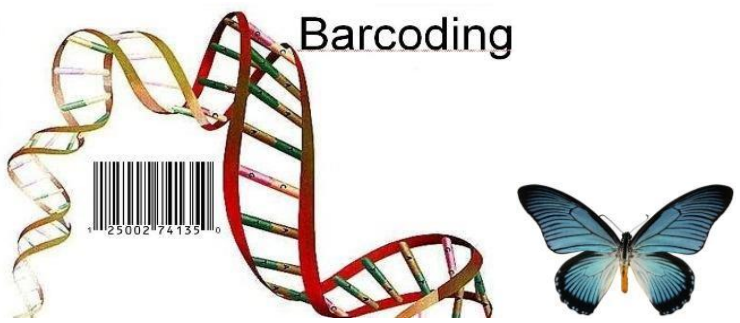
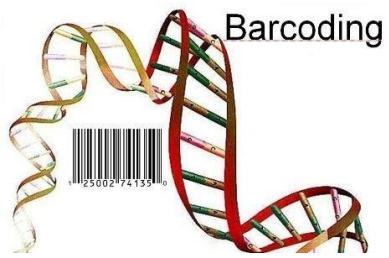


Работа с нуклеотидными последовательностями (*теоретически-практическая часть*)

Нина Владимировна Воронова
кандидат биологических наук, доцент
Белорусский государственный университет

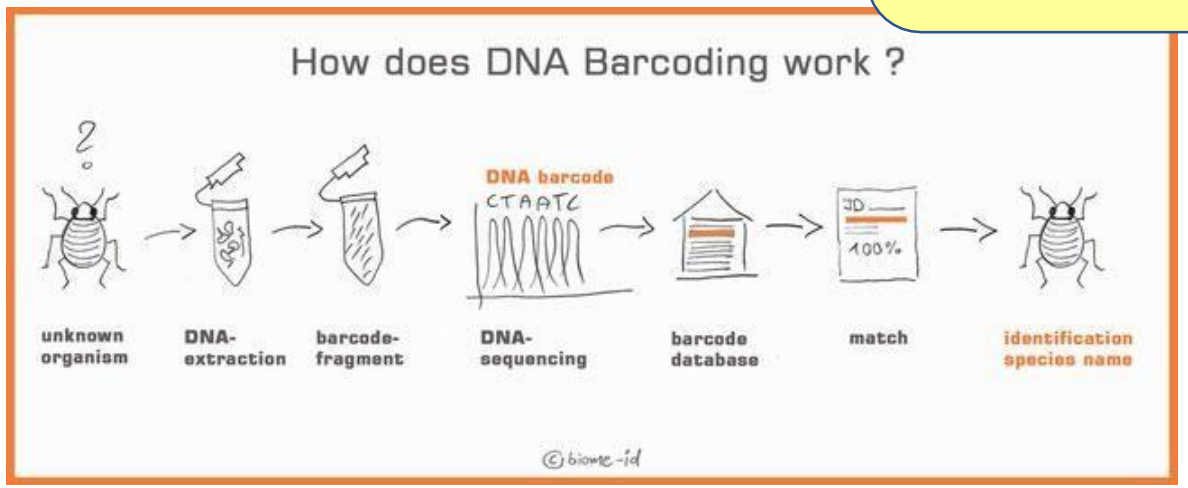


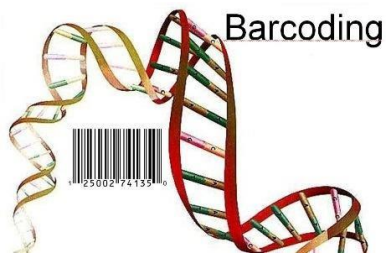
ДНК-штрихкодирование. Суть метода

Выделение ДНК

Секвенирование нужного участка ДНК

Определение с использованием базы данных





Стандарты системы BOLD

Идентификация животных – COI

Идентификация растений - *matK* и *rbcL*

Идентификация грибов – ITS-регион

Идентификация протистов - ?

Идентификация бактерий - ? (16SrRNA)

ДНК-штрихкодирование. Выбор маркера

Molecular Marine Biology and Biotechnology
(1994) 3(5), 294-299

Molecular Marine Biology and Biotechnology (1994) 3(5), 294-299

DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates

O. Folmer, M. Black, W. Hoeh,* R. Lutz, and R. Vrijenhoek+
Center for Theoretical and Applied Genetics, and Institute of Marine and Coastal Science, Rutgers University, New Brunswick, New Jersey 08903-231

Abstract

We describe "universal" DNA primers for polymerase chain reaction (PCR) amplification of a 710-bp fragment of the mitochondrial cytochrome *c* oxidase subunit I gene (*COI*) from 11 invertebrate phyla: Echinodermata, Mollusca, Annelida, Pogonophora, Arthropoda, Nemertinea, Echiura, Sipuncula, Platyhelminthes, Tardigrada, and Coelenterata, as well as the putative phylum Vestimentifera. Preliminary comparisons revealed that these *COI* primers generate informative sequences for phylogenetic analyses at the species and higher taxonomic levels.

Introduction

The purpose of this short communication is to describe "universal" DNA primers for the polymerase chain reaction (PCR) amplification of a 710-bp fragment of the mitochondrial cytochrome *c* oxidase subunit I gene (*COI*). This study was motivated by the recent discoveries of more than 230 new invertebrate species, comprising new genera, families, classes, orders, and potentially a new phylum, from deep-sea hydrothermal vent and cold-water sulfide or methane seep communities (Tunnicliffe, 1991). Our goal was to develop molecular techniques for phylogenetic studies of these diverse organisms. We focused on the mitochondrial cytochrome *c* oxidase subunit I (*COI*) gene because it appears to be among the most conservative protein-coding genes in the mitochondrial genome of animals (Brown, 1985), which was preferable for the evolutionary

*Present address: Department of Biology, Dalhousie University, Halifax, Nova Scotia, Canada.
+Correspondence should be sent to this author.

294

time depths likely to be found in our studies. We quickly became aware of the broad utility of these *COI* primers for broader systematic studies of metazoan invertebrates, including acoelomates, pseudocoelomates, and coelomate protostomes and deuterostomes.

Results

To design candidate primers, we compared published DNA sequences from the following species: blue mussel, *Mytilus edulis*; fruitfly, *Drosophila grimshawi*; honeybee, *Apis mellifera*; mosquito, *Anopheles gambiae*; brine shrimp, *Artemia franciscana*; nematodes, *Ascaris suum* and *Caenorhabditis elegans*; sea urchin, *Strongylocentrotus purpuratus*; carp, *Cyprinus carpio*; frog, *Xenopus laevis*; chicken, *Gallus gallus*; mouse, *Mus musculus*; cow, *Bos taurus*; fin whale, *Balaenoptera physalus*; and human, *Homo sapiens* (Figure 1). Several highly conserved regions of these *COI* genes were used as the targets for primer designs.

Altogether, three coding-strand and six anti-coding-strand primers were tested (Table 1) for amplification efficiency. The following primer pair consistently amplified a 710-bp fragment of *COI* across the broadest array of invertebrates:

LCO1490 5'-ggtcaacaaatcataaagatattgg-3'

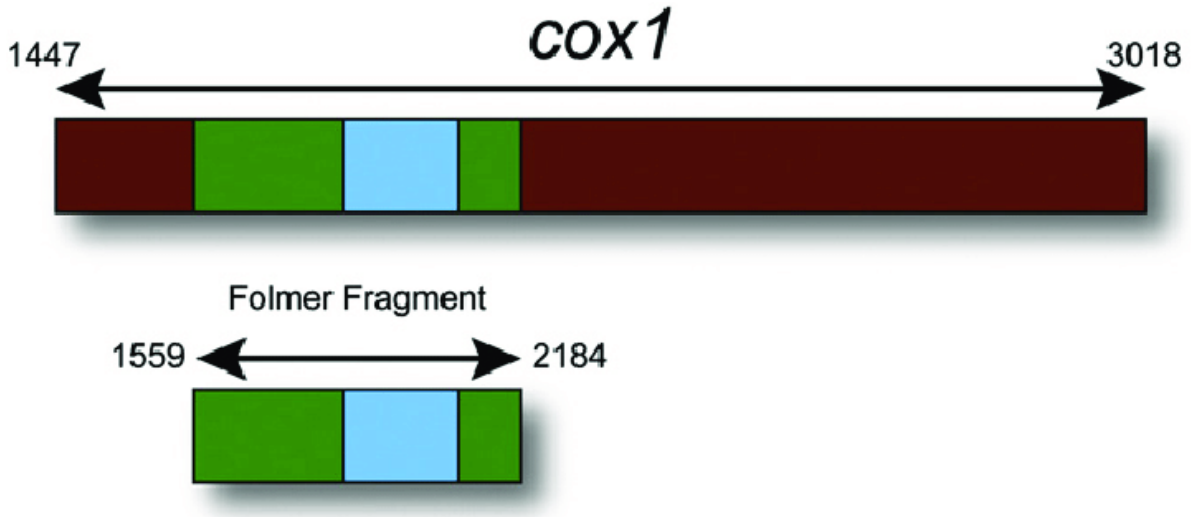
HC02198: 5'-taaacttcagggtgaccaaaaaatca-3'

In the code names above, L and H refer to light and heavy DNA strands, CO refers to cytochrome oxidase, and the numbers (1490 and 2198) refer to the position of the *D. yakuba* 5' nucleotide.

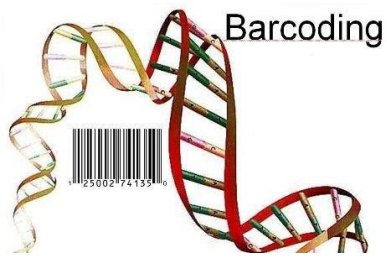
We also present the primers as coding-strand sequences, along with their inferred amino acids (Figure 1). The usefulness of these primers results from the high degree of sequence conservation in their respective 3' ends across the 15 taxa. The 3' end of each primer is on a second-position nucleotide. All other pairwise primer combinations amplified fewer taxa or gave additional nonspecific products under less stringent amplification conditions.

The LCO1490 and HC02198 amplified DNA from more than 80 invertebrate species from 11

Фрагмент Фолмера
658 п.н. в 5´-области гена COI



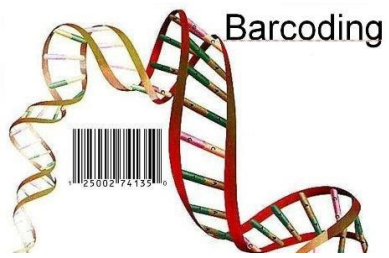
Ратгерский университет
Нью-Джерси, США



Особенности маркеров, используемых для ДНК-штрихкодирования

Растения	Животные	Грибы
matK + *rbcL*	COI	ITS1 + ITS2
Хлоропластные	Митохондриальные	Ядерные
Белок-кодирующие	Белок-кодирующие	Некодирующие
Гаплоидные	Гаплоидные	Гапло/диплоидные

Размер фрагмента ПЦР, используемого для ДНК-штрихкодирования, не должен превышать 800 bp



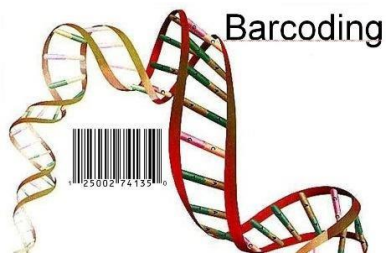
ДНК-штрихкодирование. Анализ данных

1. Получение данных с секвенатора

2. Подготовка данных для анализа

3. Автоматическая идентификация вида

4. Интерпретация результатов



1. Получение данных с секвенатора

«Классическое» ДНК-штрихкодирование

Выделение ДНК

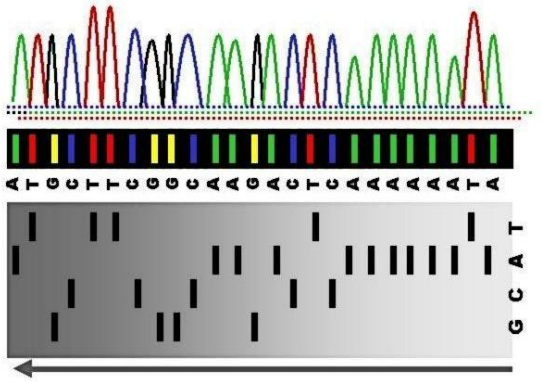
↓

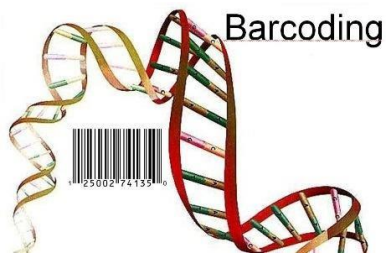
ПЦР с универсальными праймерами

↓

Секвенирование по Сэнгеру

Анализ данных

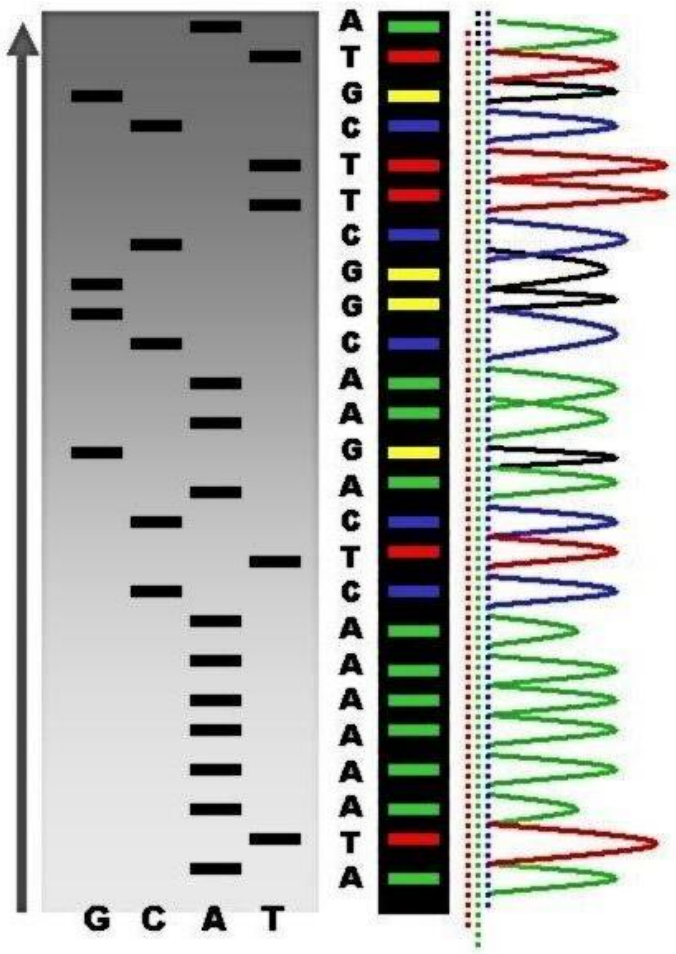


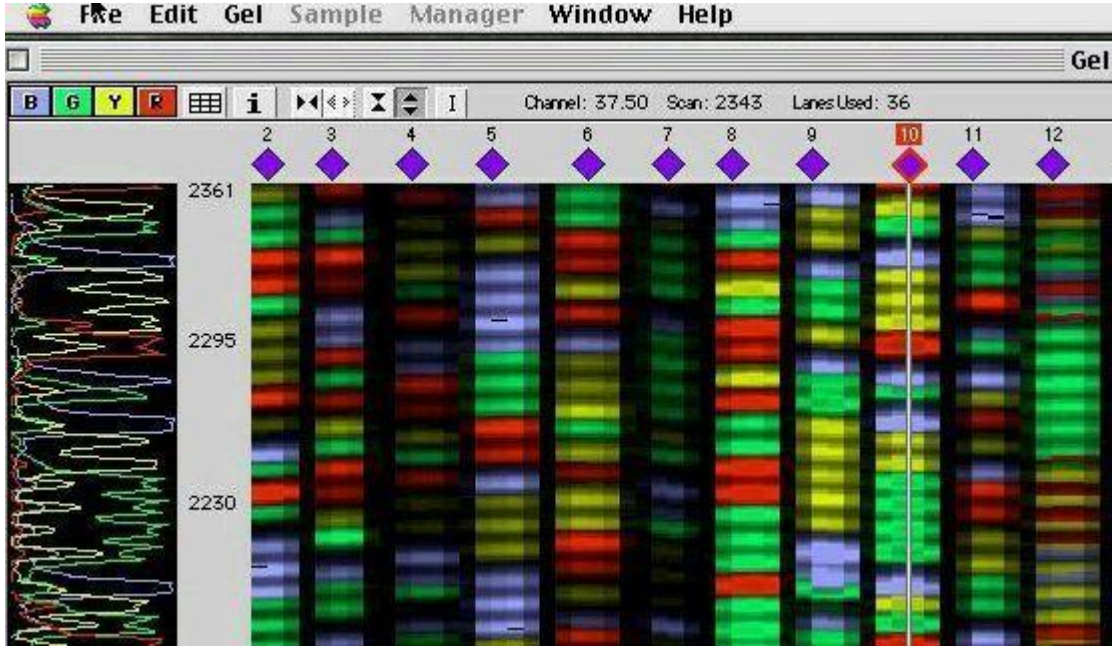


1. Получение данных с секвенатора

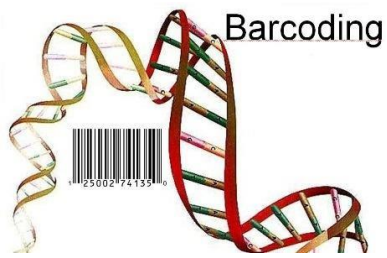
Секвенирование по Сэнгеру

Анализ данных





Секвенирование может быть проведено в одном (1 реакция, 1 праймер) или двух (2 реакции, 1 праймер в каждой реакции) направлениях



1. Получение данных с секвенатора

Environmental DNA barcoding и метабаркодинг

Секвенирование по Сэнгеру

Секвенирование нового поколения (NGS)

Забор «сложных» проб, концентрация проб, выделение общей ДНК

ПЦР с видоспецифичными праймерами

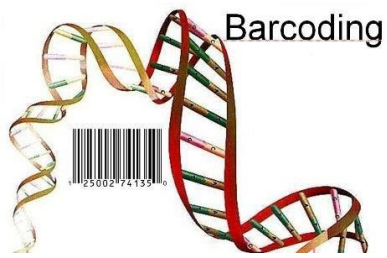
Секвенирование

Приготовление ПЦР ДНК-библиотеки

NGS

Сборка ДНК-штрихкодов

Анализ данных



1. Получение данных с секвенатора

NGS (Next Generation Sequencing) при метабаркодинге

Выделение ДНК → ПЦР с универсальными праймерами

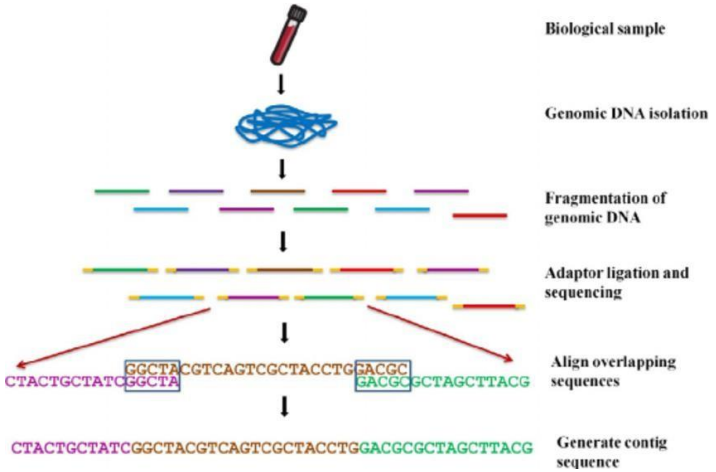
↓

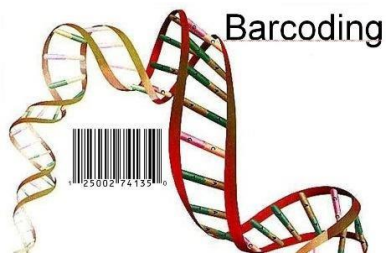
Приготовление библиотеки и «пришивание» адаптеров → NGS

↓

Сборка индивидуальных ДНК-баркодов

Анализ данных





2. Подготовка данных для анализа

Алгоритм действия

Данные с секвенатора

Оценка качества секвенирования

Редактирование сиквенса

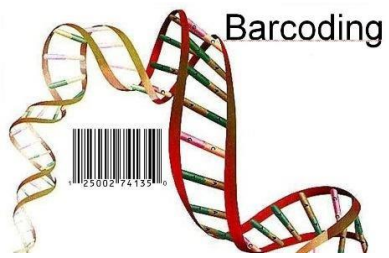
Получение консенсусной последовательности

При необходимости

Получение последовательности в текстовом формате

ДНК-штрихкод

Анализ данных



2. Подготовка данных для анализа

Данные, приходящие с секвенатора

Анализ данных

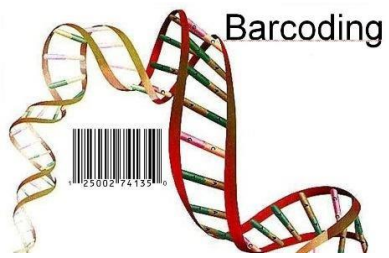
Секвенирование по Сэнгеру

Одна буквенная последовательность длиной 650-700 bp (при использовании прямого праймера)

Две последовательности длиной 650-700 bp : прямая и обратная (при использовании двух праймеров)

NGS

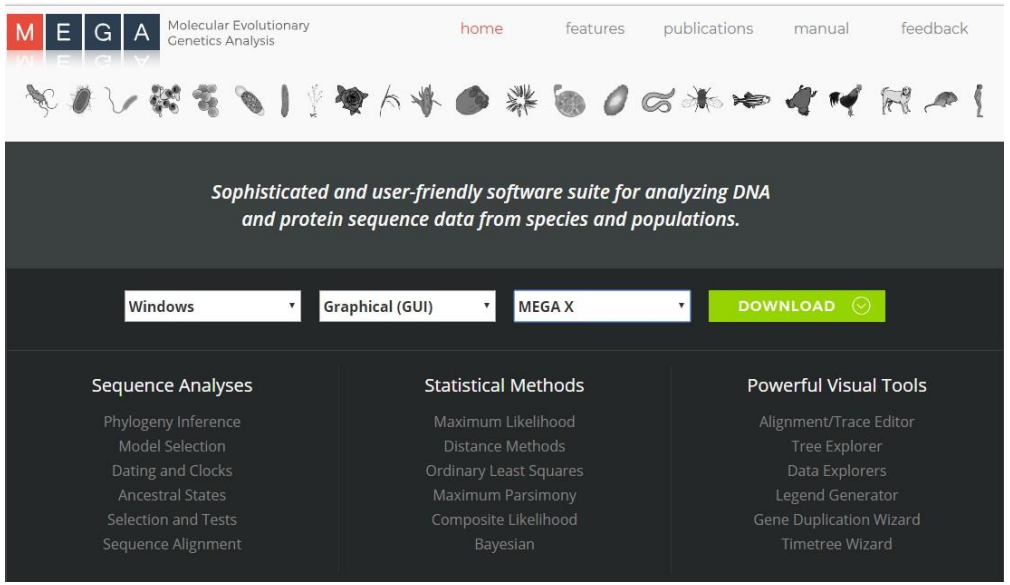
Миллионы буквенных последовательностей длиной 60-120 bp, представляющих собой неупорядоченные фрагменты ДНК-штрихкодов



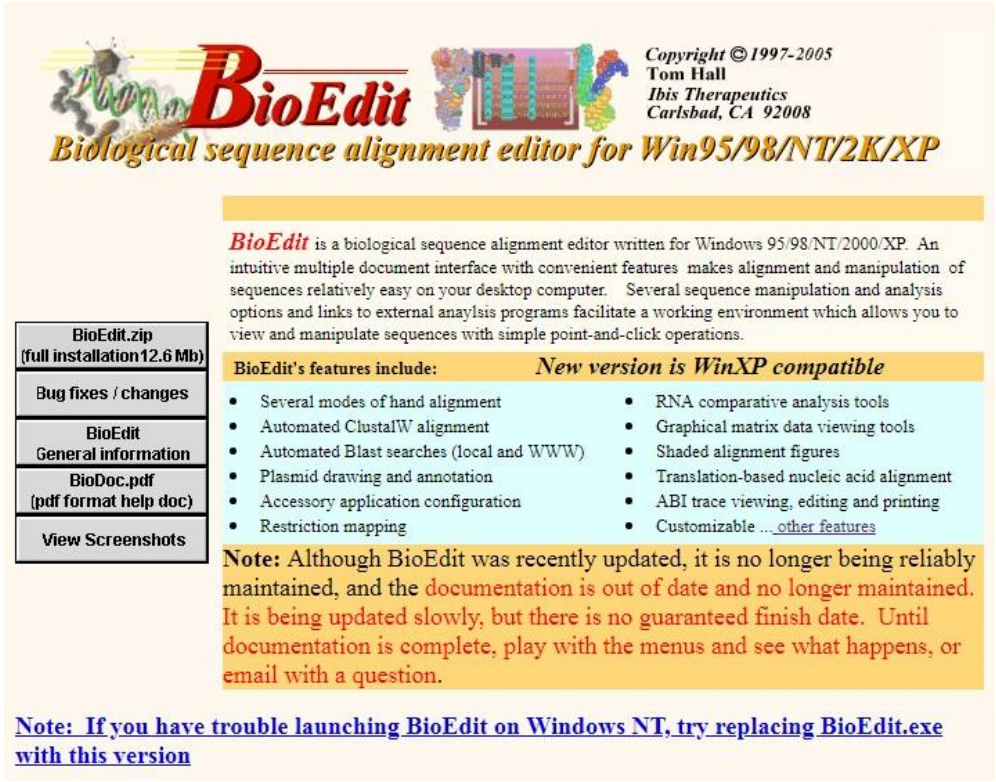
2. Подготовка данных для анализа

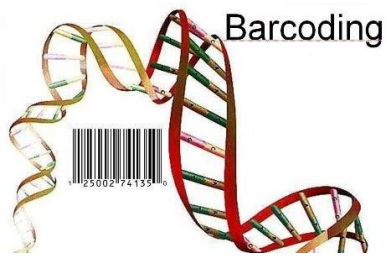
Наиболее часто используемые программы, работающие с данными секвенирования

Анализ данных



https://www.megasoftware.net/





2. Подготовка данных для анализа

Ошибки секвенирования. Причины:

- Случайная вставка ошибочного нуклеотида
- Присутствие неспецифичного продукта в ПЦП-образце
- Присутствие посторонних образцов ДНК в первоначальной ДНК-пробе

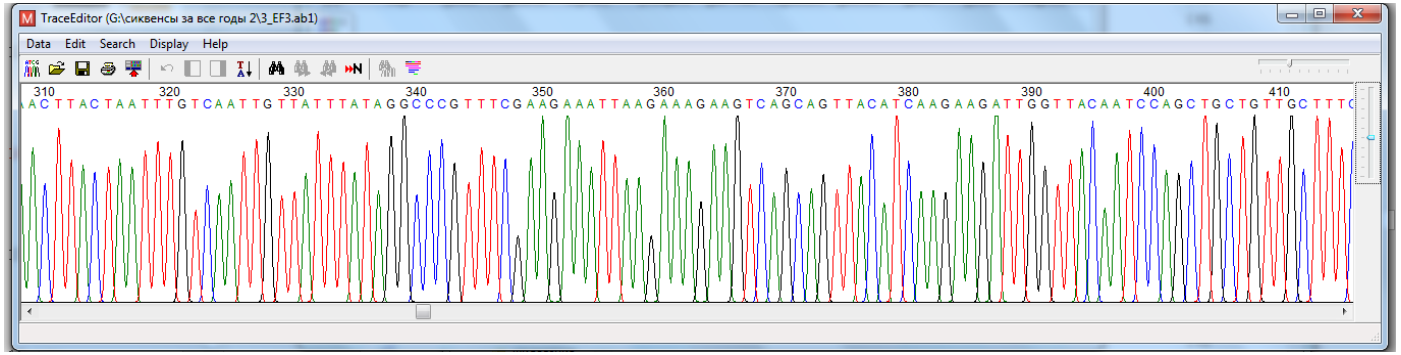
- Внутриорганизменный полиморфизм
- NUMTs (Nuclear Mitochondrial DNA segments)
- Псевдогены

Анализ данных

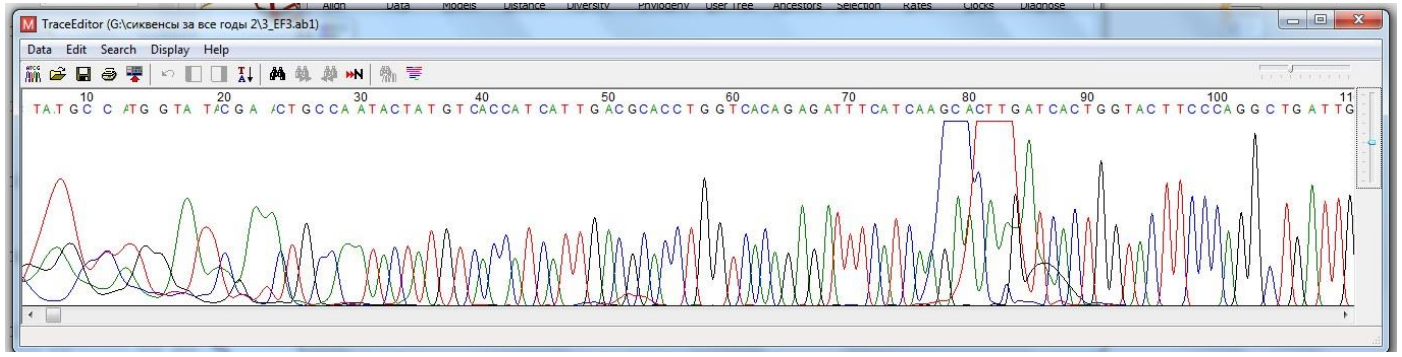
2. Подготовка данных для анализа

Анализ качества сиквенса

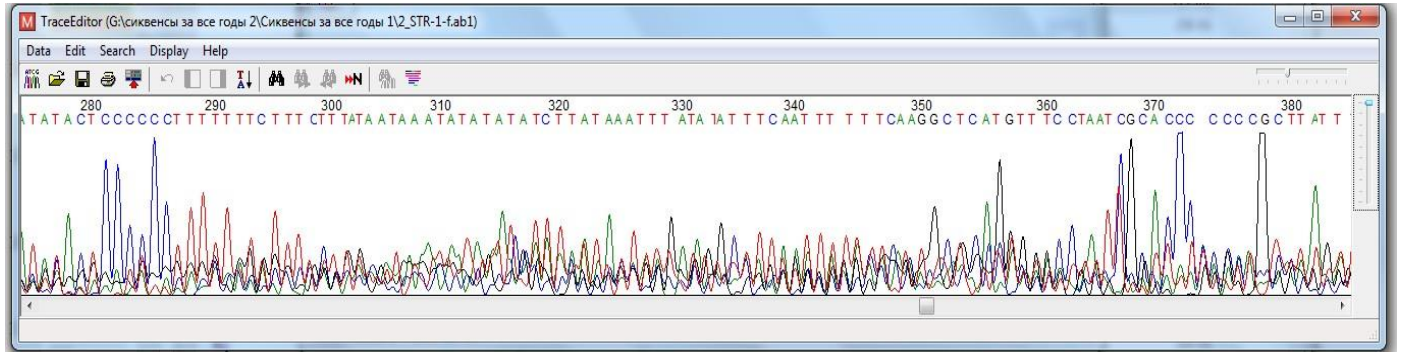
Анализ данных



Высокое качество



Качество, достаточное для редактирования



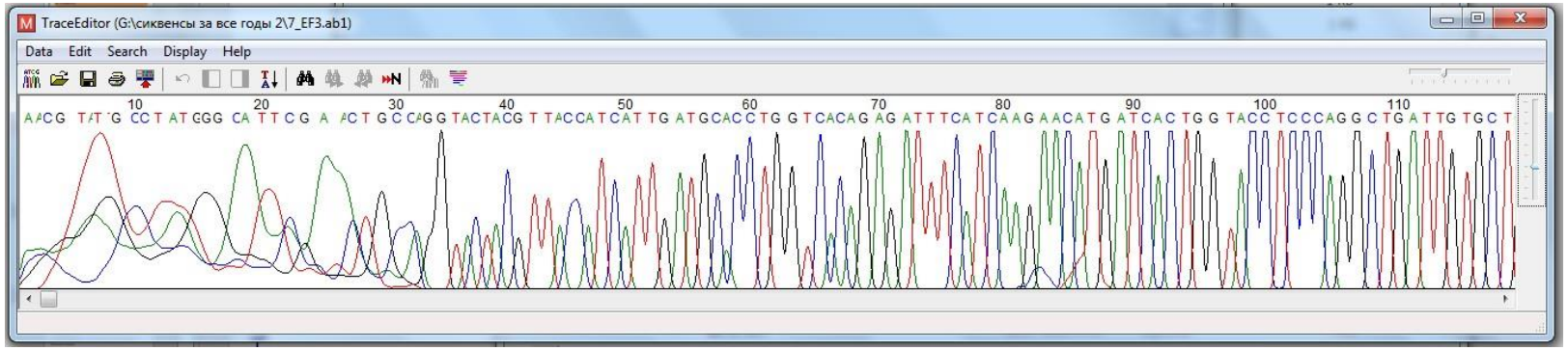
Низкое качество

Barcoding

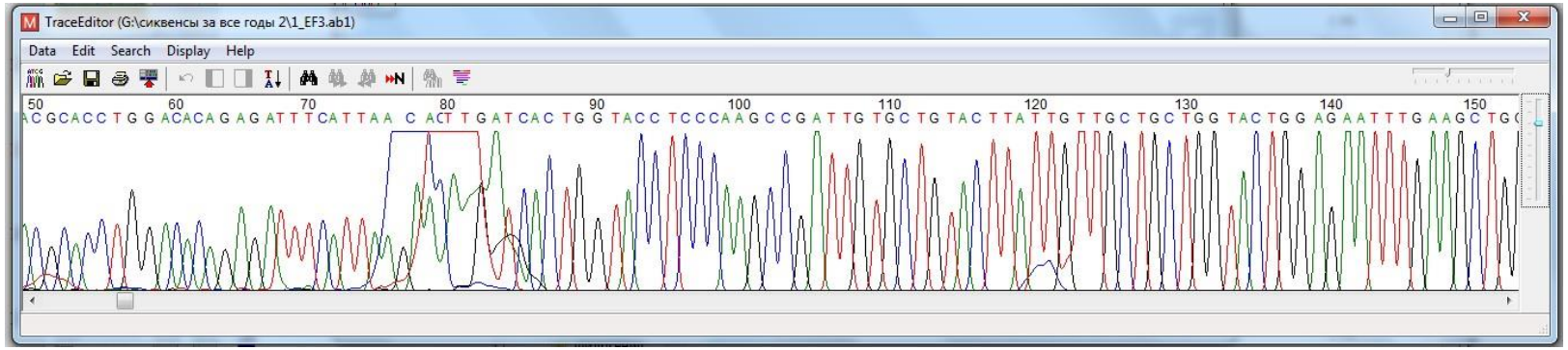
2. Подготовка данных для анализа

Редактирование сиквенсов

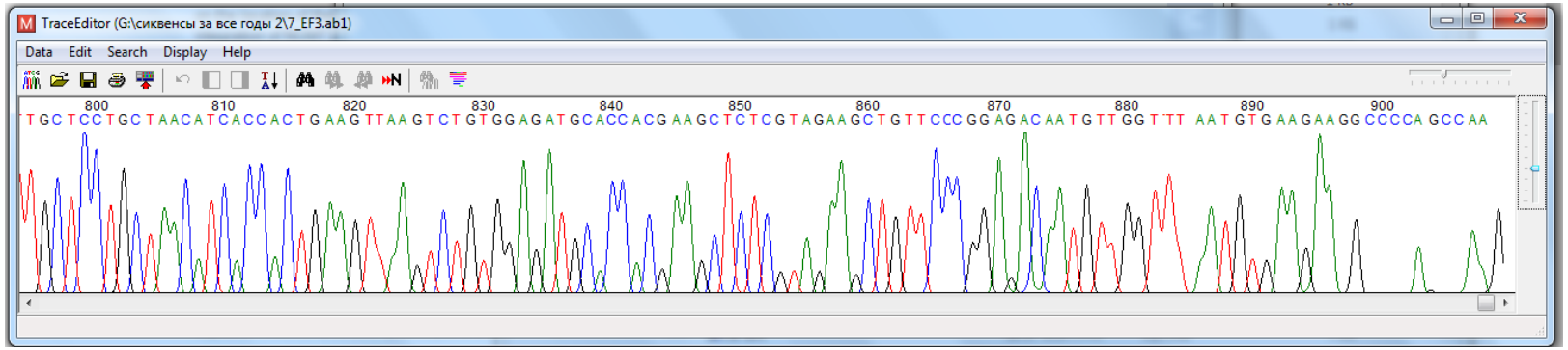
Анализ данных



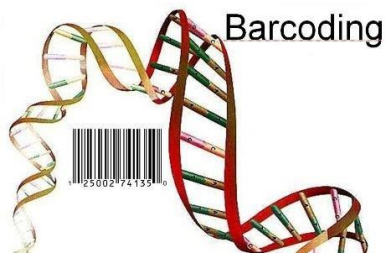
Удаление некачественного участка в начале



Редактирование коротких некачественных участков



Удаление некачественного участка в конце



2. Подготовка данных для анализа

Формат файла, приходящего с секвенатора → Формат файла для дальнейшего анализа

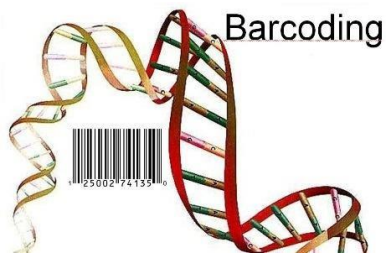
.ab1

Хроматограмма и буквенная последовательность

.fas

Только буквенная последовательность

Анализ данных



2. Подготовка данных для анализа

Получение консенсусной последовательности (при необходимости)

3´-attccgaatcATGCCTAGCTAGGCT-5´ Прямая последовательность

3´-cctaggcattAGCCTAGCTAGGCAT-5´ Обратная последовательность

3´-attccgaatcATGCCTAGCTAGGCTaatgcctagg-5´ Консенсусная последовательность

Анализ данных

Получение обратной комплементарной последовательности → Выравнивание последовательностей → Получение консенсусной последовательности

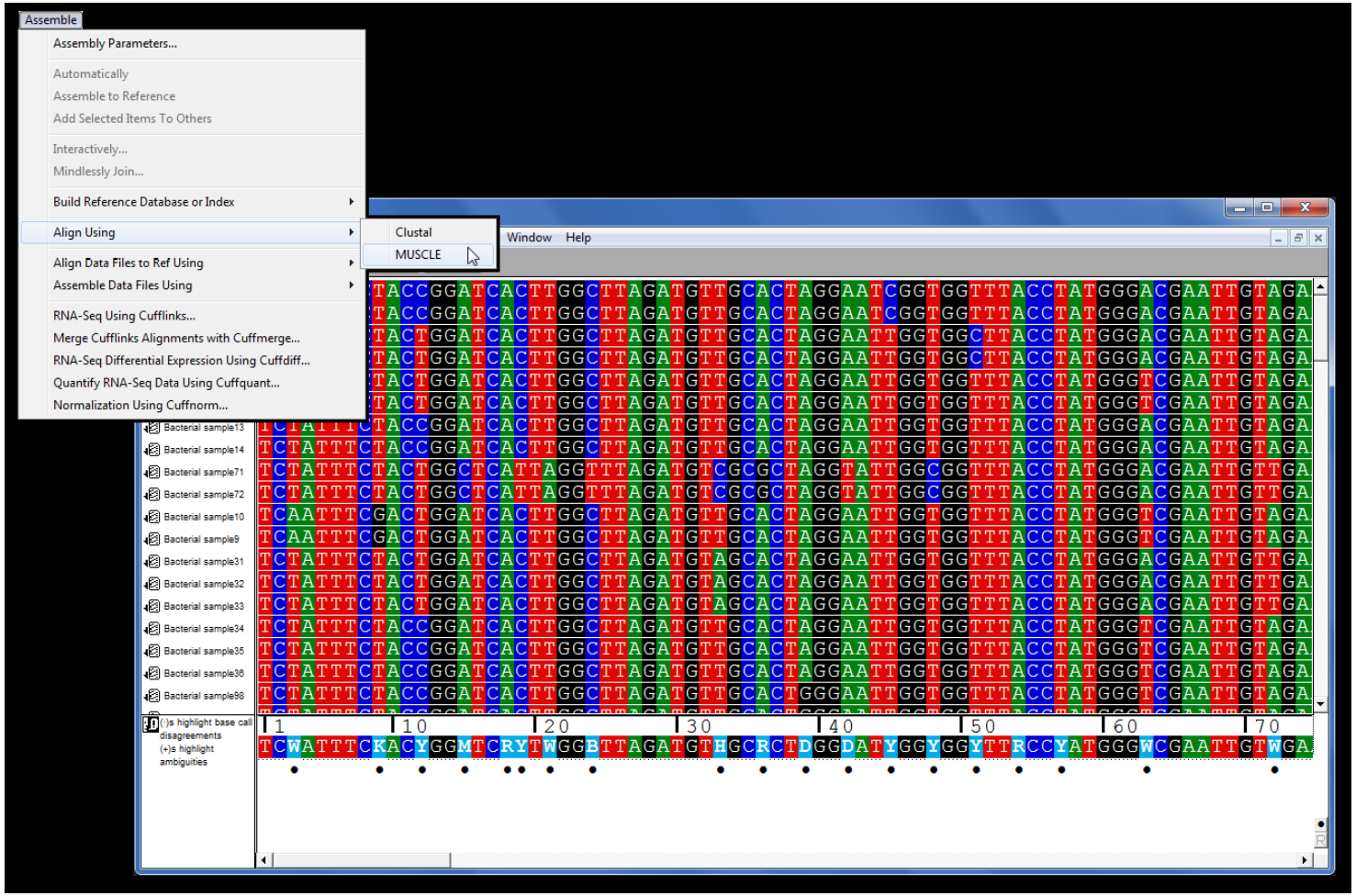
Для последовательности, полученной с обратным праймером

Консенсусная последовательность – «обобщающая» последовательность, полученная из нескольких последовательностей

2. Подготовка данных для анализа

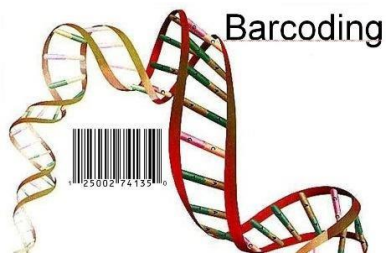
Получение консенсусной последовательности (при необходимости)

Анализ данных



Выравнивание последовательностей

Получение консенсусной последовательности



2. Идентификация вида

Алгоритм действия

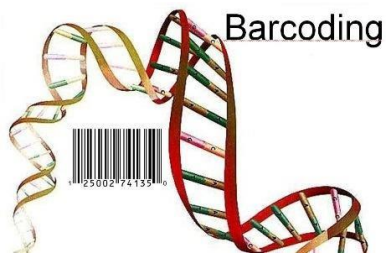
ДНК-штрихкод

Выравнивание последовательности в системах BOLD или BLAST

Оценка качества выравнивания

Вид

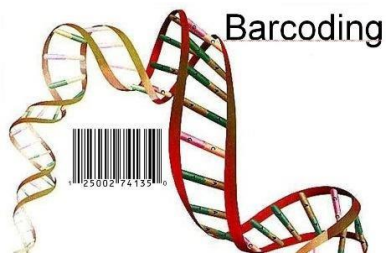
Анализ данных



3. Идентификация вида

Выравнивание последовательностей

Анализ данных

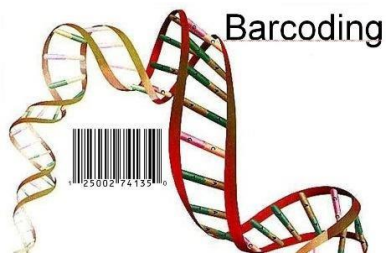


3. Идентификация вида

Основные алгоритмы выравнивания последовательностей

MUSCLE	CLUSTAL
Допускает множественное выравнивание	Допускает множественное выравнивание
Быстрый	Медленный
Расслабленный	Строгий

Анализ данных



3. Идентификация вида

Оценка результатов выравнивания

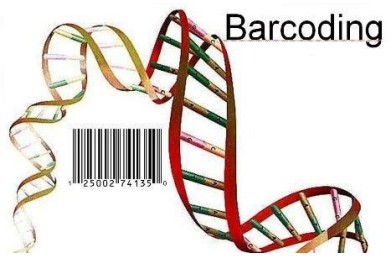
Визуальная оценка

Анализ данных

Nucleotide identity - процент совпадающих нуклеотидов

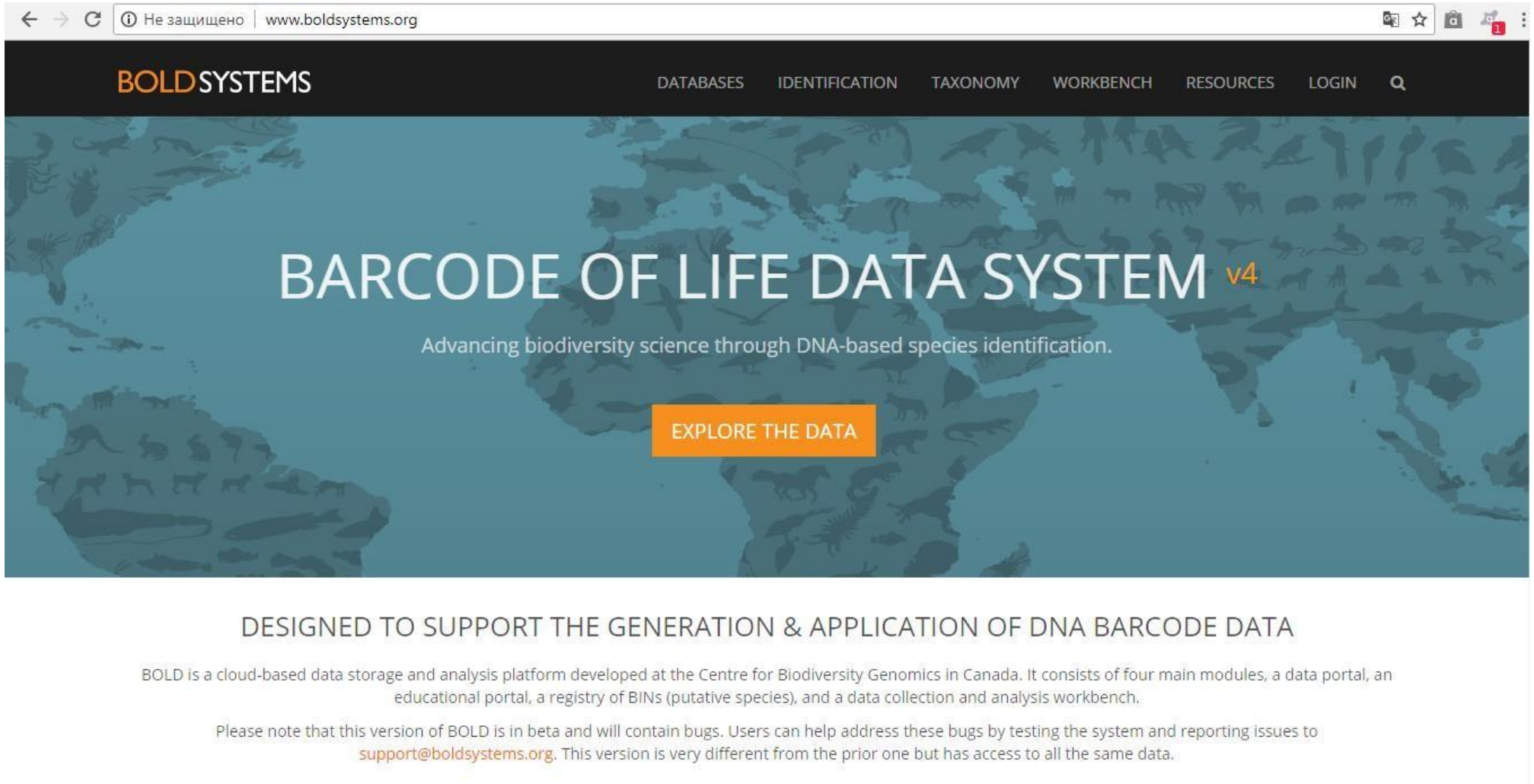
Query coverage - процент перекрытия анализируемой последовательности с последовательностями сравнения

Параметральная оценка



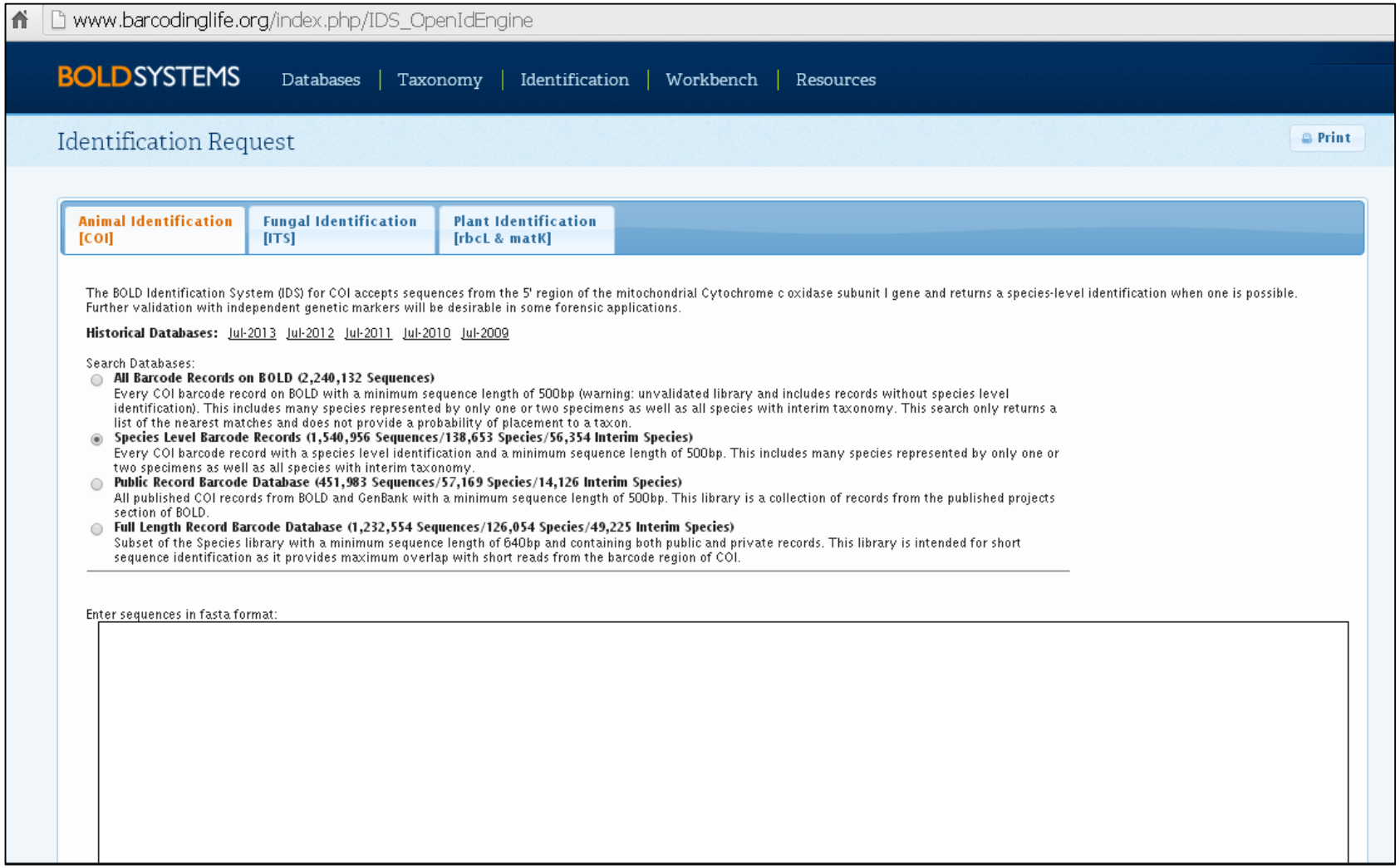
3. Идентификация вида BOLD

Анализ данных



3. Идентификация вида BOLD

Анализ данных



3. Идентификация вида BOLD

Анализ данных

www.barcodinglife.org/index.php/IDS_IdentificationRequest

BOLDSYSTEMS | Databases | Taxonomy | Identification | Workbench | Resources

Specimen Identification Request

Print

Query: unlabeled_sequence — Top Hit: Arthropoda - Hemiptera - Brachycaudus sp. A rgf-2008 (100%)

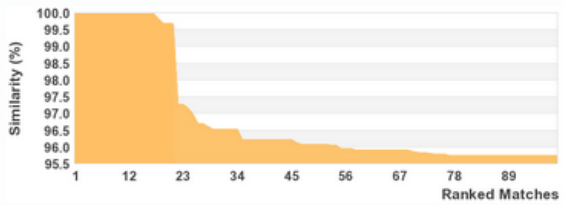
Search Request:

Type : COI FULL DATABASE (includes records without species designation)

Search Result:

Tree Based Identification

Similarity scores of the top 99 matches:



TOP 20 Matches :

Display option: Top 20

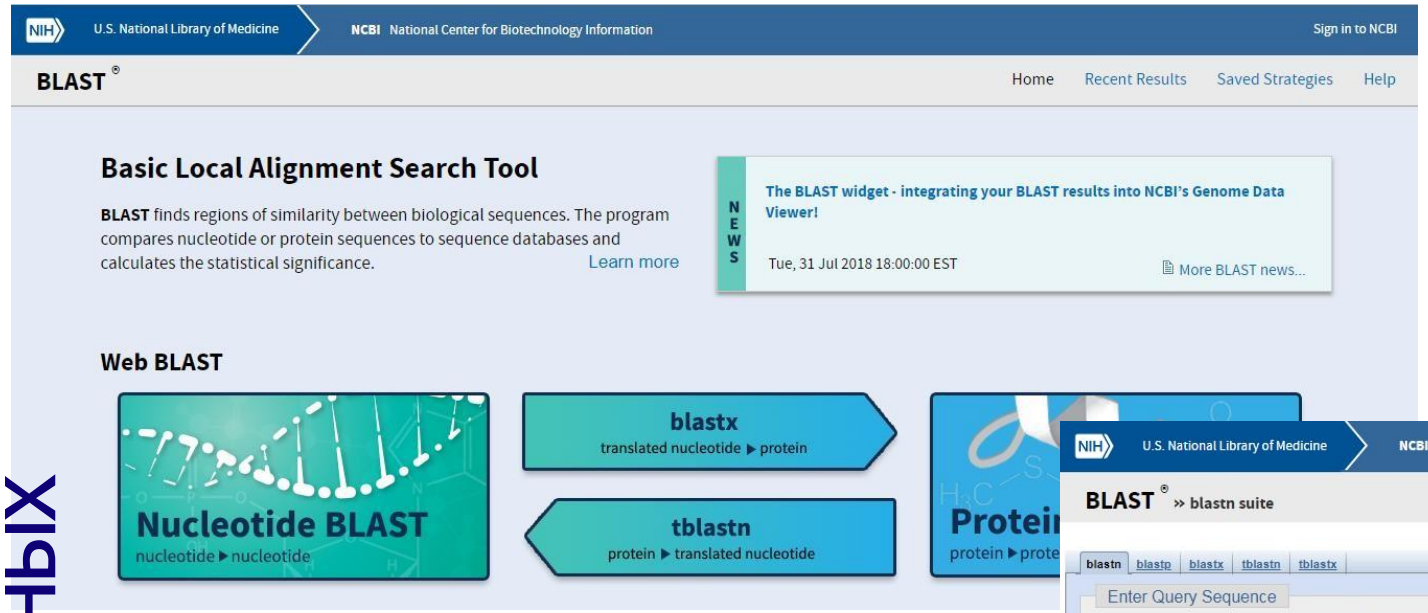
Phylum	Class	Order	Family	Genus	Species	Similarity (%)	Status
Arthropoda	Insecta	Hemiptera	Aphididae	Brachycaudus	*sp. A rgf-2008*	100	Published
Arthropoda	Insecta	Hemiptera	Aphididae	Brachycaudus	*tragopogonis*	100	Early-Release
Arthropoda	Insecta	Hemiptera	Aphididae	Brachycaudus	*tragopogonis*	100	Early-Release
Arthropoda	Insecta	Hemiptera	Aphididae	Brachycaudus	*tragopogonis*	100	Early-Release
Arthropoda	Insecta	Hemiptera	Aphididae	Brachycaudus	*tragopogonis*	100	Published
Arthropoda	Insecta	Hemiptera	Aphididae	Brachycaudus	*prunicola*	100	Published
Arthropoda	Insecta	Hemiptera	Aphididae	Brachycaudus	*schwartzi*	100	Published
Arthropoda	Insecta	Hemiptera	Aphididae	Brachycaudus	*schwartzi*	100	Published
Arthropoda	Insecta	Hemiptera	Aphididae	Brachycaudus	*schwartzi*	100	Published
Arthropoda	Insecta	Hemiptera	Aphididae	Brachycaudus		100	Early-Release

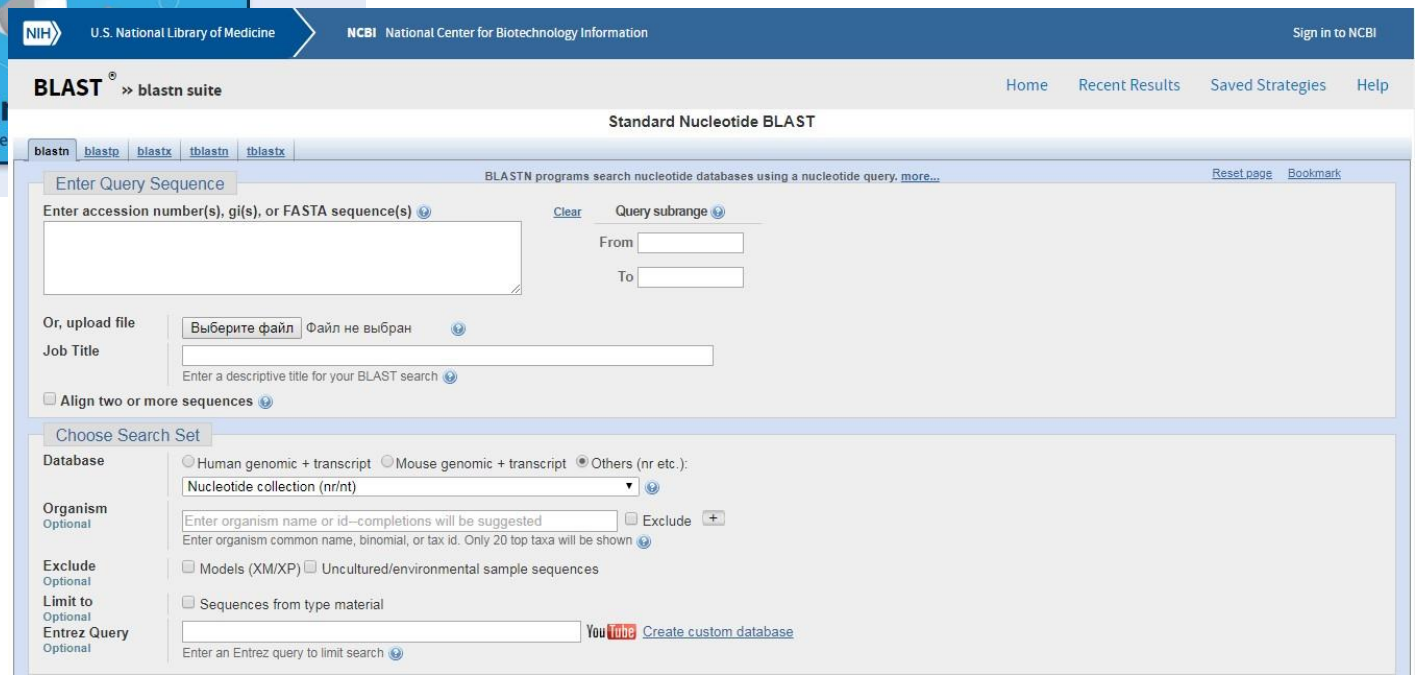
Barcoding

3. Идентификация вида

BLAST

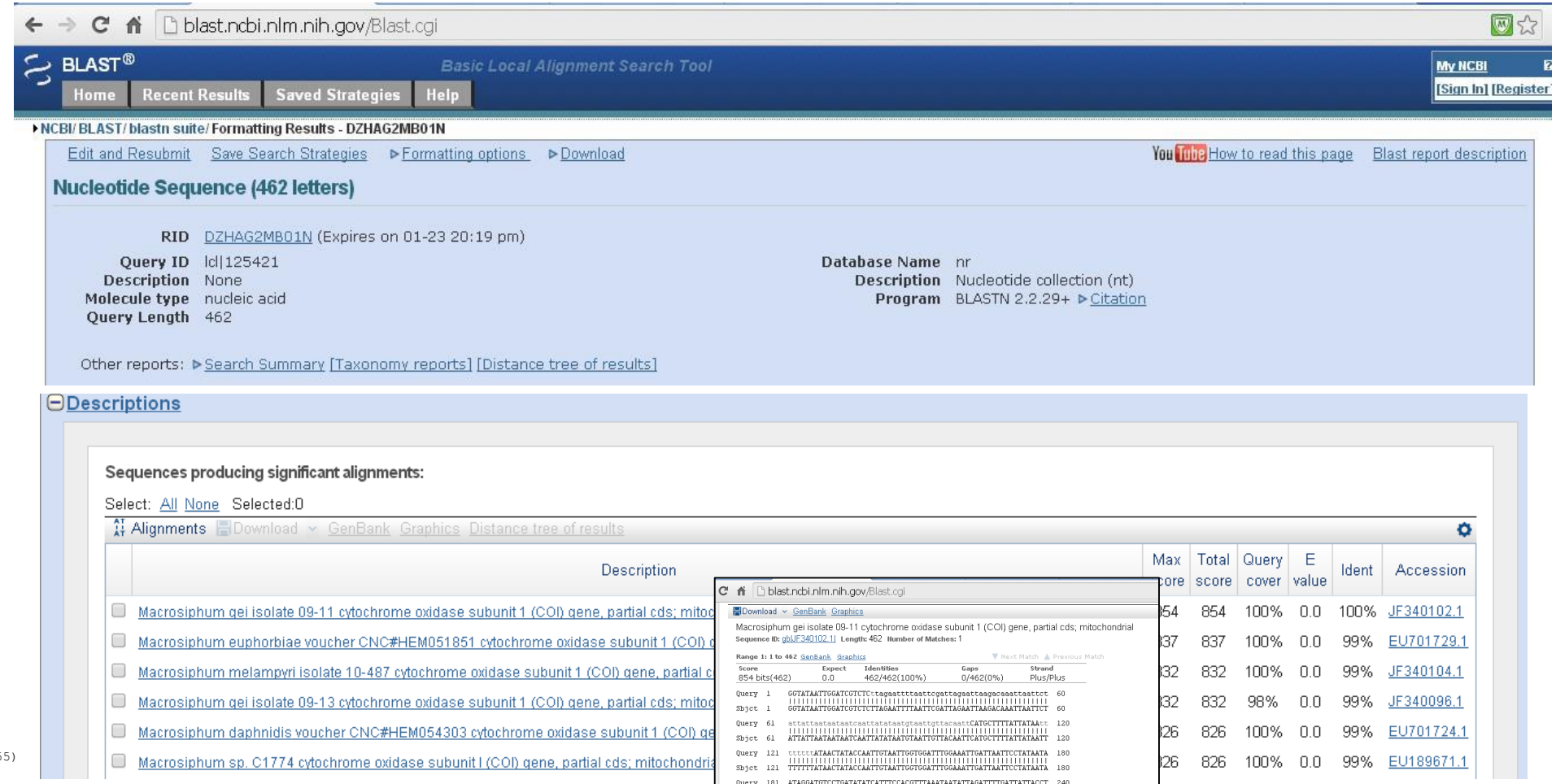
Анализ данных





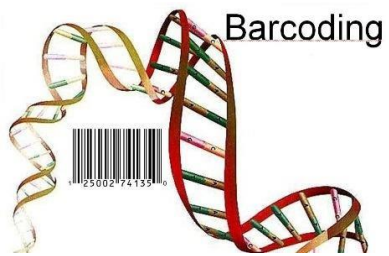
3. Идентификация вида. BLAST

Анализ данных



Macrosiphum (*Macrosiphum*) *gei* (Koch, 1855)
462 bp. Код доступа GenBank: JF340102

```
  1 ggtataattg gatcgtctct tagaatttta attcgattag aattaagaca  aattaattct
 61 attattaata ataatcaatt atataatgta attgttacaa ttcatgcttt  tattataatt
121 ttttttataa ctataccaat tgtaattggt ggatttggaa attgattaat  tcctataata
181 ataggatgtc ctgatatatc atttccacgt ttaaataata ttagattttg  attattacct
241 ccatcattaa taataataat ttgtagattt ttaattaata atggaacagg  aacaggatga
301 acaatttacc ccccttatc aaacaatatt gcacataata acatttcagt  tgatttaact
361 attttttctt tacatttagc aggaatctca tcaattttag gagcaattaa  ctttatttgt
421 acaattctta atataatacc aaacaatata aaattaaatc aa
```



3. Идентификация вида. BLAST

Анализ данных

www.ncbi.nlm.nih.gov/nuccore/JF340102

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide Search

Limits Advanced Help

Display Settings: GenBank Send:

Macrosiphum gei isolate 09-11 cytochrome oxidase subunit 1 (COI) gene, partial cds; mitochondrial

GenBank: JF340102.1

FASTA Graphics PopSet

Go to:

```
LOCUS       JF340102                462 bp    DNA     linear   INV 14-MAR-2011
DEFINITION  Macrosiphum gei isolate 09-11 cytochrome oxidase subunit 1 (COI)
            gene, partial cds; mitochondrial.
ACCESSION   JF340102
VERSION     JF340102.1  GI:325560635
KEYWORDS    .
SOURCE      mitochondrion Macrosiphum gei
  ORGANISM  Macrosiphum gei
            Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta;
            Pterygota; Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha;
            Aphidiformes; Aphidoidea; Aphididae; Macrosiphini; Macrosiphum.
REFERENCE   1  (bases 1 to 462)
  AUTHORS   Voronova,N.V., Kurchenko,V.P. and Buga,S.V.
  TITLE     The nucleotide partial sequences of cytochrome c oxidase subunit I
            (COI) gene of some aphid species from Belarus region
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 462)
  AUTHORS   Voronova,N.V., Kurchenko,V.P. and Buga,S.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-FEB-2011) Zoology, BSU, Kurchatova 10, Minsk, Minsk
            Region 220000, Belarus
FEATURES             Location/Qualifiers
```

Change region shown

Customize view

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

Related information

Related Sequences

PopSet

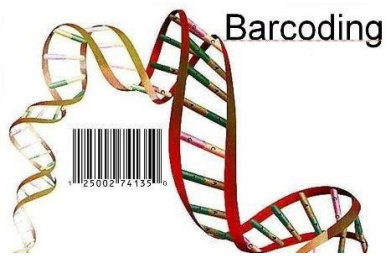
Protein

Taxonomy

Recent activity

Turn Off Clear

Macrosiphum gei isolate 09-11 cytochrome oxidase subunit 1 (COI) gene, partia Nucleotide



2. Идентификация вида. Алгоритм действия

Анализ данных

Определение вида по морфологии, ДНК-штрихкодирование, помещение штрихкода в базу данных

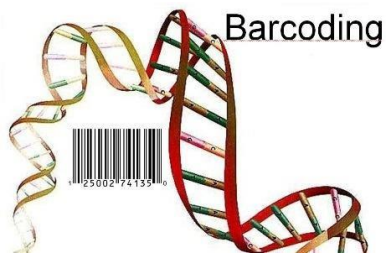
Работа других исследователей

ДНК-штрихкод

Выравнивание последовательности в системах BOLD или BLAST

Оценка качества выравнивания

Вид

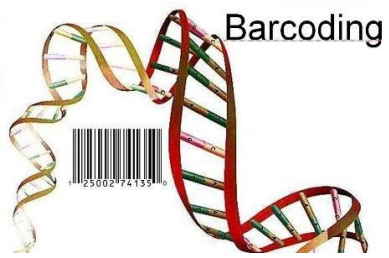


3. Идентификация вида

В базе нет референса

Что делать ?!!

Анализ данных



3. Идентификация вида

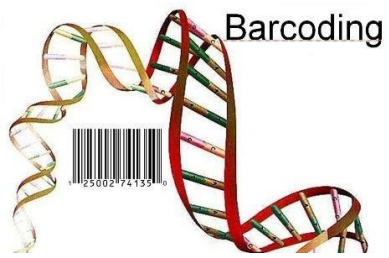
В базе нет референса?

Иди к систематикам!!

Анализ данных

1. Таксономист
2. Ваучерный образец
3. Хорошая этикетка

} решают ВСЕ проблемы

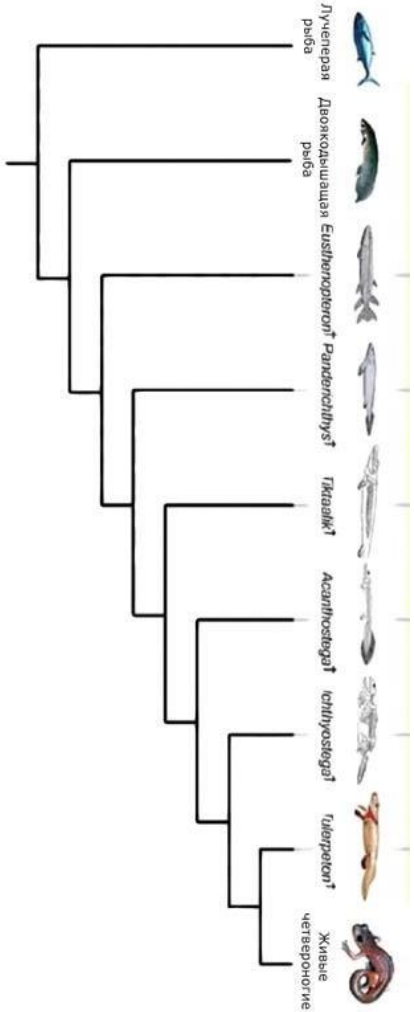


3. Идентификация вида
Использование филогенетического анализа

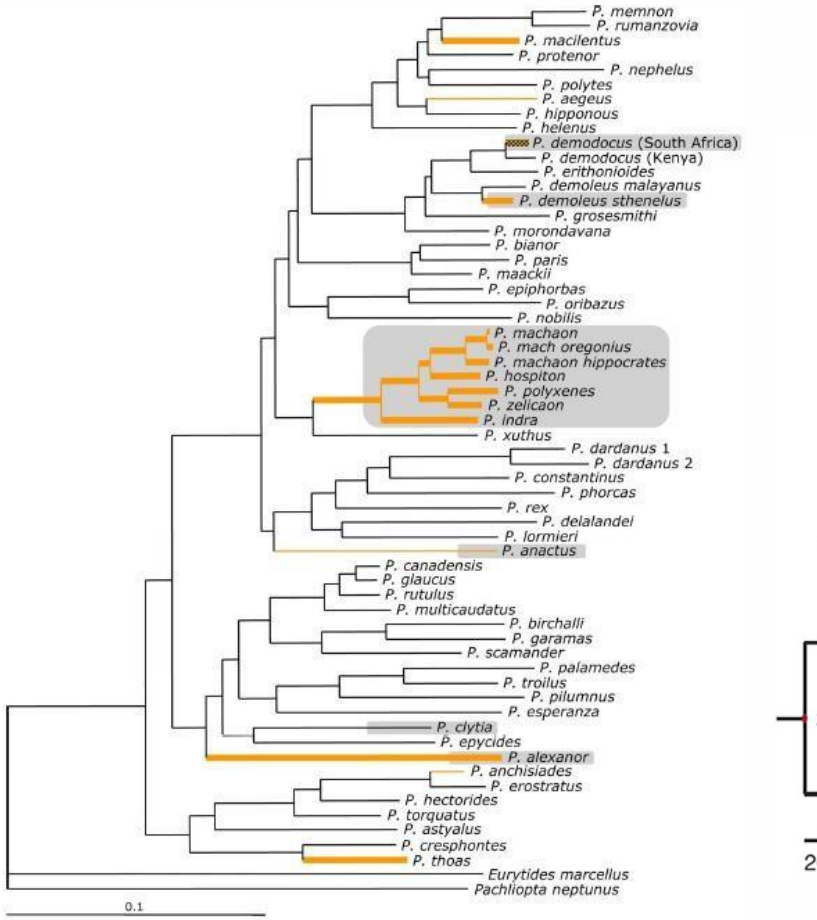
Суть метода

Анализ данных

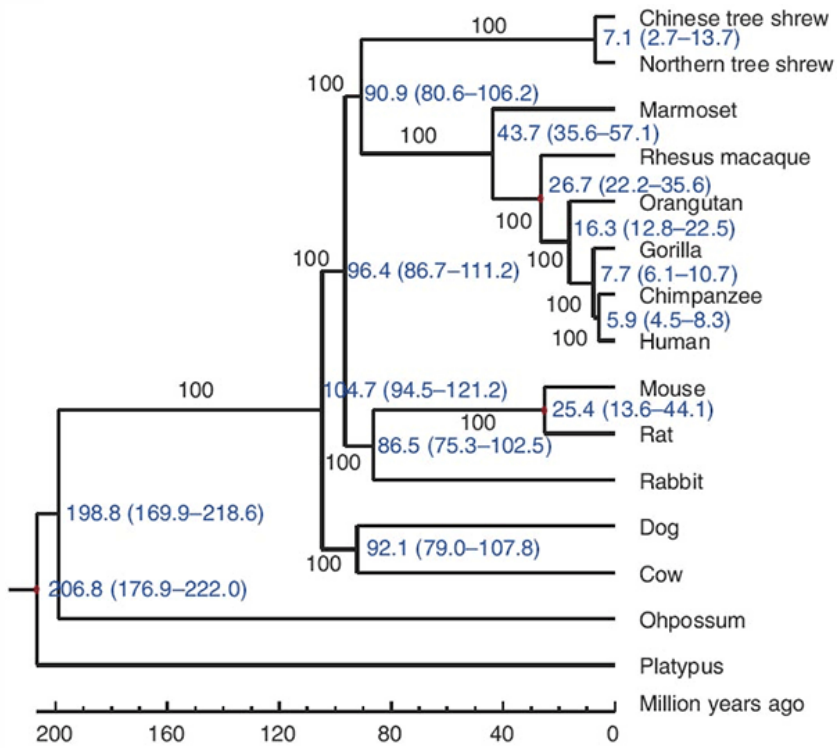
Кладограмма

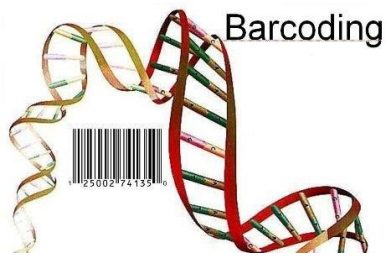


Филогенетическое дерево



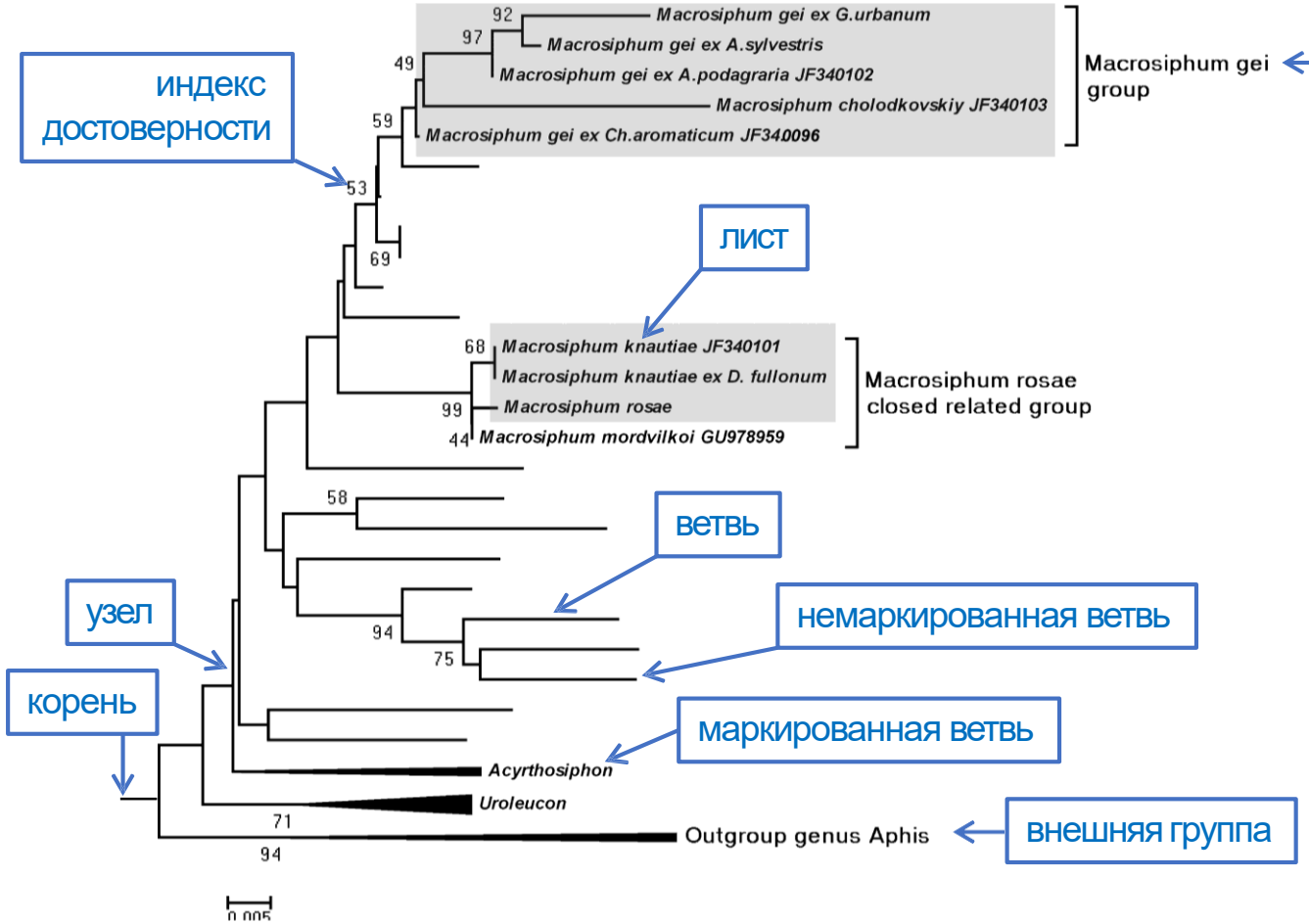
Хронограмма



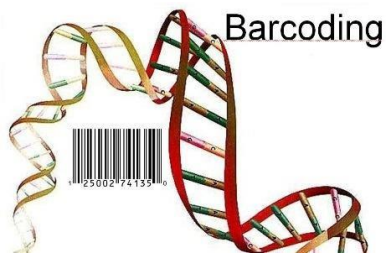


3. Идентификация вида
Использование филогенетического анализа

Анализ данных

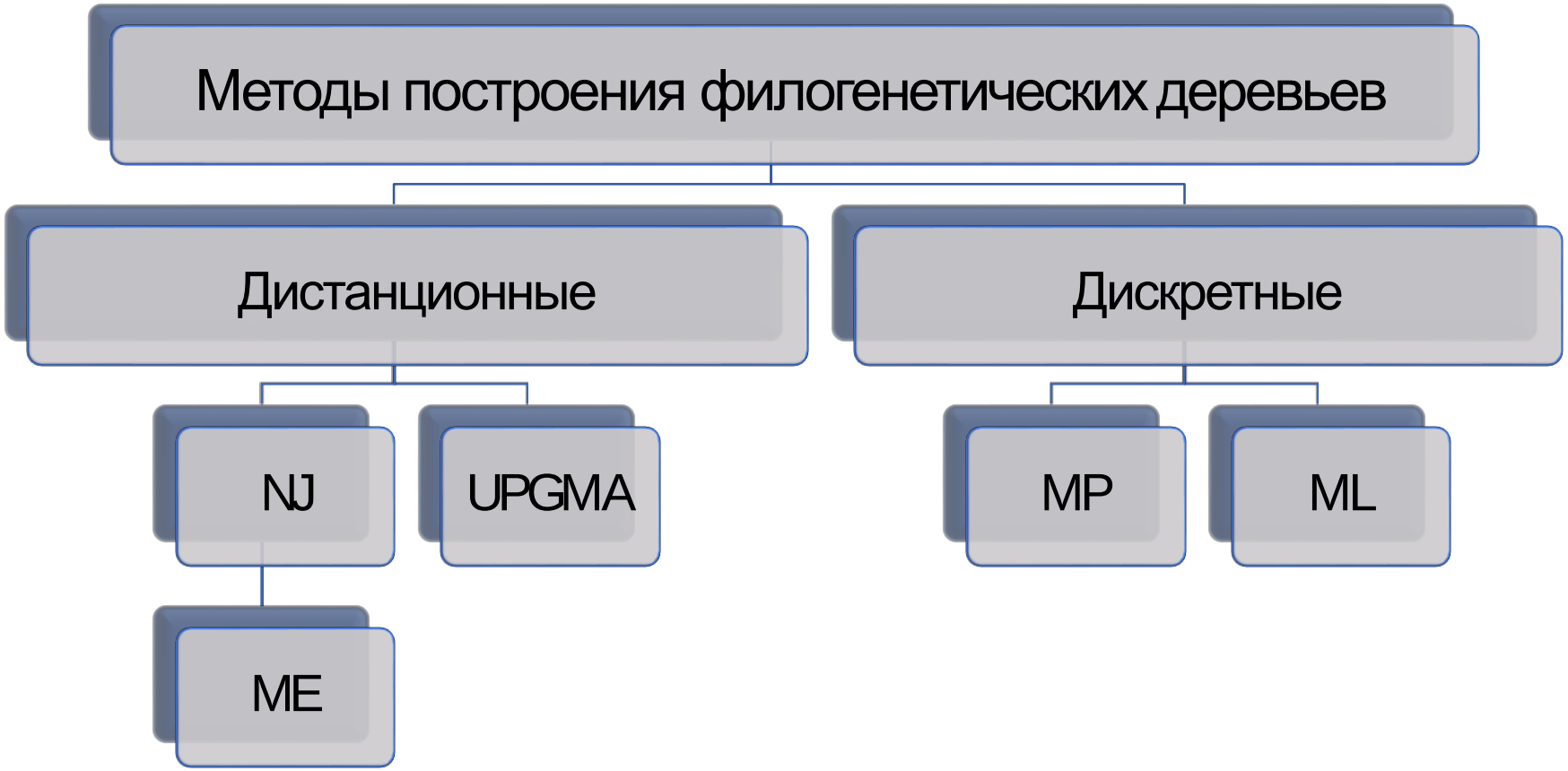


Филогенетическое дерево

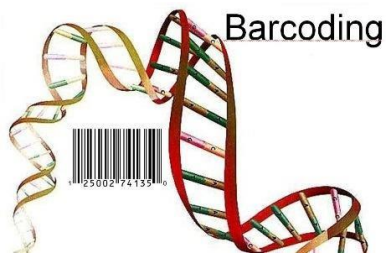


3. Идентификация вида
Использование филогенетического анализа

Анализ данных



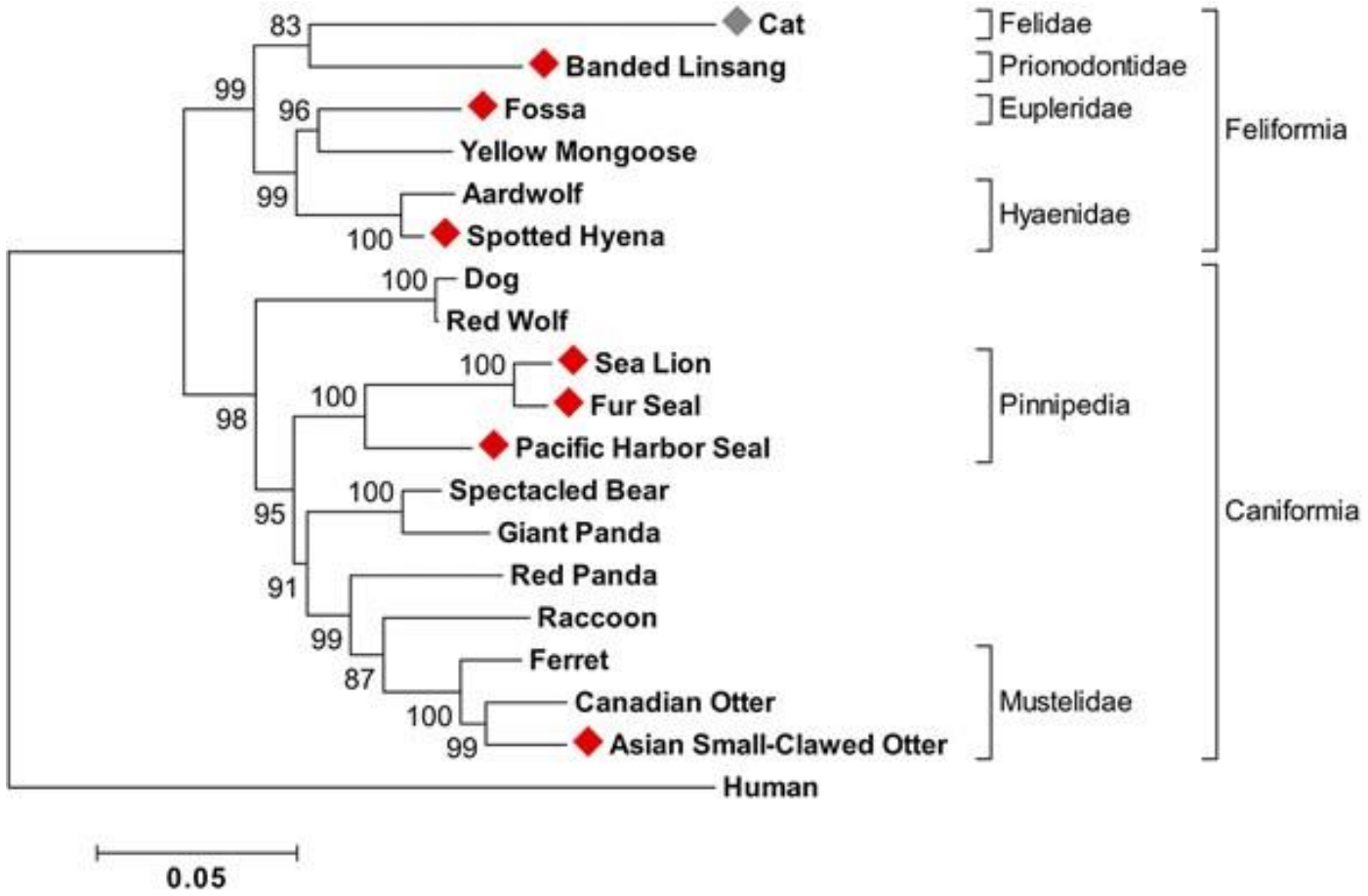
Основные методы построения деревьев



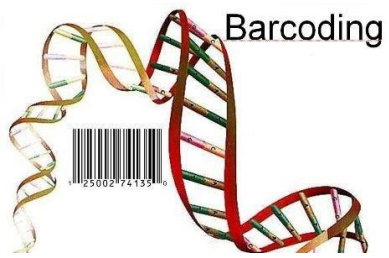
3. Идентификация вида
Использование филогенетического анализа

Статистическая проверка топологии дерева

Анализ данных



Бутстрэп-анализ — практический компьютерный метод определения статистик вероятностных распределений, основанный на многократной генерации выборок на базе имеющейся выборки



3. Идентификация вида
Использование филогенетического анализа

Анализ данных

Исходные последовательности

```
 00000000011111111122
 12345678901234567901
1AATCGTTCGATGACTTCGGAG
2AAGGACTCGTCGACCTCAGAG
```

```
 00000000011111111122
 12567878012345678801
1AAGTTCTCATGTGACTTCGGAG
2AAACTCTCTCGCGACCTCAAAG
```

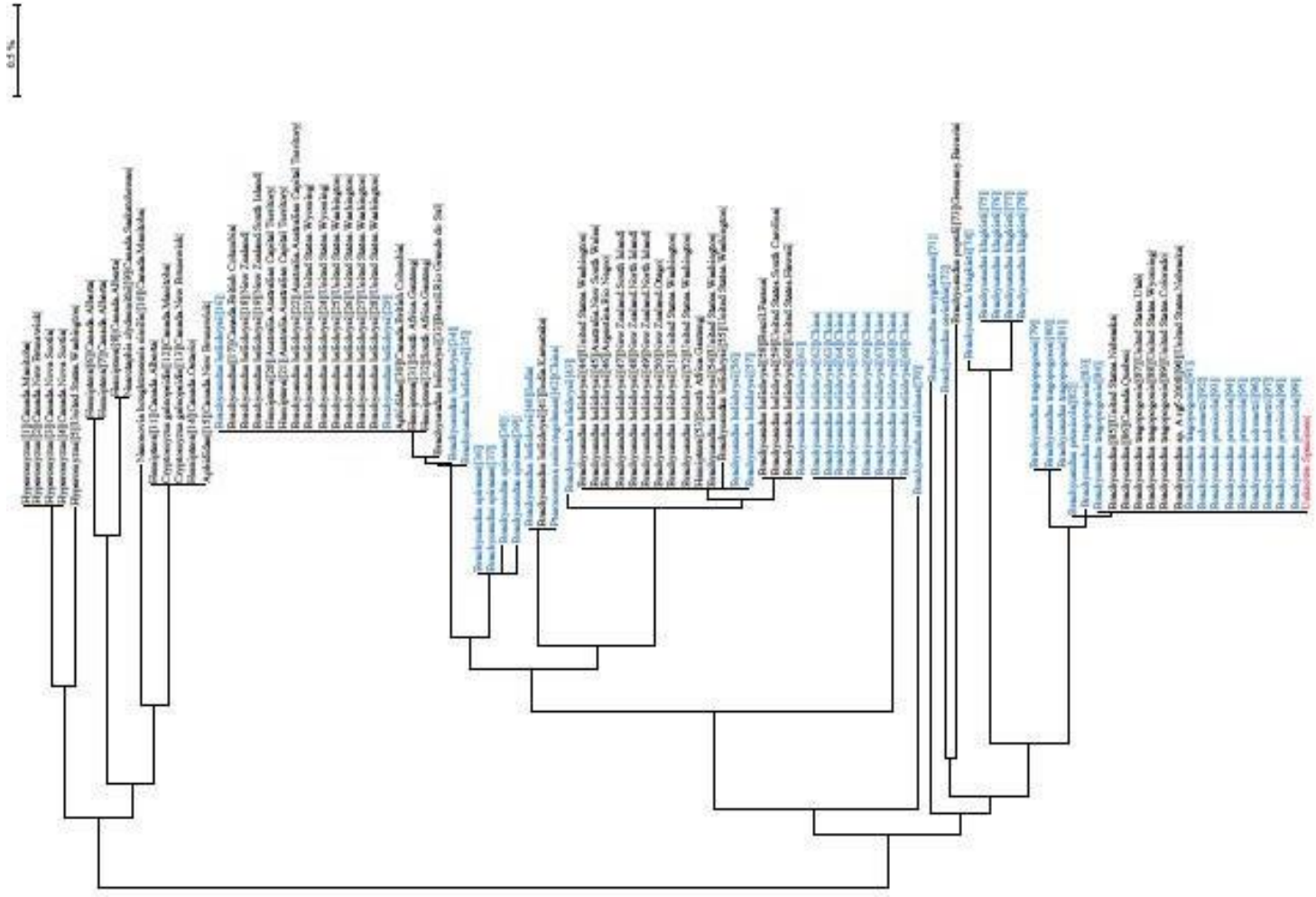
```
 00000000011111122122
 45678978901454590190 1
1CGTTCGTCGATCTCTGAGGAG
2GACTCGTCGTCCCCGAGGAG
```

Симулированные последовательности
(реплики)

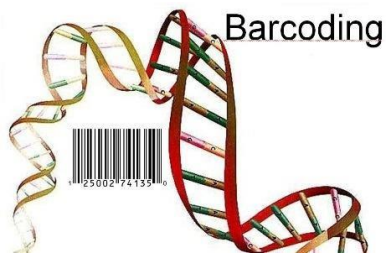
Статистическая проверка топологии дерева

3. Идентификация вида
Использование филогенетического анализа

Филогенетическое дерево, построенное в BOLD



Анализ данных



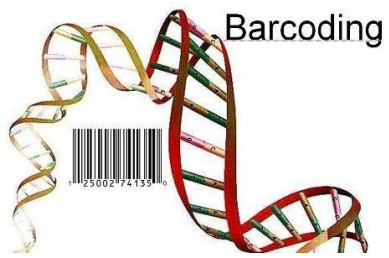
Выделение нового вида на основе ДНК-штрихкода

Криптические виды и механизмы их возникновения

Близкие виды – виды, обладающие определенным сходством морфологии и экологии вследствие общности происхождения

Сестринские виды – обычно молодые виды, обладающие высоким сходством морфологии вследствие единства происхождения

Криптические (скрытые) виды – виды, не различимые по морфологическим признакам, но имеющие различия в маркерных областях геномов



Выделение нового вида на основе ДНК-штрихкода

Генетическая (геномная) концепция вида

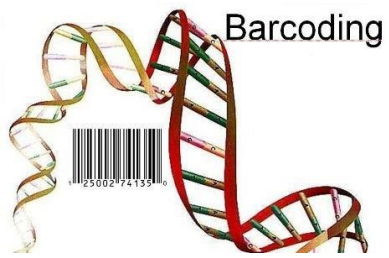
Генетический вид – группа особей, генетически изолированная от других таких же групп

↓

Свидетельство существования генетической изоляции: наличие «генетической дистанции» между группами

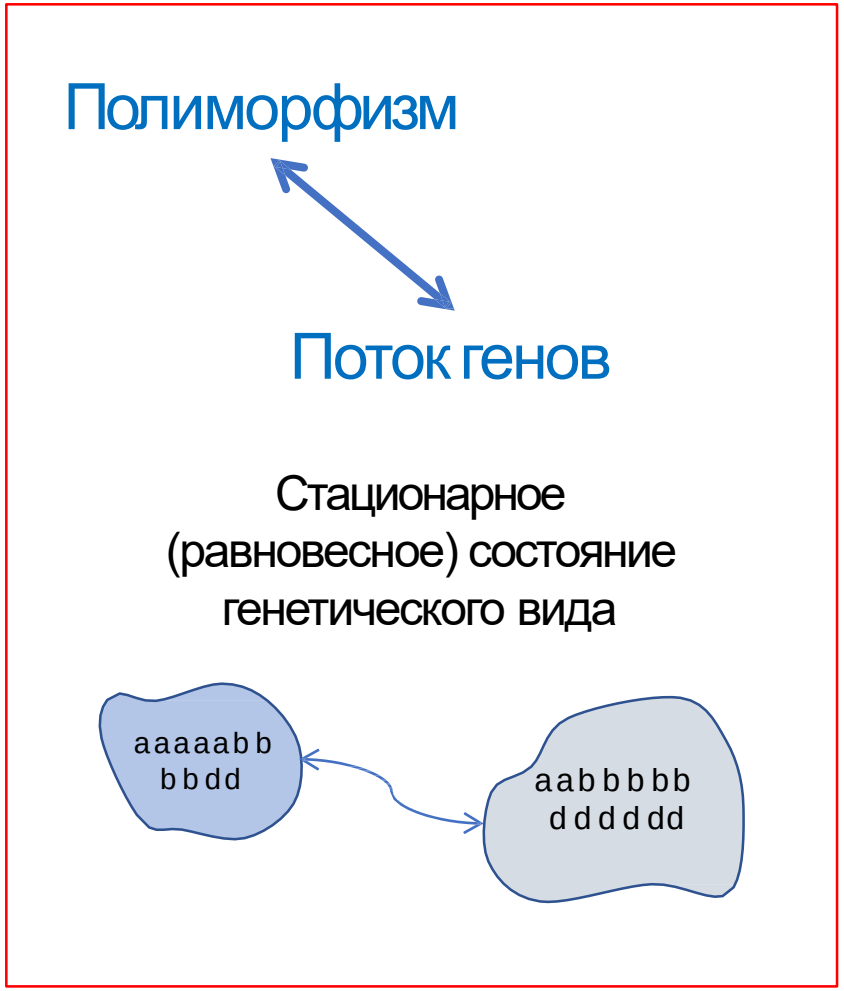
↓

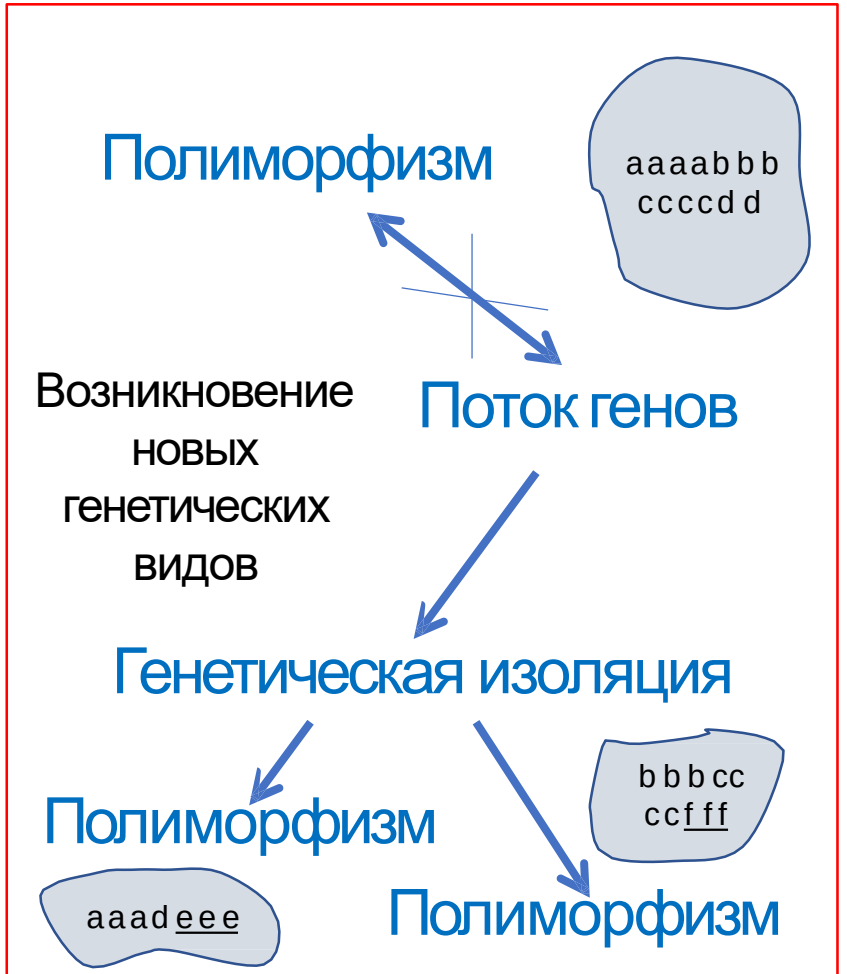
Генетическая дистанция – мера любых генетических различий между двумя особями

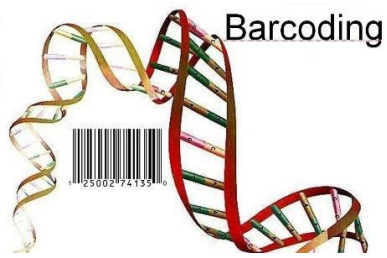


Выделение нового вида на основе ДНК-штрихкода

Генетические различия между видами







Выделение нового вида на основе ДНК-штрихкода

Криптические виды и механизмы их возникновения

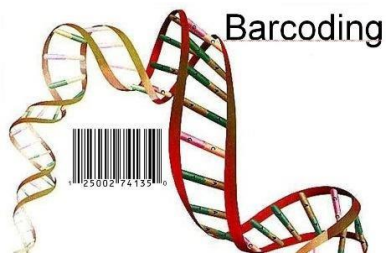
Генетический вид – группа особей, генетически изолированная от других таких же групп

↓

Свидетельство существования генетической изоляции : наличие «генетической дистанции» между группами

↓

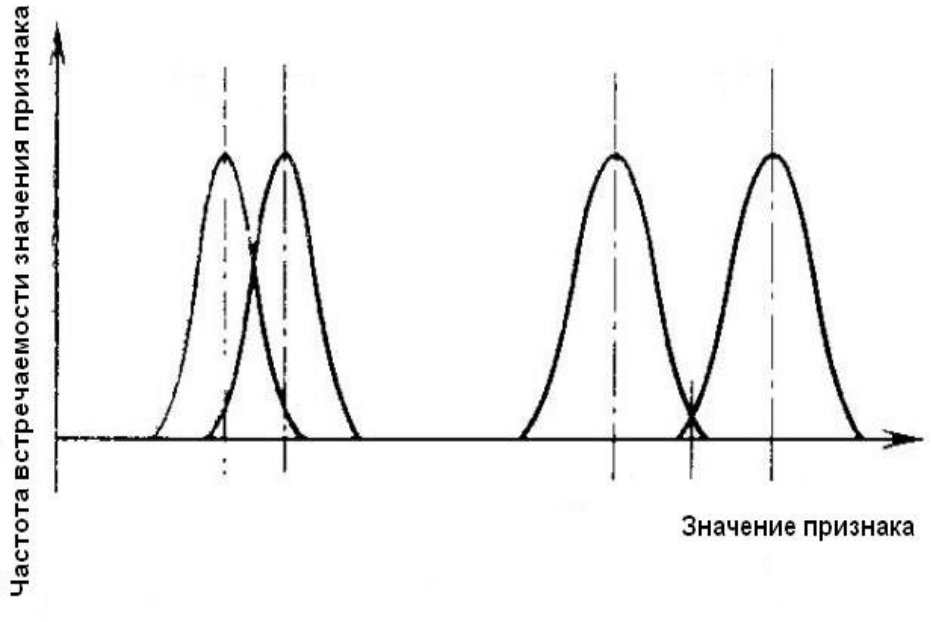
Генетическая дистанция – мера любых генетических различий между двумя особями

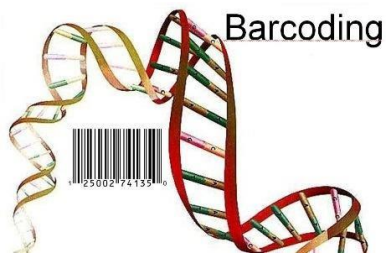


Выделение нового вида на основе ДНК-штрихкода

Концепция вида. Существование хиатусов между видами в значении или состоянии признака

Длина тела	
A	B
2	4 min
1	5
2	5
max 3	6
1	7
2	6
2	6
1	5
2	4
2	6
1	6





Выделение нового вида на основе ДНК-штрихкода

Генетическая дистанция

SNP – Single Nucleotide Polymorphism

GD=Np/N

$$d_{MTN} = -\frac{2\pi_A\pi_G}{\pi_R}\log\left(1 - \frac{\pi_R}{f_{AG}}p_{AG} - \frac{1}{2\pi_R}p_{RY}\right)$$
$$- \frac{2\pi_T\pi_C}{\pi_Y}\log\left(1 - \frac{\pi_Y}{f_{TC}}p_{TC} - \frac{1}{2\pi_Y}p_{RY}\right)$$
$$- 2\left(\pi_R\pi_Y - \frac{\pi_A\pi_G\pi_Y}{\pi_R} - \frac{\pi_T\pi_C\pi_R}{\pi_Y}\right)$$
$$\times \log\left(1 - \frac{1}{f_{RY}}p_{RY}\right)$$

Aphis fabae group

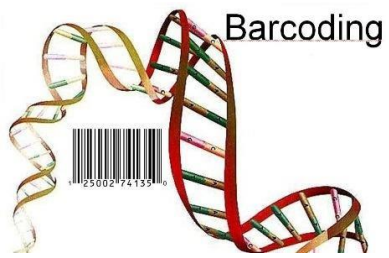
```
                                  190        200        210
Aphis_fabae_fabae           181 ATTATAATTA TTTTTATAAC TATACCAATT
Aphis_fabae_philadelfi      181 .........T .......... ..........
Aphis_fabae_cirsiiacanthoidis 181 .........T .......... ..........
Aphis_fabae_solanella       181 .........T .......... ..........

                                  310        320        330
Aphis_fabae_fabae           301 TTATTACCAC CATCACTAAT AATAATGATT
Aphis_fabae_philadelfi      301 .......... .......... ......A...
Aphis_fabae_cirsiiacanthoidis 301 .......... .......... ......A...
Aphis_fabae_solanella       301 .......... .......... ......A...

                                  400        410        420
Aphis_fabae_fabae           391 AATAATATTG CGCATAATAA TGTTTCAGTT
Aphis_fabae_philadelfi      391 .......... .C........ .A........
Aphis_fabae_cirsiiacanthoidis 391 .......... .C........ .A........
Aphis_fabae_solanella       391 .......... .C........ .A........
```

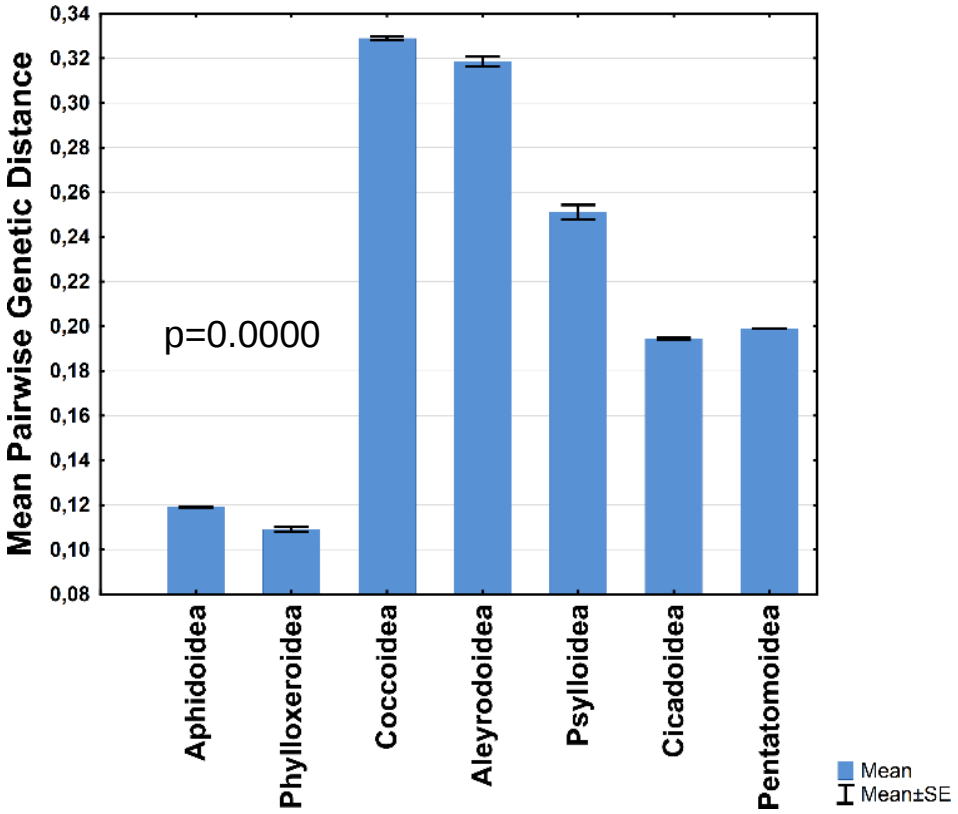
Генетическая дистанция между двумя последовательностями – доля полиморфных сайтов в последовательностях заданной длины

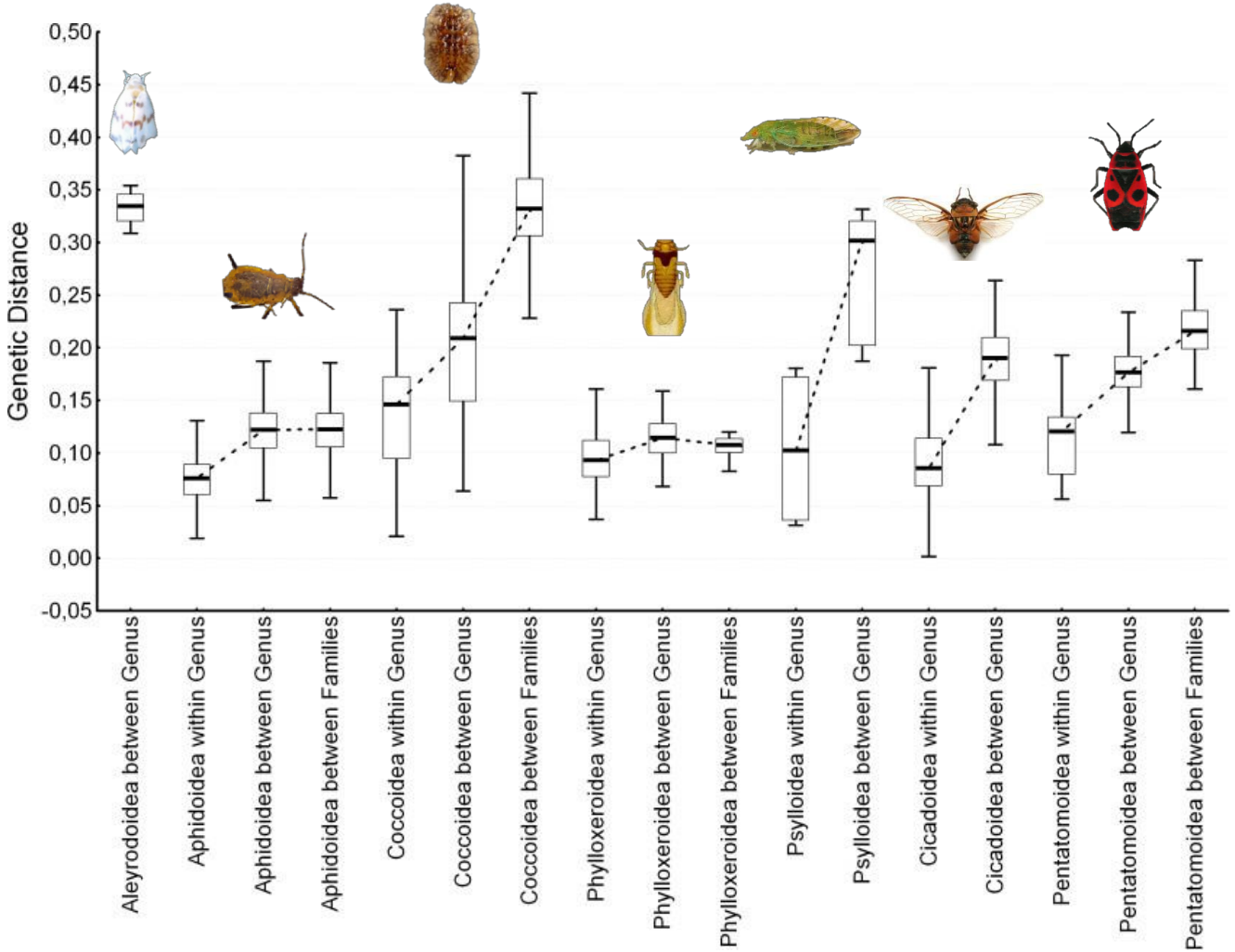
«Золотой стандарт» молекулярной таксономии – генетическая дистанция между близкими видами приблизительно равна 2 %

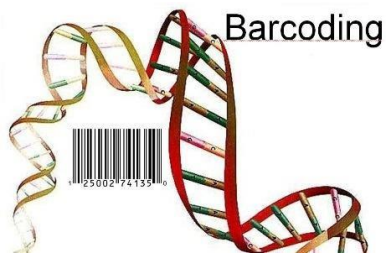


Выделение нового вида на основе ДНК-штрихкода

Особенности изменчивости COI в родственных таксонах насекомых







Выделение нового вида на основе ДНК-штрихкода

Алгоритм действия

Выделение в том же роде «хороших» видов с полученными ДНК-штрихкодами

→

Расчет значения генетических дистанций между ними

↓

Расчет значения генетических дистанций между потенциально новым видом и сестринскими

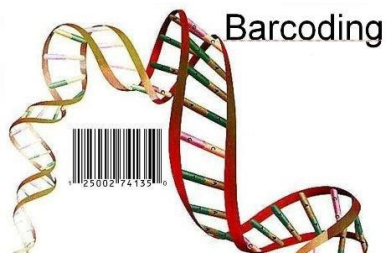
←

Сравнение значений генетических дистанций

↓

Заключение о существовании криптического вида (*о соответствии различий межвидовому уровню*)

→ ?



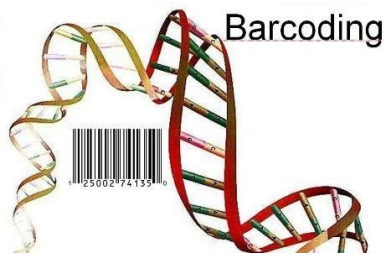
Возможные проблемы, которые могут возникнуть в процессе ДНК-штрихкодирования

1. Отсутствие референсной последовательности для определения (отсутствие ДНК-штрихкода в базах данных)
2. Ошибочно определенный референс (ДНК-штрихкод первоначально был получен для образца, определенного неверно)

Построение дендрограммы способно выявить проблему

3. Контаминация образца ДНК

Решение - только перевыделение ДНК и повторение процедуры

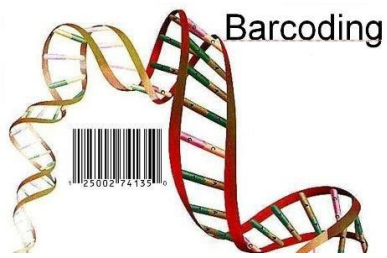


Возможные проблемы, которые могут возникнуть в процессе ДНК-штрихкодирования

Типы контаминации образцов :

- Паразиты (особенно, при выделении из тотального образца)
- Смешанные колонии
- Механическое загрязнение посторонним биоматериалом или «перенос» миниколичеств ДНК из образца в образец

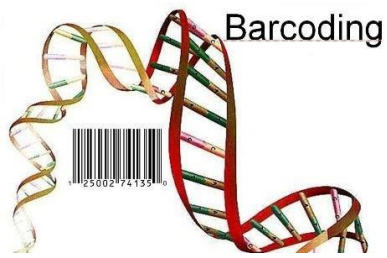
- Перепутанные (плохо маркированные) образцы



Возможные проблемы, которые могут возникнуть в процессе ДНК-штрихкодирования

Особые случаи

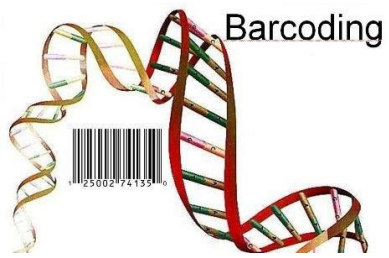
- Разные маркеры определяют один образец как разные виды
- Стандартные маркеры не работают для организмов целевого таксона
- Получить качественный сиквенс не удается, несмотря на многократные попытки



Возможные проблемы, которые могут возникнуть в процессе ДНК-штрихкодирования

Разные маркеры определяют один и тот же образец как разные виды. Причины

- Перепутанные образцы
- Контаминированный образец
- Гибрид



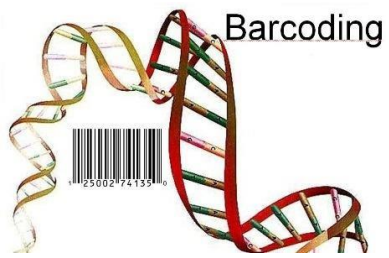
Возможные проблемы, которые могут возникнуть в процессе ДНК-штрихкодирования

Изучение гибридов

Обеспечение достаточно большой выборки → Использование митохондриальных и/или пластидных и ядерных маркеров

↓

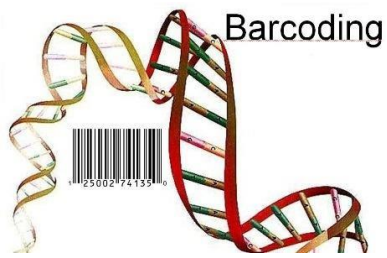
Анализ воспроизводимости результатов → ?



Возможные проблемы, которые могут возникнуть в процессе ДНК-штрихкодирования

Стандартные маркеры не работают для организмов целевого таксона

- Слишком консервативный маркер – выявляемых различий не достаточно
- Слишком вариабельный маркер – выявляется много внутривидовых различий, которые снижают уверенность в получаемых результатах
- Универсальные праймеры не работают

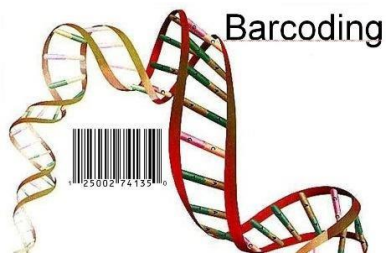


Возможные проблемы, которые могут возникнуть в процессе ДНК-штрихкодирования

Слишком консервативный маркер

Как выглядит : *филогенетическое дерево остается «не разрешенным»*

1. Оценка вариабельности разных митохондриальных маркеров
2. Смена белок-кодирующего маркера на некодирующий

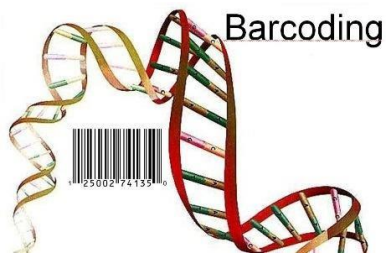


Возможные проблемы, которые могут возникнуть в процессе ДНК-штрихкодирования

Слишком вариабельный маркер

Как выглядит : *неправдоподобно большие значения генетических дистанций между видами, последовательности на дереве «перемешиваются»*

1. Использование ядерного маркера
2. Смена митохондриального маркера или анализируемого региона
3. Работа с аминокислотной последовательностью

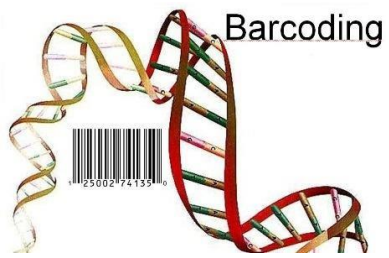


Возможные проблемы, которые могут возникнуть в процессе ДНК-штрихкодирования

Универсальные праймеры не работают

Как выглядит : *ПЦР-продукт отсутствует, в то время как реагенты остаются рабочими*

1. Смена универсальных праймеров
2. Секвенирование мтДНК и дизайн новых праймеров



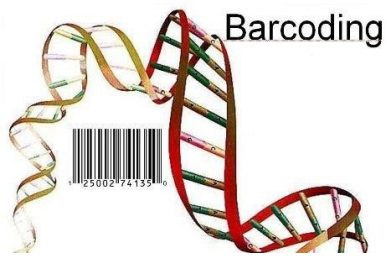
Возможные проблемы, которые могут возникнуть в процессе ДНК-штрихкодирования

Получить качественный сиквенс не удается

Как выглядит : *Несмотря на усилия, каждый сиквенс целевого региона оказывается «плохим»*

Причины :

1. NUMTs
2. Другие псевдогены



Возможные проблемы, которые могут возникнуть в процессе ДНК-штрихкодирования

Псевдогены (англ. *pseudogenes*) - нефункциональные аналоги структурных генов, утратившие способность кодировать белок и не экспрессирующиеся в клетке. Происходят от обычных функциональных генов, утрачивают способность экспрессии в результате нонсенс-мутаций, отсутствия функциональных участков и т.п.

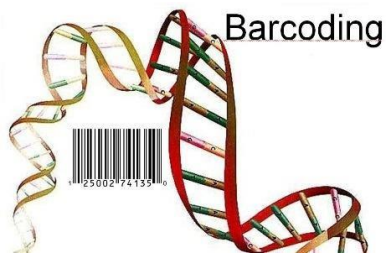
Numt (от “nuclear mitochondrial DNA”), псевдогены митохондриального происхождения, предположительно перенесенные транспосознами в ядерный геном

Species	mt genome size (bp)	Database size (bp)[b]	% of genome checked	No. of Numts[b]	Total length of Numts (bp)	Numt % of genome[b]
Saccharomyces cerevisiae	85 779	12 069 247	100	17	1389	0.012
Plasmodium falciparum	5967	28 718 804	50	3	228	0.0008
Caenorhabditis elegans	13 794	106 660 070	~100	2	212	~0.0002
Drosophila melanogaster	19 496	122 655 632	70	3	724	0.0006
Homo sapiens	16 569	2 853 531 108	84	354	418 552	0.012

[a]Abbreviations: bp, base pairs; Numt, nuclear mitochondrial pseudogenes.
[b]Summary of Numts in GenBank genome projects. Mitochondrial genomes were queried against nuclear mapped genome databases using the NCBI BLAST search facilities (http://www.ncbi.nlm.nih.gov/BLAST, http://www.ncbi.nlm.nih.gov/Malaria/plasmodiumblcus, http://www.ncbi.nlm.nih.gov/genome/seq/HsBlast) set at default parameters (16 February 2001). Only data with < 0.0001 probability of occurring by chance were included.

Количество NUMTs в геномах разных видов

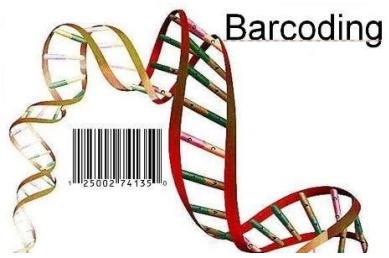
Bensasson et al. 2001



Правило «разумной аккуратности» применения ДНК-штрихкодирования

1. Оценка разрешающей способности метода в конкретном таксоне живых организмов
2. Оценка необходимости включения в идентификацию дополнительных ДНК-маркеров
3. Оценка достоверности определения видов референсных ДНК-штрихкодов

Отклонение от правила 2 %различий по ДНК-штрихкоду между близкими видами



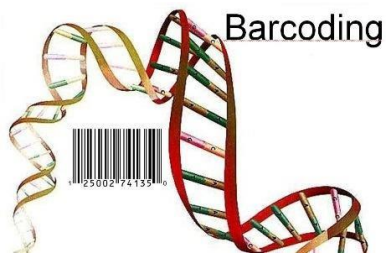
Практическая часть занятия. Порядок действий

Получение данных с секвенатора

Подготовка данных для анализа

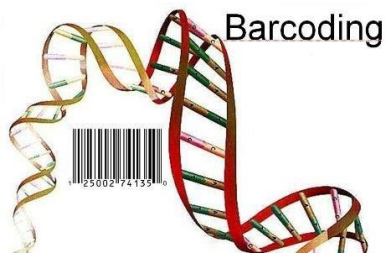
Автоматическая идентификация вида

Интерпретация результатов



Практическая часть занятия
Блок 1. Идентификация

1. Отобрать из набора качественные сиквенсы
2. Провести редактирование сиквенсов и перевести их в текстовый формат
3. Провести идентификацию видов с использованием системы BOLD и BLAST. Сравнить получаемые результаты
4. Построить филогенетические деревья в BOLD, оценить достоверность результатов идентификации



Практическая часть занятия
Блок 2. Анализ полиморфизма

1. Скачать из BOLD выборку нуклеотидных последовательностей целевого таксона
2. Провести очистку данных
3. Провести выравнивание последовательностей, оценить его результаты
4. Рассчитать генетические дистанции между последовательностями
5. Выбрать лучшую модель нуклеотидных замещений
6. Построить филогенетические деревья разными методами с оценкой статистической значимости топологии, с задействованием разных позиций в кодоне