

# The new method of genetic barcoding



Olga Kiryanova<sup>1</sup>, Irek Gubaydullin<sup>1, 2</sup>, Alexey Chemeris<sup>3</sup>

<sup>1</sup> *Ufa State Petroleum Technological University, Ufa, Russia*

<sup>2</sup> *Institute of Petrochemistry and Catalysis, Ufa Federal Research Center, Ufa, Russia*

<sup>3</sup> *Institute of Biochemistry and Genetics, Ufa Federal Research Center, Ufa, Russia*

The study was funded by Russian Foundation for Basic Research (17-44-020120).

# Hebert DNA barcoding

Nucleotides: 660 bp

```

CTGCGGAAGGATCATTGTCGTGACCCTGAACAAAACAGACCGTGCATGCGTCATCCAATC
CGTCGGCGATGGCACCCTCCATCGCCCCGGCCAATGCCTCGACCACCTCCCCTCCTCGGAG
CGGGTGGGGGCTCGGGGTAAGAACCACGGCGTCAAGGCGTCAAGGAATACTGTGCC
TAACCCGGGGGCATGGCTAGCTTGTAGCCGTCCCTCGTGTGCAAAGCTATTTAATCCA
CACGACTCTCGCAACGGATATCTCGGCTCTCGCATCAATGAAGAACGTAGCGAAATGCG
ATACCTGGTGTGAATTGCAGAATCCACGAACAATCGAGTCTTTGAACGCAAGTTGCGCC
CGAGGCCACTCGGCCGAGGGCAGCCTGCCTGGGCGTCACGCCAAAACACGCTCCCAACC
ACCCTCAATGGGAATCGGGATGCGGCATCTGATCCCTCATCTCGCAAGGGGCGGTGGACC
GAAGATCAGGCTGCCGGCCTACTGCGCCGGACACAGCGCATGGTGGGCGTCTCGCTTTA
TCAATGCAGTGCATCCGAAGCGTAGCTGGCATTATGGCCTCTAAACGACCGAACAAACGA
AGCGCACGTTGCTTCGACC GCGACCCAGGTCAGGCGGGACTACCCGCTGAGTTTAAGCA
  
```

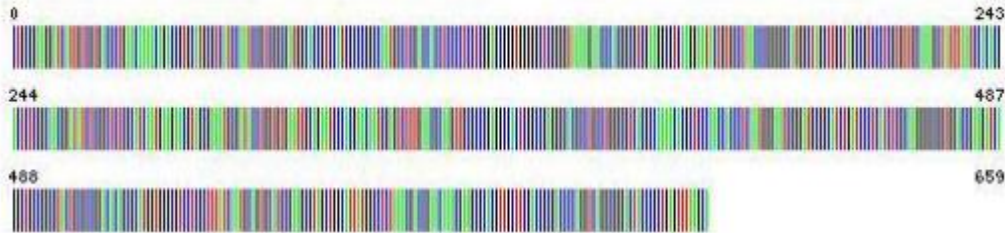
← Small reference DNA fragment ( 600-800 bp)

- not universal for all genomes

The sequence of nucleotides must be exclusive in individuals of one species (less than 1% difference) and occur in individuals of different species.

Illustrative Barcode:

A T C G



- Barcoding within individual families does not exceed 50%

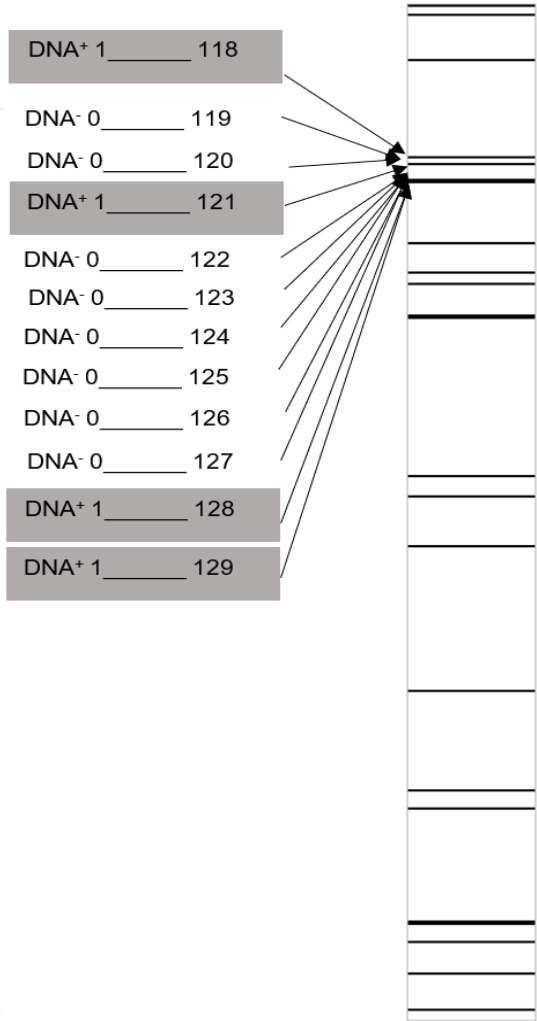
- Difficult selection of primers

## The aim

Development of the new genetic barcoding method on the base of DNA fragments of the whole genome via computer modelling of multiplex RAPD PCR for any organism classification



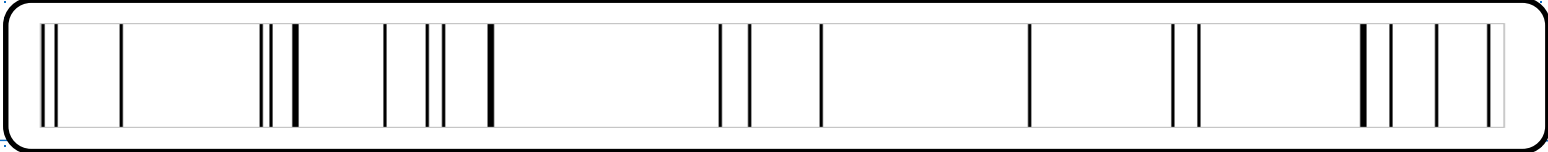
# The principle of constructing a genetic barcode



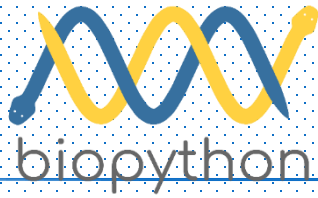
51, 55, 75, 118, 121, 128, 129, 156, 169, 174, 188, 189, **259, 259, 259**, 268, 290, 354, 398, 406, 456, 457, 465, **479, 479**, 495 – amplicons sizes



```
1000100000000000000000001000000000000000000000000000000000000000000000000011  
0000000000000000000000001000000000001000010000000000001100000000000000000  
000000000000000000000000000000000000000000000000000000010000000010000000000000  
000100000000000000000000000000000000000000000000000000000000000000000000001000000000  
0000000000000000000000000000000000001000000100000000000000000000000000000000  
0000000000000110000000100000000000010000000000000100000
```



**Important!** Uniquely identifies the genome of the data set: primers + barcode.  
Different primers correspond to different barcodes.



# Computer analysis

nucleotide sequence of the genome

Primers set

1. Search for primer annealing sites, determination of amplicon sizes

Amplicon sizes

2. Converting amplicon sizes to barcode format

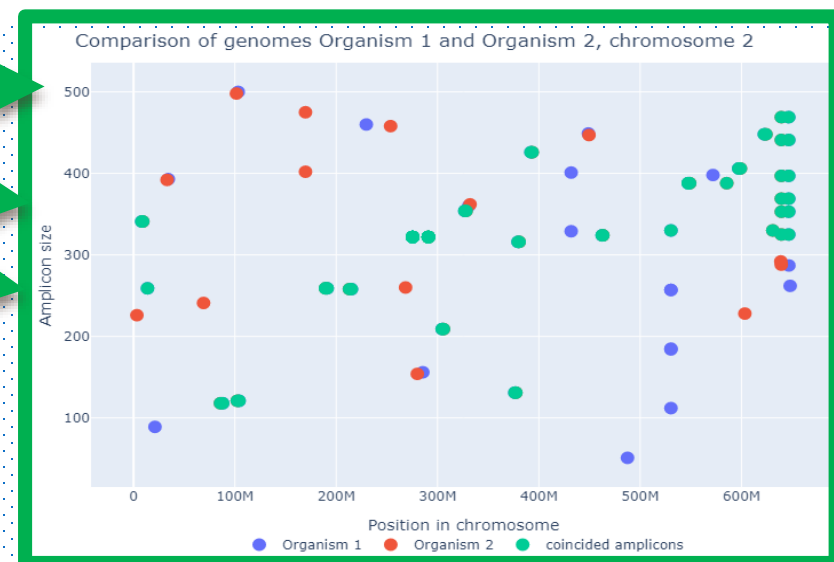
Binary sequence

Comparative analysis of data from item 1.

Primer annealing positions

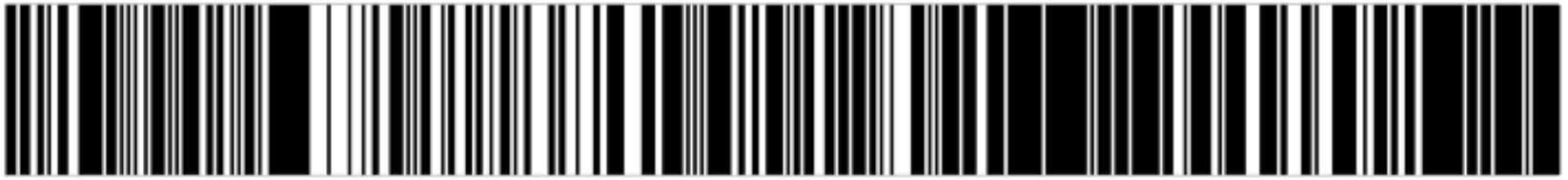

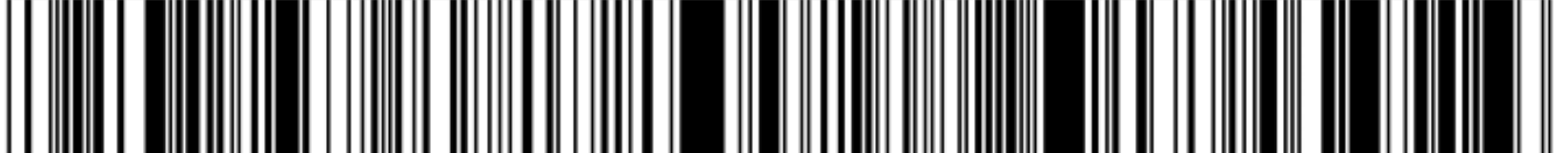


Genome similarity/difference data

filename/chromosome	direct prime	positio	inverse prin	positio	Amplicon leng	contig num
fasta/GCA_021605145.1_bicornis	AGCCTCACTC	693743886	GTGGTGAGCA	693743953	67	2
fasta/GCA_021605145.1_bicornis	AAAACGCCAA	777767307	TTTTGGCGTT	777767379	72	2
fasta/GCA_021605145.1_bicornis	AAGGGACAAA	803170261	GTGGAGAGCT	803170371	110	2
fasta/GCA_021605145.1_bicornis	AGCTCTCCAC	515364239	TTTGTCCTT	515364349	110	2
fasta/GCA_021605145.1_bicornis	TGCTCACCAC	216101900	GAGTGAGGCT	216102038	138	2
fasta/GCA_021605145.1_bicornis	TGCTCACCAC	832607397	GAGTGAGGCT	832607536	139	2
fasta/GCA_021605145.1_bicornis	AGCCTCACTC	693743886	GTGGTGAGCA	693744028	142	2
fasta/GCA_021605145.1_bicornis	AGCCTCACTC	654393975	GAGTGAGGCT	654394120	145	2
fasta/GCA_021605145.1_bicornis	AGCTCCACTC	590788854	GTGTGTGGCA	590788999	145	2
fasta/GCA_021605145.1_bicornis	AGCTCTCCAC	340762878	GTGGTGAGCA	340763058	180	2
fasta/GCA_021605145.1_bicornis	TGCTCACCAC	832607326	GAGTGAGGCT	832607536	210	2
fasta/GCA_021605145.1_bicornis	AGCCTCACTC	693743886	GTGGTGAGCA	693744099	213	2
fasta/GCA_021605145.1_bicornis	AACCGAACAA	457974279	GAGTGGAGCT	457974501	222	2
fasta/GCA_021605145.1_bicornis	AGCCTCACTC	722915989	GTGGAGAGCT	722916213	224	2
fasta/GCA_021605145.1_bicornis	AACCGAACAA	754571262	TTTTGGCGTT	754571516	254	2
fasta/GCA_021605145.1_bicornis	TGCTCACCAC	832607255	GAGTGAGGCT	832607536	281	2
fasta/GCA_021605145.1_bicornis	AGCCTCACTC	693743886	GTGGTGAGCA	693744170	284	2
fasta/GCA_021605145.1_bicornis	TGCTCACCAC	614791519	TTGTCGGTT	614791811	292	2
fasta/GCA_021605145.1_bicornis	AACGCCAAAA	357683896	GAGTGAGGCT	357684204	308	2
fasta/GCA_021605145.1_bicornis	AGCTCCACTC	590788854	GAGTGAGGCT	590789188	334	2
fasta/GCA_021605145.1_bicornis	TGCCACACAC	6536239	GTGTGTGGCA	6536585	346	2



# Results. Examples of barcodes

Primer set: TGCCACACAC, AGCCTCACTC, TGCTCACCAC, C GACTCTCAC, AGCTCTCCAC, AGCTCCACTC  
AACCAGACAA, AAGGGACAAA, AACCGAACAA, AACGCACAAA, AAAACGCCAA, AACGCCAAAA

Species / genome	Genetic Barcodes
<i>Triticum aestivum</i> / <b>BAD</b> Genome size approx. 15Gb	
<i>Triticum turgidum</i> / <b>BA</b> Genome size approx. 10 Gb	
<i>Triticum dicoccoides</i> / <b>BA</b> Genome size approx. 10 Gb	
<i>Aegilops tauschii</i> / <b>D</b> Genome size approx. 4.2 Gb	
<i>Triticum urartu</i> / <b>A</b> Genome size approx. 5 Gb	

- 
- ❖ A new approach to the certification of various groups of plants and other organisms based on the fragments (PCR amplicons) of the whole genome is proposed.
  - ❖ Unique genetic barcodes have been proposed for plant varieties based on the detected DNA polymorphism.
  - ❖ New software allows computer simulation of multiplex RAPD PCR, to convert PCR results into binary format and genetic barcode
  - ❖ New software can be used as tool for planning experimental PCR.
-