accgccttct actctgttac ccatatctgc cgagacgtaa attacggctg aatcctacgt
•      241 tacctccatg ccaacggagc atccatattc ttcatctgcc tatttataca tgtaggccga
•      301 ggcatctatt acggctcata cctattcaca gaaacatgaa acattggcat tatccttcta
•      361 ttcgccgtaa tagcaacagc attcataggc tatgtcctcc ca
•//
```

SOURCE

```
•LOCUS       EF443167                 402 bp    DNA     linear   MAM 15-OCT-2007
•DEFINITION  Rhinopoma hardwickei haplotype 2949 cytochrome b gene, partial cds;
•            mitochondrial.
•ACCESSION   EF443167
•VERSION     EF443167.1  GI:148832288
•KEYWORDS    .
•SOURCE      mitochondrion Rhinopoma hardwickii (Lesser mouse-tailed bat)
•  ORGANISM  Rhinopoma hardwickii
•            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
•            Mammalia; Eutheria; Laurasiatheria; Chiroptera; Microchiroptera;
•            Rhinopomatidae; Rhinopoma.
•REFERENCE   1  (bases 1 to 402)
•  AUTHORS   Hulva,P., Horacek,I. and Benda,P.
•  TITLE     Molecules, morphometrics and new fossils provide an integrated view
•            of the evolutionary history of Rhinopomatidae (Mammalia:
•            Chiroptera)
•  JOURNAL   BMC Evol. Biol. 7, 165 (2007)
•   PUBMED   17868440
•  REMARK    Publication Status: Online-Only
•REFERENCE   2  (bases 1 to 402)
•  AUTHORS   Hulva,P.
•  TITLE     Direct Submission
•  JOURNAL   Submitted (18-FEB-2007) Zoology, Charles University, Vinicna 7,
•            Prague, 2 12843, Czech Republic
•FEATURES             Location/Qualifiers
•     source          1..402
•                     /organism="Rhinopoma hardwickii"
•                     /organelle="mitochondrion"
•                     /mol_type="genomic DNA"
•                     /db_xref="taxon:124756"
•                     /haplotype="2949"
•     CDS             1..>402
•                     /codon_start=1
•                     /transl_table=2
•                     /product="cytochrome b"
•                     /protein_id="ABR14084.1"
•                     /db_xref="GI:148832289"
•                     /translation="MTHIRKSHPLFKIINDSFIDLPAPSNISSWWNFGSLLGICLAVQ
•                     ILTGLFLAMHYTSDTTTAFYSVTHICRDVNYGWILRYLHANGASMFFICLFMHVGRGI
•                     YYGSYLFTETWNIGIILLFAVMATAFMGYVLP"
•ORIGIN
•        1 atgacccaca tccgaaaatc ccacccctta ttcaaaatta tcaacgactc attcatcgac
•       61 ctaccagctc catcaaacat ttcctcctga tgaaattttg ggtccctact aggtatttgt
•      121 ttagctgtac aaatcttaac aggactgttc ctagcaatac attatacatc agataccaca
•      181 accgccttct actctgttac ccatatctgc cgagacgtaa attacggctg aatcctacgt
•      241 tacctccatg ccaacggagc atccatattc ttcatctgcc tatttataca tgtaggccga
•      301 ggcatctatt acggctcata cctattcaca gaaacatgaa acattggcat tatccttcta
•      361 ttcgccgtaa tagcaacagc attcataggc tatgtcctcc ca
•//
```

REFERENCE

```
•LOCUS       EF443167                 402 bp    DNA     linear   MAM 15-OCT-2007
•DEFINITION  Rhinopoma hardwickei haplotype 2949 cytochrome b gene, partial cds;
•            mitochondrial.
•ACCESSION   EF443167
•VERSION     EF443167.1  GI:148832288
•KEYWORDS    .
•SOURCE      mitochondrion Rhinopoma hardwickii (Lesser mouse-tailed bat)
•  ORGANISM  Rhinopoma hardwickii
•            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
•            Mammalia; Eutheria; Laurasiatheria; Chiroptera; Microchiroptera;
•            Rhinopomatidae; Rhinopoma.
•REFERENCE   1  (bases 1 to 402)
•  AUTHORS   Hulva,P., Horacek,I. and Benda,P.
•  TITLE     Molecules, morphometrics and new fossils provide an integrated view
•            of the evolutionary history of Rhinopomatidae (Mammalia:
•            Chiroptera)
•  JOURNAL   BMC Evol. Biol. 7, 165 (2007)
•   PUBMED   17868440
•  REMARK    Publication Status: Online-Only
•REFERENCE   2  (bases 1 to 402)
•  AUTHORS   Hulva,P.
•  TITLE     Direct Submission
•  JOURNAL   Submitted (18-FEB-2007) Zoology, Charles University, Vinicna 7,
•            Prague, 2 12843, Czech Republic
•FEATURES             Location/Qualifiers
•     source          1..402
•                     /organism="Rhinopoma hardwickii"
•                     /organelle="mitochondrion"
•                     /mol_type="genomic DNA"
•                     /db_xref="taxon:124756"
•                     /haplotype="2949"
•     CDS             1..>402
•                     /codon_start=1
•                     /transl_table=2
•                     /product="cytochrome b"
•                     /protein_id="ABR14084.1"
•                     /db_xref="GI:148832289"
•                     /translation="MTHIRKSHPLFKIINDSFIDLPAPSNISSWWNFGSLLGICLAVQ
•                     ILTGLFLAMHYTSDTTTAFYSVTHICRDVNYGWILRYLHANGASMFFICLFMHVGRGI
•                     YYGSYLFTETWNIGIILLFAVMATAFMGYVLP"
•ORIGIN
•        1 atgacccaca tccgaaaatc ccacccctta ttcaaaatta tcaacgactc attcatcgac
•       61 ctaccagctc catcaaacat ttcctcctga tgaaattttg ggtccctact aggtatttgt
•      121 ttagctgtac aaatcttaac aggactgttc ctagcaatac attatacatc agataccaca
•      181 accgccttct actctgttac ccatatctgc cgagacgtaa attacggctg aatcctacgt
•      241 tacctccatg ccaacggagc atccatattc ttcatctgcc tatttataca tgtaggccga
•      301 ggcatctatt acggctcata cctattcaca gaaacatgaa acattggcat tatccttcta
•      361 ttcgccgtaa tagcaacagc attcataggc tatgtcctcc ca
•//
```

FEATURES

Detailed description of all genes, position, translation …

Position of the gene in the DNA sequence

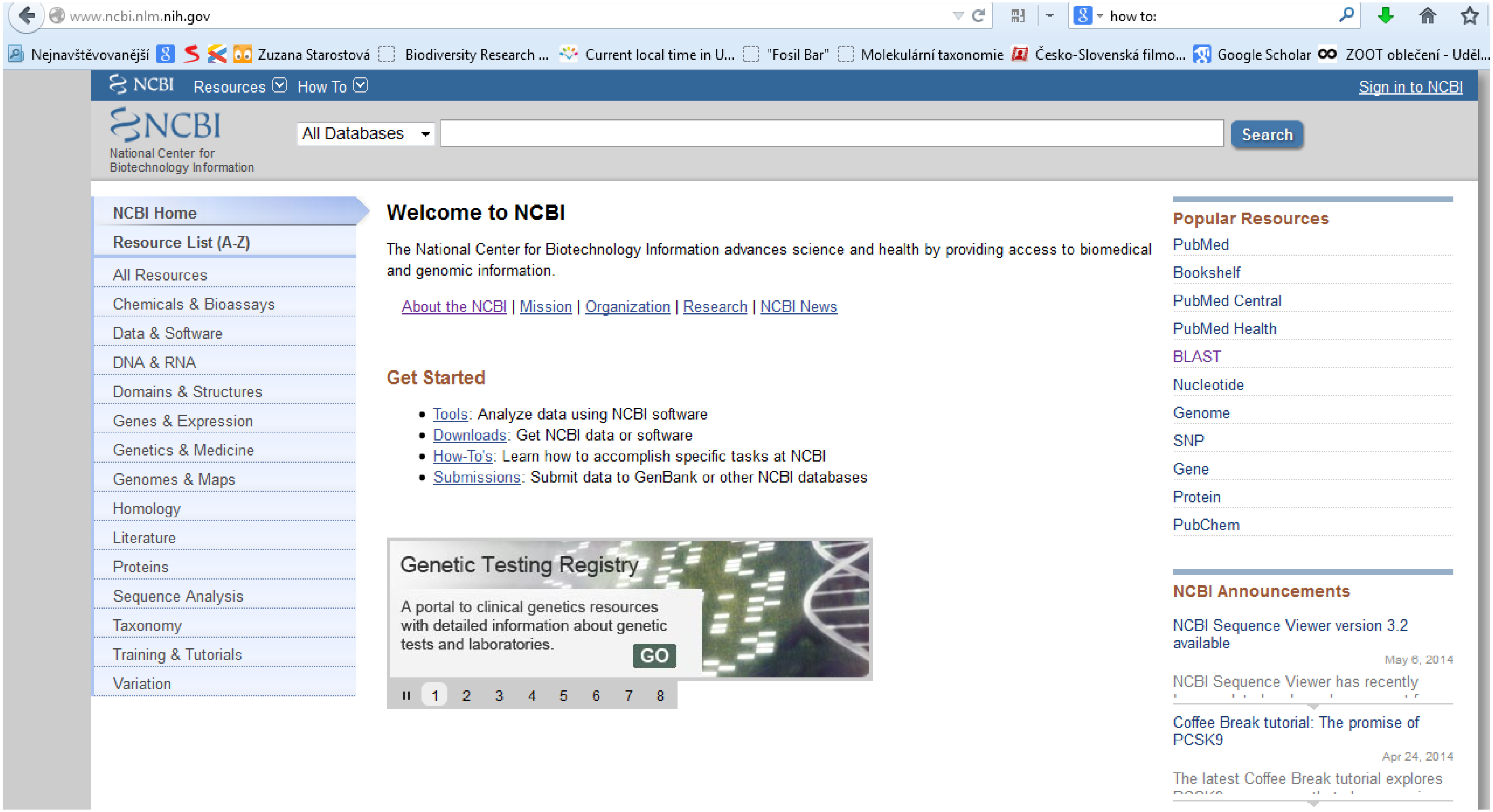
```
•LOCUS       EF443167                 402 bp    DNA     linear   MAM 15-OCT-2007
•DEFINITION  Rhinopoma hardwickei haplotype 2949 cytochrome b gene, partial cds;
•            mitochondrial.
•ACCESSION   EF443167
•VERSION     EF443167.1  GI:148832288
•KEYWORDS    .
•SOURCE      mitochondrion Rhinopoma hardwickii (Lesser mouse-tailed bat)
•  ORGANISM  Rhinopoma hardwickii
•            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
•            Mammalia; Eutheria; Laurasiatheria; Chiroptera; Microchiroptera;
•            Rhinopomatidae; Rhinopoma.
•REFERENCE   1  (bases 1 to 402)
•  AUTHORS   Hulva,P., Horacek,I. and Benda,P.
•  TITLE     Molecules, morphometrics and new fossils provide an integrated view
•            of the evolutionary history of Rhinopomatidae (Mammalia:
•            Chiroptera)
•  JOURNAL   BMC Evol. Biol. 7, 165 (2007)
•   PUBMED   17868440
•  REMARK    Publication Status: Online-Only
•REFERENCE   2  (bases 1 to 402)
•  AUTHORS   Hulva,P.
•  TITLE     Direct Submission
•  JOURNAL   Submitted (18-FEB-2007) Zoology, Charles University, Vinicna 7,
•            Prague, 2 12843, Czech Republic
•FEATURES             Location/Qualifiers
•     source          1..402
•                     /organism="Rhinopoma hardwickii"
•                     /organelle="mitochondrion"
•                     /mol_type="genomic DNA"
•                     /db_xref="taxon:124756"
•                     /haplotype="2949"
•     CDS             1..>402
•                     /codon_start=1
•                     /transl_table=2
•                     /product="cytochrome b"
•                     /protein_id="ABR14084.1"
•                     /db_xref="GI:148832289"
•                     /translation="MTHIRKSHPLFKIINDSFIDLPAPSNISSWWNFGSLLGICLAVQ
•                     ILTGLFLAMHYTSDTTTAFYSVTHICRDVNYGWILRYLHANGASMFFICLFMHVGRGI
•                     YYGSYLFTETWNIGIILLFAVMATAFMGYVLP"
•ORIGIN
•        1 atgacccaca tccgaaaatc ccacccctta ttcaaaatta tcaacgactc attcatcgac
•       61 ctaccagctc catcaaacat ttcctcctga tgaaattttg ggtccctact aggtatttgt
•      121 ttagctgtac aaatcttaac aggactgttc ctagcaatac attatacatc agataccaca
•      181 accgccttct actctgttac ccatatctgc cgagacgtaa attacggctg aatcctacgt
•      241 tacctccatg ccaacggagc atccatattc ttcatctgcc tatttataca tgtaggccga
•      301 ggcatctatt acggctcata cctattcaca gaaacatgaa acattggcat tatccttcta
•      361 ttcgccgtaa tagcaacagc attcataggc tatgtcctcc ca
•//
```

ORIGIN

sequence

```
•LOCUS       EF443167                 402 bp    DNA     linear   MAM 15-OCT-2007
•DEFINITION  Rhinopoma hardwickei haplotype 2949 cytochrome b gene, partial cds;
•            mitochondrial.
•ACCESSION   EF443167
•VERSION     EF443167.1  GI:148832288
•KEYWORDS    .
•SOURCE      mitochondrion Rhinopoma hardwickii (Lesser mouse-tailed bat)
•  ORGANISM  Rhinopoma hardwickii
•            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
•            Mammalia; Eutheria; Laurasiatheria; Chiroptera; Microchiroptera;
•            Rhinopomatidae; Rhinopoma.
•REFERENCE   1  (bases 1 to 402)
•  AUTHORS   Hulva,P., Horacek,I. and Benda,P.
•  TITLE     Molecules, morphometrics and new fossils provide an integrated view
•            of the evolutionary history of Rhinopomatidae (Mammalia:
•            Chiroptera)
•  JOURNAL   BMC Evol. Biol. 7, 165 (2007)
•   PUBMED   17868440
•  REMARK    Publication Status: Online-Only
•REFERENCE   2  (bases 1 to 402)
•  AUTHORS   Hulva,P.
•  TITLE     Direct Submission
•  JOURNAL   Submitted (18-FEB-2007) Zoology, Charles University, Vinicna 7,
•            Prague, 2 12843, Czech Republic
•FEATURES             Location/Qualifiers
•     source          1..402
•                     /organism="Rhinopoma hardwickii"
•                     /organelle="mitochondrion"
•                     /mol_type="genomic DNA"
•                     /db_xref="taxon:124756"
•                     /haplotype="2949"
•     CDS             1..>402
•                     /codon_start=1
•                     /transl_table=2
•                     /product="cytochrome b"
•                     /protein_id="ABR14084.1"
•                     /db_xref="GI:148832289"
•                     /translation="MTHIRKSHPLFKIINDSFIDLPAPSNISSWWNFGSLLGICLAVQ
•                     ILTGLFLAMHYTSDTTTAFYSVTHICRDVNYGWILRYLHANGASMFFICLFMHVGRGI
•                     YYGSYLFTETWNIGIILLFAVMATAFMGYVLP"
•ORIGIN
•        1 atgacccaca tccgaaaatc ccacccctta ttcaaatta tcaacgactc attcatcgac
•       61 ctaccagctc catcaaacat ttcctcctga tgaaattttg ggtccctact aggtatttgt
•      121 ttagctgtac aaatcttaac aggactgttc ctagcaatac attatacatc agataccaca
•      181 accgccttct actctgttac ccatatctgc cgagacgtaa attacggctg aatcctacgt
•      241 tacctccatg ccaacggagc atccatattc ttcatctgcc tatttataca tgtaggccga
•      301 ggcatctatt acggctcata cctattcaca gaaacatgaa acattggcat tatccttcta
•      361 ttcgccgtaa tagcaacagc attcataggc tatgtcctcc ca
•//
```

http://www.ncbi.nlm.nih.gov/



Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

About the NCBI | Mission | Organization | Research | NCBI News

Get Started

- Tools: Analyze data using NCBI software
- Downloads: Get NCBI data or software
- How-To's: Learn how to accomplish specific tasks at NCBI
- Submissions: Submit data to GenBank or other NCBI databases

Genetic Testing Registry

A portal to clinical genetics resources with detailed information about genetic tests and laboratories.

Popular Resources

- PubMed
- Bookshelf
- PubMed Central
- PubMed Health
- BLAST
- Nucleotide
- Genome
- SNP
- Gene
- Protein
- PubChem

NCBI Announcements

NCBI Sequence Viewer version 3.2 available

May 6, 2014

NCBI Sequence Viewer has recently

Coffee Break tutorial: The promise of PCSK9

Apr 24, 2014

The latest Coffee Break tutorial explores

Exercise 1

- Search for the cytochrome b sequence of all mammoths which were sequenced
- Export the protein-coding part to FASTA and save

How-to:

- GenBank on the webpage of the NCBI – search in "*Nucleotide*" - GenBank + RefSeq
- Search for "*Mammuthus*"
- Too many sequences – restrict it to the non-redundant database RefSeq

RefSeq: NCBI Reference Sequence Database

A comprehensive, integrated, non-redundant, well-annotated set of reference sequences including genomic, transcript, and protein.

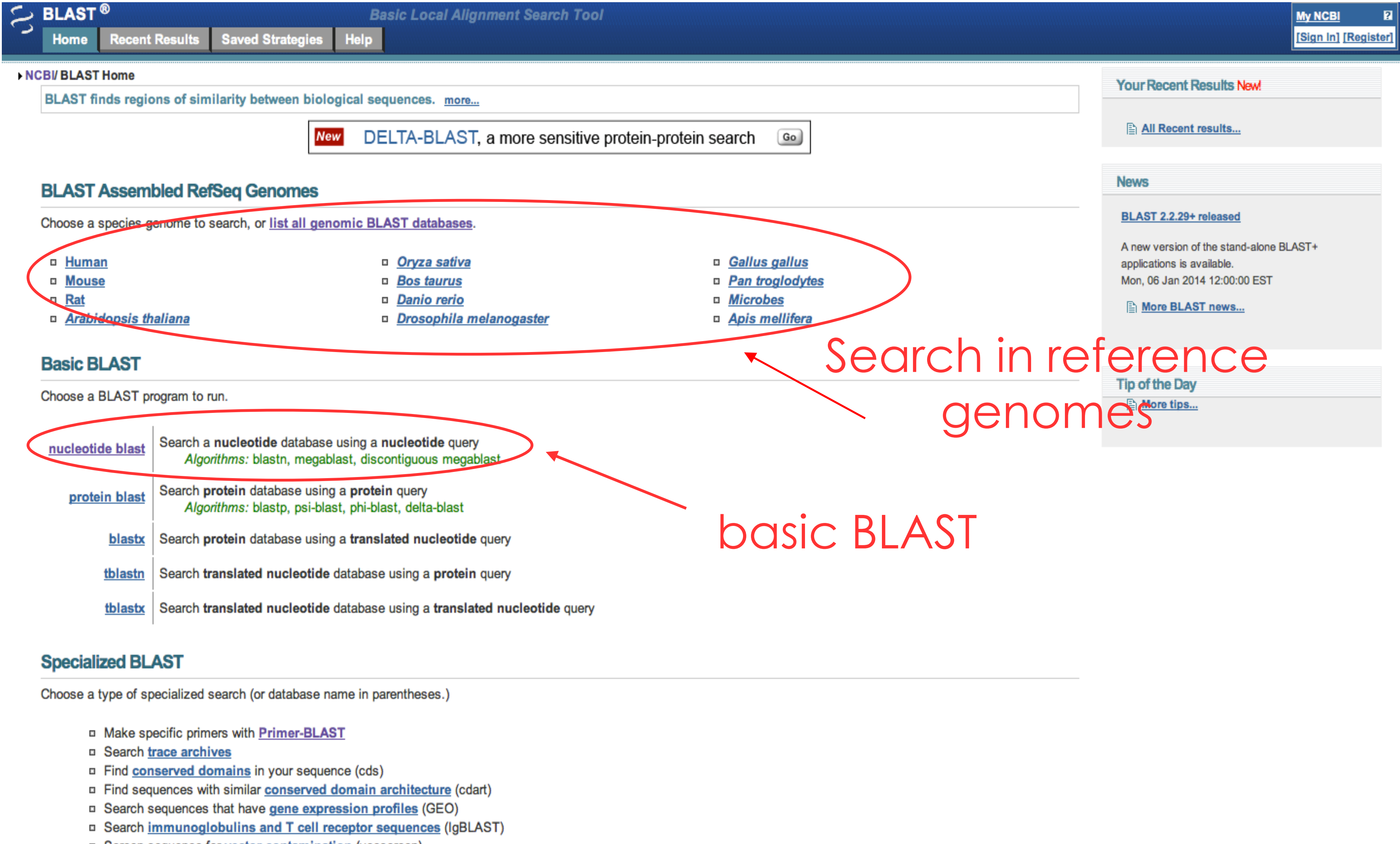
- Whole mtgenome – GenBank format - SOURCE – CDS – list cytochrome b

BLAST

Basic Local Alignment Search Tool

- finds regions of local similarity between sequences
- The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches.
- BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

BLAST



NCBI/ BLAST Home

BLAST finds regions of similarity between biological sequences. more...

DELTA-BLAST, a more sensitive protein-protein search

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.

nucleotide blast	Search a **nucleotide** database using a **nucleotide** query *Algorithms:* blastn, megablast, discontiguous megablast
protein blast	Search **protein** database using a **protein** query *Algorithms:* blastp, psi-blast, phi-blast, delta-blast
blastx	Search **protein** database using a **translated nucleotide** query
tblastn	Search **translated nucleotide** database using a **protein** query
tblastx	Search **translated nucleotide** database using a **translated nucleotide** query

Specialized BLAST

Choose a type of specialized search (or database name in parentheses.)

- Make specific primers with Primer-BLAST
- Search trace archives
- Find conserved domains in your sequence (cds)
- Find sequences with similar conserved domain architecture (cdart)
- Search sequences that have gene expression profiles (GEO)
- Search immunoglobulins and T cell receptor sequences (IgBLAST)

Your Recent Results New!

All Recent results...

News

BLAST 2.2.29+ released

A new version of the stand-alone BLAST+ applications is available.
Mon, 06 Jan 2014 12:00:00 EST

More BLAST news...

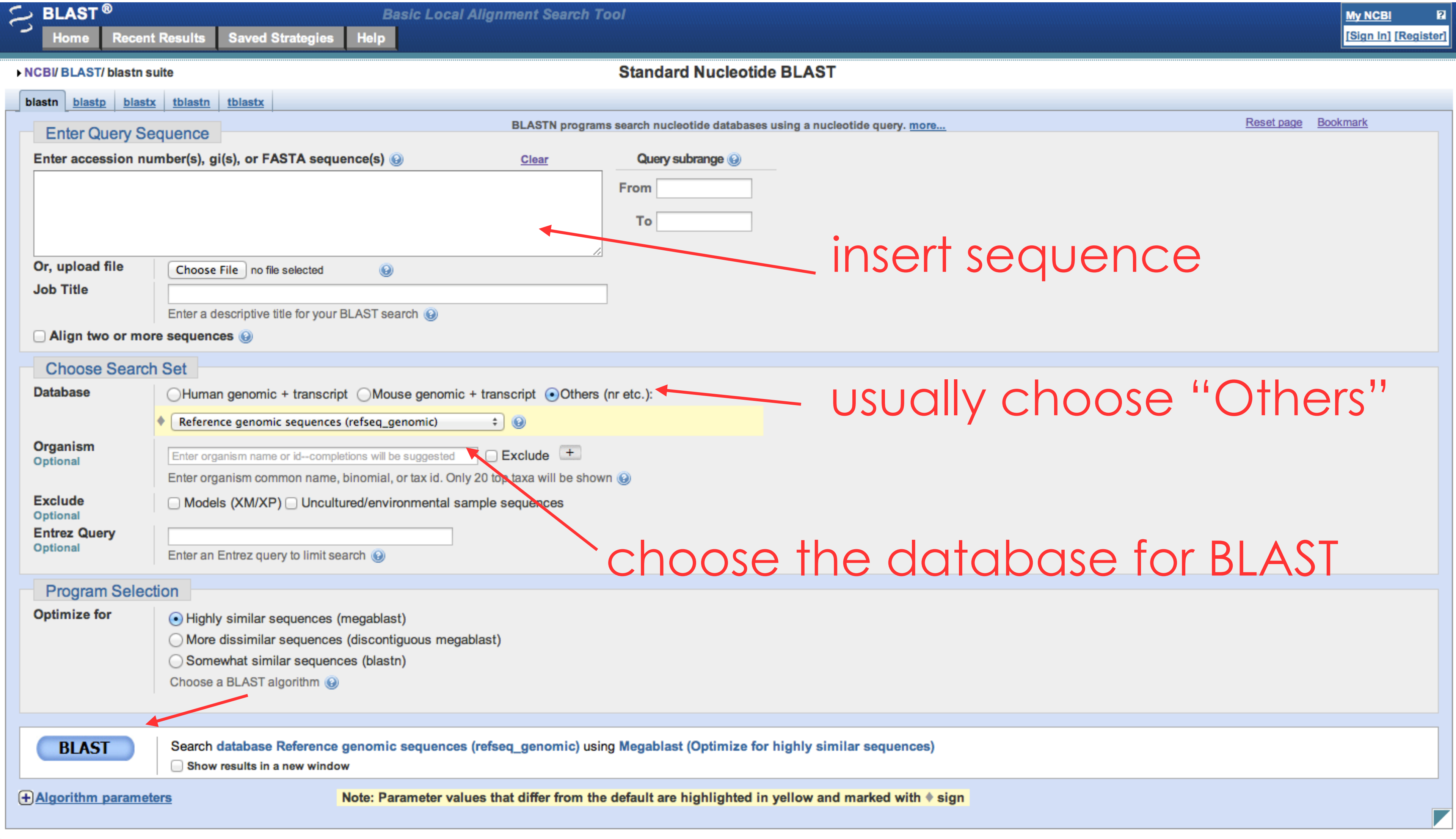
Tip of the Day

More tips...

Search in reference genomes

basic BLAST

BLAST



Standard Nucleotide BLAST

BLASTN programs search nucleotide databases using a nucleotide query. more...

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

Query subrange

From

To

Or, upload file: no file selected

Job Title

Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

Database: ○ Human genomic + transcript ○ Mouse genomic + transcript ● Others (nr etc.):

Reference genomic sequences (refseq_genomic)

Organism (Optional): Enter organism name or id--completions will be suggested ☐ Exclude

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown

Exclude (Optional): ☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Entrez Query (Optional)

Enter an Entrez query to limit search

Program Selection

Optimize for:

● Highly similar sequences (megablast)
○ More dissimilar sequences (discontiguous megablast)
○ Somewhat similar sequences (blastn)

Choose a BLAST algorithm

Search database Reference genomic sequences (refseq_genomic) using Megablast (Optimize for highly similar sequences)

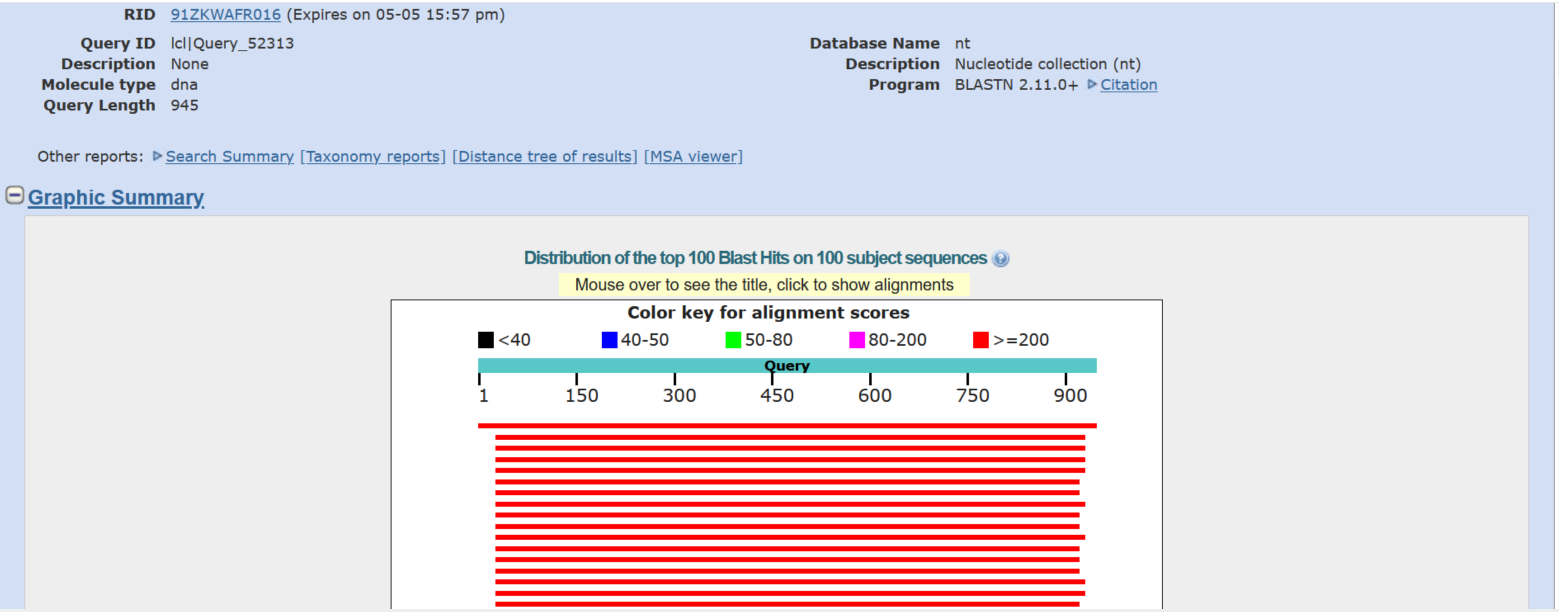
☐ Show results in a new window

Note: Parameter values that differ from the default are highlighted in yellow and marked with ♦ sign

insert sequence

usually choose “Others”

choose the database for BLAST



RID 91ZKWAFR016 (Expires on 05-05 15:57 pm)

Query ID lcl|Query_52313
Description None
Molecule type dna
Query Length 945

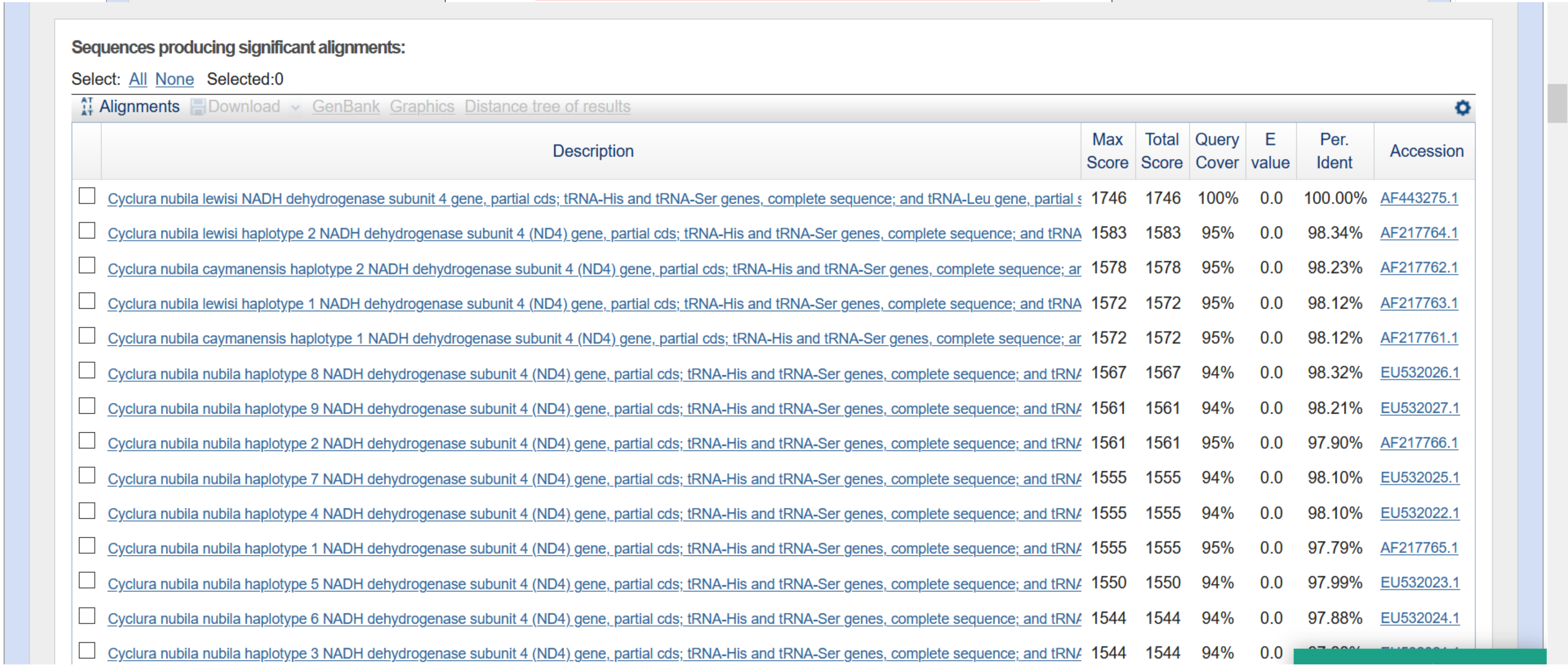
Database Name nt
Description Nucleotide collection (nt)
Program BLASTN 2.11.0+ ▷ Citation

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

Graphic Summary

Distribution of the top 100 Blast Hits on 100 subject sequences

Mouse over to see the title, click to show alignments



Sequences producing significant alignments:

Select: All None Selected:0

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
☐	Cyclura nubila lewisi NADH dehydrogenase subunit 4 gene, partial cds; tRNA-His and tRNA-Ser genes, complete sequence; and tRNA-Leu gene, partial s	1746	1746	100%	0.0	100.00%	AF443275.1
☐	Cyclura nubila lewisi haplotype 2 NADH dehydrogenase subunit 4 (ND4) gene, partial cds; tRNA-His and tRNA-Ser genes, complete sequence; and tRNA	1583	1583	95%	0.0	98.34%	AF217764.1
☐	Cyclura nubila caymanensis haplotype 2 NADH dehydrogenase subunit 4 (ND4) gene, partial cds; tRNA-His and tRNA-Ser genes, complete sequence; ar	1578	1578	95%	0.0	98.23%	AF217762.1
☐	Cyclura nubila lewisi haplotype 1 NADH dehydrogenase subunit 4 (ND4) gene, partial cds; tRNA-His and tRNA-Ser genes, complete sequence; and tRNA	1572	1572	95%	0.0	98.12%	AF217763.1
☐	Cyclura nubila caymanensis haplotype 1 NADH dehydrogenase subunit 4 (ND4) gene, partial cds; tRNA-His and tRNA-Ser genes, complete sequence; ar	1572	1572	95%	0.0	98.12%	AF217761.1
☐	Cyclura nubila nubila haplotype 8 NADH dehydrogenase subunit 4 (ND4) gene, partial cds; tRNA-His and tRNA-Ser genes, complete sequence; and tRN/	1567	1567	94%	0.0	98.32%	EU532026.1
☐	Cyclura nubila nubila haplotype 9 NADH dehydrogenase subunit 4 (ND4) gene, partial cds; tRNA-His and tRNA-Ser genes, complete sequence; and tRN/	1561	1561	94%	0.0	98.21%	EU532027.1
☐	Cyclura nubila nubila haplotype 2 NADH dehydrogenase subunit 4 (ND4) gene, partial cds; tRNA-His and tRNA-Ser genes, complete sequence; and tRN/	1561	1561	95%	0.0	97.90%	AF217766.1
☐	Cyclura nubila nubila haplotype 7 NADH dehydrogenase subunit 4 (ND4) gene, partial cds; tRNA-His and tRNA-Ser genes, complete sequence; and tRN/	1555	1555	94%	0.0	98.10%	EU532025.1
☐	Cyclura nubila nubila haplotype 4 NADH dehydrogenase subunit 4 (ND4) gene, partial cds; tRNA-His and tRNA-Ser genes, complete sequence; and tRN/	1555	1555	94%	0.0	98.10%	EU532022.1
☐	Cyclura nubila nubila haplotype 1 NADH dehydrogenase subunit 4 (ND4) gene, partial cds; tRNA-His and tRNA-Ser genes, complete sequence; and tRN/	1555	1555	95%	0.0	97.79%	AF217765.1
☐	Cyclura nubila nubila haplotype 5 NADH dehydrogenase subunit 4 (ND4) gene, partial cds; tRNA-His and tRNA-Ser genes, complete sequence; and tRN/	1550	1550	94%	0.0	97.99%	EU532023.1
☐	Cyclura nubila nubila haplotype 6 NADH dehydrogenase subunit 4 (ND4) gene, partial cds; tRNA-His and tRNA-Ser genes, complete sequence; and tRN/	1544	1544	94%	0.0	97.88%	EU532024.1
☐	Cyclura nubila nubila haplotype 3 NADH dehydrogenase subunit 4 (ND4) gene, partial cds; tRNA-His and tRNA-Ser genes, complete sequence; and tRN/	1544	1544	94%	0.0	[illegible]	[illegible]

Exercise 2

- Blast this sequence:

```
TGACTACCAAAAGCTCATGTAGAAGCCCCAATTGCCGGCTCCATAGTACTAGCCGCCATCCTACTAAAAC
TGGGAGGCTACGGAATCATCCGAATCACCCTAACATTAATGCCAATAACACAAAAACTATACTACTCCTT
CATGATCTTAGCCCTATGAGGAATCGTAATGACCAGCTCAATCTGCATACGACAAACGGACCTAAAATCG
CTCATCGCCTACTCCTCAGTCAGCCACATAGGACTAGTCATCGCTGCCTGCCTAATCCAAACACCATGAA
GCATTACAGGAGCTATAACCCTAATAATTGCCCATGGTTTAACATCTTCCATAATATTCTGCCTAGCCAA
CACCAACTACGAACGAACCCACACCCGAACCATAATCCTAACACGAGGCCTCCAAATCATCCTTCCCCTA
ATAACCTCATGATGGCTACTAGCCATCCTCACAAACATAGCCCTGCCCCCGACTATTAACATAATAGGAG
AAATTATGATCATCTCCGCCCTATTCAATTGATCCTCACCAACAATTATCTTAACAGGCCTAGGAACCCT
AATCACAGCCGTGTACTCATTACACATATTCCTAACAACACAACGAAACATCATGCCACTTCACATCATC
ACCACAGACCCAACACACACCCGAGAACATCTCATCATAGCCCTACACACTCTACCCCTCGTTCTACTAA
TCCTTAAACCCGATCTAATCTCAAACTACTTCGCCTGTTAGCATAGTTTAATAAAAACATTAGGATGTGG
CCCTAAAAAAAGAAGTGCAACCCTTCTTGCGAACCGGAGAGGTGTTTAGAACACTAAGAACTGCTAATTC
TTATCCCTGAAGTTAAATTCCTCAGACCCCTCACTTTTAAAGGATAAAAGCAACCATTGGTTTTAGGCAC
CAAAAATCTTGGTGCAAATCCAAGTAAAAGTAATG
```

- create multiple FASTA file from selected sequences

How-to:

- BLAST at the NCBI – „nucleotide blast" option - "standard databases"
- Select sequences and download FASTA (aligned sequences)

Species identification

- **DNA barcoding**

  – taxon identification
  using a standardized DNA region

Pixmac.cz 66281685

Genebank (NCBI)
http://www.ncbi.nlm.nih.gov/genbank/

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**.

SEQUENCE: COI-5P [Funding Source: N/A]

Sequence ID:	ABCMA692-07.COI-5P	GenBank Accession:	JF445285
Last Updated:	2013-02-12	Genome:	Mitochondrial
Locus:	Cytochrome Oxidase Subunit 1 5' Region		
Nucleotides:	657 bp		

```
ACCCTCTATCTATTATTTGGTGCCTGAGCAGGAATAGTAGGAACAGCCTTGAGCATTCTAATTCGAGCTGAACTA
GGACAACCAGGAGCACTCCTAGGCGATGACCAAATTTATAATGTCATTGTTACAGCCCATGCATTCGTAATAATT
TTCTTTATAGTTATGCCTATGATAATCGGAGGCTTCGGAAACTGGCTTGTACCACTAATGATTGGAGCCCCTGAT
ATAGCATTCCCACGAATAAACAATATAAGCTTTTGATTGCTTCCCCCATCATTTTTACTCCTTTTAGCATCATCT
ATAGTAGAAGCCGGAGCCGGAACAGGATGAACAGTATACCCACCCTTAGCCGGTAACCTAGCCCATGCCGGAGCA
TCCGTTGACCTAACCATTTTCTCCCTTCACCTAGCTGGTGTATCCTCTATCTTAGGAGCTATTAATTTTATCACC
ACTATCATCAACATAAAACCCCCTGCTATAACCCAATATCAGACCCCTCTATTTGTGTGATCCGTATTAATTACA
GCTGTACTTCTACTTCTTTCACTACCAGTTTTAGCAGCAGGCATTACCATACTCCTCACAGATCGAAACCTAAAT
ACTACTTTTTTTGATCCTGCTGGAGGCGGAGATCCAATTCTCTATCAACATCTATTT
```

Amino Acids:

```
TLYLLFGAWAGMVGTALSILIRAELGQPGALLGDDQIYNVIVTAHAFVMIFFMVMPMMIGGFGNWLVPLMIGAPD
MAFPRMNNMSFWLLPPSFLLLLASSMVEAGAGTGWTVYPPLAGNLAHAGASVDLTIFSLHLAGVSSILGAINFIT
TIINMKPPAMTQYQTPLFVWSVLITAVLLLLSLPVLAAGITMLLTDRNLNTTFFDPAGGGDPILYQHLF
```

Illustrative Barcode:

Exercise 3:

- I was at the sushi restaurant yesterday
- some sushi tasted funny
- I thought of my lecture from the course Molecular Applications in Zoology and took samples
- I used DNA barcoding (COI from mtDNA)

I obtained 3 different sequences. What did I eat?

Since I sequenced COI, we can use specialized barcoding database **BOLD.**

http://www.boldsystems.org/

```
>sample1
GCAGGAGCATCCGTCGACTTAACTATCTTCTCCCTTCATTTAGCTGGA
ATCTCATCAATTTTAGGGGCCATTAATTTTATTACGACCATTATCAACAT
AAAACCACCGGCAATCTCTCAGTACCAAACCCCACTTTTTGTTTGAG
CTGTGCTAATCACTGCTGTACTTCTACTACTATCCCTCCCCGTTCTGG
CAGCAGGTATCACTATGTTGCTCACGGACCGAAATTTAAACACTACT
TTCTTTGACCCAGCGGGGGGCGGAGATCCAATTTTATACCAACACC
TCTTTTGATTCTTCGGTCACCCAGAAGTGTATATTCTTATCCTTCCAGGC
TTTGGCATAATTTCACACATCGTTGCATACTACTCCGGTAAGAAAGAA
CCCTTCGGGTACATGGGAATAGTATGAGCTATAATGGCCATCGGCT
TGTTAGGATTTATCGTTTGAGCCCACCACATGTTCACTGTCGGGATGG

>sample2
GAACGGGATGAACCGTATACCCCCCACTGGCTGGCAATCTGGCC
CATGCAGGAGCATCCGTTGACCTTACAATTTTCTCCTTACACTTAGCC
GGAGTCTCTTCTATTTTAGGGGCAATTAATTTCATCACTACTATTATCAA
CATAAAACCCCCTGCAATATCCCAGTATCAAACTCCCCTGTTTGTAT
GATCAGTACTAATTACAGCAGTTCTACTCTTACTATCCCTGCCTGTACT
GGCTGCTGGAATTACAATACTTTTAACAGACCGGAATCTTAATACAA
CATTTTTTGATCCCGCTGGAGGAGGAGACCCTATCCTATATCAACAC
CTATTC

>sample3
GAACAGGATGAACAGTATATCCCCCCTTAGCCGGAAACCTAGCC
CATGCTGGGGCATCCGTAGATTTAACTATTTTTTCCCTCCACCTAGCC
GGGGTGTCTTCTATCTTAGGAGCTATCAACTTTATCACCACTATCATTA
ATATAAAACCCCCTGCTATAACCCAATATCAGACACCTCTCTTTGTAT
GATCCGTACTAATTACAGCCGTCCTACTACTTCTCTCACTGCCAGTAT
TAGCAGCAGGTATCACTATACTCCTTACAGACCGAAATCTAAATACT
ACTTTCTTCGACCCCCGCTGGAGGTGGAGACCCAATTCTTTATCAACA
CCTATTC
```

How to make phylogenetic trees?

Workflow:

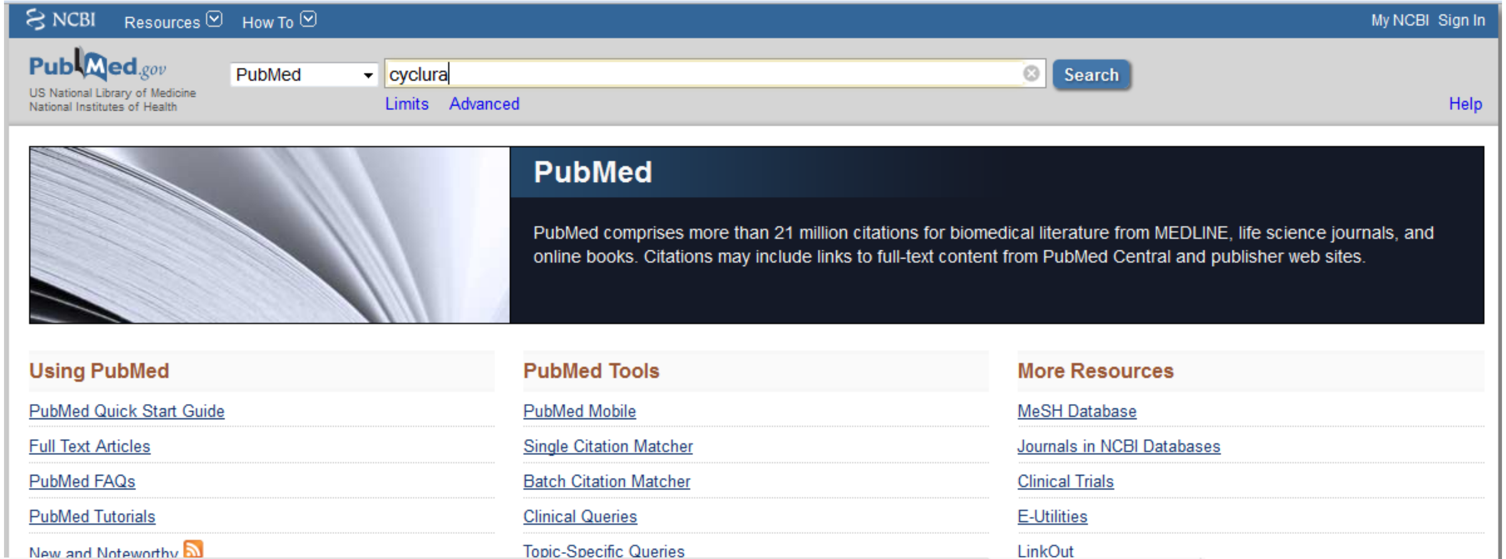
- obtain DNA sequence
- quality check
- sequence alignment
- calculating genetic distances
- phylogeny estimation – topology and branch length
- reliability test (bootstrap)
- tree visualization

Sequence search in the GenBank

- I would like to study rock iguanas (genus *Cyclura*). Are there some sequences in the GenBank?

NCBI

V NCBI http://www.ncbi.nlm.nih.gov/



PubMed

PubMed comprises more than 21 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full-text content from PubMed Central and publisher web sites.

Using PubMed

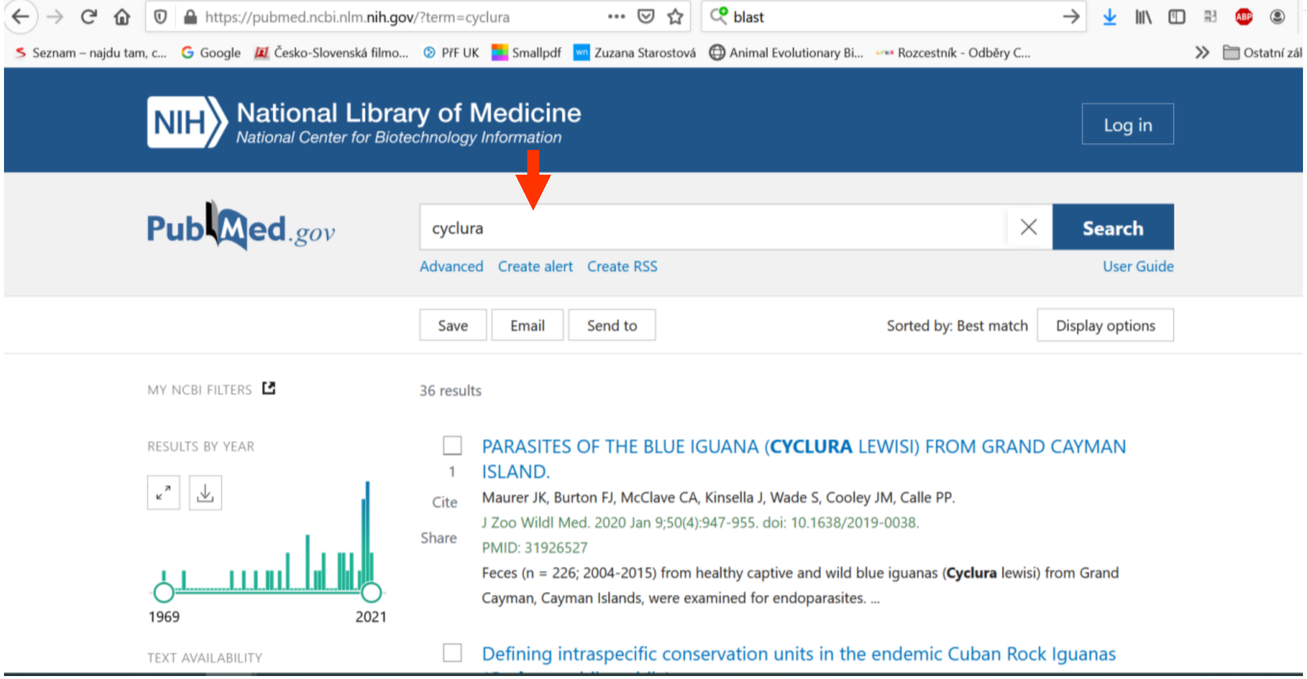
- PubMed Quick Start Guide
- Full Text Articles
- PubMed FAQs
- PubMed Tutorials
- New and Noteworthy

PubMed Tools

- PubMed Mobile
- Single Citation Matcher
- Batch Citation Matcher
- Clinical Queries
- Topic-Specific Queries

More Resources

- MeSH Database
- Journals in NCBI Databases
- Clinical Trials
- E-Utilities
- LinkOut

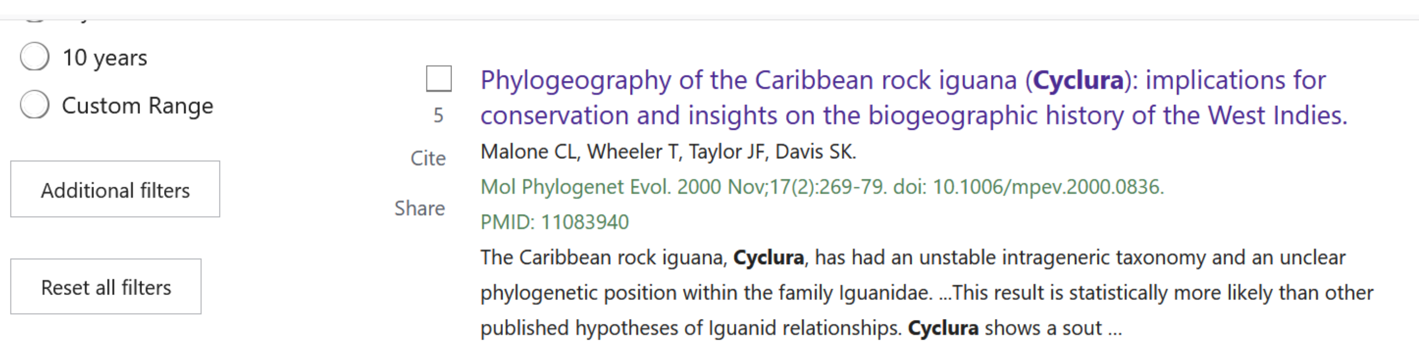


36 results

☐ 1 PARASITES OF THE BLUE IGUANA (**CYCLURA** LEWISI) FROM GRAND CAYMAN ISLAND.
Maurer JK, Burton FJ, McClave CA, Kinsella J, Wade S, Cooley JM, Calle PP.
J Zoo Wildl Med. 2020 Jan 9;50(4):947-955. doi: 10.1638/2019-0038.
PMID: 31926527
Feces (n = 226; 2004-2015) from healthy captive and wild blue iguanas (**Cyclura** lewisi) from Grand Cayman, Cayman Islands, were examined for endoparasites. ...

☐ Defining intraspecific conservation units in the endemic Cuban Rock Iguanas

Cyclura



○ 10 years

○ Custom Range

Additional filters

Reset all filters

☐ 5 Cite Share

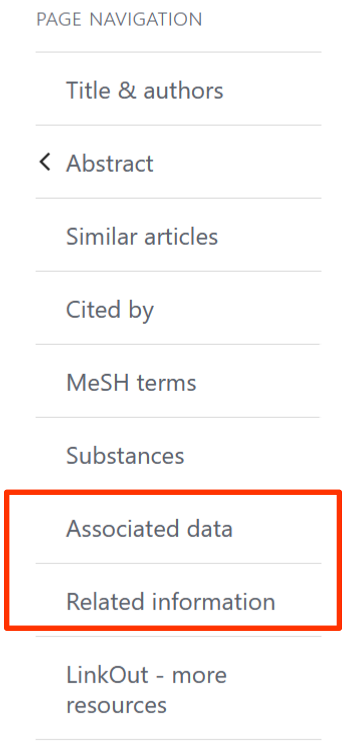
Phylogeography of the Caribbean rock iguana (**Cyclura**): implications for conservation and insights on the biogeographic history of the West Indies.

Malone CL, Wheeler T, Taylor JF, Davis SK.

Mol Phylogenet Evol. 2000 Nov;17(2):269-79. doi: 10.1006/mpev.2000.0836.

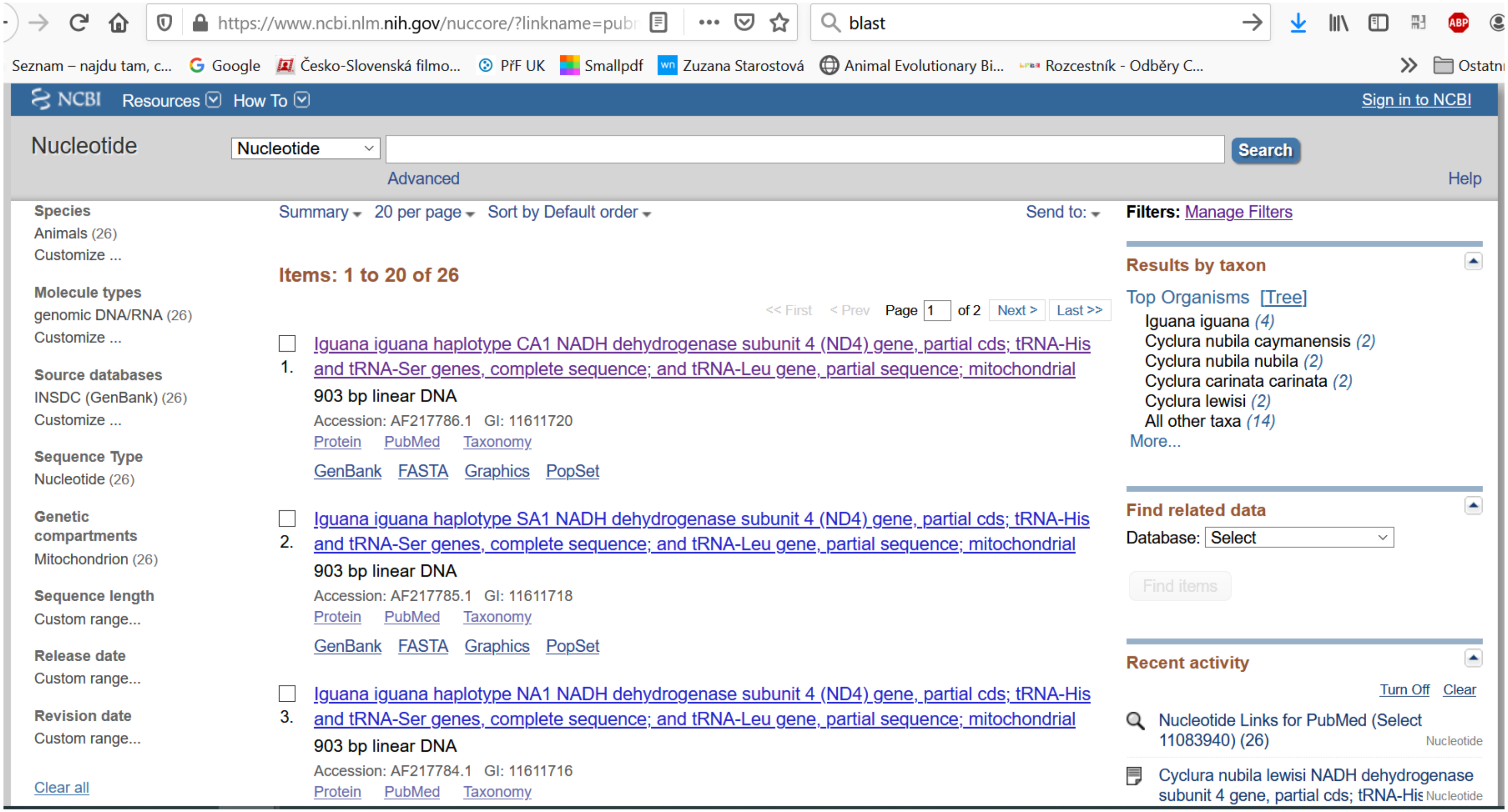
PMID: 11083940

The Caribbean rock iguana, **Cyclura**, has had an unstable intrageneric taxonomy and an unclear phylogenetic position within the family Iguanidae. ...This result is statistically more likely than other published hypotheses of Iguanid relationships. **Cyclura** shows a sout ...



PAGE NAVIGATION

- Title & authors
- Abstract
- Similar articles
- Cited by
- MeSH terms
- Substances
- Associated data
- Related information
- LinkOut - more resources



Items: 1 to 20 of 26

☐ 1. Iguana iguana haplotype CA1 NADH dehydrogenase subunit 4 (ND4) gene, partial cds; tRNA-His and tRNA-Ser genes, complete sequence; and tRNA-Leu gene, partial sequence; mitochondrial
903 bp linear DNA
Accession: AF217786.1 GI: 11611720
Protein PubMed Taxonomy
GenBank FASTA Graphics PopSet

☐ 2. Iguana iguana haplotype SA1 NADH dehydrogenase subunit 4 (ND4) gene, partial cds; tRNA-His and tRNA-Ser genes, complete sequence; and tRNA-Leu gene, partial sequence; mitochondrial
903 bp linear DNA
Accession: AF217785.1 GI: 11611718
Protein PubMed Taxonomy
GenBank FASTA Graphics PopSet

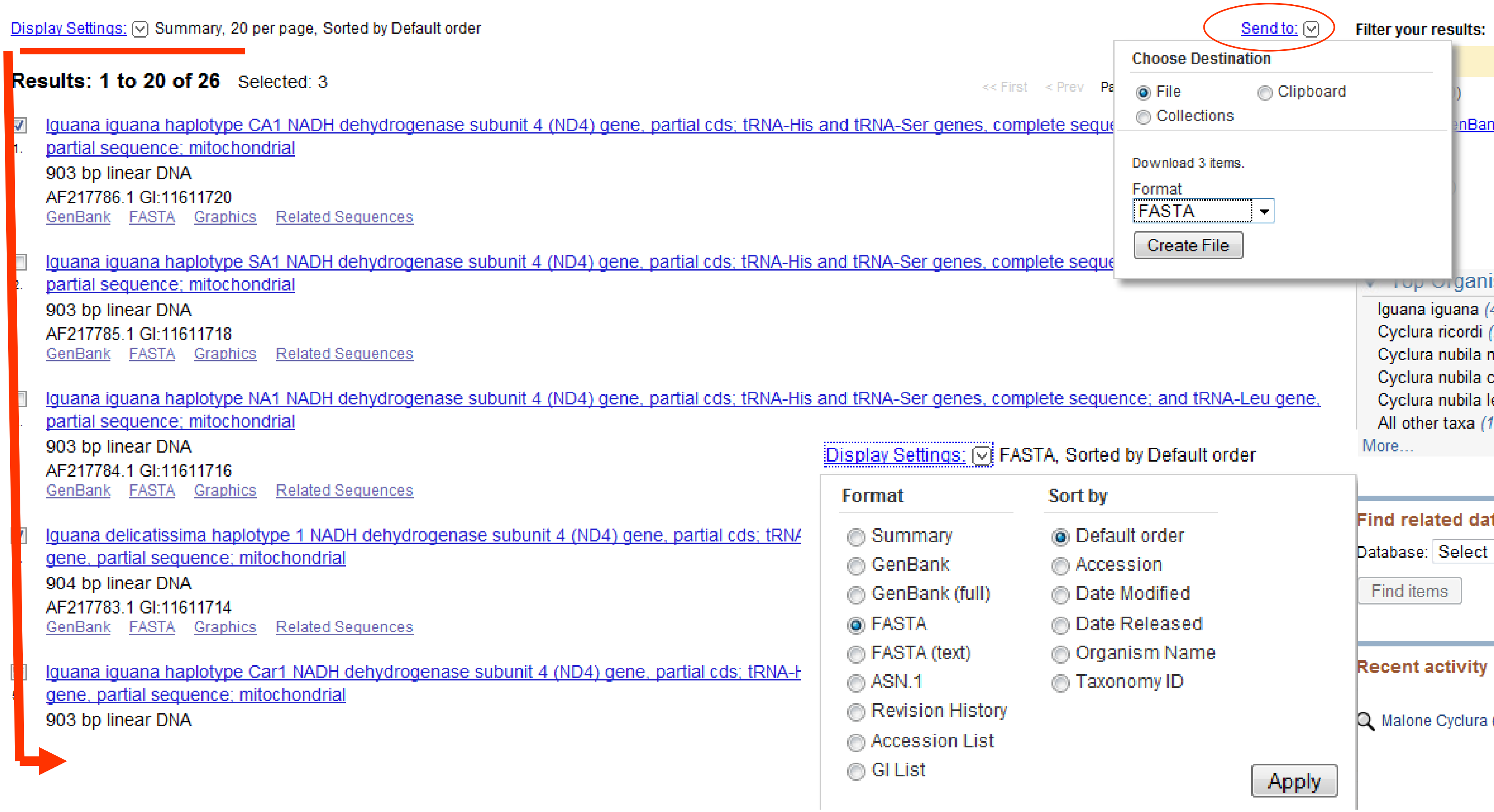
☐ 3. Iguana iguana haplotype NA1 NADH dehydrogenase subunit 4 (ND4) gene, partial cds; tRNA-His and tRNA-Ser genes, complete sequence; and tRNA-Leu gene, partial sequence; mitochondrial
903 bp linear DNA
Accession: AF217784.1 GI: 11611716
Protein PubMed Taxonomy

Top Organisms [Tree]
Iguana iguana (4)
Cyclura nubila caymanensis (2)
Cyclura nubila nubila (2)
Cyclura carinata carinata (2)
Cyclura lewisi (2)
All other taxa (14)
More...

Nucleotide Links for PubMed (Select 11083940) (26) Nucleotide

Cyclura nubila lewisi NADH dehydrogenase subunit 4 gene, partial cds; tRNA-His Nucleotide

Choose 7 sequences of the *Cyclura* species + 1 *Iguana iguana*



Display Settings: Summary, 20 per page, Sorted by Default order

Send to:

Filter your results:

Choose Destination

● File ○ Clipboard

○ Collections

Download 3 items.

Format

FASTA

Results: 1 to 20 of 26 Selected: 3

☑ 1. Iguana iguana haplotype CA1 NADH dehydrogenase subunit 4 (ND4) gene, partial cds; tRNA-His and tRNA-Ser genes, complete seque
partial sequence; mitochondrial
903 bp linear DNA
AF217786.1 GI:11611720
GenBank FASTA Graphics Related Sequences

☐ Iguana iguana haplotype SA1 NADH dehydrogenase subunit 4 (ND4) gene, partial cds; tRNA-His and tRNA-Ser genes, complete seque
partial sequence; mitochondrial
903 bp linear DNA
AF217785.1 GI:11611718
GenBank FASTA Graphics Related Sequences

☐ Iguana iguana haplotype NA1 NADH dehydrogenase subunit 4 (ND4) gene, partial cds; tRNA-His and tRNA-Ser genes, complete sequence; and tRNA-Leu gene, partial sequence; mitochondrial
903 bp linear DNA
AF217784.1 GI:11611716
GenBank FASTA Graphics Related Sequences

☑ Iguana delicatissima haplotype 1 NADH dehydrogenase subunit 4 (ND4) gene, partial cds; tRN
gene, partial sequence; mitochondrial
904 bp linear DNA
AF217783.1 GI:11611714
GenBank FASTA Graphics Related Sequences

☐ Iguana iguana haplotype Car1 NADH dehydrogenase subunit 4 (ND4) gene, partial cds; tRNA-H
gene, partial sequence; mitochondrial
903 bp linear DNA

Top Organi
Iguana iguana (
Cyclura ricordi (
Cyclura nubila n
Cyclura nubila c
Cyclura nubila l
All other taxa (1
More...

Display Settings: FASTA, Sorted by Default order

Format

○ Summary
○ GenBank
○ GenBank (full)
● FASTA
○ FASTA (text)
○ ASN.1
○ Revision History
○ Accession List
○ GI List

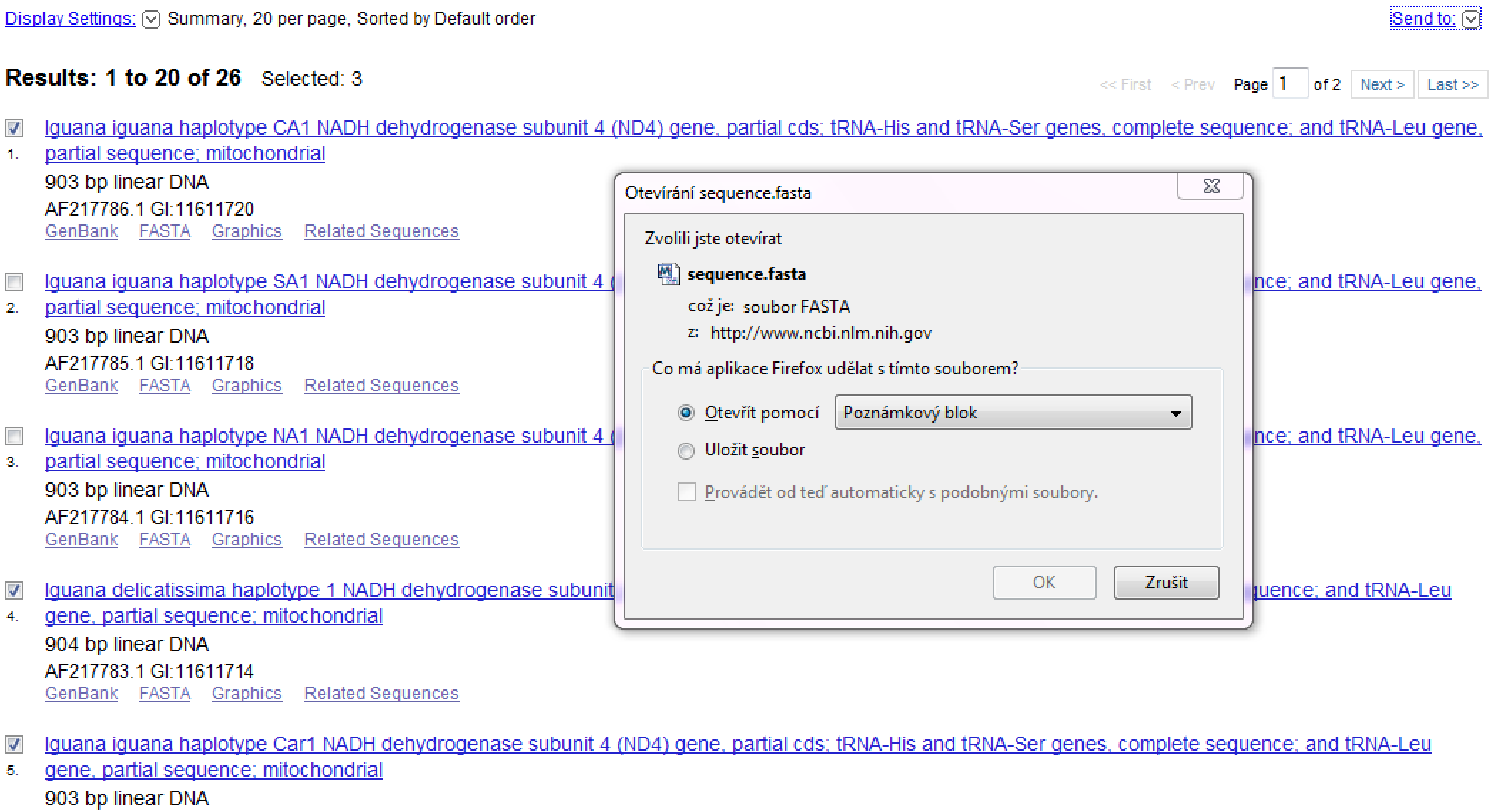
Sort by

● Default order
○ Accession
○ Date Modified
○ Date Released
○ Organism Name
○ Taxonomy ID

Find related dat
Database: Select

Recent activity
Malone Cyclura

Select sequences - choose destination Send to = File and FASTA format



Display Settings: Summary, 20 per page, Sorted by Default order

Send to:

Results: 1 to 20 of 26 Selected: 3

<< First < Prev Page 1 of 2 Next > Last >>

☑ 1. Iguana iguana haplotype CA1 NADH dehydrogenase subunit 4 (ND4) gene, partial cds; tRNA-His and tRNA-Ser genes, complete sequence; and tRNA-Leu gene, partial sequence; mitochondrial
903 bp linear DNA
AF217786.1 GI:11611720
GenBank FASTA Graphics Related Sequences

☐ 2. Iguana iguana haplotype SA1 NADH dehydrogenase subunit 4 (… nce; and tRNA-Leu gene, partial sequence; mitochondrial
903 bp linear DNA
AF217785.1 GI:11611718
GenBank FASTA Graphics Related Sequences

☐ 3. Iguana iguana haplotype NA1 NADH dehydrogenase subunit 4 (… nce; and tRNA-Leu gene, partial sequence; mitochondrial
903 bp linear DNA
AF217784.1 GI:11611716
GenBank FASTA Graphics Related Sequences

☑ 4. Iguana delicatissima haplotype 1 NADH dehydrogenase subunit … uence; and tRNA-Leu gene, partial sequence; mitochondrial
904 bp linear DNA
AF217783.1 GI:11611714
GenBank FASTA Graphics Related Sequences

☑ 5. Iguana iguana haplotype Car1 NADH dehydrogenase subunit 4 (ND4) gene, partial cds; tRNA-His and tRNA-Ser genes, complete sequence; and tRNA-Leu gene, partial sequence; mitochondrial
903 bp linear DNA

Otevírání sequence.fasta

Zvolili jste otevírat

sequence.fasta

což je: soubor FASTA

z: http://www.ncbi.nlm.nih.gov

Co má aplikace Firefox udělat s tímto souborem?

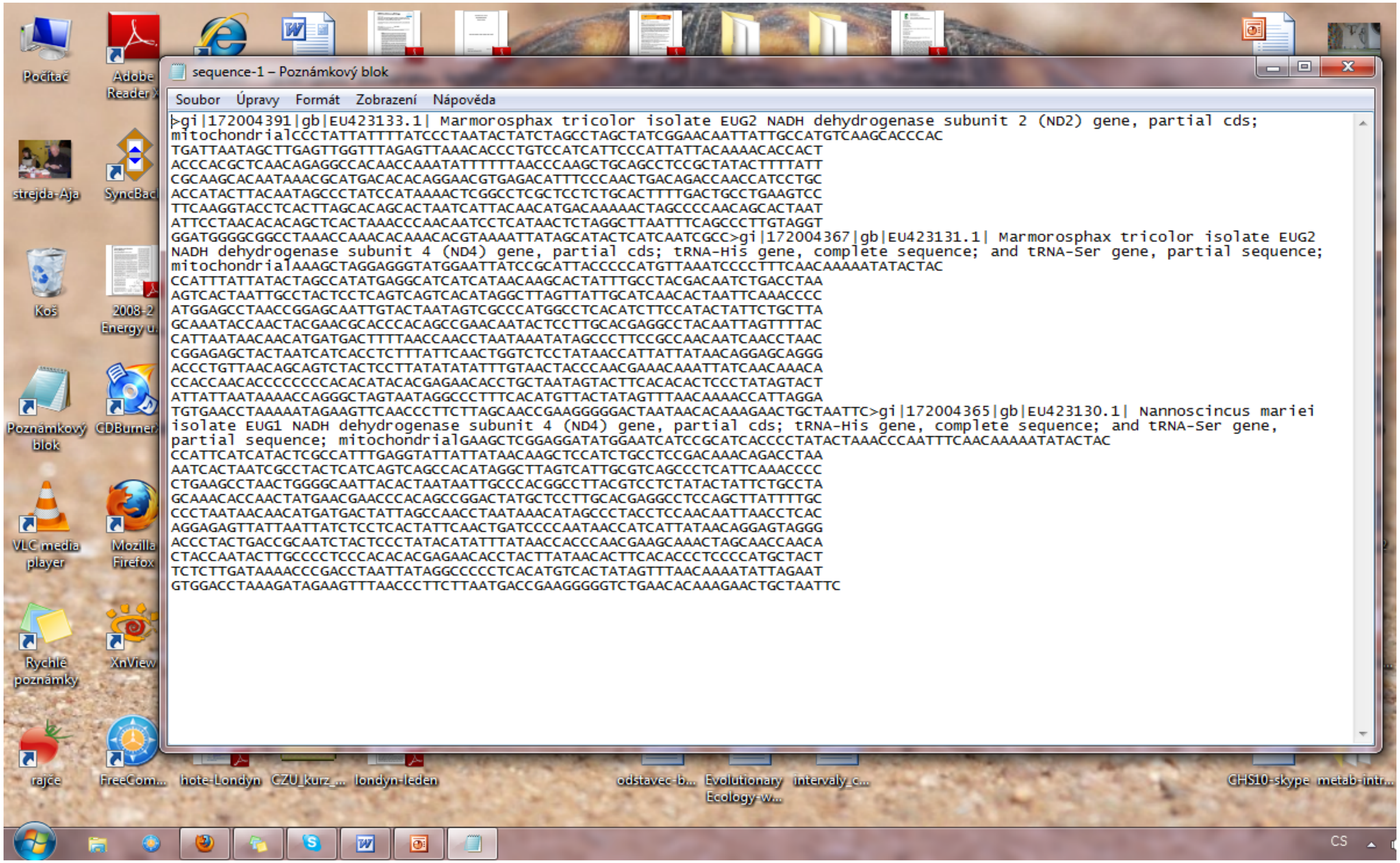
● Otevřít pomocí: Poznámkový blok

○ Uložit soubor

☐ Provádět od teď automaticky s podobnými soubory.

OK Zrušit

Open as FASTA program Notepad (possible directly in BioEdit)



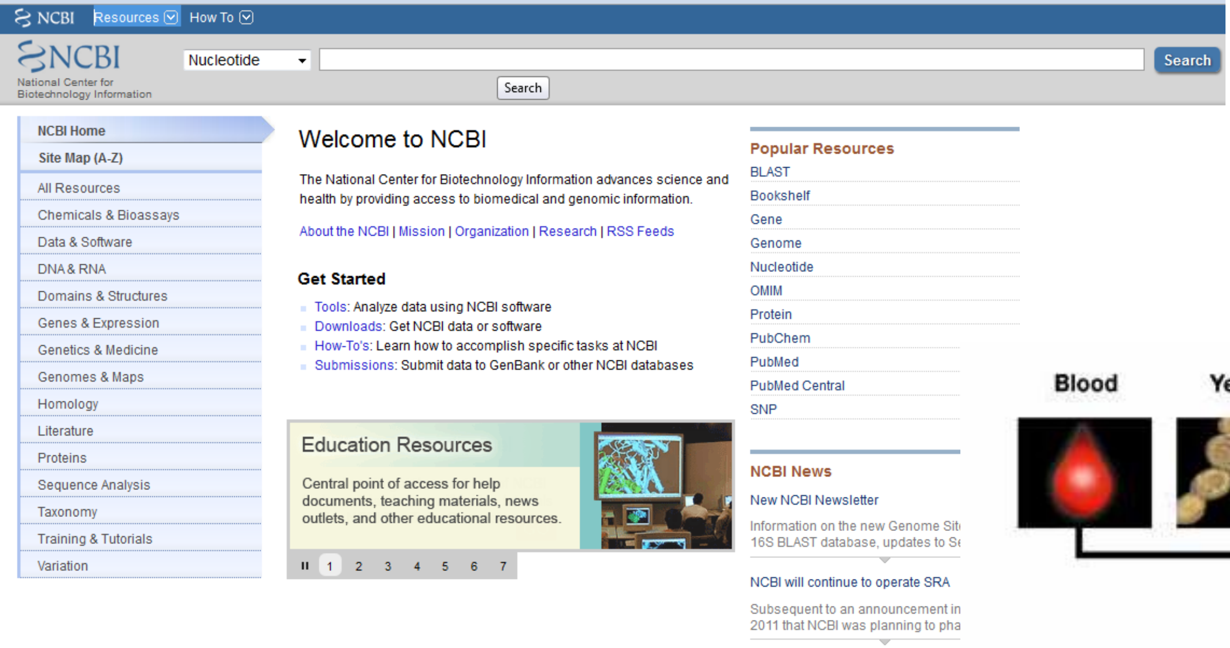
```
>gi|172004391|gb|EU423133.1| Marmorosphax tricolor isolate EUG2 NADH dehydrogenase subunit 2 (ND2) gene, partial cds;
mitochondrialCCCTATTATTTTATCCCTAATACTATCTAGCCTAGCTATCGGAACAATTATTGCCATGTCAAGCACCCAC
TGATTAATAGCTTGAGTTGGTTTAGAGTTAAACACCCTGTCCATCATTCCCATTATTACAAAACACCACT
ACCCACGCTCAACAGAGGCCACAACCAAATATTTTTTAACCCAAGCTGCAGCCTCCGCTATACTTTTATT
CGCAAGCACAATAAACGCATGACACACAGGAACGTGAGACATTTCCCAACTGACAGACCAACCATCCTGC
ACCATACTTACAATAGCCCTATCCATAAAACTCGGCCTCGCTCCTCTGCACTTTTGACTGCCTGAAGTCC
TTCAAGGTACCTCACTTAGCACAGCACTAATCATTACAACATGACAAAAACTAGCCCCAACAGCACTAAT
ATTCCTAACACACAGCTCACTAAACCCAACAATCCTCATAACTCTAGGCTTAATTTCAGCCCTTGTAGGT
GGATGGGGCGGCCTAAACCAAACACAAACACGTAAAATTATAGCATACTCATCAATCGCC>gi|172004367|gb|EU423131.1| Marmorosphax tricolor isolate EUG2
NADH dehydrogenase subunit 4 (ND4) gene, partial cds; tRNA-His gene, complete sequence; and tRNA-Ser gene, partial sequence;
mitochondrialAAAGCTAGGAGGGTATGGAATTATCCGCATTACCCCCATGTTAAATCCCCTTTCAACAAAAATATACTAC
CCATTTATTATACTAGCCATATGAGGCATCATCATAACAAGCACTATTTGCCTACGACAATCTGACCTAA
AGTCACTAATTGCCTACTCCTCAGTCAGTCACATAGGCTTAGTTATTGCATCAACACTAATTCAAACCCC
ATGGAGCCTAACCGGAGCAATTGTACTAATAGTCGCCCATGGCCTCACATCTTCCATACTATTCTGCTTA
GCAAATACCAACTACGAACGCACCCACAGCCGAACAATACTCCTTGCACGAGGCCTACAATTAGTTTTAC
CATTAATAACAACATGATGACTTTTAACCAACCTAATAAATATAGCCCTTCCGCCAACAATCAACCTAAC
CGGAGAGCTACTAATCATCACCTCTTTATTCAACTGGTCTCCTATAACCATTATTATAACAGGAGCAGGG
ACCCTGTTAACAGCAGTCTACTCCTTATATATATTTGTAACTACCCAACGAAACAAATTATCAACAAACA
CCACCAACACCCCCCCCACACATACACGAGAACACCTGCTAATAGTACTTCACACACTCCCTATAGTACT
ATTATTAATAAAACCAGGGCTAGTAATAGGCCCTTTCACATGTTACTATAGTTTAACAAAACCATTAGGA
TGTGAACCTAAAAATAGAAGTTCAACCCTTCTTAGCAACCGAAGGGGGACTAATAACACAAAGAACTGCTAATTC>gi|172004365|gb|EU423130.1| Nannoscincus mariei
isolate EUG1 NADH dehydrogenase subunit 4 (ND4) gene, partial cds; tRNA-His gene, complete sequence; and tRNA-Ser gene,
partial sequence; mitochondrialGAAGCTCGGAGGATATGGAATCATCCGCATCACCCCTATACTAAACCCAATTTCAACAAAAATATACTAC
CCATTCATCATACTCGCCATTTGAGGTATTATTATAACAAGCTCCATCTGCCTCCGACAAACAGACCTAA
AATCACTAATCGCCTACTCATCAGTCAGCCACATAGGCTTAGTCATTGCGTCAGCCCTCATTCAAACCCC
CTGAAGCCTAACTGGGGCAATTACACTAATAATTGCCCACGGCCTTACGTCCTCTATACTATTCTGCCTA
GCAAACACCAACTATGAACGAACCCACAGCCGGACTATGCTCCTTGCACGAGGCCTCCAGCTTATTTTGC
CCCTAATAACAACATGATGACTATTAGCCAACCTAATAAACATAGCCCTACCTCCAACAATTAACCTCAC
AGGAGAGTTATTAATTATCTCCTCACTATTCAACTGATCCCCAATAACCATCATTATAACAGGAGTAGGG
ACCCTACTGACCGCAATCTACTCCCTATACATATTTATAACCACCCAACGAAGCAAACTAGCAACCAACA
CTACCAATACTTGCCCCTCCCACACACGAGAACACCTACTTATAACACTTCACACCCTCCCCATGCTACT
TCTCTTGATAAAACCCGACCTAATTATAGGCCCCCTCACATGTCACTATAGTTTAACAAAATATTAGAAT
GTGGACCTAAAGATAGAAGTTTAACCCTTCTTAATGACCGAAGGGGGTCTGAACACAAAGAACTGCTAATTC
```

Change sequence names – ideally 1-word name

>iguana

CTACCTAAATGGCTAGCC

Save as iguanas.fas



- NCBI Home
- Site Map (A-Z)
- All Resources
- Chemicals & Bioassays
- Data & Software
- DNA & RNA
- Domains & Structures
- Genes & Expression
- Genetics & Medicine
- Genomes & Maps
- Homology
- Literature
- Proteins
- Sequence Analysis
- Taxonomy
- Training & Tutorials
- Variation

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

About the NCBI | Mission | Organization | Research | RSS Feeds

Get Started

- Tools: Analyze data using NCBI software
- Downloads: Get NCBI data or software
- How-To's: Learn how to accomplish specific tasks at NCBI
- Submissions: Submit data to GenBank or other NCBI databases

Education Resources

Central point of access for help documents, teaching materials, news outlets, and other educational resources.

Popular Resources

- BLAST
- Bookshelf
- Gene
- Genome
- Nucleotide
- OMIM
- Protein
- PubChem
- PubMed
- PubMed Central
- SNP

NCBI News

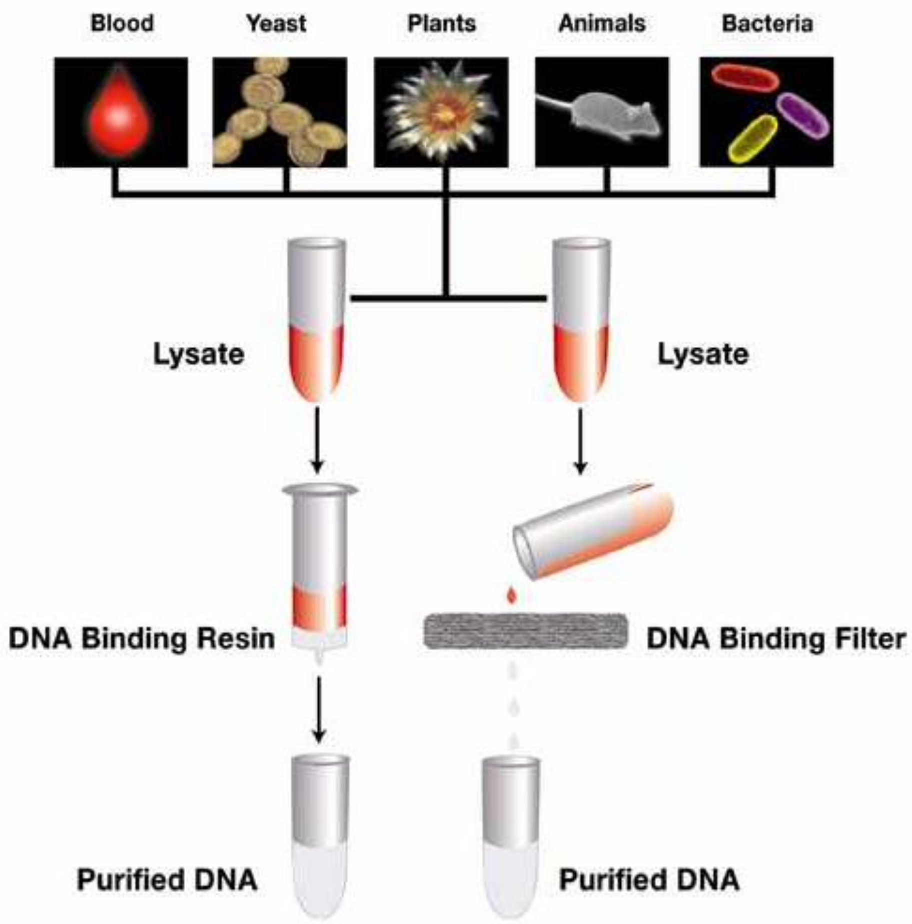
New NCBI Newsletter

Information on the new Genome Sit 16S BLAST database, updates to Se

NCBI will continue to operate SRA

Subsequent to an announcement in 2011 that NCBI was planning to pha

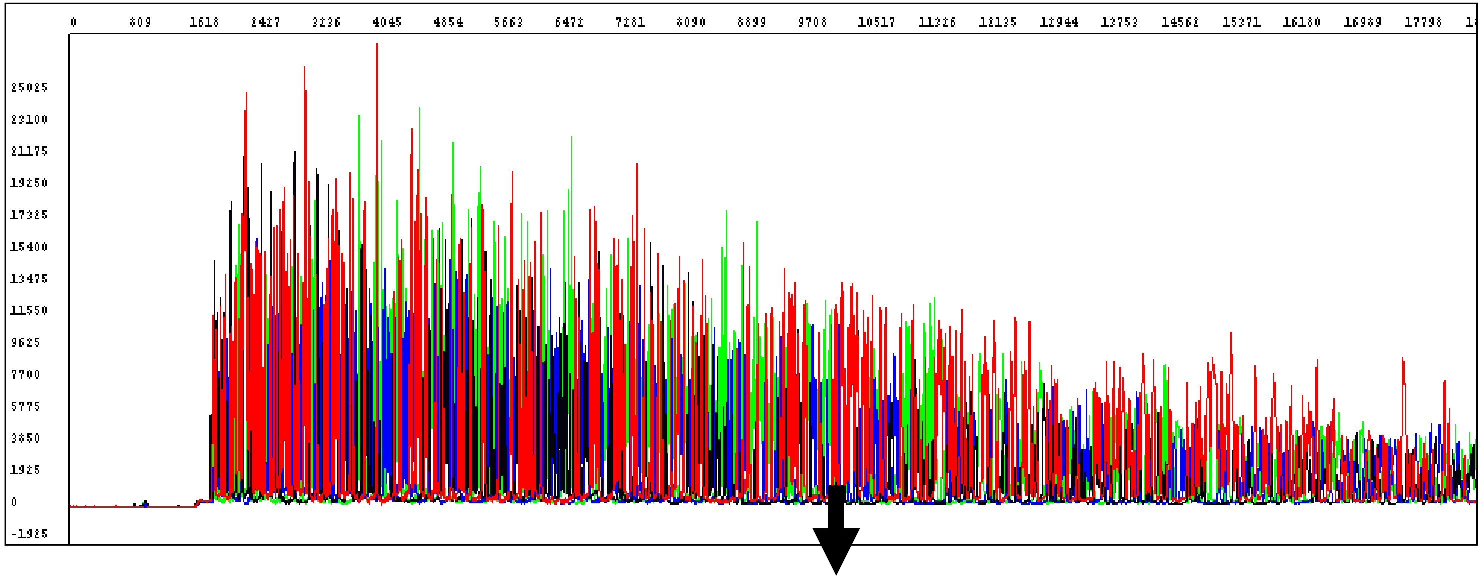
+

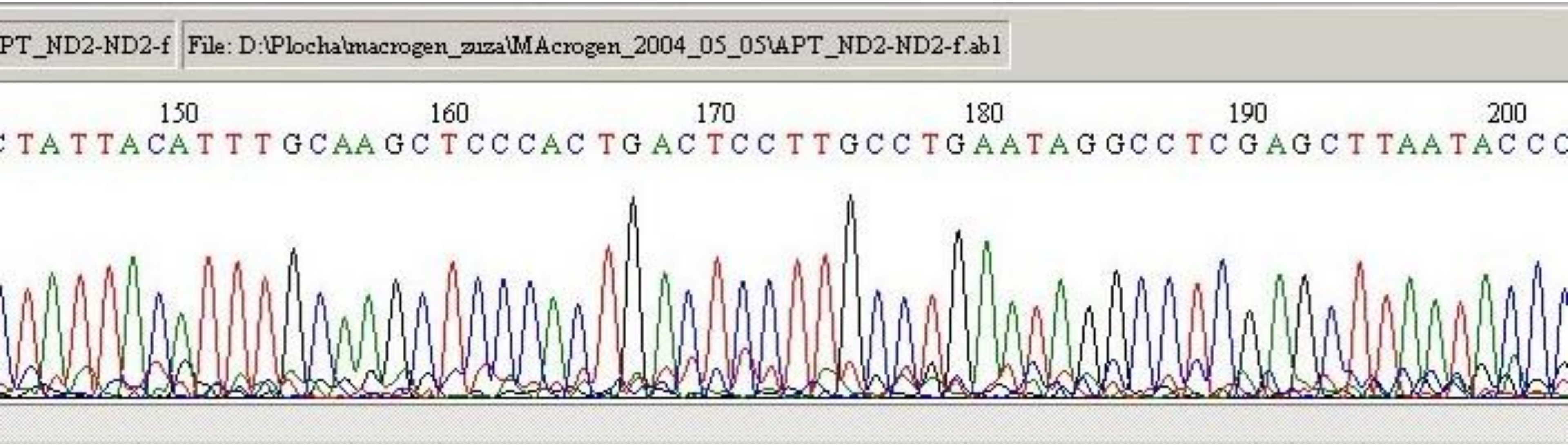


http://www.ncbi.nlm.nih.gov/

Chromatogram files:

Can be opened in different software- Chromas, BioEdit, DNASTAR , …

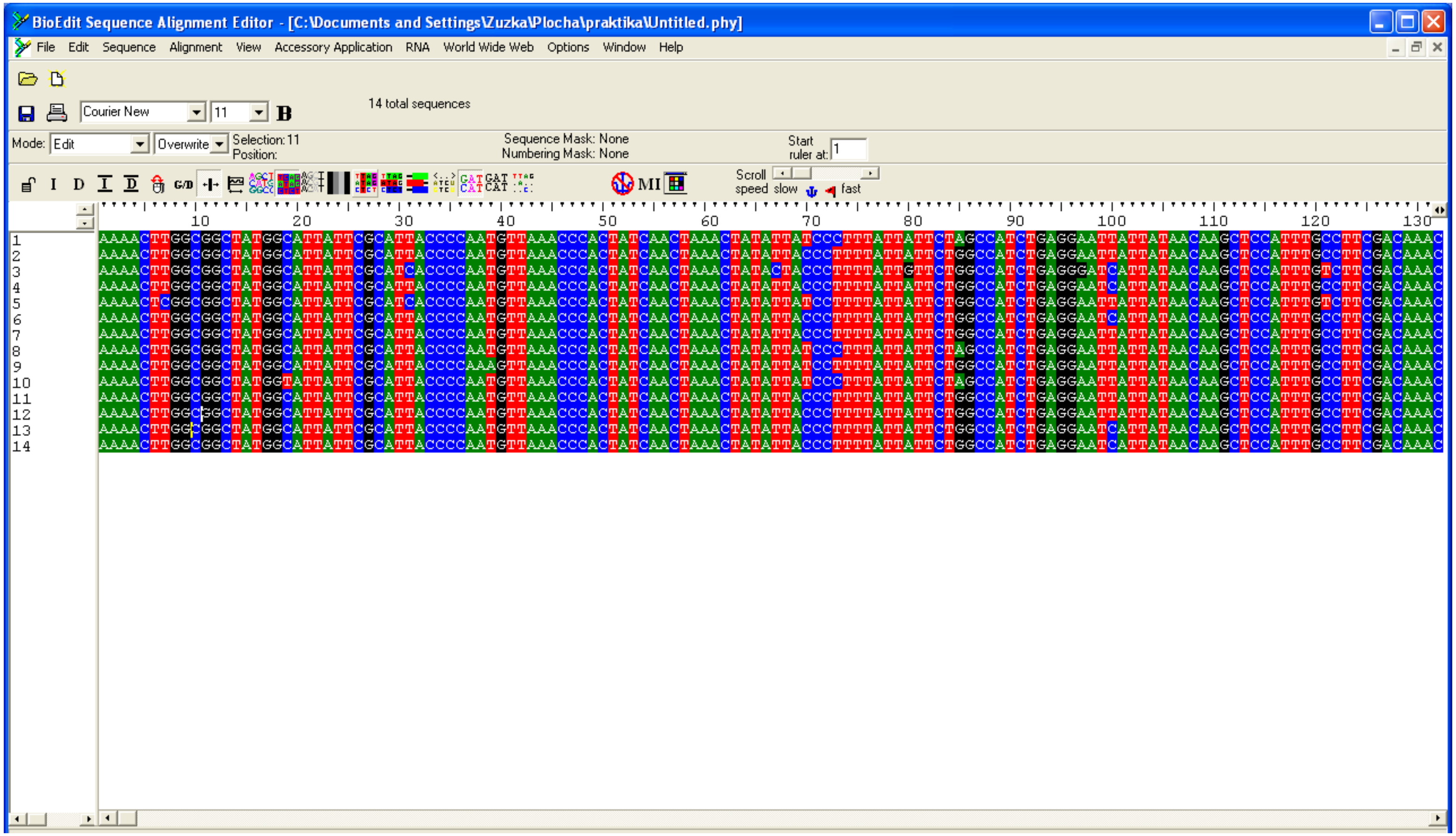



PT_ND2-ND2-f

File: D:\Plocha\macrogen_zuza\MAcrogen_2004_05_05\APT_ND2-ND2-f.abl

TATTACATTTGCAAGCTCCCACTGACTCCTTGCCTGAATAGGCCTCGAGCTTAATACC

http://www.mbio.ncsu.edu/BioEdit/bioedit.html



Courier New 11

14 total sequences

Mode: Edit Overwrite

Selection: 11

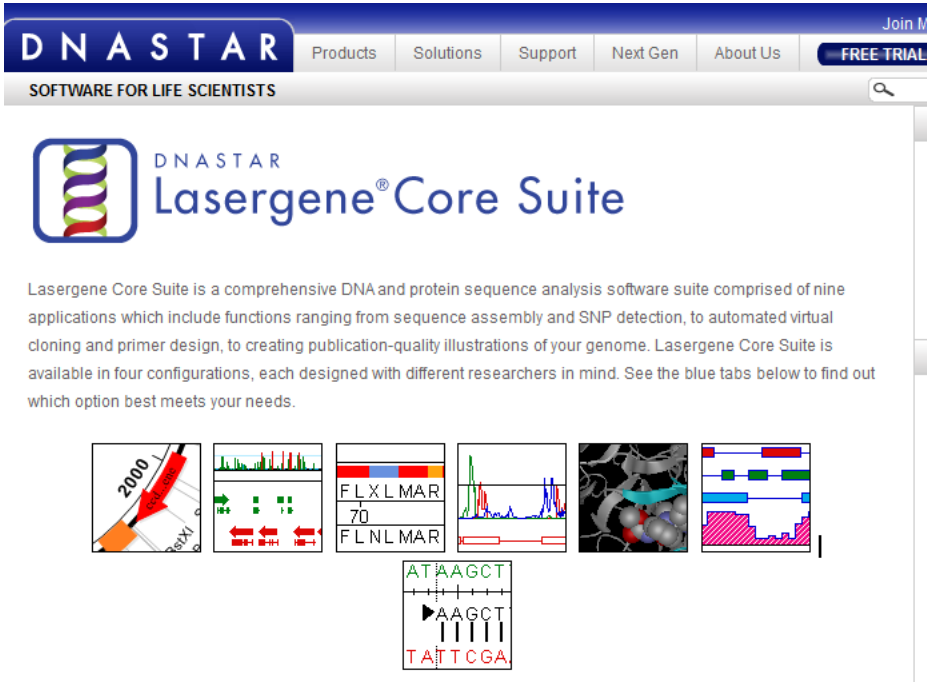
Position:

Sequence Mask: None

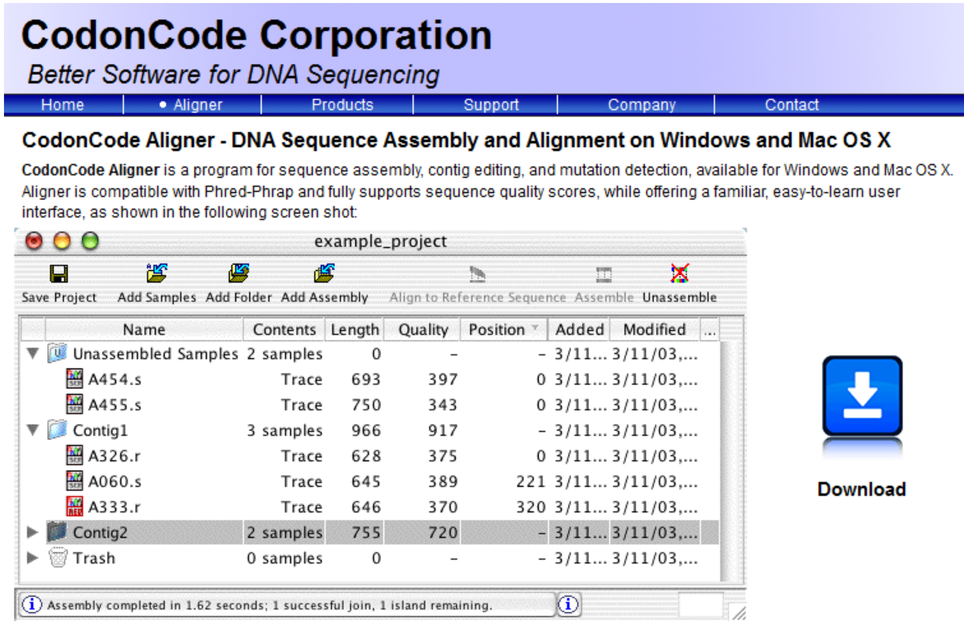
Numbering Mask: None

Start ruler at: 1

```
1  AAAACTTGGCGGCTATGGCATTATTCGCATTACCCCAATGTTAAACCCACTATCAACTAAACTATATTATCCCTTTATTATTCTAGCCATCTGAGGAATTATTATAACAAGCTCCATTTGCCTTCGACAAAC
2  AAAACTTGGCGGCTATGGCATTATTCGCATTACCCCAATGTTAAACCCACTATCAACTAAACTATATTACCCTTTTATTATTCTGGCCATCTGAGGAATTATTATAACAAGCTCCATTTGCCTTCGACAAAC
3  AAAACTTGGCGGCTATGGCATTATTCGCATCACCCCAATGTTAAACCCACTATCAACTAAACTATACTACCCTTTTATTGTTCTGGCCATCTGAGGGATCATTATAACAAGCTCCATTTGTCTTCGACAAAC
4  AAAACTTGGCGGCTATGGCATTATTCGCATTACCCCAATGTTAAACCCACTATCAACTAAACTATATTACCCTTTTATTATTCTGGCCATCTGAGGAATCATTATAACAAGCTCCATTTGCCTTCGACAAAC
5  AAAACTCGGCGGCTATGGCATTATTCGCATCACCCCAATGTTAAACCCACTATCAACTAAACTATATTATCCTTTTATTATTCTGGCCATCTGAGGAATTATTATAACAAGCTCCATTTGTCTTCGACAAAC
6  AAAACTTGGCGGCTATGGCATTATTCGCATTACCCCAATGTTAAACCCACTATCAACTAAACTATATTACCCTTTTATTATTCTGGCCATCTGAGGAATCATTATAACAAGCTCCATTTGCCTTCGACAAAC
7  AAAACTTGGCGGCTATGGCATTATTCGCATTACCCCAATGTTAAACCCACTATCAACTAAACTATATTACCCTTTTATTATTCTGGCCATCTGAGGAATTATTATAACAAGCTCCATTTGCCTTCGACAAAC
8  AAAACTTGGCGGCTATGGCATTATTCGCATTACCCCAATGTTAAACCCACTATCAACTAAACTATATTATCCCTTTATTATTCTAGCCATCTGAGGAATTATTATAACAAGCTCCATTTGCCTTCGACAAAC
9  AAAACTTGGCGGCTATGGCATTATTCGCATTACCCCAAAGTTAAACCCACTATCAACTAAACTATATTATCCTTTTATTATTCTGGCCATCTGAGGAATTATTATAACAAGCTCCATTTGCCTTCGACAAAC
10 AAAACTTGGCGGCTATGGTATTATTCGCATTACCCCAATGTTAAACCCACTATCAACTAAACTATATTATCCCTTTATTATTCTAGCCATCTGAGGAATTATTATAACAAGCTCCATTTGCCTTCGACAAAC
11 AAAACTTGGCGGCTATGGCATTATTCGCATTACCCCAATGTTAAACCCACTATCAACTAAACTATATTACCCTTTTATTATTCTGGCCATCTGAGGAATTATTATAACAAGCTCCATTTGCCTTCGACAAAC
12 AAAACTTGGCGGCTATGGCATTATTCGCATTACCCCAATGTTAAACCCACTATCAACTAAACTATATTACCCTTTTATTATTCTGGCCATCTGAGGAATTATTATAACAAGCTCCATTTGCCTTCGACAAAC
13 AAAACTTGGCGGCTATGGCATTATTCGCATTACCCCAATGTTAAACCCACTATCAACTAAACTATATTACCCTTTTATTATTCTGGCCATCTGAGGAATCATTATAACAAGCTCCATTTGCCTTCGACAAAC
14 AAAACTTGGCGGCTATGGCATTATTCGCATTACCCCAATGTTAAACCCACTATCAACTAAACTATATTACCCTTTTATTATTCTGGCCATCTGAGGAATCATTATAACAAGCTCCATTTGCCTTCGACAAAC
```



Lasergene Core Suite is a comprehensive DNA and protein sequence analysis software suite comprised of nine applications which include functions ranging from sequence assembly and SNP detection, to automated virtual cloning and primer design, to creating publication-quality illustrations of your genome. Lasergene Core Suite is available in four configurations, each designed with different researchers in mind. See the blue tabs below to find out which option best meets your needs.



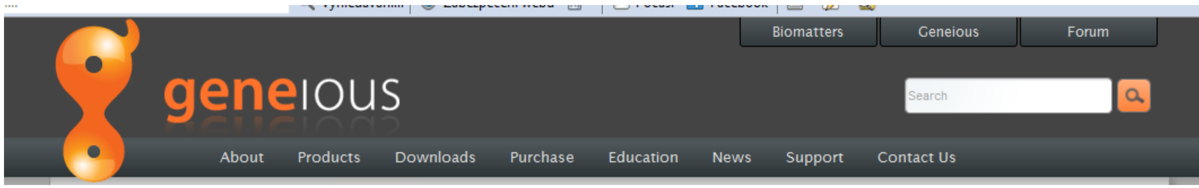
CodonCode Aligner - DNA Sequence Assembly and Alignment on Windows and Mac OS X

CodonCode Aligner is a program for sequence assembly, contig editing, and mutation detection, available for Windows and Mac OS X. Aligner is compatible with Phred-Phrap and fully supports sequence quality scores, while offering a familiar, easy-to-learn user interface, as shown in the following screen shot:

Name	Contents	Length	Quality	Position	Added	Modified
Unassembled Samples	2 samples	0	-	-	3/11...	3/11/03,...
A454.s	Trace	693	397	0	3/11...	3/11/03,...
A455.s	Trace	750	343	0	3/11...	3/11/03,...
Contig1	3 samples	966	917	-	3/11...	3/11/03,...
A326.r	Trace	628	375	0	3/11...	3/11/03,...
A060.s	Trace	645	389	221	3/11...	3/11/03,...
A333.r	Trace	646	370	320	3/11...	3/11/03,...
Contig2	2 samples	755	720	-	3/11...	3/11/03,...
Trash	0 samples	0	-	-	3/11...	3/11/03,...

Assembly completed in 1.62 seconds; 1 successful join, 1 island remaining.

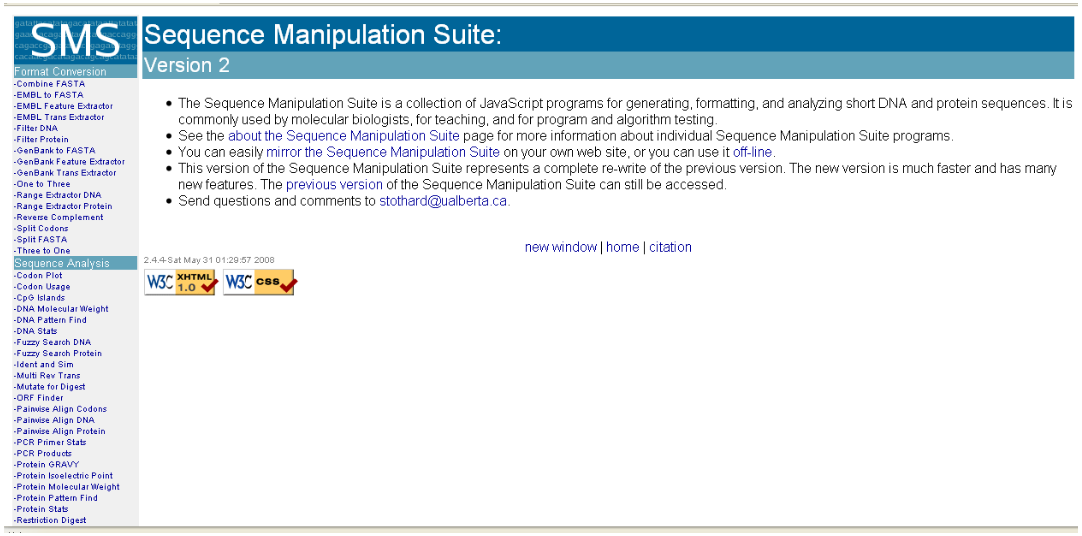
Lasergene
Geneious
CodonCode Aligner
SeqScape
Sequencher



Online work with sequence data

SMS- sequence manipulation suite

http://www.bioinformatics.org/sms2/



Sequence Manipulation Suite:

Version 2

- The Sequence Manipulation Suite is a collection of JavaScript programs for generating, formatting, and analyzing short DNA and protein sequences. It is commonly used by molecular biologists, for teaching, and for program and algorithm testing.
- See the about the Sequence Manipulation Suite page for more information about individual Sequence Manipulation Suite programs.
- You can easily mirror the Sequence Manipulation Suite on your own web site, or you can use it off-line.
- This version of the Sequence Manipulation Suite represents a complete re-write of the previous version. The new version is much faster and has many new features. The previous version of the Sequence Manipulation Suite can still be accessed.
- Send questions and comments to stothard@ualberta.ca.

new window | home | citation

2.4.4-Sat May 31 01:29:57 2008

FaBox

HTTP://www.birc.au.dk/fabox

FaBox

http://www.birc.au.dk/~biopv/php/fabox/

ABI Chromatogram: C:\Users\Zuzka\Desktop\MAZ-praktika\computers\samplend4.ab1

Selected: none

Sample: cwnd4

File: C:\Users\Zuzka\Desktop\MAZ-praktika\computers\samplend4.ab1

DNA sequence from C:\Users\Zuzka\Desktop\MAZ-praktika\computers\samplend4.ab1

Courier New 11

1 total sequences

Mode: Select / Slide

Selection: null
Position: 108

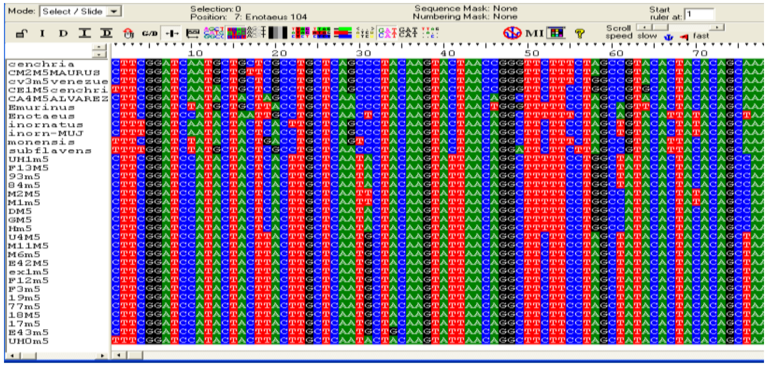
Sequence Mask: None
Numbering Mask: None

Start ruler at: 1

```
cwnd4  CCCANTTGCTGGCTCCATAGTACTAGCCGCCATCCTACTAAAACTGGGAGGCTACGGAATCATCCGAATCACCCTAACATTAATGCCAATAACCCAAAAACTATACTACCCCTTCATGATCTTAGCCCTATG
```

- open **cnF.ab1 in BioEdit**
- open sample.fas in BioEdit

Where sequences differ and where are the same?



Mode: Select / Slide

Selection: 0

Position: 7: Enotaeus 104

Sequence Mask: None

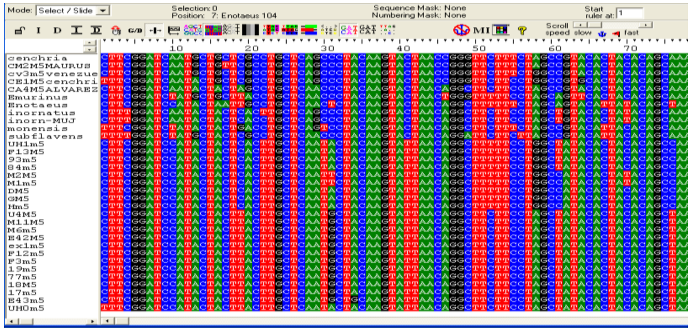
Numbering Mask: None

Start ruler at: 1

Alignment

- a way of arranging the sequences of DNA, (RNA, protein) to identify regions of similarity
- start of every phylogenetic analysis
- assessing of position homology of each base in the sequence
- each position (column in the alignment) in the sequence represents character potentially useful for the phylogenetic analysis
- different programs for calculating and editing alignments
  - manual: BioEdit, Macaw
  - automatic – different algorithms

Clustal X, PileUp, Multalin, Mafft – often online



Mode: Select / Slide

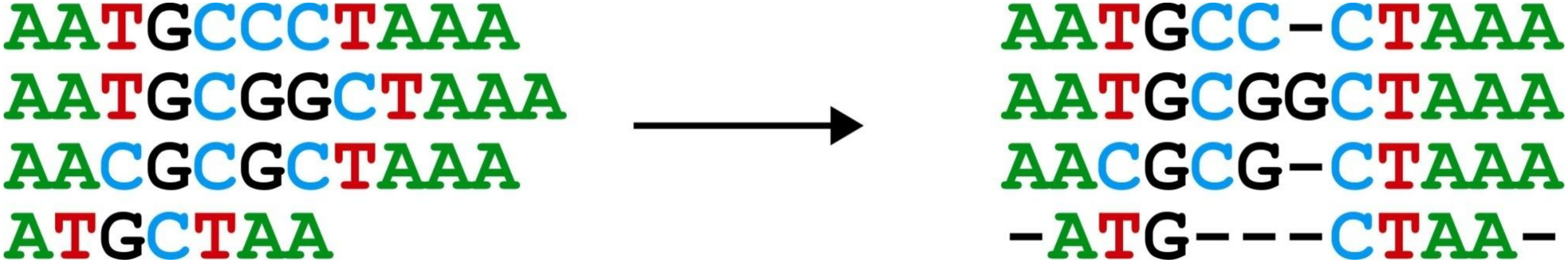
Selection: 0
Position: 7: Enotaeus 104

Sequence Mask: None
Numbering Mask: None

Start ruler at: 1

Alignment - pairwise alignment (two sequences)

- multiple alignment (more sequences)



```
AATGCCCTAAA
AATGCGGCTAAA
AACGCGCTAAA
ATGCTAA
```

```
AATGCC-CTAAA
AATGCGGCTAAA
AACGCG-CTAAA
-ATG---CTAA-
```

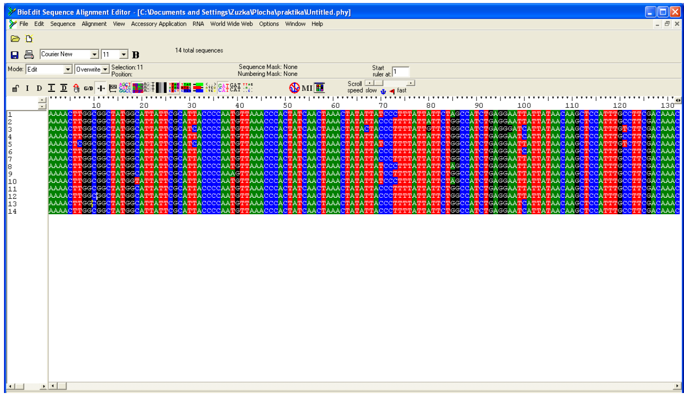
gaps are inserted between the bases so that identical or similar characters are aligned in successive columns

http://www.mbio.ncsu.edu/BioEdit/bioedit.html

- Open your **iguanas.fas** file (in the current BioEdit window)
- Copy sample.fas to **iguanas.fas**
- Create multiple alignment

(BioEdit – Accessory Applications – ClustalW Multiple Alignment)

- Cut both ends of all sequences to be of the same length
- Save alignment as dataset.fas


BioEdit Sequence Alignment Editor - [C:\Documents and Settings\Zuzka\Plocha\praktika\Untitled.phy]

14 total sequences

Mode: Edit

Overwrite

Selection: 11

Position:

Sequence Mask: None

Numbering Mask: None

Start ruler at: 1

Phylogeny estimation

character based (maximum parsimony, maximum likelihood, Bayesian analysis)

two types of methods

distance based (Neighbour-joining, UPGMA)

Two different approaches:

algorithm – number of specific steps resulting in one best tree
methods: UPGMA, Neighbour-joining

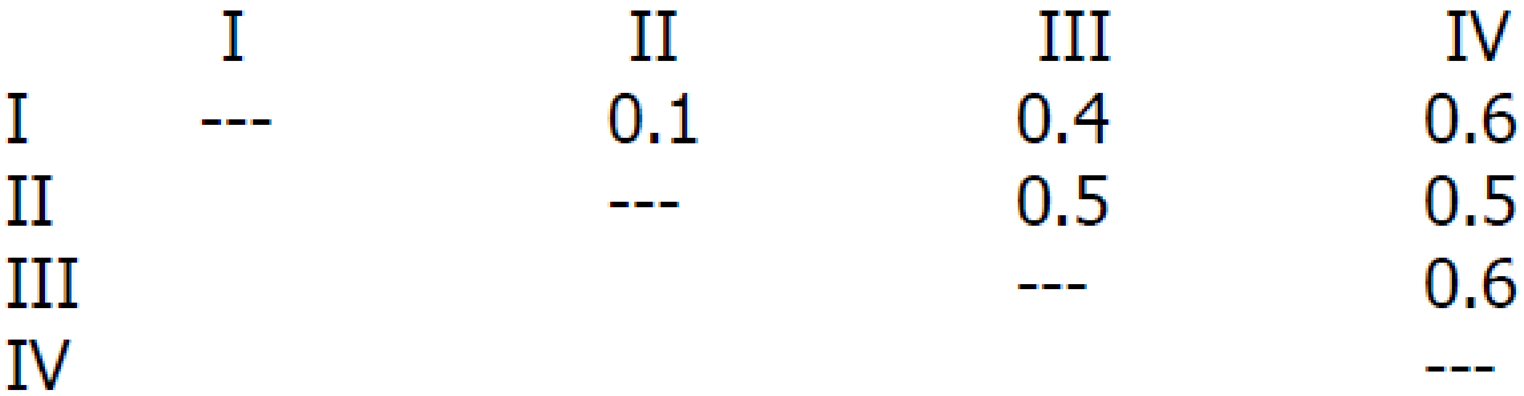
optimality criterion – consider and compare all theoretically possible trees based on selected criteria- number of evolutionary steps, likelihood value

distances

input is a matrix of distances between species

taxon

taxon



	I	II	III	IV
I	---	0.1	0.4	0.6
II		---	0.5	0.5
III			---	0.6
IV				---

Distance based methods (like NJ)

proportional (p) distance

number of substitutions between sequences

p= total number of base differences/total no. of available sites

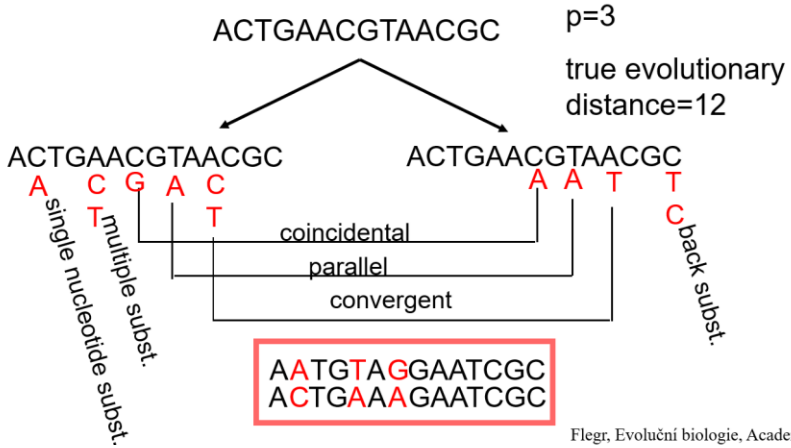
$$p = n_d/n$$

GATCATTAATGCGATAT
GACCGTTATTGCCATAT

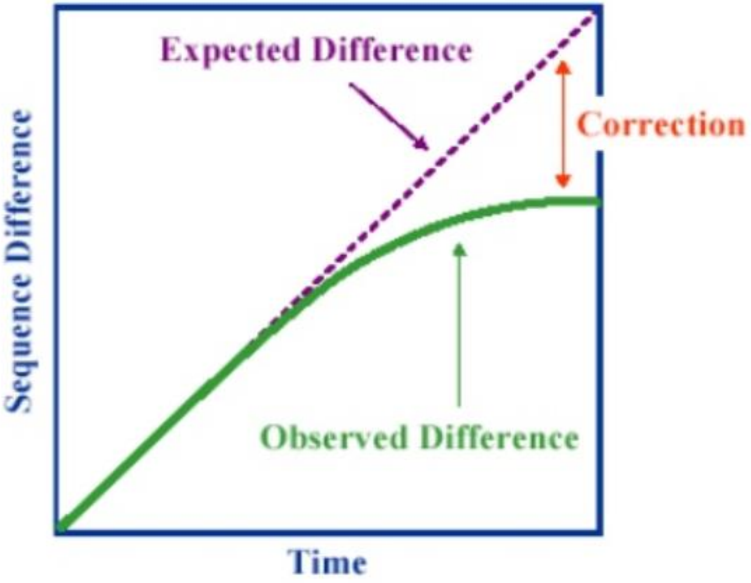
$$p = 4/17 = 0.23$$

real number of substitutions in the sequence over time is usually higher than observed p distance

we can see just 3 differences (*p*), but in fact there was 12 substitutions

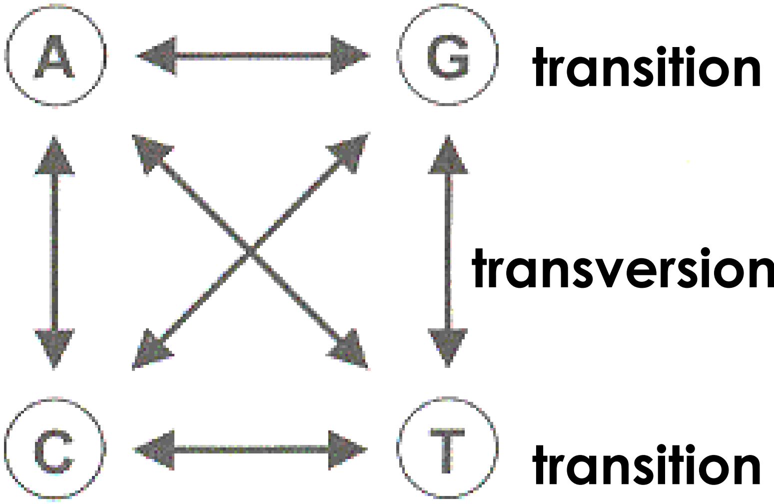


Flegr, Evoluční biologie, Academi



Jukes-Cantor model (distance)

all substitution types and base frequencies are presumed equal



transition

transversion

transition

JC distance

$$d_{JC} = -\frac{3}{4}\ln(1-\frac{4}{3}p)$$

Kimura 2-parameter model (K2P):

transitions are more likely than transversions, equal base frequencies

K2P distance

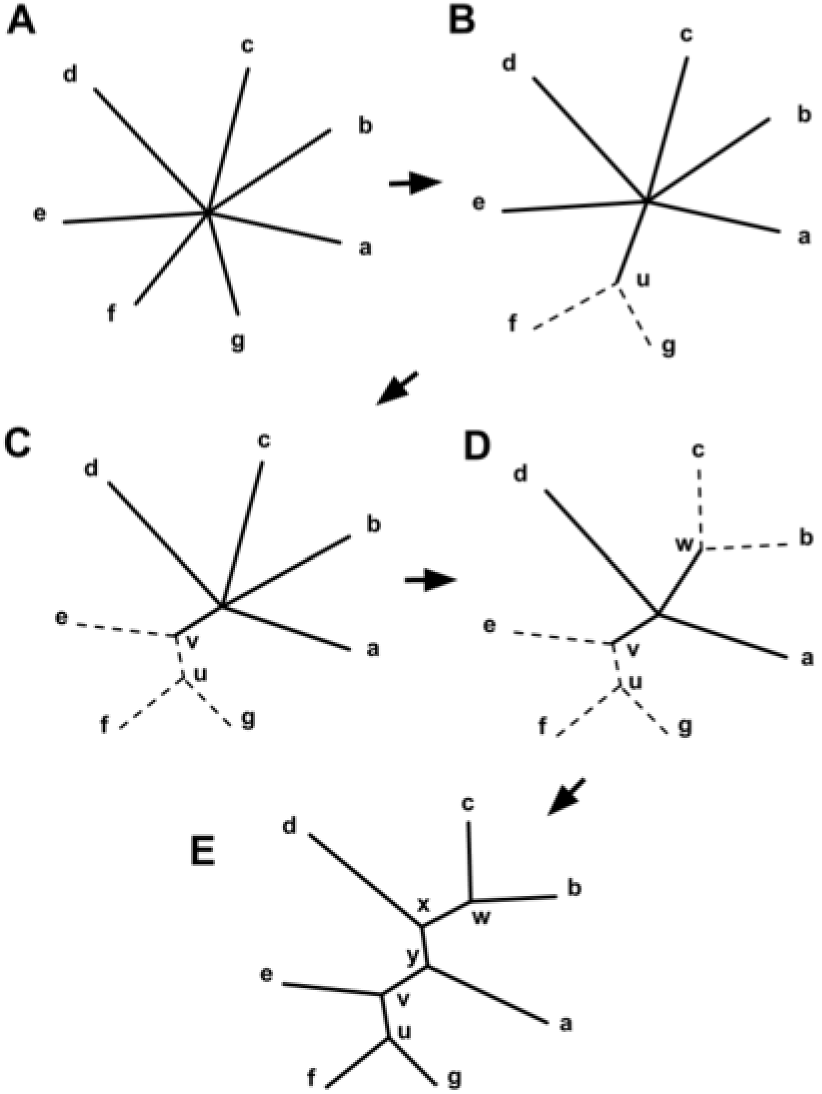
$P = n_{TS} / n$

$Q = n_{TV} / n$

$$d_{K2P} = 0.5\ln\left(\frac{1}{1-2P-Q}\right) + 0.25\ln\left(\frac{1}{1-2Q}\right)$$

Neighbour-joining (NJ) - the fully resolved tree is "decomposed" from a fully unresolved "star" tree by **successively inserting branches between a pair of closest neighbors and the remaining terminals** in the tree

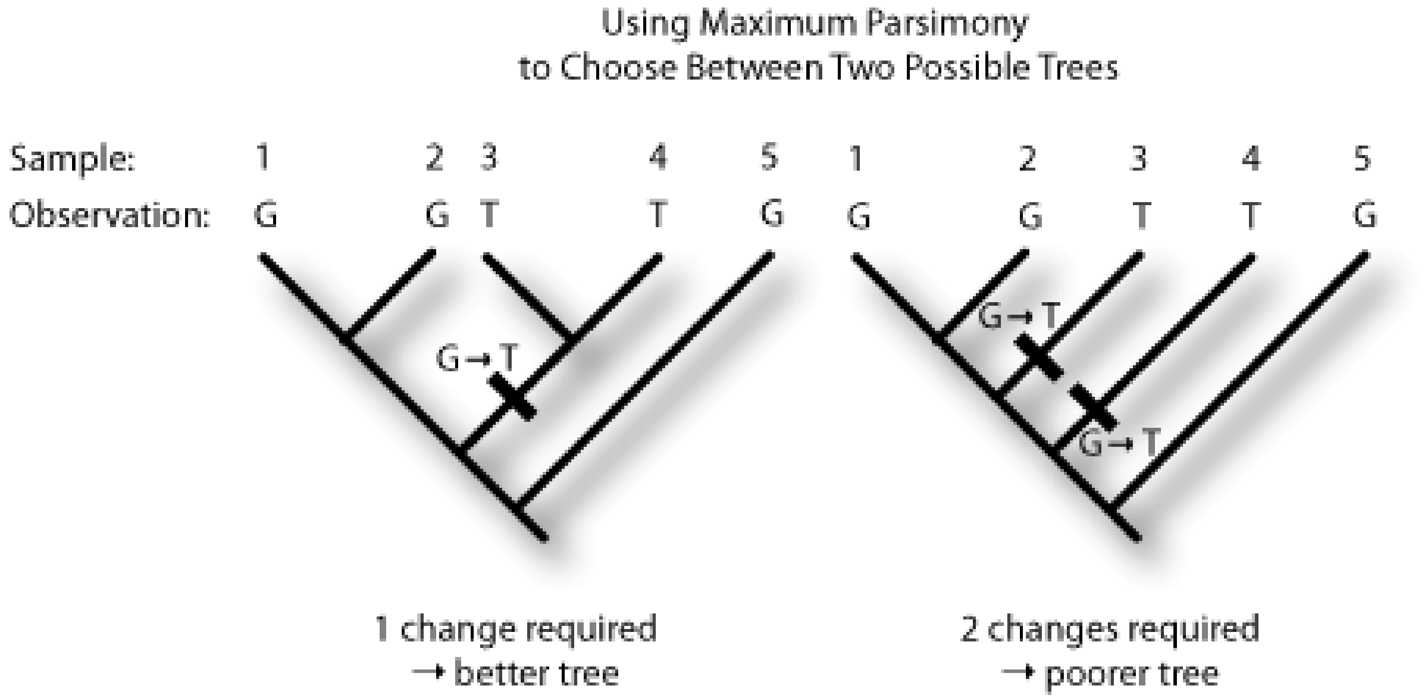
- result is one tree



Maximum parsimony:

optimality criterion - parsimony score = minimum number of events (steps) required by a tree to explain the variation in the data

search for topologies that minimize the total tree length assuming a minimum number of base changes
“Occam’s Razor” – “keep it simple“



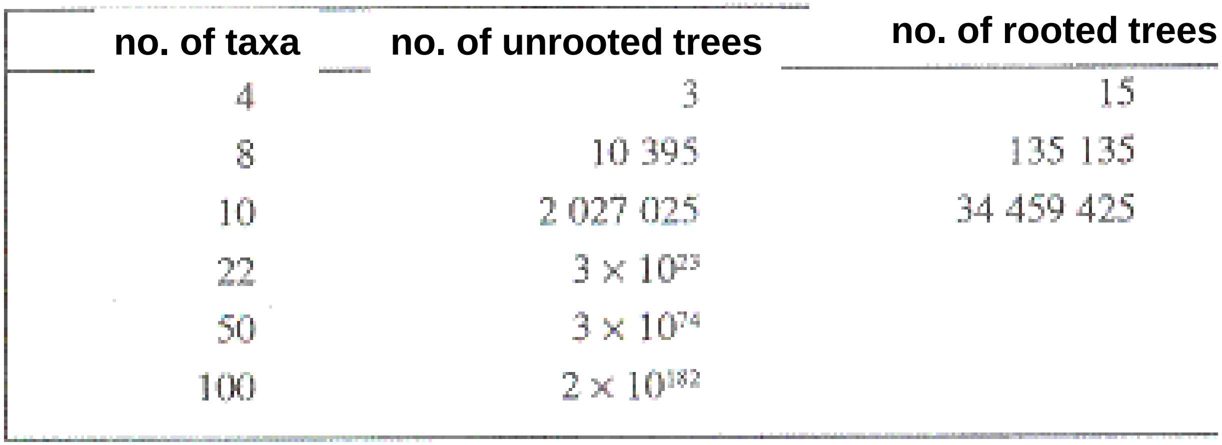
Maximum parsimony:

optimality criterion - parsimony score = minimum number of events (steps) required by a tree to explain the variation in the data

search for topologies that minimize the total tree length assuming a minimum number of base changes
“Occam’s Razor” – “keep it simple“

$$\frac{(2n-3)!}{2^{n-2}(n-2)!}$$

We know that there are a lot of possible trees- in most cases we can not compare all of them

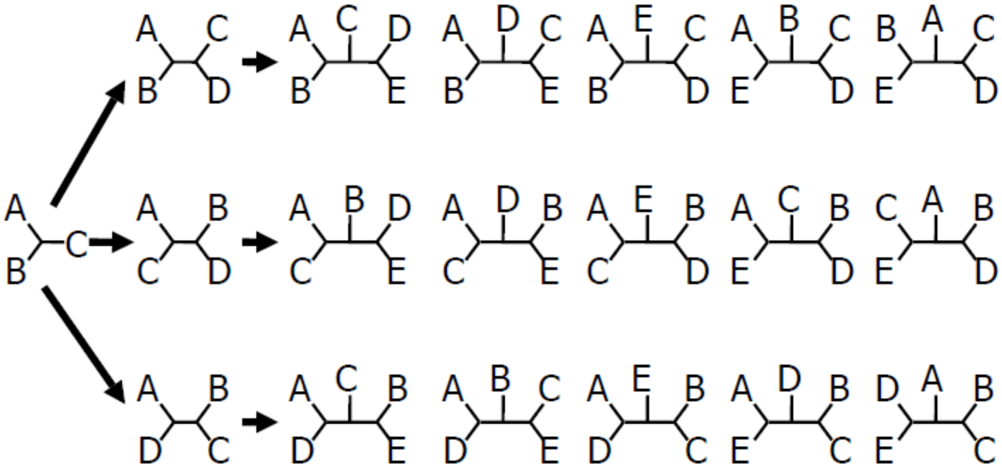


no. of taxa	no. of unrooted trees	no. of rooted trees
4	3	15
8	10 395	135 135
10	2 027 025	34 459 425
22	3×10^{23}	
50	3×10^{74}	
100	2×10^{182}	

no. of trees exponentially increases

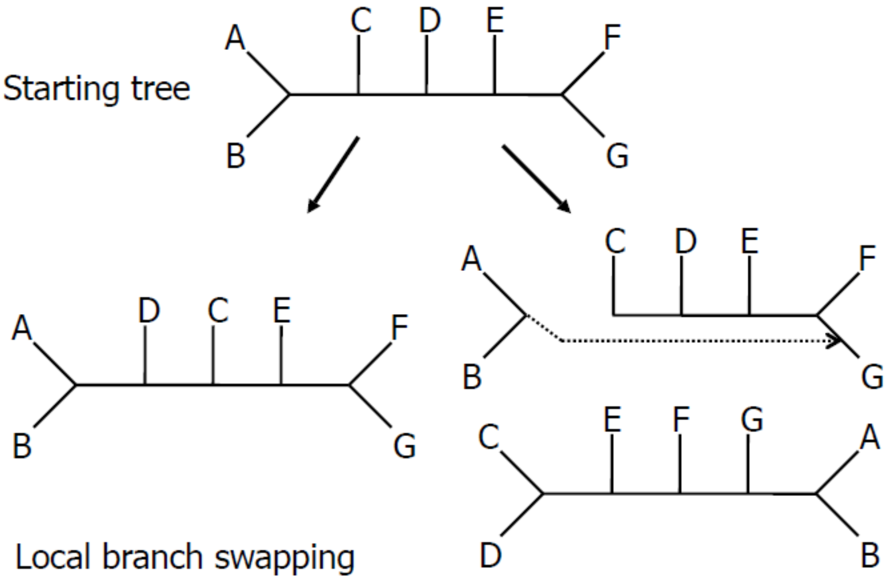
Tree searching

Exhaustive Searching



Branch and Bound Searching

Heuristic Searching



Local branch swapping
Swofford et al. (1996)

Global branch swapping

Maximum parsimony

in most cases we can not compare all trees

⇒ **heuristic search**

- create random tree
- calculate parsimony score
- rearranging of the tree,
- calculate parsimony score
- further the method works with the better (shorter) tree
- repeated rearranging and calculating scores
- at the end shortest tree

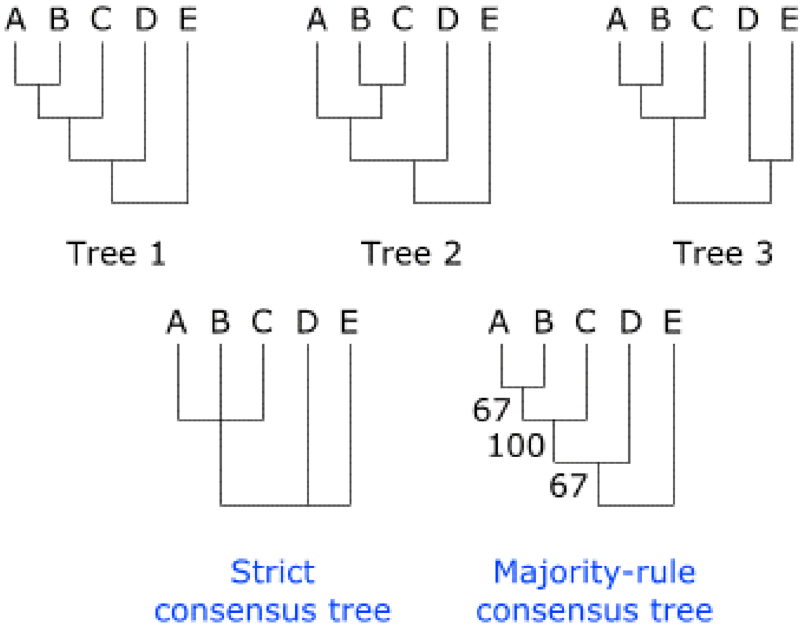
Sometimes (quite often) we find more equal trees →

Consensus tree:

when multiple phylogenies are supported - a consensus tree shows only those relationships common to all trees (based on our settings)

- **strict consensus** (only relationships common to all trees)

- **majority-rule consensus** (relationships common to 50 or 70% of trees are shown)



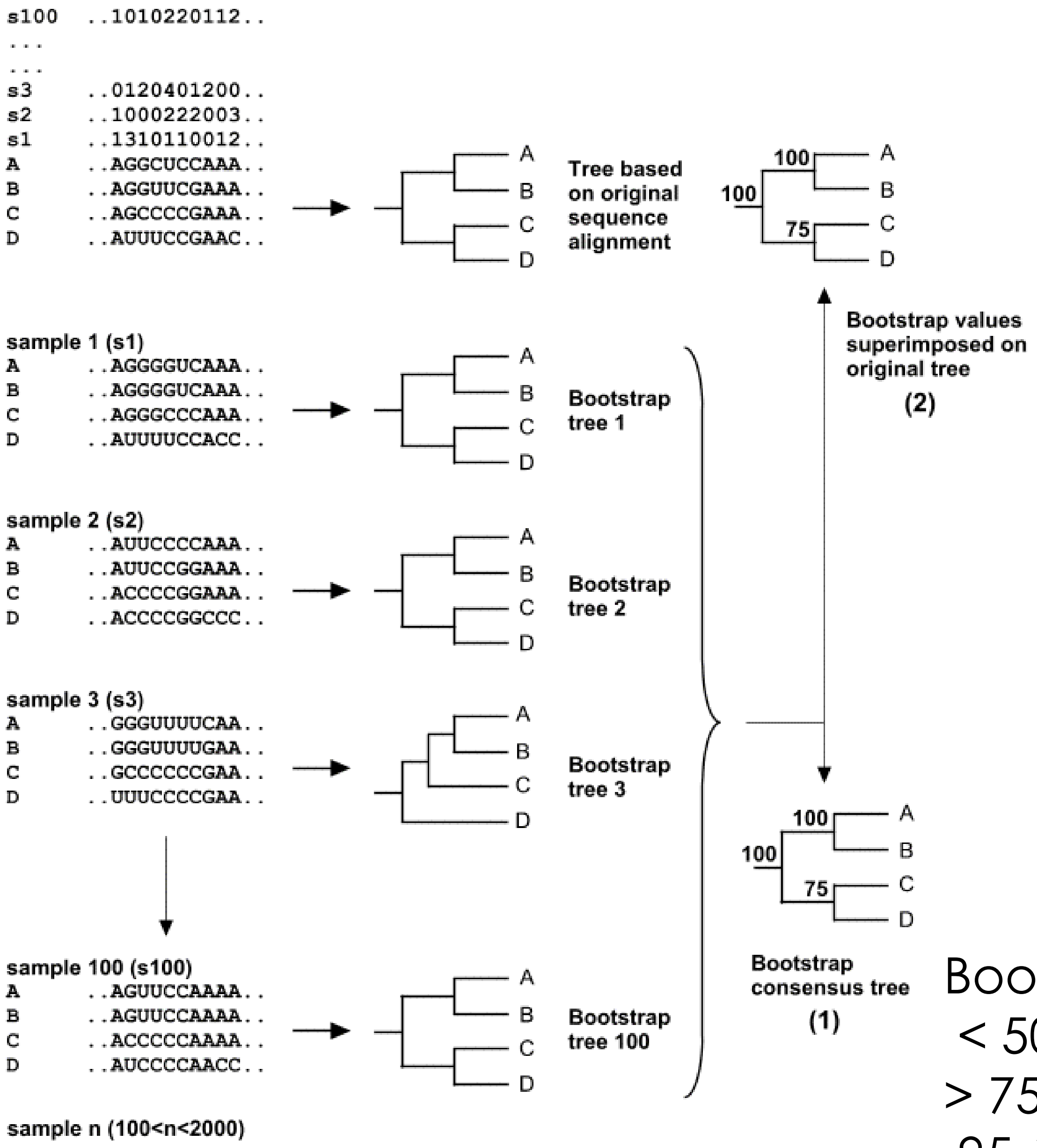
reliability tests

-nonparametric resampling methods - bootstrapping, jackknifing

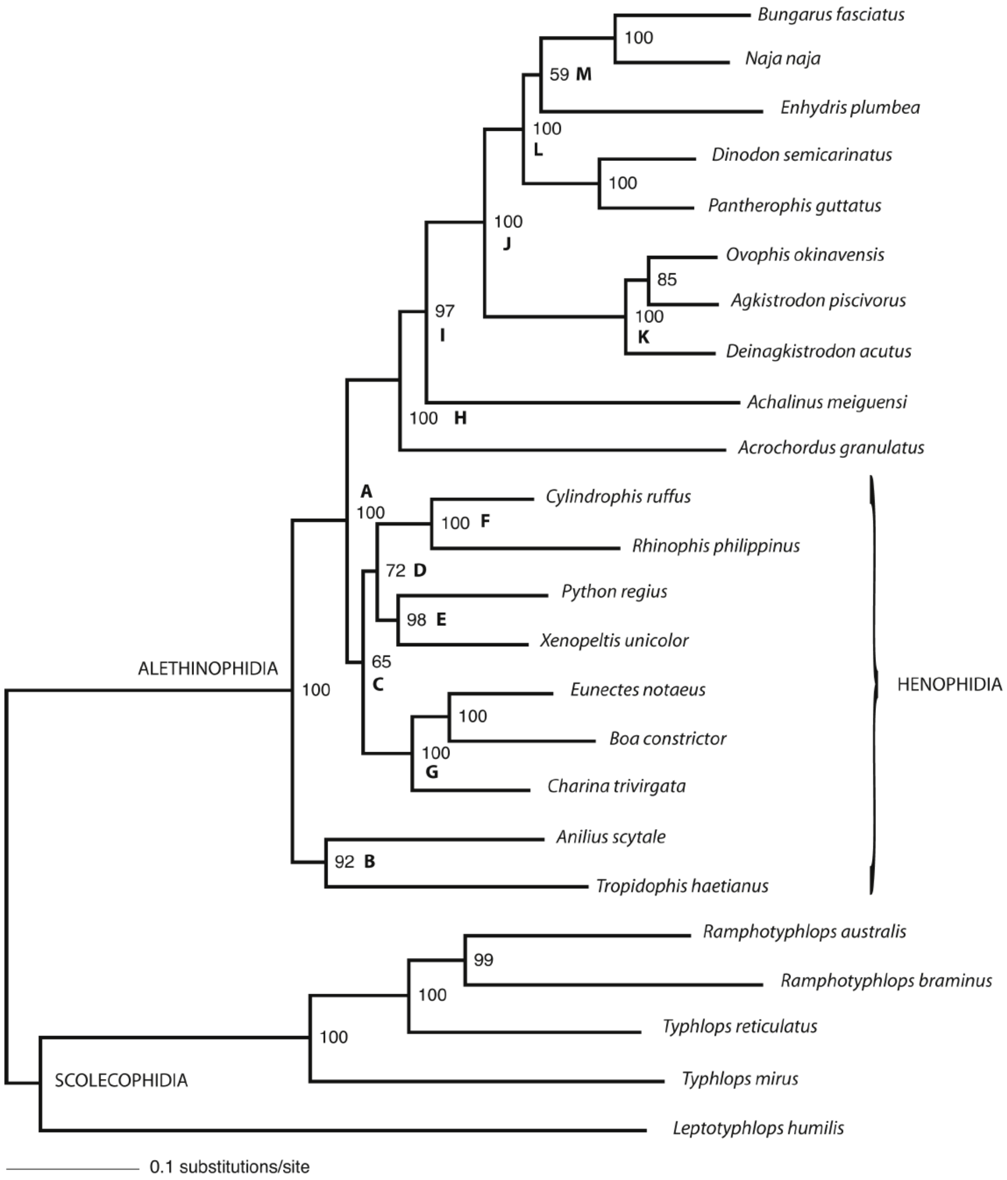
→ new data sets are created from the original data set by sampling columns of characters by random with replacement

- each site can be sampled again with the same probability as any of the other sites

Box 3
Bootstrap Analysis (Felsenstein, 1985)



Bootstrap values:
< 50% - no - just by chance
> 75% MP ok
95-100% great



Bootstrap values:

< 50% - no - just by chance

> 75% ok

95-100% great

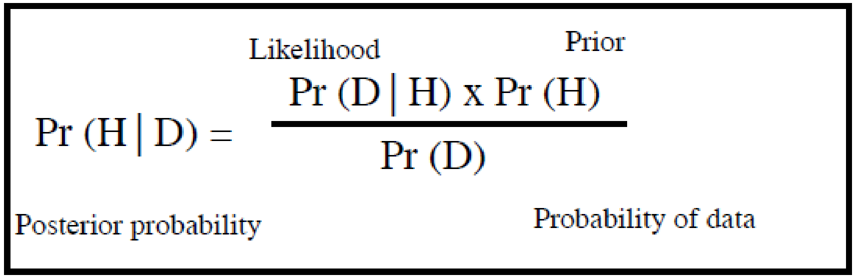
Douglas and Gower BMC Genomics 2010, 11:14

Bayesian inference/analysis

Bayesian inference of phylogeny uses a likelihood function to create a quantity called the **posterior probability** of trees using a model of evolution (substitution model), based on some prior probabilities (priors), producing the most likely phylogenetic tree for the given data
- uses Markov chain Monte Carlo (MCMC) algorithms

Based on theorem of Thomas Bayes (18. century) – Bayesian theorem
- describes the probability of an event, based on prior knowledge of conditions that might be related to the event

Bayesian inference of phylogeny uses a likelihood function to create a quantity called the **posterior probability** of trees using a model of evolution (substitution model), based on some prior probabilities (priors), producing the most likely phylogenetic tree for the given data



$$\underset{\text{Posterior probability}}{\Pr(H|D)} = \underset{\text{Probability of data}}{\frac{\overset{\text{Likelihood}}{\Pr(D|H)} \times \overset{\text{Prior}}{\Pr(H)}}{\Pr(D)}}$$

Pr (D) is not possible to calculate as this is the $\Sigma_H \Pr(D|H) \times \Pr(H)$. Too many different hypothesis.

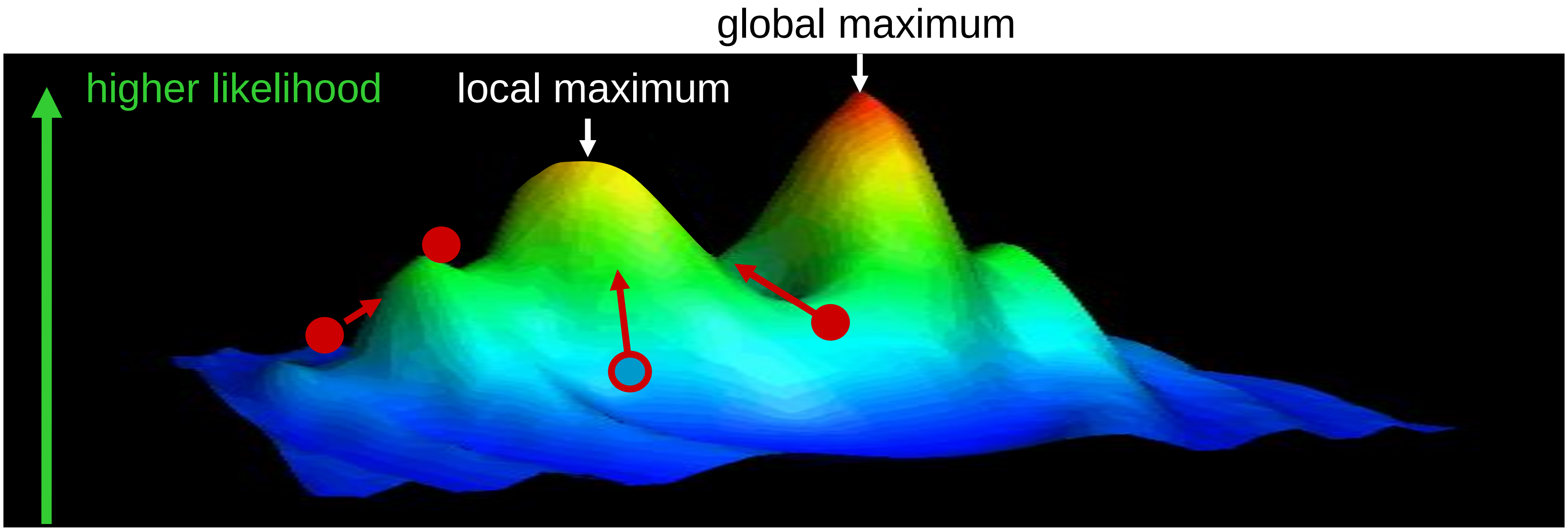
- the hypothesis H is a combination of topology of branches, branch length and parameter of the substitution model

- we may approximate the posterior distribution for H using Marcov Chain Monte Carlo (MCMC) methods

Bayesian analysis

Bayesian analysis step-by-step:

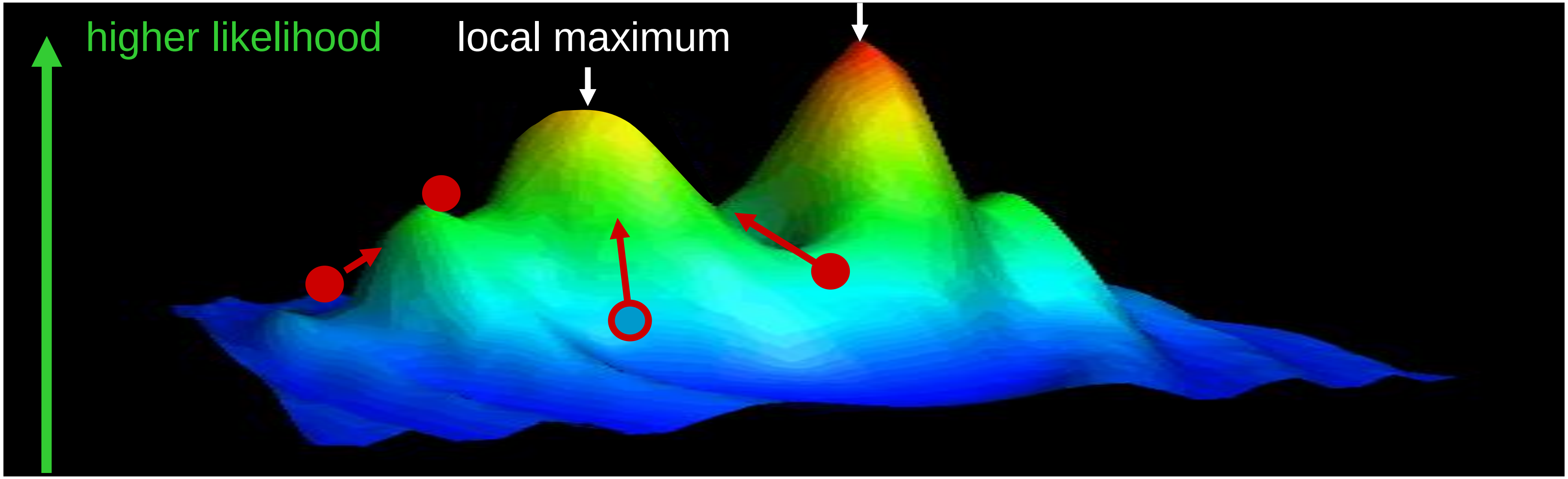
- 4 chains
- 3D space (area) with all possible trees
- find (built) first tree, compute likelihood (L)
- second tree, compute L
- if L is better, jump to the second tree, if not, stay with the first one

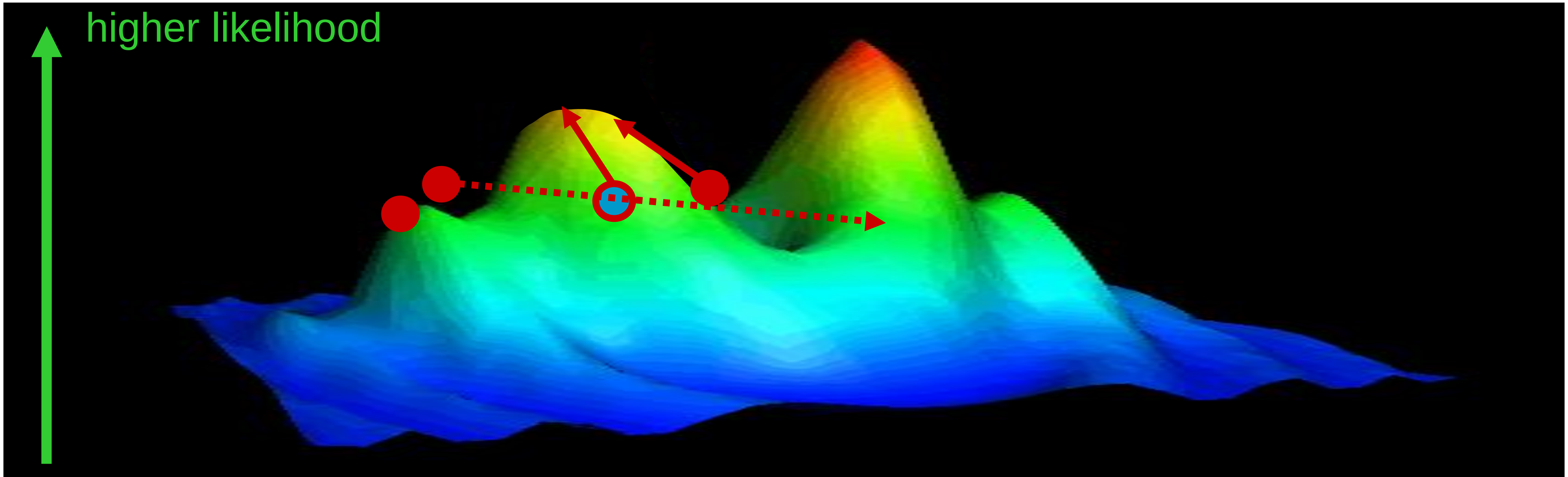


Two types of chains:

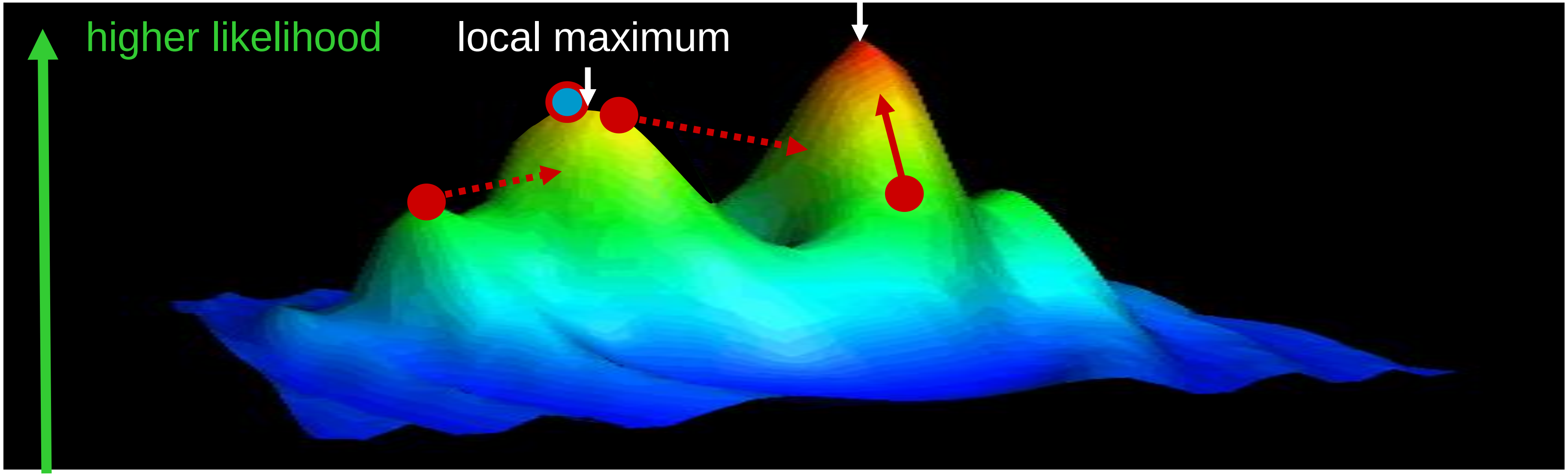
Cold – conservative one, can jump only upwards, if finds better L value

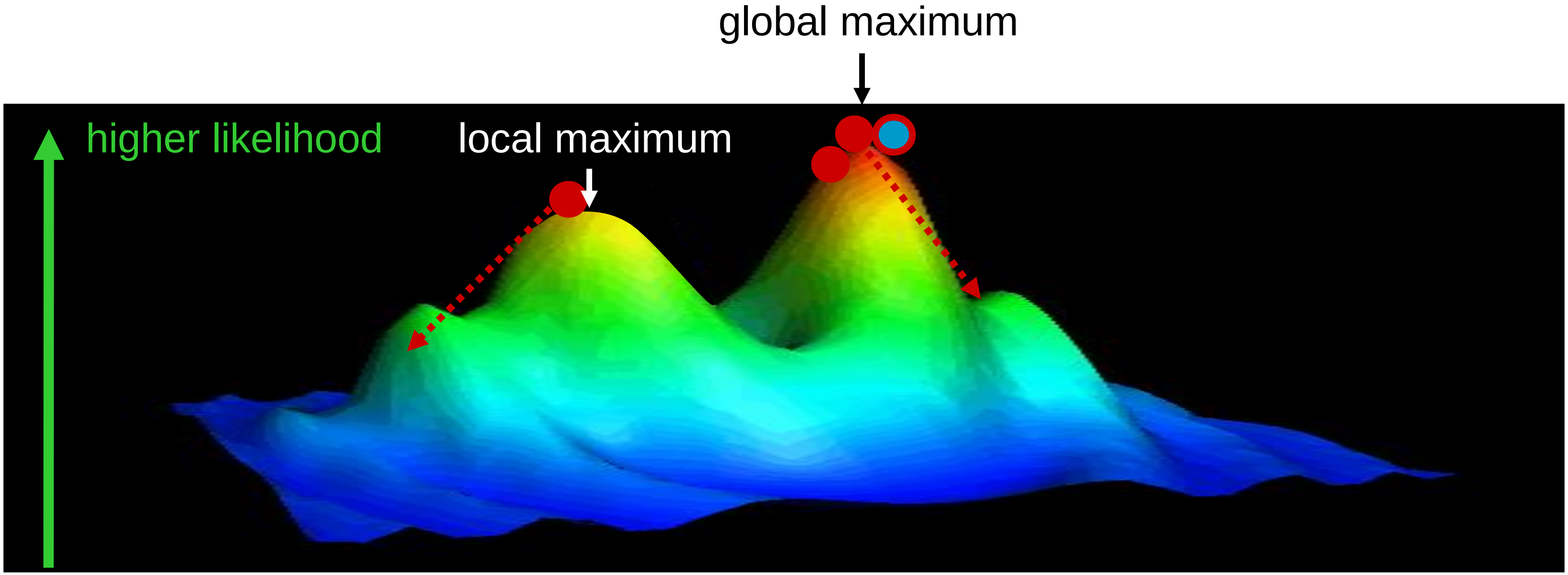
Warm – three chains – can jump also downwards + jump accidentaly + call cold one if find better topology





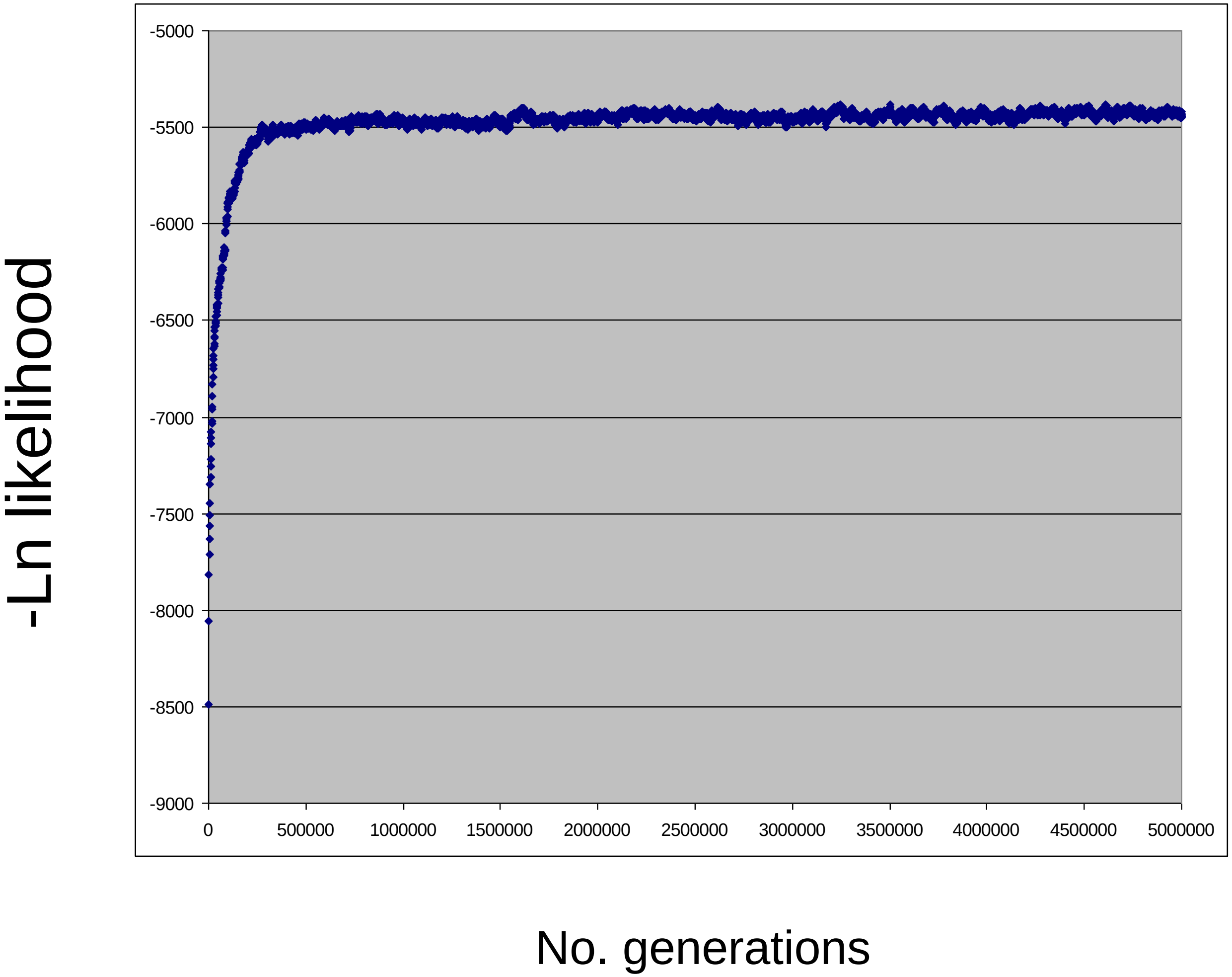
global maximum





- If there are enough generations (i.e. search steps) cold chain finds the highes global L

MrBayes run



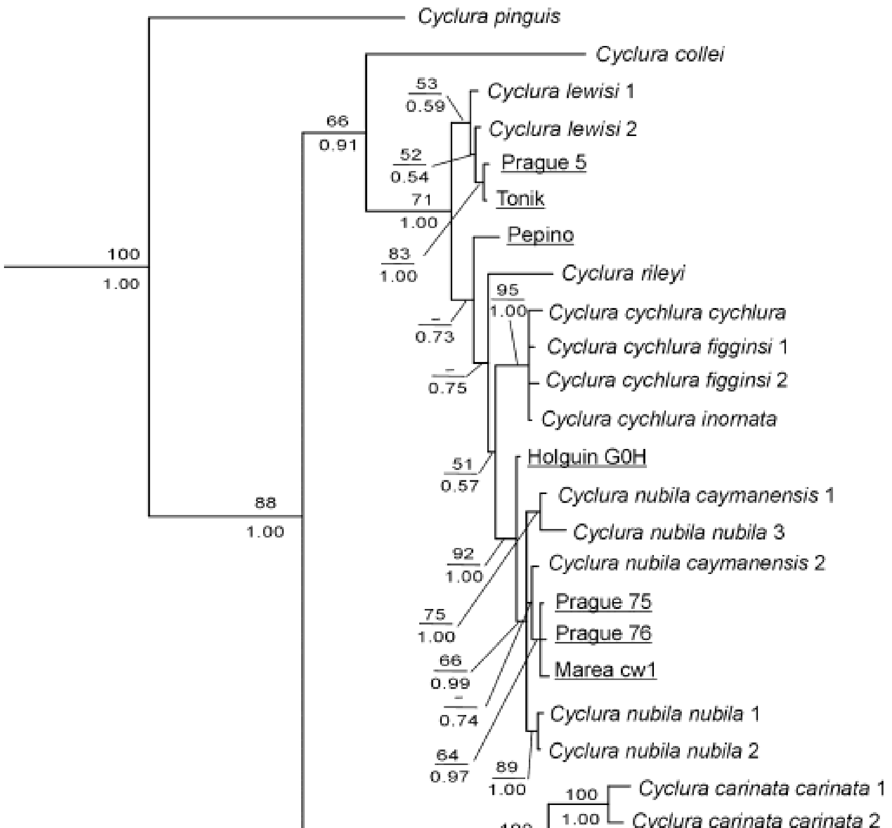
- Output of **MrBayes** is file with all trees found by cold chain during the procedure Usually every 100th tree from milions generation is saved
- Usually we have two runs

Trees at the beginning of run are not OK – we have to cut them (burnin)
Programu AWTY http://king2.scs.fsu.edu/CEBProjects/awty/awty_start.php

Posterior probabilty

BPP (PP) is parameter of Bayesian analysis – instead of bootstraps

- BPP: represent the probability that the corresponding clade is true conditional on the model, the priors, and the data
- **below 0.95 – 0.9 topology is considered unreliable**



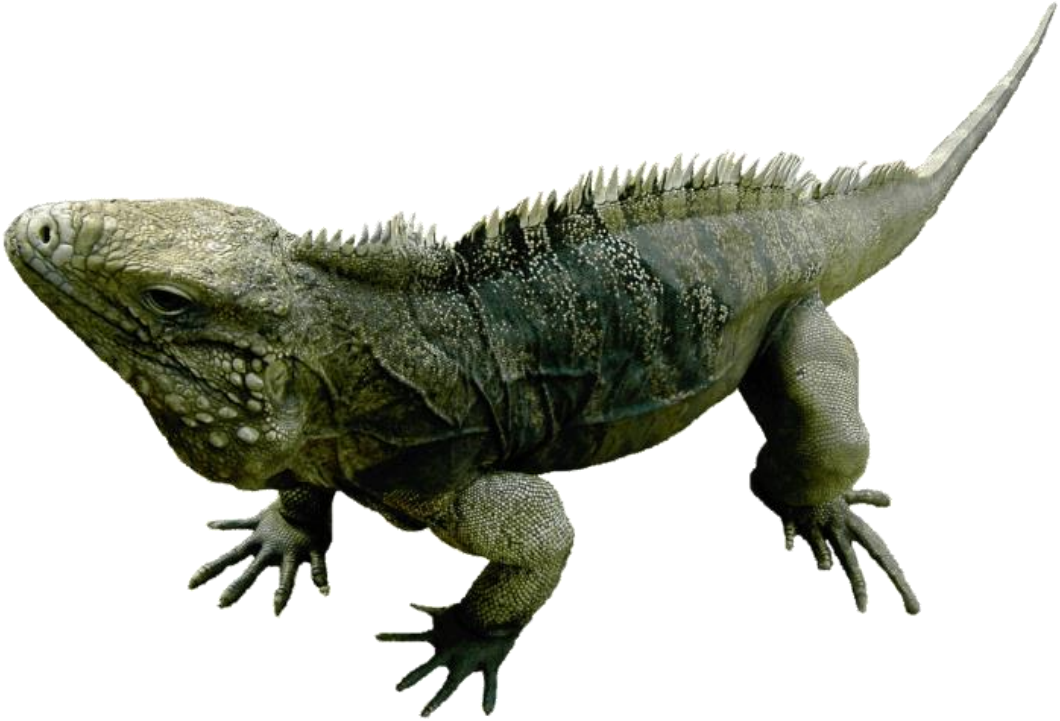
How to make phylogenetic trees?

Workflow:

- ✓ obtain DNA sequence
- ✓ quality check
- ✓ sequence alignment
- ✓ calculating genetic distances
- ✓ phylogeny estimation – topology and branch length
- reliability test (bootstrap)
- tree visualization

Program MEGA - http://www.megasoftware.net/

- open alignment
- calculate p-distances and e.g.K2P
- NJ
- MP
- do not forget about bootstrap
- What is our sample?



fasta

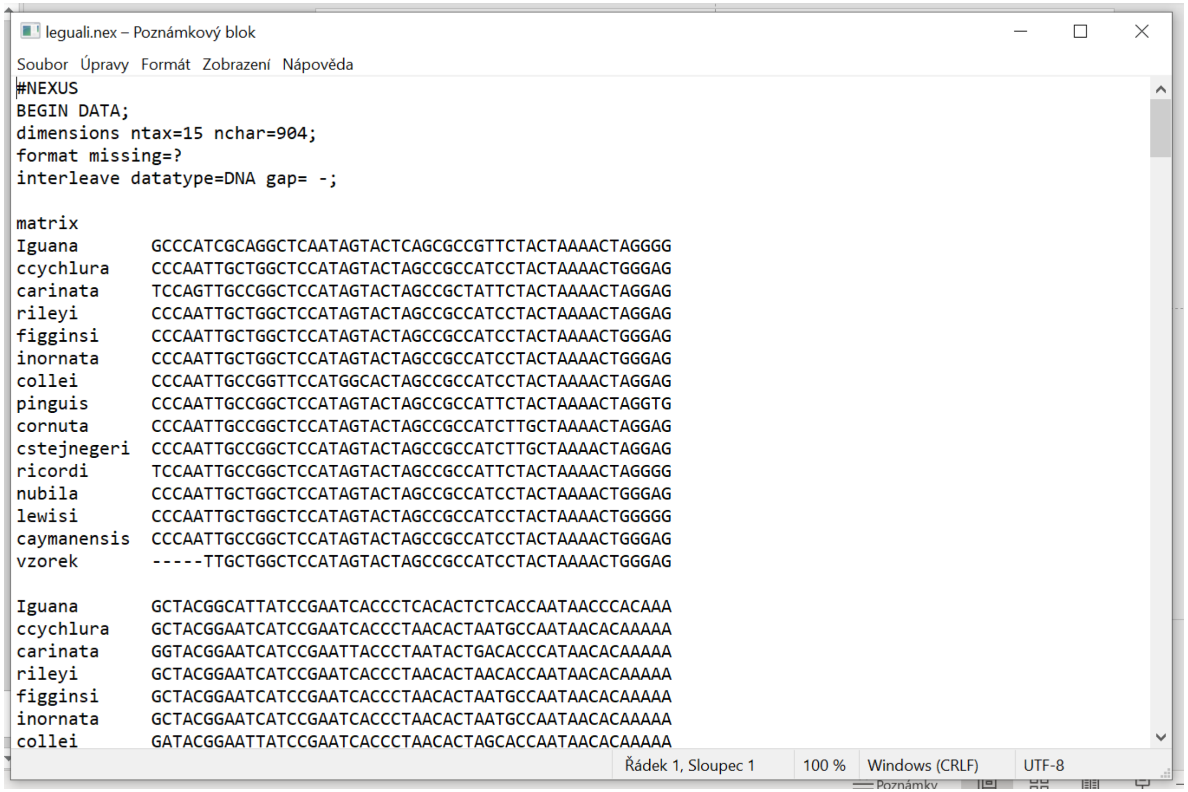
```
>gi|148832288|gb|EF443167.1| Rhinopoma hardwickei haplotype 2949
cytochrome b gene, partial cds; mitochondrial

ATGACCCACATCCGAAAATCCCACCCCTTATTCAAAATTATCAACGACTCATTCATCGACCTACC
AGCTCCATCAAACATTTCCTCCTGATGAAATTTTGGGTCCCTACTAGGTATTTGTTTAGCTGTAC
AAATCTTAACAGGACTGTTCCTAGCAATACATTATACATCAGATACCACAACCGCCTTCTACTCT
GTTACCCATATCTGCCGAGACGTAAATTACGGCTGAATCCTACGTTACCTCCATGCCAACGGAGC
ATCCATATTCTTCATCTGCCTATTTATACATGTAGGCCGAGGCATCTATTACGGCTCATACCTAT
TCACAGAAACATGAAACATTGGCATTATCCTTCTATTCGCCGTAATAGCAACAGCATTCATAGGC
TATGTCCTCCCA

>gi|...

ATGA...
```

nexus



```
#NEXUS
BEGIN DATA;
dimensions ntax=15 nchar=904;
format missing=?
interleave datatype=DNA gap= -;

matrix
Iguana       GCCCATCGCAGGCTCAATAGTACTCAGCGCCGTTCTACTAAAACTAGGGG
ccychlura    CCCAATTGCTGGCTCCATAGTACTAGCCGCCATCCTACTAAAACTGGGAG
carinata     TCCAGTTGCCGGCTCCATAGTACTAGCCGCTATTCTACTAAAACTAGGAG
rileyi       CCCAATTGCTGGCTCCATAGTACTAGCCGCCATCCTACTAAAACTAGGAG
figginsi     CCCAATTGCTGGCTCCATAGTACTAGCCGCCATCCTACTAAAACTGGGAG
inornata     CCCAATTGCTGGCTCCATAGTACTAGCCGCCATCCTACTAAAACTGGGAG
collei       CCCAATTGCCGGTTCCATGGCACTAGCCGCCATCCTACTAAAACTAGGAG
pinguis      CCCAATTGCCGGCTCCATAGTACTAGCCGCCATTCTACTAAAACTAGGTG
cornuta      CCCAATTGCCGGCTCCATAGTACTAGCCGCCATCTTGCTAAAACTAGGAG
cstejnegeri  CCCAATTGCCGGCTCCATAGTACTAGCCGCCATCTTGCTAAAACTAGGAG
ricordi      TCCAATTGCCGGCTCCATAGTACTAGCCGCCATTCTACTAAAACTAGGGG
nubila       CCCAATTGCTGGCTCCATAGTACTAGCCGCCATCCTACTAAAACTGGGAG
lewisi       CCCAATTGCTGGCTCCATAGTACTAGCCGCCATCCTACTAAAACTGGGGG
caymanensis  CCCAATTGCCGGCTCCATAGTACTAGCCGCCATCCTACTAAAACTGGGAG
vzorek       -----TTGCTGGCTCCATAGTACTAGCCGCCATCCTACTAAAACTGGGAG

Iguana       GCTACGGCATTATCCGAATCACCCTCACACTCTCACCAATAACCCACAAA
ccychlura    GCTACGGAATCATCCGAATCACCCTAACACTAATGCCAATAACACAAAAA
carinata     GGTACGGAATCATCCGAATTACCCTAATACTGACACCCATAACACAAAAA
rileyi       GCTACGGAATCATCCGAATCACCCTAACACTAACACCAATAACACAAAAA
figginsi     GCTACGGAATCATCCGAATCACCCTAACACTAATGCCAATAACACAAAAA
inornata     GCTACGGAATCATCCGAATCACCCTAACACTAATGCCAATAACACAAAAA
collei       GATACGGAATTATCCGAATCACCCTAACACTAGCACCAATAACACAAAAA
```

Bayesian inference/analysis

program MrBayes – uses commands

Important commands:
execute iguanna.nex (upload of the file)
lset nst=6 rates=invgamma (simplified model of the sequence evolution – from Modeltest) … setting GTR model of sequence evolution
mcmcp ngen=500000 samplefreq=100; (parameter settings, number of generations, frequency of sampling=saving trees)
mcmc (start)

After the run “average standard deviation of split frequencies“ - should be lower than 0.01
check files –lnL (Tracer, Excel)

If everything seems ok, then:
sump burnin=1250
sumt burnin=1250 (this will delete first 25% of saved trees and create file xxx.con) add .tre and open it in TreeView

Molecular Phylogenetics and Evolution
Vol. 17, No. 2, November, pp. 269–279, 2000
doi:10.1006/mpev.2000.0836, available online at http://www.idealibrary.com on IDEAL®

Phylogeography of the Caribbean Rock Iguana (*Cyclura*): Implications for Conservation and Insights on the Biogeographic History of the West Indies[1]

Catherine L. Malone,[2] Tana Wheeler, Jeremy F. Taylor, and Scott K. Davis