Barre-codes ADN, biodiversité, échantillons environnementaux, et nouvelles techniques de séquençage

Eric Coissac, François Pompanon, Christian Miquel, Ludovic Gielly, Alice Valentini, Gilles Rayé, Wasim Shehzad, Tiayyba Riaz, Pierre Taberlet

Laboratoire d'Ecologie Alpine, CNRS UMR 5553

Université Joseph Fourier, Grenoble, France

Outline

- The current revolution in DNA sequencing
- DNA barcoding: concept and definitions
- DNA barcoding in ecology
  - The *trn*L approach
  - Analysis of environmental samples
  - Diet analysis
- DNA barcoding and bioinformatics
  - How to choose the most efficient DNA barcode?
  - How to deal with sequencing errors?
  - How to analyse the output of next generation sequencers?

Outline

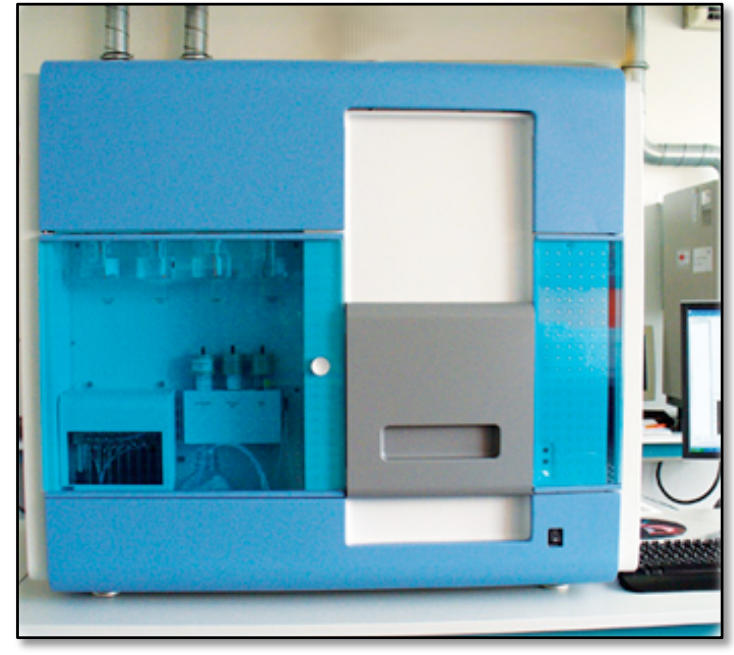
- The current revolution in DNA sequencing
- DNA barcoding: concept and definitions
- DNA barcoding in ecology
  - The *trn*L approach
  - Analysis of environmental samples
  - Diet analysis
- DNA barcoding and bioinformatics
  - How to choose the most efficient DNA barcode?
  - How to deal with sequencing errors?
  - How to analyse the output of next generation sequencers?

454 GS FLX™

- Company: Roche Diagnostic®
- Website: https://roche-applied-science.com/sis/sequencing/flx/index/jsp
- Fragment length: 400 bases
- Number of reads per run: $1\ 10^6$
- Total output per run: 0.4 Gb per run
- Time per run: 8 hours

Genetic Analyzer™/Solexa™

- Company: Illumina®
- Website: www.solexa.com
- Fragment length: 35-105 bases
- Number of reads per run: 60 10^6
- Total output per run: 6 Gb
- Time per run: 3.5 days

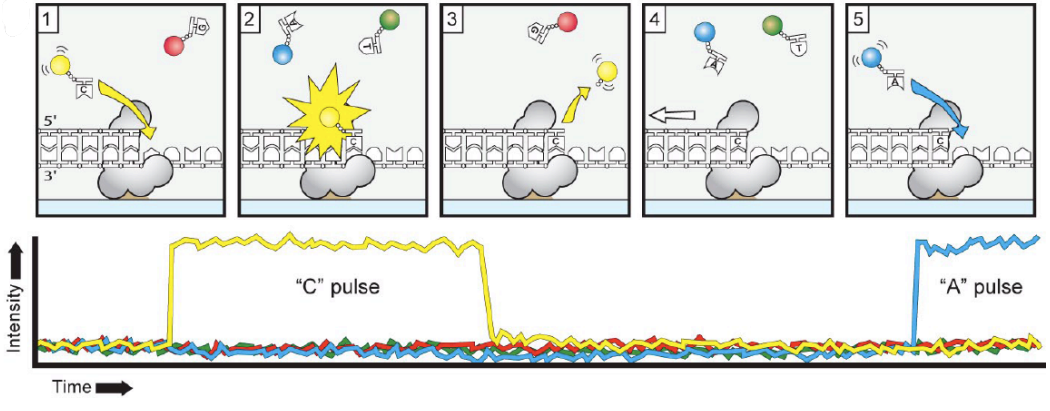


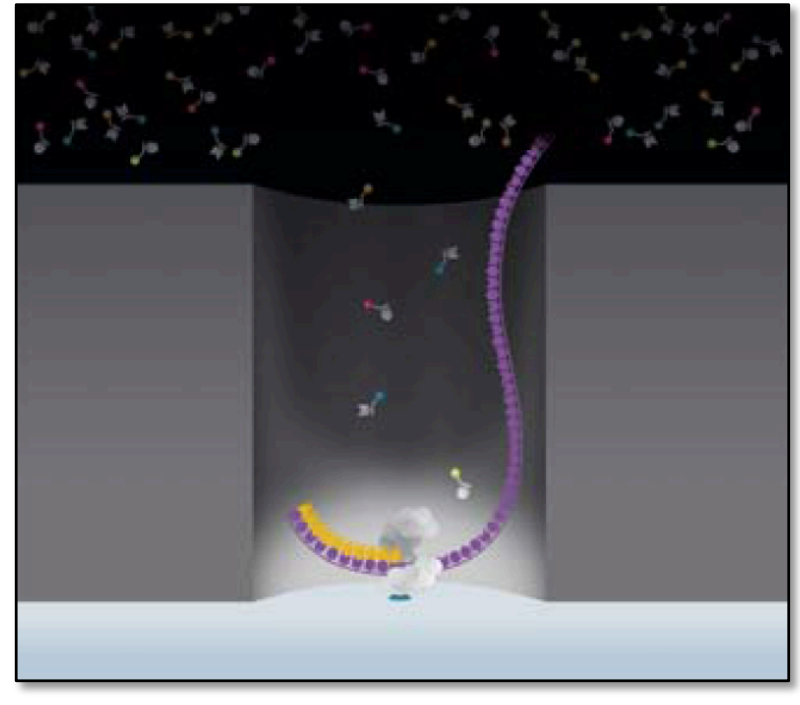
SOLiD™ DNA Sequencer

- Company: Applied Biosystem®
- Website: solid.appliedbiosystems.com
- Fragment length: 50 bases
  (possibility of "paired-end" sequencing: from 0.6 to 10 kb)
- Number of reads per run: 400 10^6
- Total output per run: 20 Gb
- Time per run: 6 days

Pacific Biosciences

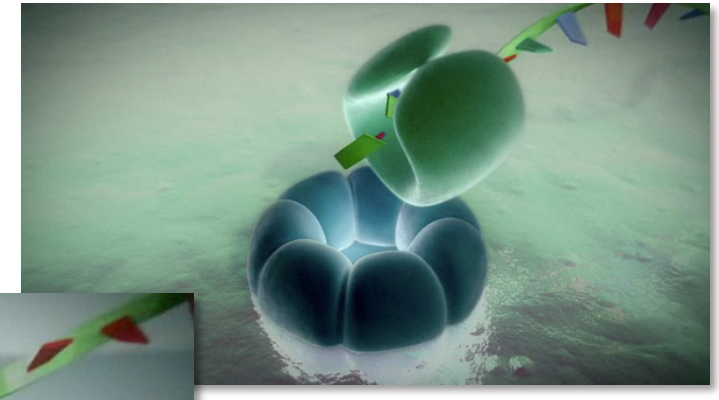
- Website: www.pacificbiosciences.com
- De-novo synthesis of single molecules
- Output speed: One genome in minutes
- Read length: 10 kb
- Cost: Less than 100 $

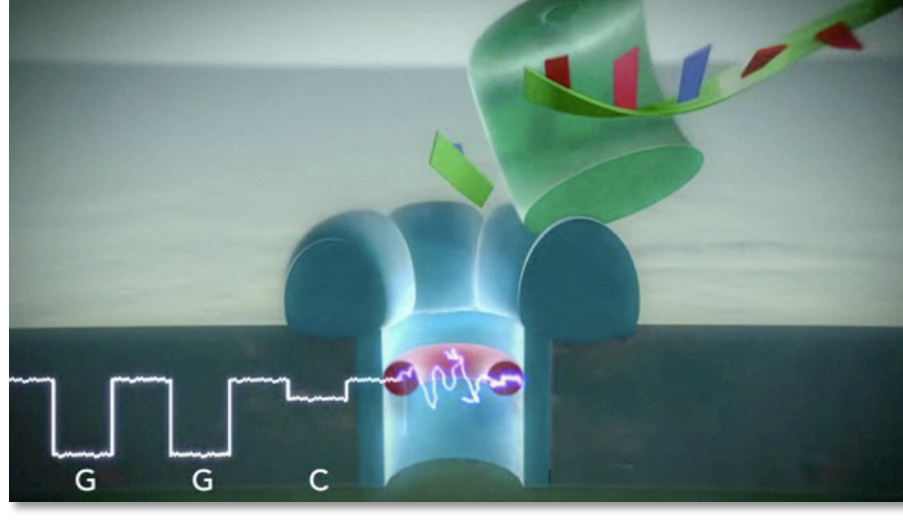




Oxford Nanopore Technology

- Website: www.nanoporetech.com
- *De novo* sequencing of single molecules without fluorescence
- Output speed: ?
- Read length: ?
- Cost: ?





The current revolution in DNA sequencing

- Gigabases of data
- Sequencing of single molecules from a mixture of molecules
- Relatively short read length at the moment
- New approach for preparing the samples

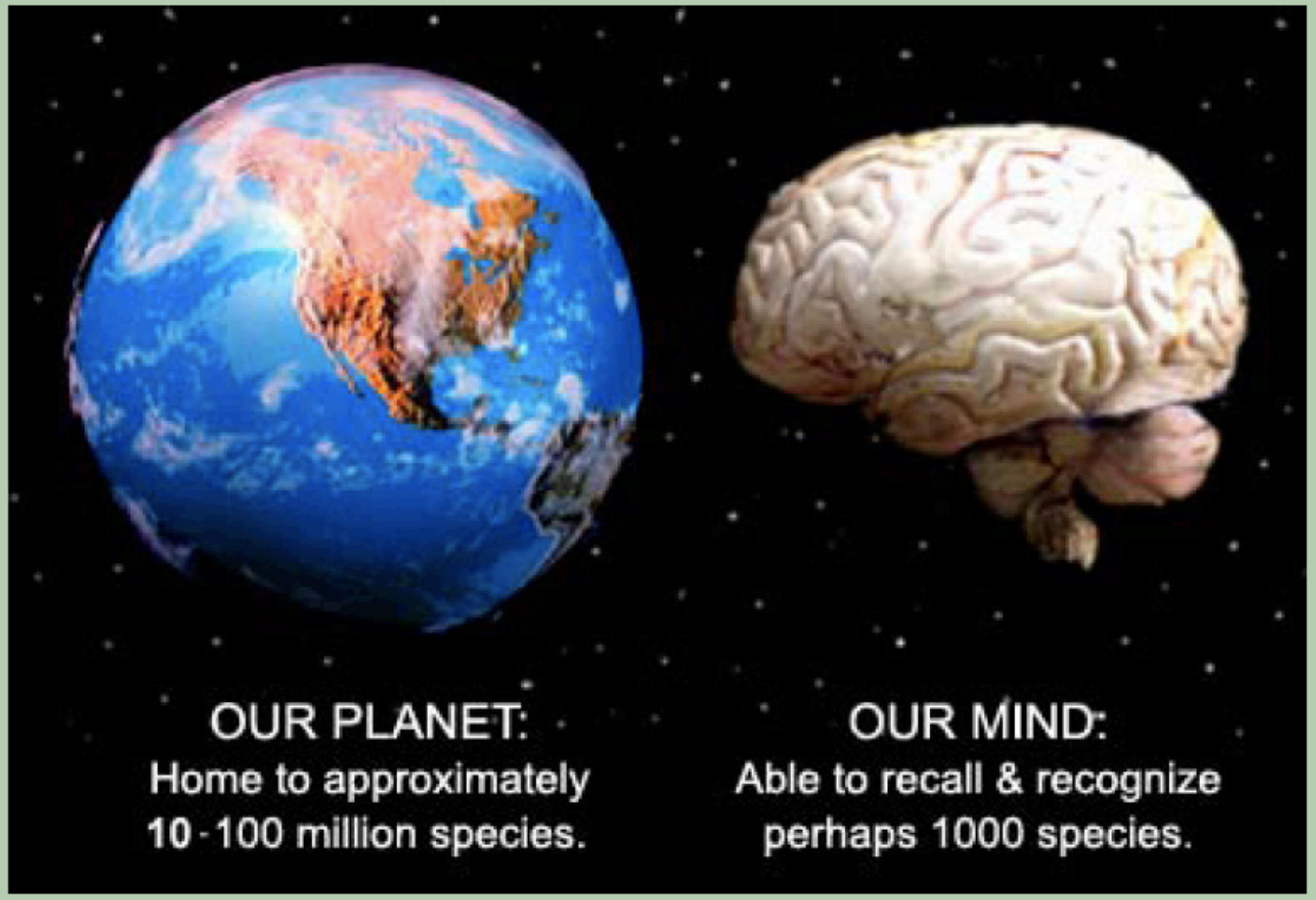
Outline

- The current revolution in DNA sequencing
- DNA barcoding: concept and definitions
- DNA barcoding in ecology
  - The *trn*L approach
  - Analysis of environmental samples
  - Diet analysis
- DNA barcoding and bioinformatics
  - How to choose the most efficient DNA barcode?
  - How to deal with sequencing errors?
  - How to analyse the output of next generation sequencers?

Consortium for the Barcode of Life

- Barcoding all eukaryotes
- Launch April 2004

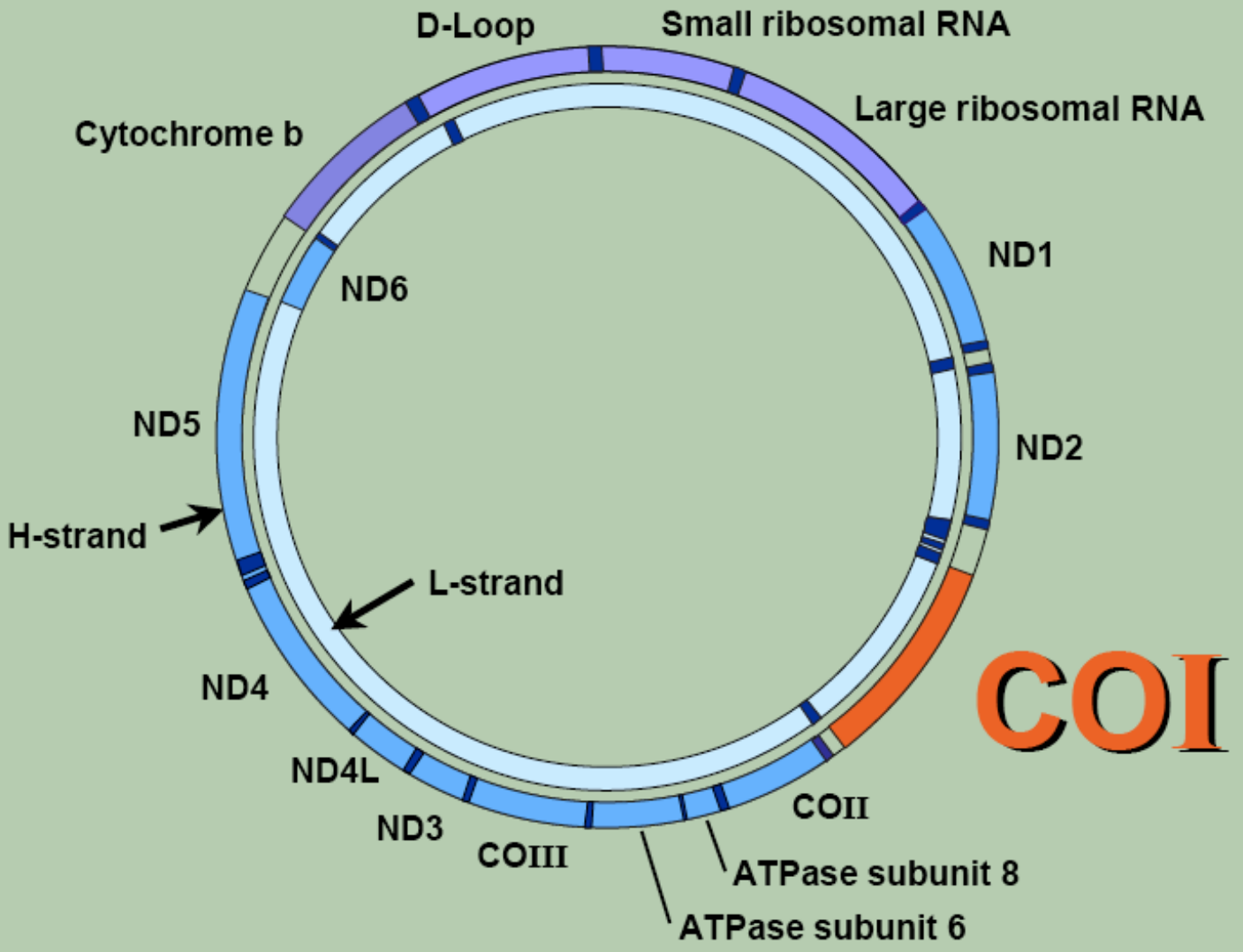
Identifying Life



DNA Barcode:

short sequence enabling species discrimination

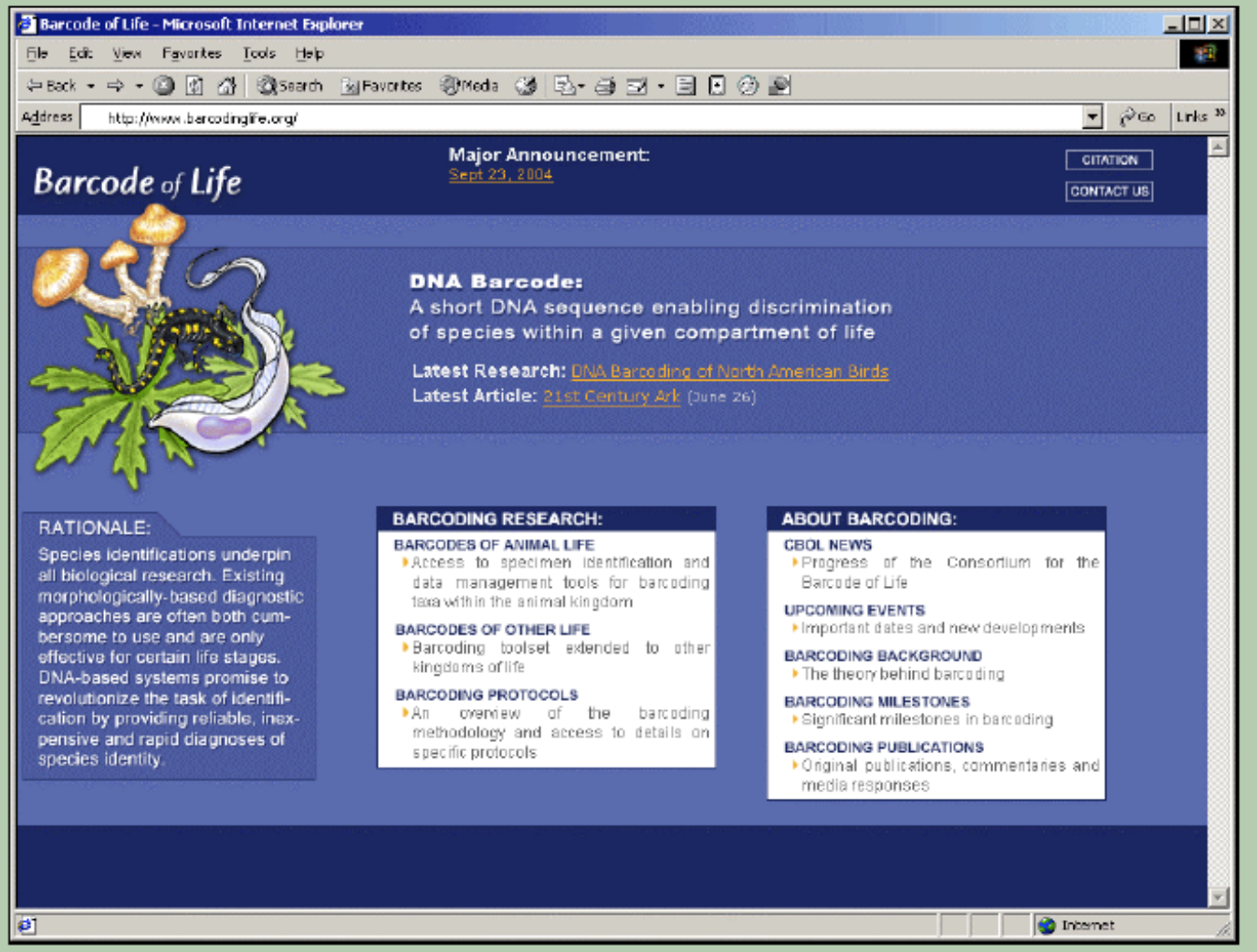
The Mitochondrial Genome



Barcoding Animals (10M species)

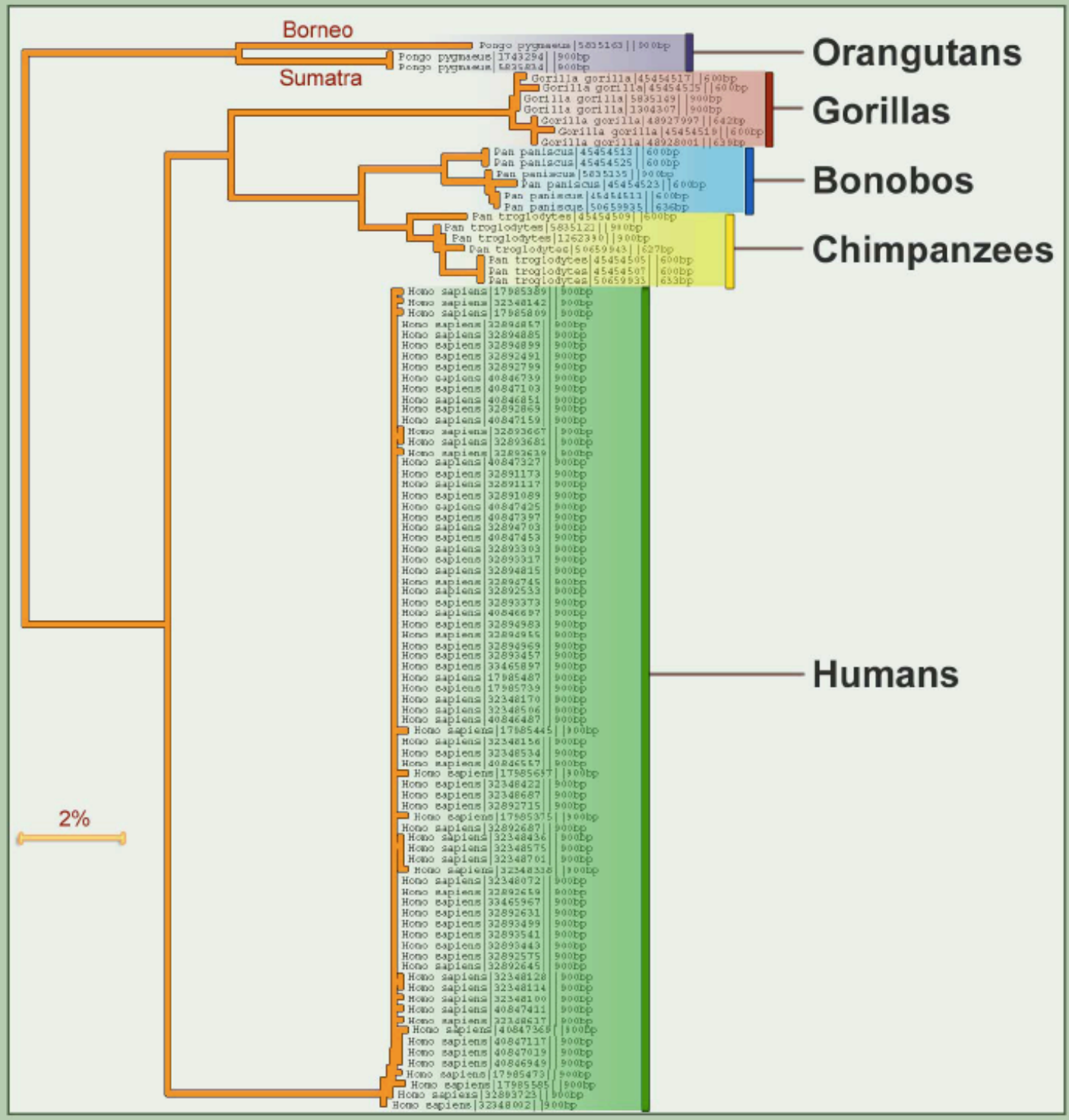
- Effective in varied geographic settings
- Effective in varied taxonomic groups
- >99.99% resolution

Barcode of Life Database (BOLD)



10 million species files, 50 million specimen files

Barcoding - The Future


Borneo
Sumatra
Orangutans
Gorillas
Bonobos
Chimpanzees
Humans
2%

DNA barcoding: definitions

- DNA barcoding *sensu stricto*: identification of the species level using a single standardized DNA fragment; (= CBoL view)
- DNA barcoding *sensu lato*: identification of any taxonomical level using any DNA fragment

Valentini A, Pompanon F, Taberlet P (2009) DNA barcoding for ecologists. *Trends in Ecology and Evolution*, **24**, 110-117.

The ideal DNA barcoding marker

- Variability: the gene region sequenced should be nearly identical among individuals of the same species, but different between species
- Standardization: the same DNA region used for different taxonomic groups
- Phylogenetic information: to easily assign unknown or not yet "barcoded" species to their taxonomic group (genus, family, etc.)
- Robustness: highly conserved priming sites, and highly reliable DNA amplifications and sequencing
- Shortness: amplification of degraded DNA

The ideal DNA barcoding marker

- Variability: the gene region sequenced should be nearly identical among individuals of the same species, but different between species
- Standardization: the same DNA region used for different taxonomic groups
- Phylogenetic information: to easily assign unknown or not yet "barcoded" species to their taxonomic group (genus, family, etc.)
- Robustness: highly conserved priming sites, and highly reliable DNA amplifications and sequencing
- Shortness: amplification of degraded DNA

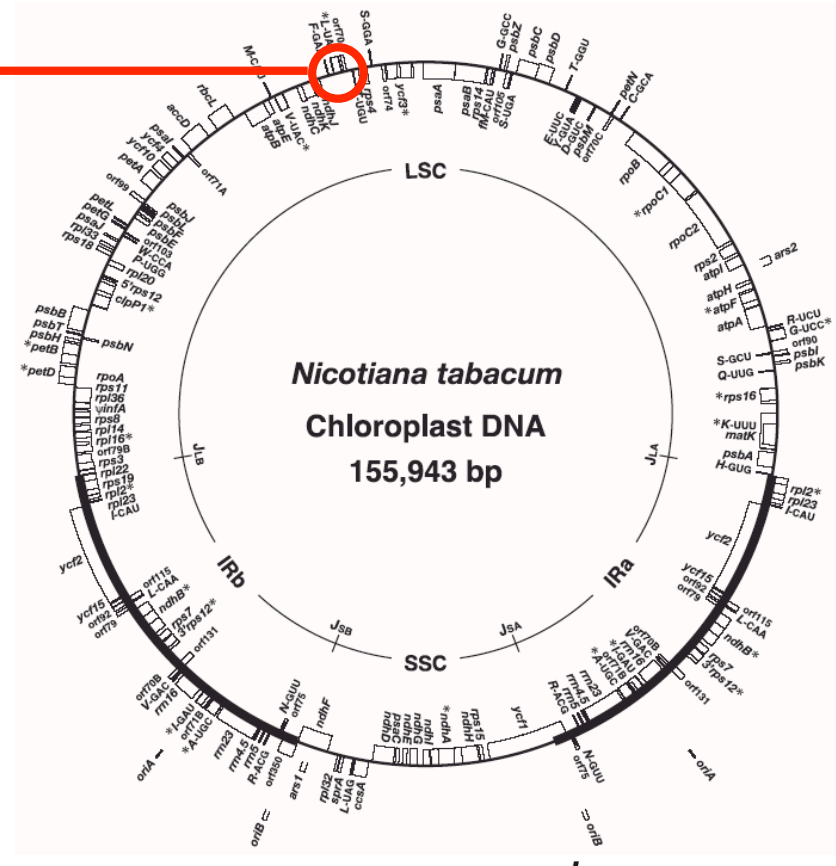
The ideal DNA barcoding marker

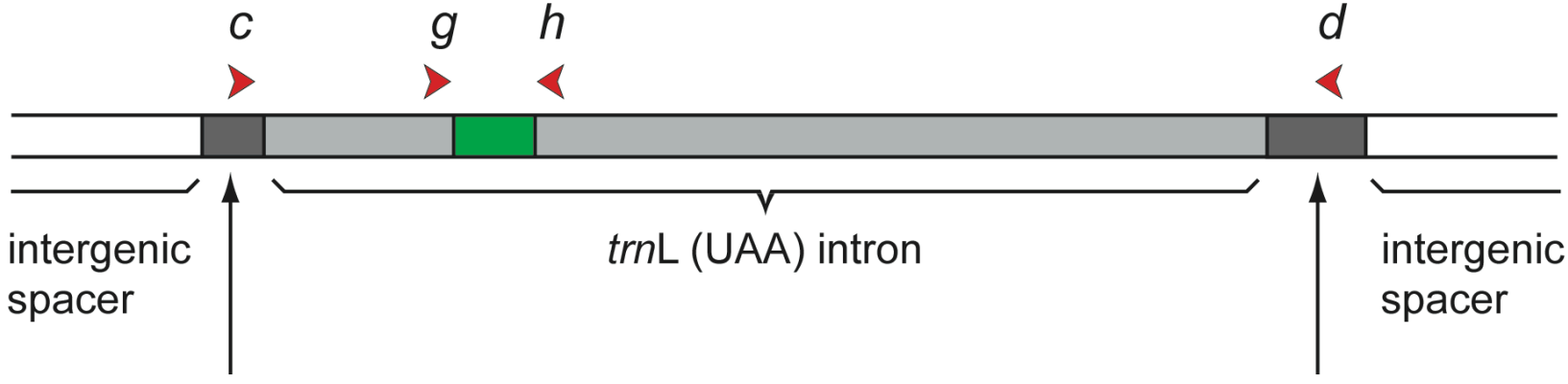
- Variability: the gene region sequenced should be nearly identical among individuals of the same species, but different between species
- Standardization: the same DNA region used for different taxonomic groups
- Phylogenetic information: to easily assign unknown or not yet "barcoded" species to their taxonomic group (genus, family, etc.)
- Robustness: highly conserved priming sites, and highly reliable DNA amplifications and sequencing
- Shortness: amplification of degraded DNA

Outline

- The current revolution in DNA sequencing
- DNA barcoding: concept and definitions
- DNA barcoding in ecology
  - The *trn*L approach
  - Analysis of environmental samples
  - Diet analysis
- DNA barcoding and bioinformatics
  - How to choose the most efficient DNA barcode?
  - How to deal with sequencing errors?
  - How to analyse the output of next generation sequencers?

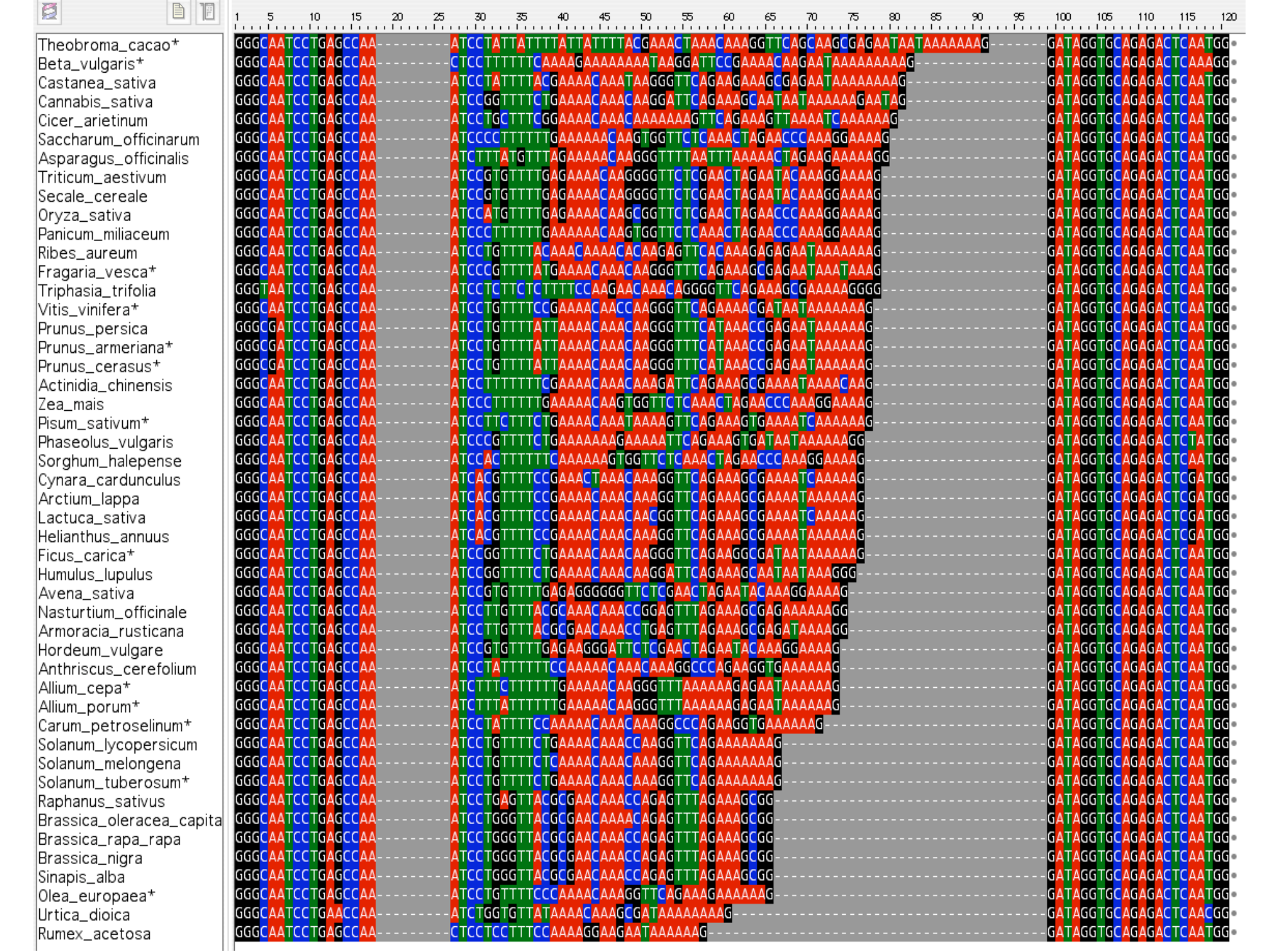
The chloroplast *trn*L(UAA) intron





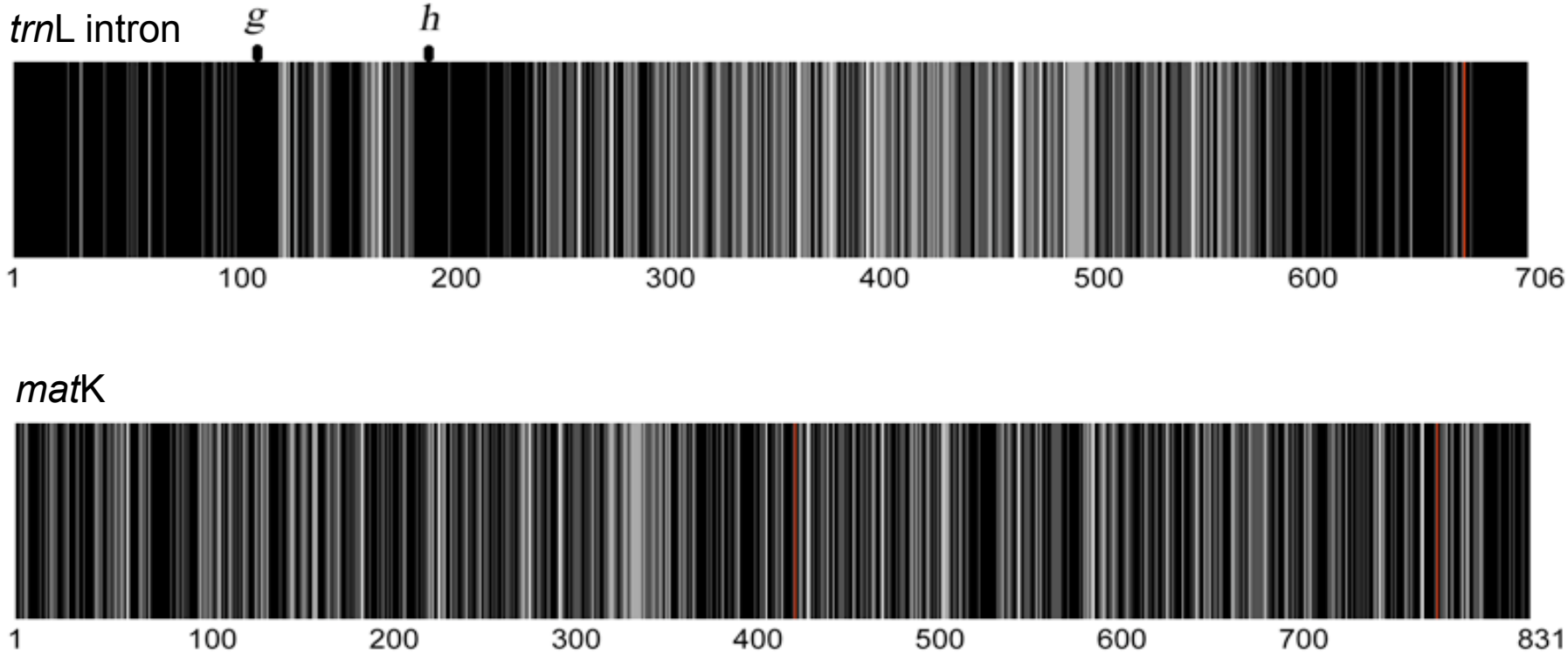
*trn*L (UAA) exon 1
positions 49296 to 49330
of tobacco chloroplast DNA

*trn*L (UAA) exon 2
positions 49833 to 49883
of tobacco chloroplast DNA


Theobroma_cacao*
Beta_vulgaris*
Castanea_sativa
Cannabis_sativa
Cicer_arietinum
Saccharum_officinarum
Asparagus_officinalis
Triticum_aestivum
Secale_cereale
Oryza_sativa
Panicum_miliaceum
Ribes_aureum
Fragaria_vesca*
Triphasia_trifolia
Vitis_vinifera*
Prunus_persica
Prunus_armeriana*
Prunus_cerasus*
Actinidia_chinensis
Zea_mais
Pisum_sativum*
Phaseolus_vulgaris
Sorghum_halepense
Cynara_cardunculus
Arctium_lappa
Lactuca_sativa
Helianthus_annuus
Ficus_carica*
Humulus_lupulus
Avena_sativa
Nasturtium_officinale
Armoracia_rusticana
Hordeum_vulgare
Anthriscus_cerefolium
Allium_cepa*
Allium_porum*
Carum_petroselinum*
Solanum_lycopersicum
Solanum_melongena
Solanum_tuberosum*
Raphanus_sativus
Brassica_oleracea_capita
Brassica_rapa_rapa
Brassica_nigra
Sinapis_alba
Olea_europaea*
Urtica_dioica
Rumex_acetosa
1 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120

The *trn*L approach

- Why the *trn*L?

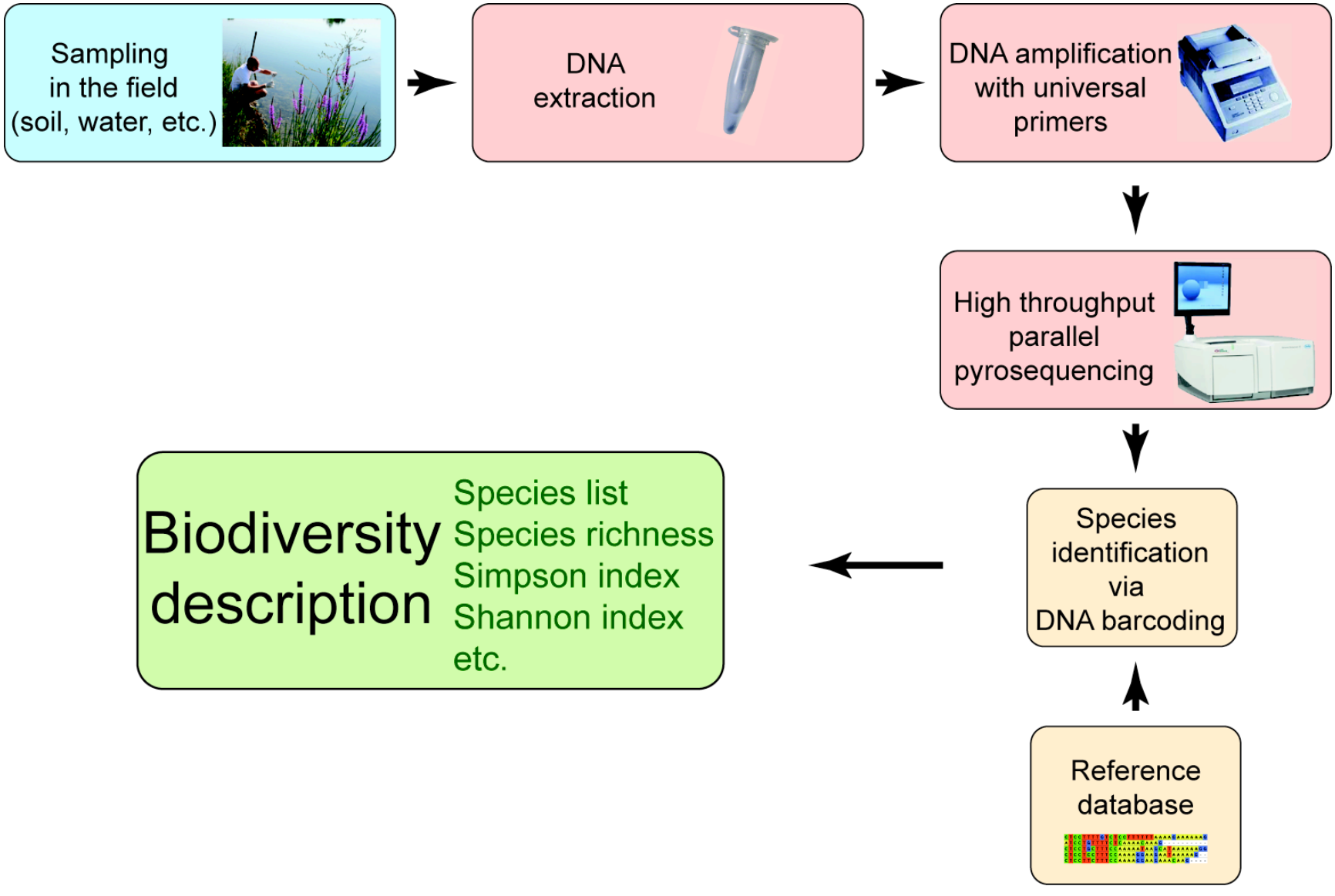


Variability of the two chloroplast regions *trn*L intron and *mat*K in 54 species (after J.H. Sønstebø)

*trn*L (UAA) intron: P6 loop resolution

	Number of species in the reference database	% of family identified	% of genera identified	% of species identified
Deosai National Park, Pakistan	91	100	95	75
Arctic flora	842	100	77	33
Bauges Regional Park, France	477	100	95	72

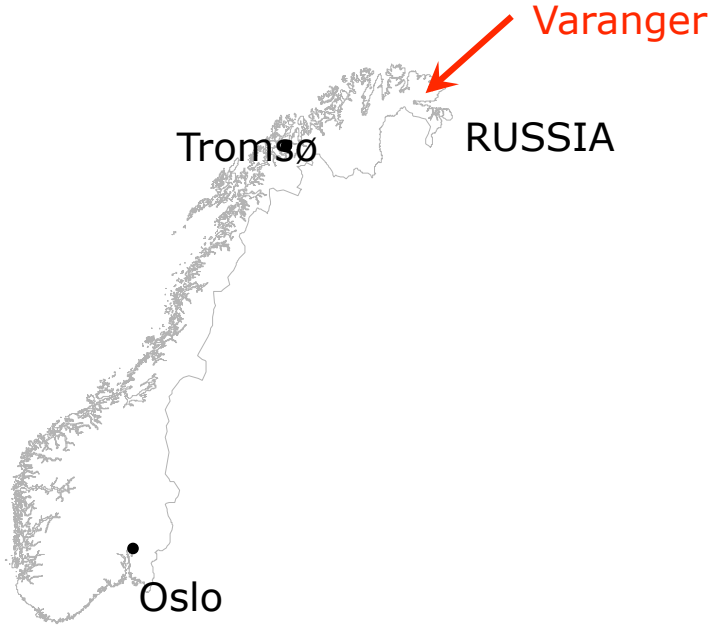
DNA-based biodiversity assessment



The *trn*L database for Arctic plant species

- 842 species was successfully sequenced for the *trn*L intron
- 709 species were sequenced twice or more
- 33.5% of the species and 77.1% of the genera could be identified by the P6 loop
- Low resolution in Salicaceae (*Salix* sp.), in Cyperaceae (*Carex* sp.), in Asteraceae, and in Poaceae.

Calibration of soil DNA with aboveground species at Varanger, Northern Norway



DNA-based soil analysis

Nb Seq	Family	Genus	Species
1614			*Bistorta vivipara*
1118			*Avenella flexuosa*
894	Salicaceae	3	*Salix sp., Chosenia arbutifolia, Populus balsamifera*
563	Asteraceae	4	*Hieracium alpinum, H.* sect. *Sylvatica, Lactuca sibirica, Saussurea alpina, S. tilesii, Taraxacum croceum, T. lacerum*
161		*Ranunculus*	*acris / subborealis / turneri*
155			*Anthoxanthum nipponicum*
133			*Alchemilla glomerulans*
89		*Deschampsia*	*alpina / anadyrensis / brevifolia / cespitosa / sukatchewii*
85		*Festuca*	*baffinensis /brachyphylla / brevissima / edlundiae / hyperborea / kolymensis / lenensis / ovina / rubra*
71		*Equisetum*	*arvense / fluviatile / sylvaticum*
26			*Viola biflora*
26			*Poa alpina*
10		*Betula*	*fruticosa / glandulosa / nana / neoalaskana / pubescens*
10		*Empetrum*	*sibiricum / subholarcticum*
9			*Trientalis europaea*
8			*Rhinanthus minor*
5			*Pyrola minor*
2	Asteraceae	3	*Endocellion glaciale, E. sibiricum, Petasites frigidus, Taraxacum alaskanum, T. arcticum, T. brachyceras, T. cratophorum, T. holmenianum, T. lacerum, T. lateritium, T. phymatocarpum, T. soczavae*
1	Poaceae	8	*Alopecurus magellanicus, Anthoxanthum arcticum, A. monticola, Arctagrostis latifolia, Beckmannia syzigachne, Hyalopoa lanatiflora, Milium effusum, Poa abbreviata, P. arctica, P. eminens, P. glauca, P. hartzii, P. laxa, P. palustris, P. paucispicula, P. pratensis, P. pseudoabbreviata, P. sibirica, Vahladea atropurpurea*
1			*Crepis chrysantha*

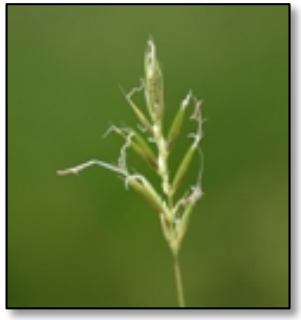
Above ground analysis

Avenella flexuosa

Poa sp.

Taraxacum sp.

Anthoxanthum nipponicum



Carex sp.

Festuca sp.

Deschampsia sp.

Rumex sp.

Calamagrostris sp.

DNA-based soil analysis

Bistorta vivipara

Salix sp.

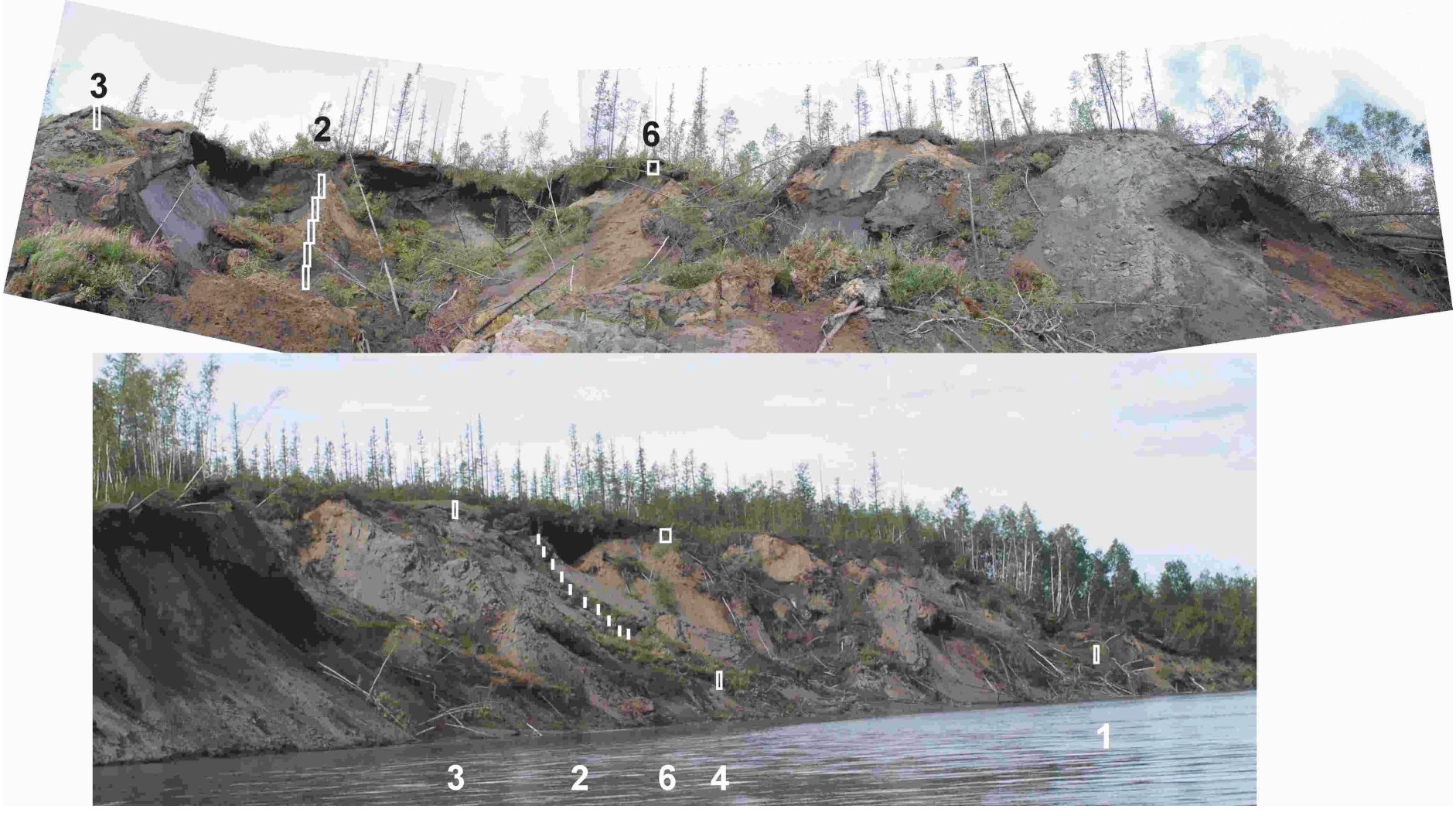
Alchemilla sp.

Viola biflora

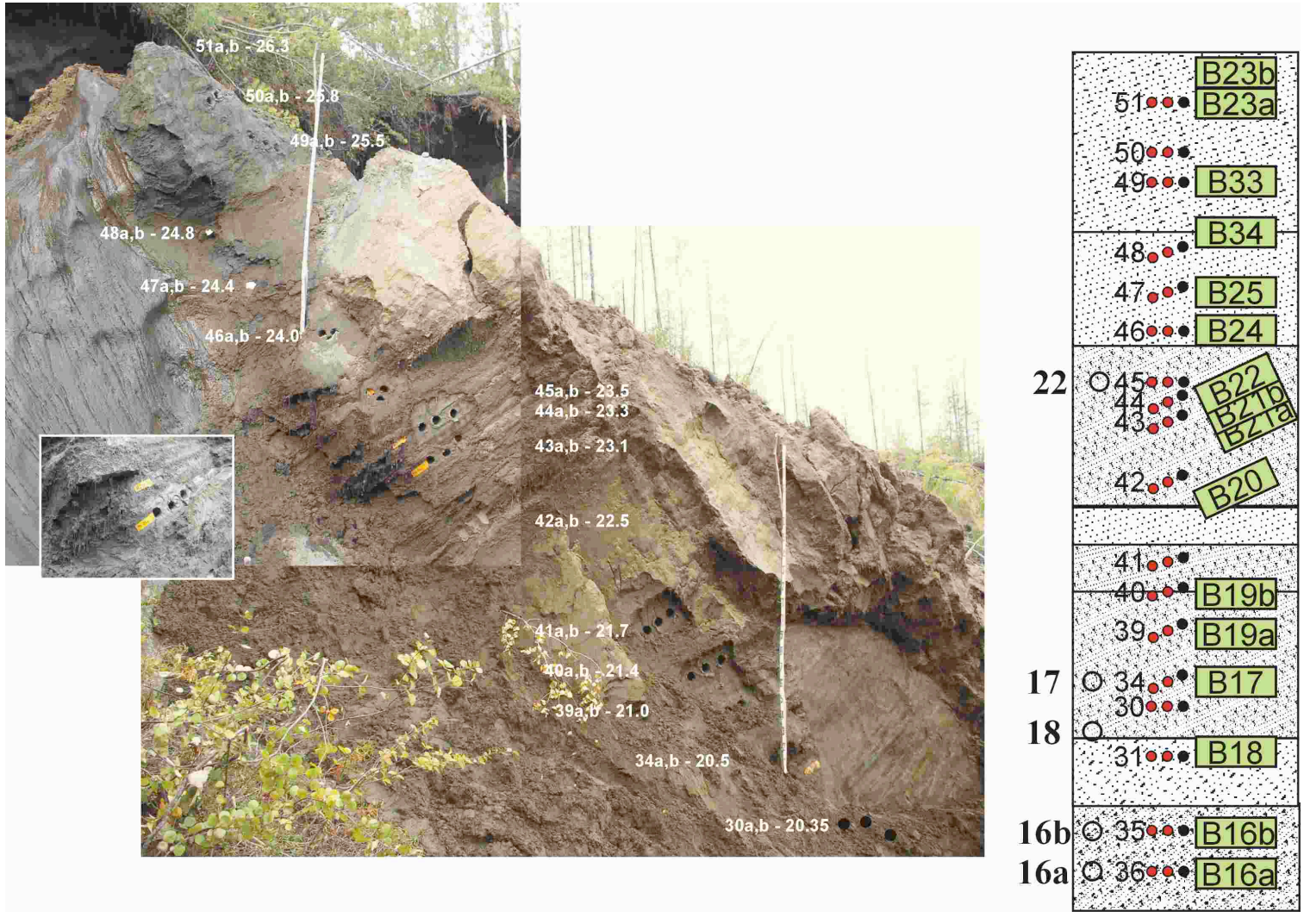
Equisetum sp.

Permafrost analysis (1)

Permafrost analysis (2)



Permafrost analysis (3)



biology letters

Biol. Lett. (2008) **4**, 423–425
doi:10.1098/rsbl.2008.0118
Published online 9 April 2008

Population genetics

Species detection using environmental DNA from water samples

Gentile Francesco Ficetola[1,2,*], Claude Miaud[2], François Pompanon[1] and Pierre Taberlet[1]

[1]*Laboratoire d'Ecologie Alpine, CNRS-UMR 5553, Université Joseph Fourier, BP 53, 38041 Grenoble Cedex 09, France*
[2]*Laboratoire d'Ecologie Alpine, CNRS-UMR 5553, Université de Savoie, 73376 Le Bourget du Lac Cedex, France*
**Author and address for correspondence: Dipartimento di Scienze dell'Ambiente e del Territorio, Università Milano Bicocca, Piazza della Scienza 1, 20126 Milano, Italy (francesco.ficetola@unimi.it).*

Experimental protocol

- Design of a PCR based system specific of *Rana catesbeiana* (79 bp of mtDNA)
- PCR experiments using water from aquariums with tadpoles
- PCR experiments using water from ponds (with and without *R. catesbeiana*)

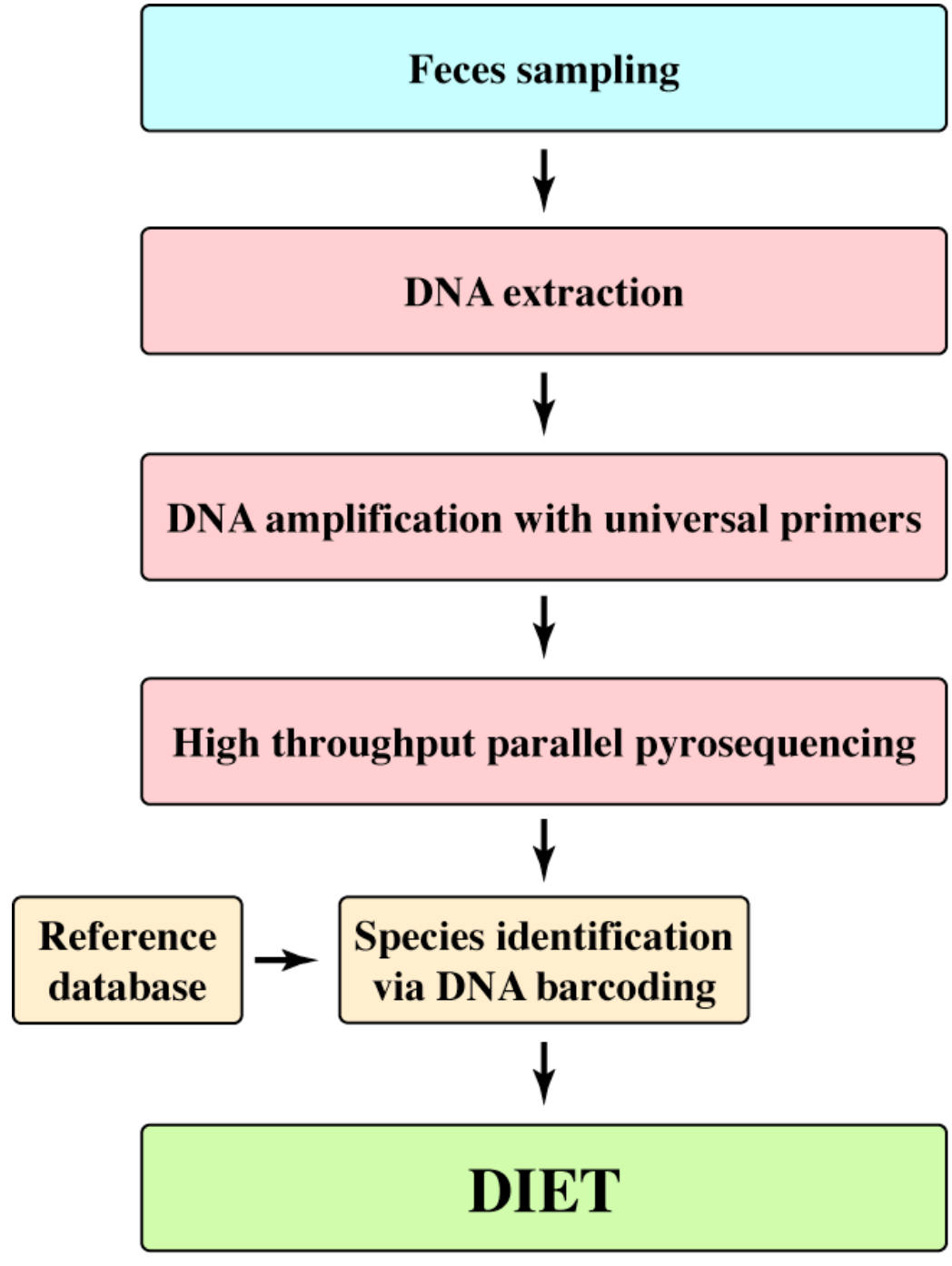
Results

- Aquariums
  - With *R. catesbeiana* tadpoles; PCR +
  - Without tadpoles; PCR -
- Ponds
  - With high densities of *R. cat*.: PCR + (22/28)
  - With low densities of *R. cat*.: PCR + (10/27)
  - Without *R. catesbeiana*: PCR -

Discussion

- Environmental DNA can be used to ascertain species presence in wetlands, even at low densities
- This technique requires several precautions (as in ancient DNA studies)
- New avenues for biodiversity assessment

DNA-based diet analysis



TECHNICAL ADVANCES

New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the *trn*L approach

ALICE VALENTINI,*† CHRISTIAN MIQUEL,* MUHAMMAD ALI NAWAZ,‡§ EVA BELLEMAIN,* ERIC COISSAC,* FRANÇOIS POMPANON,* LUDOVIC GIELLY,* CORINNE CRUAUD,¶ GIUSEPPE NASCETTI,† PATRICK WINCKER,¶ JON E. SWENSON‡** and PIERRE TABERLET*

**Laboratoire d'Ecologie Alpine, CNRS UMR 5553, Université Joseph Fourier, BP 53, F-38041 Grenoble cedex 9, France, †Dipartimento di Ecologia e Sviluppo Economico Sostenibile, Università degli Studi della Tuscia, via S. Giovanni Decollato 1, I-01100 Viterbo, Italy, ‡Department of Ecology and Natural Resource Management, Norwegian University of Life Sciences, Post Box 5003, NO-1432 Ås, Norway, §Himalayan Wildlife Foundation, 01, Park Road, Sector F-8/1 Islamabad 44000, Pakistan, ¶Genoscope — CNS, 2 rue Gaston Crémieux, BP 5706, F-91057 Evry cedex, France, **Norwegian Institute for Nature Research, NO-7485 Trondheim, Norway*

Abstract

The development of DNA barcoding (species identification using a standardized DNA sequence), and the availability of recent DNA sequencing techniques offer new possibilities in diet analysis. DNA fragments shorter than 100–150 bp remain in a much higher proportion in degraded DNA samples and can be recovered from faeces. As a consequence, by using universal primers that amplify a very short but informative DNA fragment, it is possible to reliably identify the plant taxon that has been eaten. According to our experience and using this identification system, about 50% of the taxa can be identified to species using the *trn*L approach, that is, using the P6 loop of the chloroplast *trn*L (UAA) intron. We demonstrated that this new method is fast, simple to implement, and very robust. It can be applied for diet analyses of a wide range of phytophagous species at large scales. We also demonstrated that our approach is efficient for mammals, birds, insects and molluscs. This method opens new perspectives in ecology, not only by allowing large-scale studies on diet, but also by enhancing studies on resource partitioning among competing species, and describing food webs in ecosystems.

Keywords: chloroplast DNA, diet analysis, DNA barcoding, faeces, pyrosequencing, *trn*L (UAA) intron, universal primers

Received 16 March 2008, accepted 24 March 2008



Mollusks

Deroceras reticulatum (1)
Arion rufus (1)
Helix aspera (1)


Insects

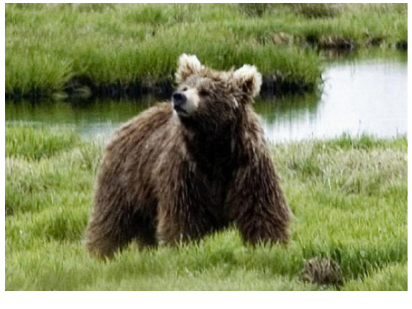
Gonfophocerippus rufus (2)

Chorthippus biguttulus (1)

Birds

Tetrao urogallus aquitanicus (4)

Tetrao urogallus major (2)




Mammals

Ursus arctos (12)

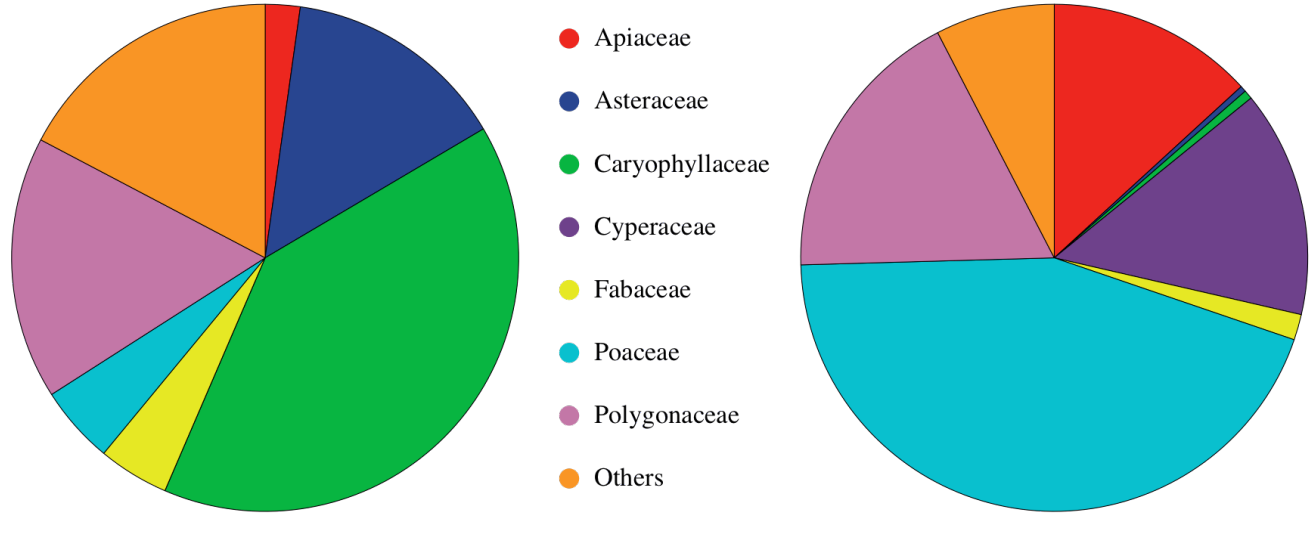
Marmota caudata (12)

Marmot and bear diets

Deosai National Park, Pakistan

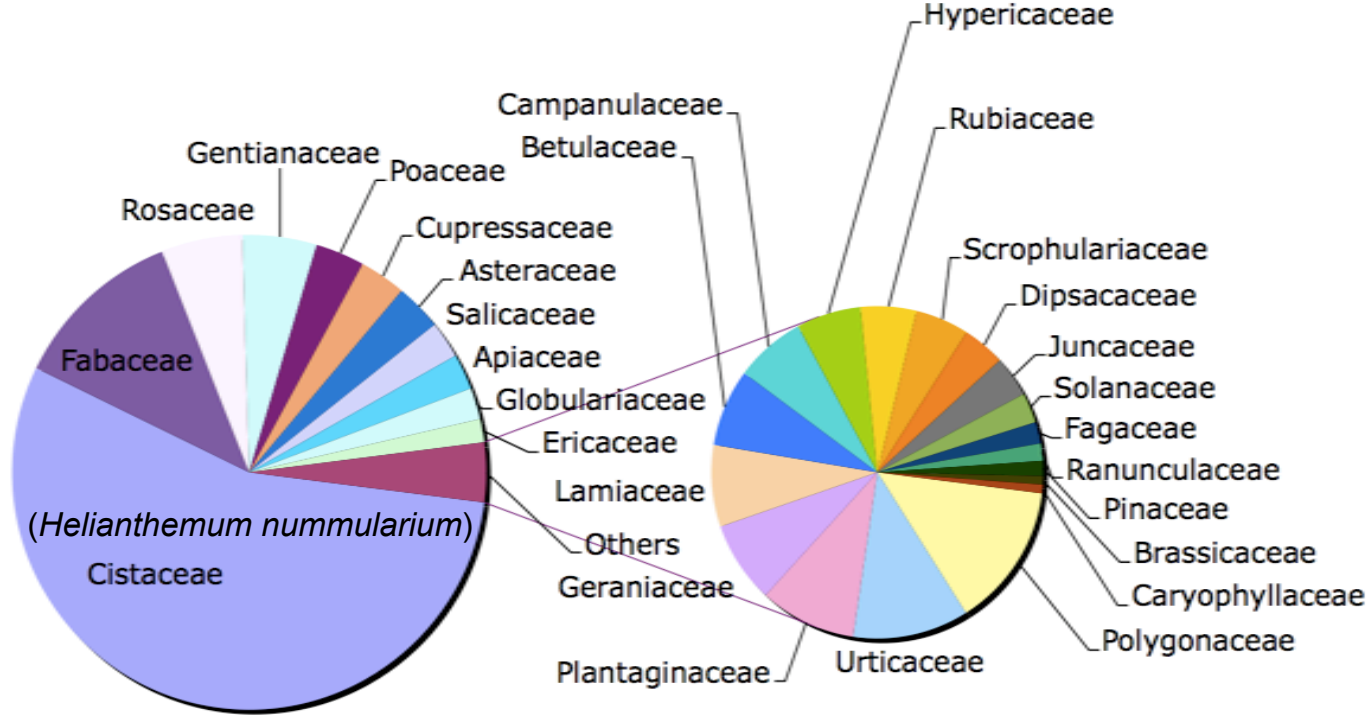
Golden marmot

Brown bear

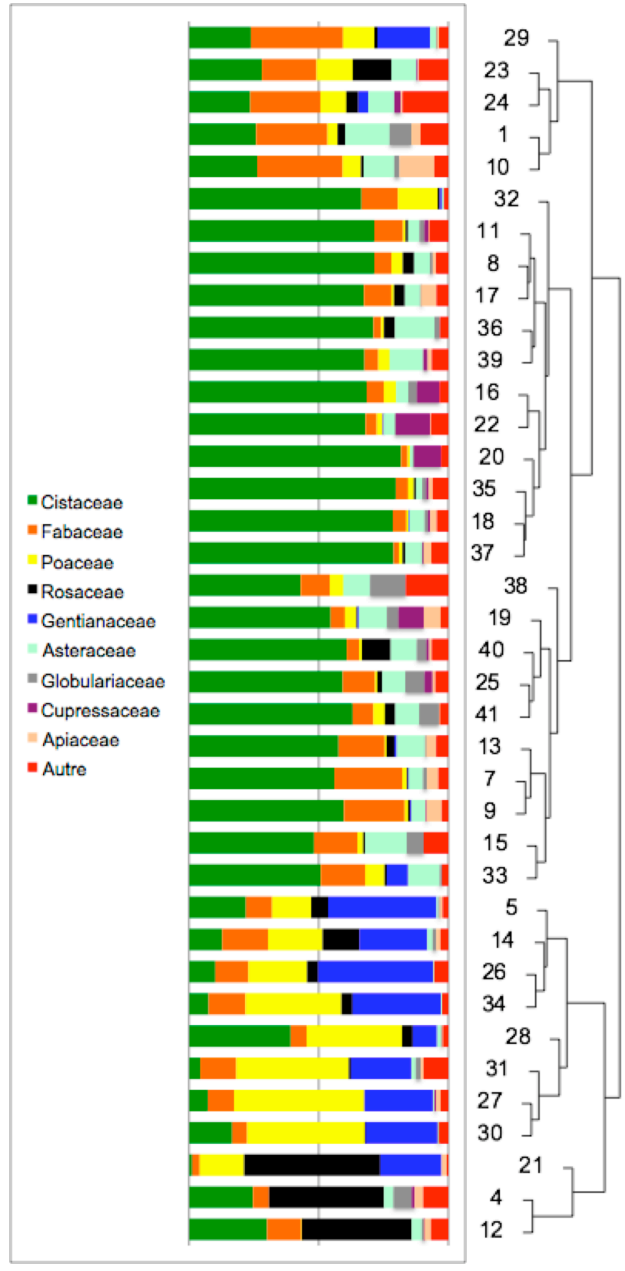


Bauges massif, France

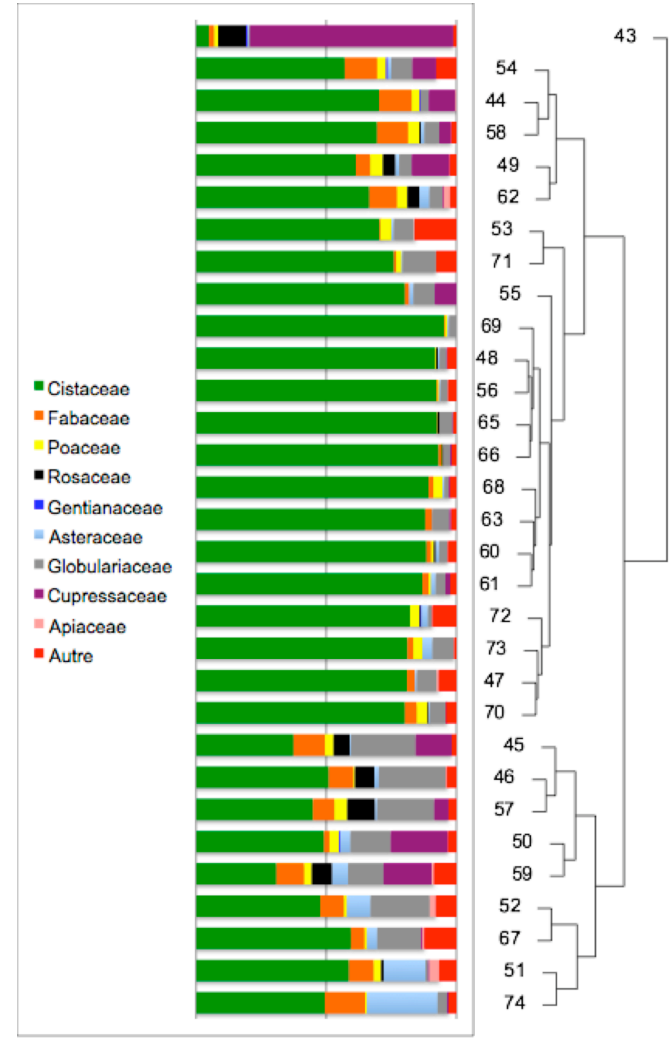
Chamois diet



Chamois diet



October



November

Outline

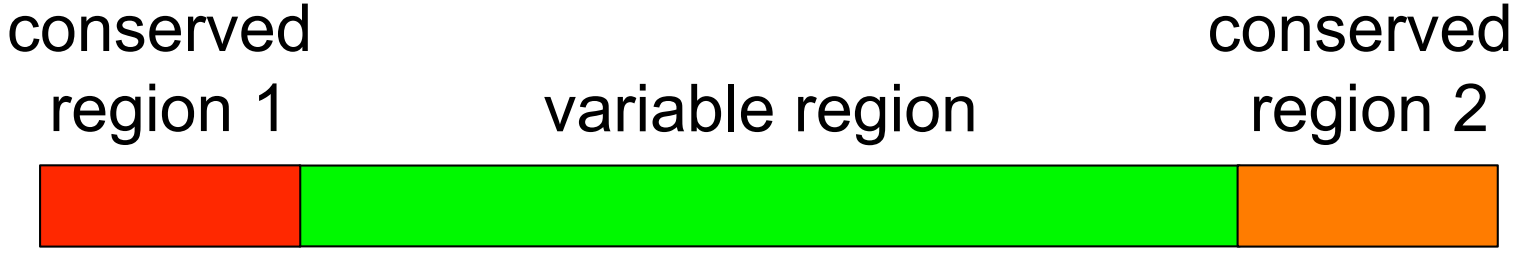
- The current revolution in DNA sequencing
- DNA barcoding: concept and definitions
- DNA barcoding in ecology
  - The *trn*L approach
  - Analysis of environmental samples
  - Diet analysis
- DNA barcoding and bioinformatics
  - How to choose the most efficient DNA barcode?
  - How to deal with sequencing errors?
  - How to analyse the output of next generation sequencers?

How to choose the most efficient DNA barcode?

- The standardized barcodes are not always the most efficient
- Bioinformatics can help to optimize the barcode according to the scientific question and to the technical constraints

Identifying new barcodes using a bioinformatic approach

- A DNA barcode is characterized by two conserved regions flanking a variable region
- The identification of conserved regions corresponds to the identification of the same DNA repeats among different genomes
- Possibility to identify barcodes that fit with one taxonomical group, and that do not amplify other groups



Number of occurrences	P6 loop (chloroplast *trn*L (UAA) intron) sequences	Identification
3103	ATCCGGTTCATGGAGAC-AATAGTTT-CTT-CTTTTATTCTCCTAAGATA-GGAAGGG	*Picea*
45	.....	*Picea* variant
42	.....	*Picea* variant
13	.....	*Picea* variant
9	.....	*Picea* variant
9	.....	*Picea* variant
6	.....	*Picea* variant
6	.....	*Picea* variant
6	.....	*Picea* variant
5	.....	*Picea* variant
5	.....	*Picea* variant
5	.....	*Picea* variant
5	.....	*Picea* variant
5	.....	*Picea* variant
5	.....	*Picea* variant
4	.....	*Picea* variant
4	.....	*Picea* variant
4	.....	*Picea* variant
4	.....	*Picea* variant
4	.....	*Picea* variant
4	.....	*Picea* variant
4	.....	*Picea* variant
4	.....	*Picea* variant
236	ATCCGGTTCATAGAGAAAAGGGTTTCTCTCCTTCTCCTAAGGAAAGG	*Abies*
4	.....	*Abies* variant

P6 loop (chloroplast *trn*L (UAA) intron) sequences obtained after high throughput pyrosequencing for a bird faeces sample (*Tetrao urogallus major*). A total of 3546 sequences were obtained with an occurrence higher than three.

Example of 454 output

```
>000032_1087_0741 length=97 uaccno=FCKM4NA01CXQBB
ACATTGGGGCAATCCTGAGCCAAATCCCGTTTATGAAACAAAGGTTTCAGAAAGCGAGAA
TAAATAAGGGATAGGTGCAGAGACTCAATGGCAATGT
>000034_1490_0543 length=98 uaccno=FCKM4NA01DW3T9
ACATTGGGGCAATCCTGAGCCAAATCCTGTTTTCAGAAAACAAAAGGGTTCAGAAAGCAA
GAATCAAAAGGATAGGTGCAGAGACTCAATGGCAATGT
>000036_1580_1030 length=59 uaccno=FCKM4NA01D40NS
ACATTGGGGCAATCCTGAGCCAAGGCAAGGCGATAGGTGCAGAGACTCAATGGCAATGT
>000048_0160_1887 length=77 uaccno=FCKM4NA01AODE5
CACGTAGGGCAATCCTGAGCCAAATCCTGCTTTCAGAAAGCAAAAAAGGGATAGGTGCAG
AGACTCAATGGTACGTG
>000052_0792_0334 length=101 uaccno=FCKM4NA01B7TNK
ACATTGGGGCAATCCTGAGCCAAATCACGTTTTCCGAAAACAAGCAAAGGTTCAGAAAGC
GAAAATCAAAAAGGATAGGTGCAGAGACTCAATGGCAATGT
>000067_1516_2260 length=51 uaccno=FCKM4NA01DZFB6
CACGTAGGGCAATCCTGAGCCAATGATAGGTGCAGAACTCAATGGTACGTG
>000075_0085_2079 length=77 uaccno=FCKM4NA01AHSI5
CACGTAGGGCAATCCTGAGCCAAATCCTGCTTTCAGAAAGCAAAAAAGGGATAGGTGCAG
AGACTCAATGGTACGTG
>000083_1104_0379 length=69 uaccno=FCKM4NA01CY7RH
CACGTAGGGCAATCCTGAGCCAAATCCTGCTTTCAGAAAGCAAAAAAGGGATAGGTGCAG
AGACTCAAT
>000088_1640_0364 length=104 uaccno=FCKM4NA01D99RY
CACGTAGGGCAATCCTGAGCCAAATCCTTGTTTTGAGAAAACAAAAGGGTTCTCGAACTA
GAATACAAAGGAAAAGGATAGGTGCAGAGACTCAATGGTACGTG
```

Example of 454 output

```
>000032_1087_0741 length=97 uaccno=FCKM4NA01CXQBB
ACATTGGGGCAATCCTGAGCCAAATCCCGTTTATGAAACAAAGGTTTCAGAAAGCGAGAA
TAAATAAGGGATAGGTGCAGAGACTCAATGGCAATGT
>000034_1490_0543 length=98 uaccno=FCKM4NA01DW3T9
ACATTGGGGCAATCCTGAGCCAAATCCTGTTTTCAGAAAACAAAAGGGTTCAGAAAGCAA
GAATCAAAAGGATAGGTGCAGAGACTCAATGGCAATGT
>000036_1580_1030 length=59 uaccno=FCKM4NA01D40NS
ACATTGGGGCAATCCTGAGCCAAGGCAAGGCGATAGGTGCAGAGACTCAATGGCAATGT
>000048_0160_1887 length=77 uaccno=FCKM4NA01AODE5
CACGTAGGGCAATCCTGAGCCAAATCCTGCTTTCAGAAAGCAAAAAAGGGATAGGTGCAG
AGACTCAATGGTACGTG
>000052_0792_0334 length=101 uaccno=FCKM4NA01B7TNK
ACATTGGGGCAATCCTGAGCCAAATCACGTTTTCCGAAAACAAGCAAAGGTTCAGAAAGC
GAAAATCAAAAAGGATAGGTGCAGAGACTCAATGGCAATGT
>000067_1516_2260 length=51 uaccno=FCKM4NA01DZFB6
CACGTAGGGCAATCCTGAGCCAATGATAGGTGCAGAACTCAATGGTACGTG
>000075_0085_2079 length=77 uaccno=FCKM4NA01AHSI5
CACGTAGGGCAATCCTGAGCCAAATCCTGCTTTCAGAAAGCAAAAAAGGGATAGGTGCAG
AGACTCAATGGTACGTG
>000083_1104_0379 length=69 uaccno=FCKM4NA01CY7RH
CACGTAGGGCAATCCTGAGCCAAATCCTGCTTTCAGAAAGCAAAAAAGGGATAGGTGCAG
AGACTCAAT
>000088_1640_0364 length=104 uaccno=FCKM4NA01D99RY
CACGTAGGGCAATCCTGAGCCAAATCCTTGTTTTGAGAAAACAAAAGGGTTCTCGAACTA
GAATACAAAGGAAAAGGATAGGTGCAGAGACTCAATGGTACGTG
```

Example of 454 output

```
>000032_1087_0741 length=97 uaccno=FCKM4NA01CXQBB
ACATTGGGGCAATCCTGAGCCAAATCCCGTTTATGAAACAAAGGTTTCAGAAAGCGAGAA
TAAATAAGGGATAGGTGCAGAGACTCAATGGCAATGT
>000034_1490_0543 length=98 uaccno=FCKM4NA01DW3T9
ACATTGGGGCAATCCTGAGCCAAATCCTGTTTTCAGAAAACAAAAGGGTTCAGAAAGCAA
GAATCAAAAGGATAGGTGCAGAGACTCAATGGCAATGT
>000036_1580_1030 length=59 uaccno=FCKM4NA01D40NS
ACATTGGGGCAATCCTGAGCCAAGGCAAGGCGATAGGTGCAGAGACTCAATGGCAATGT
>000048_0160_1887 length=77 uaccno=FCKM4NA01AODE5
CACGTAGGGCAATCCTGAGCCAAATCCTGCTTTCAGAAAGCAAAAAAGGGATAGGTGCAG
AGACTCAATGGTACGTG
>000052_0792_0334 length=101 uaccno=FCKM4NA01B7TNK
ACATTGGGGCAATCCTGAGCCAAATCACGTTTTCCGAAAACAAGCAAAGGTTCAGAAAGC
GAAAATCAAAAAGGATAGGTGCAGAGACTCAATGGCAATGT
>000067_1516_2260 length=51 uaccno=FCKM4NA01DZFB6
CACGTAGGGCAATCCTGAGCCAATGATAGGTGCAGAACTCAATGGTACGTG
>000075_0085_2079 length=77 uaccno=FCKM4NA01AHSI5
CACGTAGGGCAATCCTGAGCCAAATCCTGCTTTCAGAAAGCAAAAAAGGGATAGGTGCAG
AGACTCAATGGTACGTG
>000083_1104_0379 length=69 uaccno=FCKM4NA01CY7RH
CACGTAGGGCAATCCTGAGCCAAATCCTGCTTTCAGAAAGCAAAAAAGGGATAGGTGCAG
AGACTCAAT
>000088_1640_0364 length=104 uaccno=FCKM4NA01D99RY
CACGTAGGGCAATCCTGAGCCAAATCCTTGTTTTGAGAAAACAAAAGGGTTCTCGAACTA
GAATACAAAGGAAAAGGATAGGTGCAGAGACTCAATGGTACGTG
```

Example of 454 output

```
>000032_1087_0741 length=97 uaccno=FCKM4NA01CXQBB
ACATTGGGGCAATCCTGAGCCAAATCCCGTTTATGAAACAAAGGTTTCAGAAAGCGAGAA
TAAATAAGGGATAGGTGCAGAGACTCAATGGCAATGT
>000034_1490_0543 length=98 uaccno=FCKM4NA01DW3T9
ACATTGGGGCAATCCTGAGCCAAATCCTGTTTTCAGAAAACAAAAGGGTTCAGAAAGCAA
GAATCAAAAGGATAGGTGCAGAGACTCAATGGCAATGT
>000036_1580_1030 length=59 uaccno=FCKM4NA01D40NS
ACATTGGGGCAATCCTGAGCCAAGGCAAGGCGATAGGTGCAGAGACTCAATGGCAATGT
>000048_0160_1887 length=77 uaccno=FCKM4NA01AODE5
CACGTAGGGCAATCCTGAGCCAAATCCTGCTTTCAGAAAGCAAAAAAGGGATAGGTGCAG
AGACTCAATGGTACGTG
>000052_0792_0334 length=101 uaccno=FCKM4NA01B7TNK
ACATTGGGGCAATCCTGAGCCAAATCACGTTTTCCGAAAACAAGCAAAGGTTCAGAAAGC
GAAAATCAAAAAGGATAGGTGCAGAGACTCAATGGCAATGT
>000067_1516_2260 length=51 uaccno=FCKM4NA01DZFB6
CACGTAGGGCAATCCTGAGCCAATGATAGGTGCAGAACTCAATGGTACGTG
>000075_0085_2079 length=77 uaccno=FCKM4NA01AHSI5
CACGTAGGGCAATCCTGAGCCAAATCCTGCTTTCAGAAAGCAAAAAAGGGATAGGTGCAG
AGACTCAATGGTACGTG
>000083_1104_0379 length=69 uaccno=FCKM4NA01CY7RH
CACGTAGGGCAATCCTGAGCCAAATCCTGCTTTCAGAAAGCAAAAAAGGGATAGGTGCAG
AGACTCAAT
>000088_1640_0364 length=104 uaccno=FCKM4NA01D99RY
CACGTAGGGCAATCCTGAGCCAAATCCTTGTTTTGAGAAAACAAAAGGGTTCTCGAACTA
GAATACAAAGGAAAAGGATAGGTGCAGAGACTCAATGGTACGTG
```

Example of 454 output

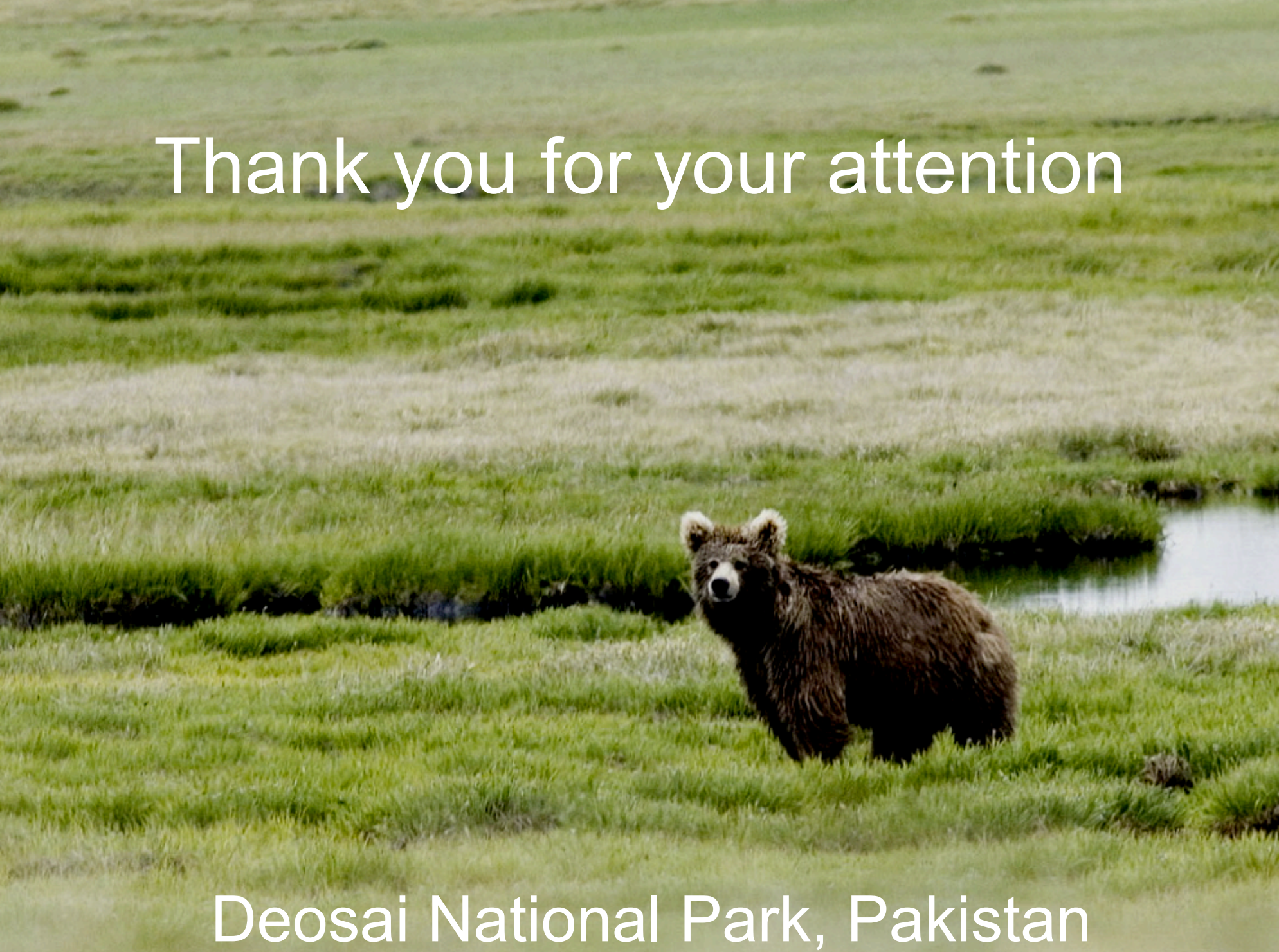
```
>000032_1087_0741 length=97 uaccno=FCKM4NA01CXQBB
ACATTGGGGCAATCCTGAGCCAAATCCCGTTTATGAAACAAAGGTTTCAGAAAGCGAGAA
TAAATAAGGGATAGGTGCAGAGACTCAATGGCAATGT
>000034_1490_0543 length=98 uaccno=FCKM4NA01DW3T9
ACATTGGGGCAATCCTGAGCCAAATCCTGTTTTCAGAAAACAAAAGGGTTCAGAAAGCAA
GAATCAAAAGGATAGGTGCAGAGACTCAATGGCAATGT
>000036_1580_1030 length=59 uaccno=FCKM4NA01D40NS
ACATTGGGGCAATCCTGAGCCAAGGCAAGGCGATAGGTGCAGAGACTCAATGGCAATGT
>000048_0160_1887 length=77 uaccno=FCKM4NA01AODE5
CACGTAGGGCAATCCTGAGCCAAATCCTGCTTTCAGAAAGCAAAAAAGGGATAGGTGCAG
AGACTCAATGGTACGTG
>000052_0792_0334 length=101 uaccno=FCKM4NA01B7TNK
ACATTGGGGCAATCCTGAGCCAAATCACGTTTTCCGAAAACAAGCAAAGGTTCAGAAAGC
GAAAATCAAAAAGGATAGGTGCAGAGACTCAATGGCAATGT
>000067_1516_2260 length=51 uaccno=FCKM4NA01DZFB6
CACGTAGGGCAATCCTGAGCCAATGATAGGTGCAGAACTCAATGGTACGTG
>000075_0085_2079 length=77 uaccno=FCKM4NA01AHSI5
CACGTAGGGCAATCCTGAGCCAAATCCTGCTTTCAGAAAGCAAAAAAGGGATAGGTGCAG
AGACTCAATGGTACGTG
>000083_1104_0379 length=69 uaccno=FCKM4NA01CY7RH
CACGTAGGGCAATCCTGAGCCAAATCCTGCTTTCAGAAAGCAAAAAAGGGATAGGTGCAG
AGACTCAAT
>000088_1640_0364 length=104 uaccno=FCKM4NA01D99RY
CACGTAGGGCAATCCTGAGCCAAATCCTTGTTTTGAGAAAACAAAAGGGTTCTCGAACTA
GAATACAAAGGAAAAGGATAGGTGCAGAGACTCAATGGTACGTG
```

Example of 454 output

>000032_1087_0741 length=97 uaccno=FCKM4NA01CXQBB
ACATTGGGGCAATCCTGAGCCAAATCCCGTTTATGAAACAAAGGTTTCAGAAAGCGAGAA
TAAATAAGGGATAGGTGCAGAGACTCAATGGCAATGT (*Alchemilla* sp.)
>000034_1490_0543 length=98 uaccno=FCKM4NA01DW3T9
ACATTGGGGCAATCCTGAGCCAAATCCTGTTTTCAGAAAACAAAAGGGTTCAGAAAGCAA
GAATCAAAAGGATAGGTGCAGAGACTCAATGGCAATGT (*Trollius* sp.)
>000036_1580_1030 length=59 uaccno=FCKM4NA01D40NS
ACATTGGGGCAATCCTGAGCCAAGGCAAGGCGATAGGTGCAGAGACTCAATGGCAATGT
>000048_0160_1887 length=77 uaccno=FCKM4NA01AODE5
CACGTAGGGCAATCCTGAGCCAAATCCTGCTTTCAGAAAGCAAAAAAGGGATAGGTGCAG
AGACTCAATGGTACGTG (*Ranunculus* sp.)
>000052_0792_0334 length=101 uaccno=FCKM4NA01B7TNK
ACATTGGGGCAATCCTGAGCCAAATCACGTTTTCCGAAAACAAGCAAAGGTTCAGAAAGC
GAAAATCAAAAAGGATAGGTGCAGAGACTCAATGGCAATGT (Asteraceae)
>000067_1516_2260 length=51 uaccno=FCKM4NA01DZFB6
CACGTAGGGCAATCCTGAGCCAATGATAGGTGCAGAACTCAATGGTACGTG
>000075_0085_2079 length=77 uaccno=FCKM4NA01AHSI5
CACGTAGGGCAATCCTGAGCCAAATCCTGCTTTCAGAAAGCAAAAAAGGGATAGGTGCAG
AGACTCAATGGTACGTG (*Ranunculus* sp.)
>000083_1104_0379 length=69 uaccno=FCKM4NA01CY7RH
CACGTAGGGCAATCCTGAGCCAAATCCTGCTTTCAGAAAGCAAAAAAGGGATAGGTGCAG
AGACTCAAT
>000088_1640_0364 length=104 uaccno=FCKM4NA01D99RY
CACGTAGGGCAATCCTGAGCCAAATCCTTGTTTTGAGAAAACAAAAGGGTTCTCGAACTA
GAATACAAAGGAAAAGGATAGGTGCAGAGACTCAATGGTACGTG (*Anthoxanthum odoratum*)

Acknowledgement

Kari Anne Bråthen
Christian Brochmann
Anne Krag Brysting
Eric Coissac
Corinne Cruaud
Mary Edwards
James Haile
Christian Miquel
François Pompanon
Delphine Rioux
Jorn Henrik Sønstebø
Alice Valentini
Eske Willerslev
Patrick Wincker
Nigel Yoccoz


Thank you for your attention
Deosai National Park, Pakistan