



INSTITUTO
SUPERIOR DE
AGRONOMIA

Identificação de espécies por DNA e determinação do seu estatuto de conservação

Filipa Monteiro

UC: Flora e Vegetação

28 Maio 2025

Roadmap to species ID



1



3

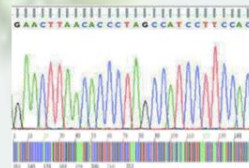


5

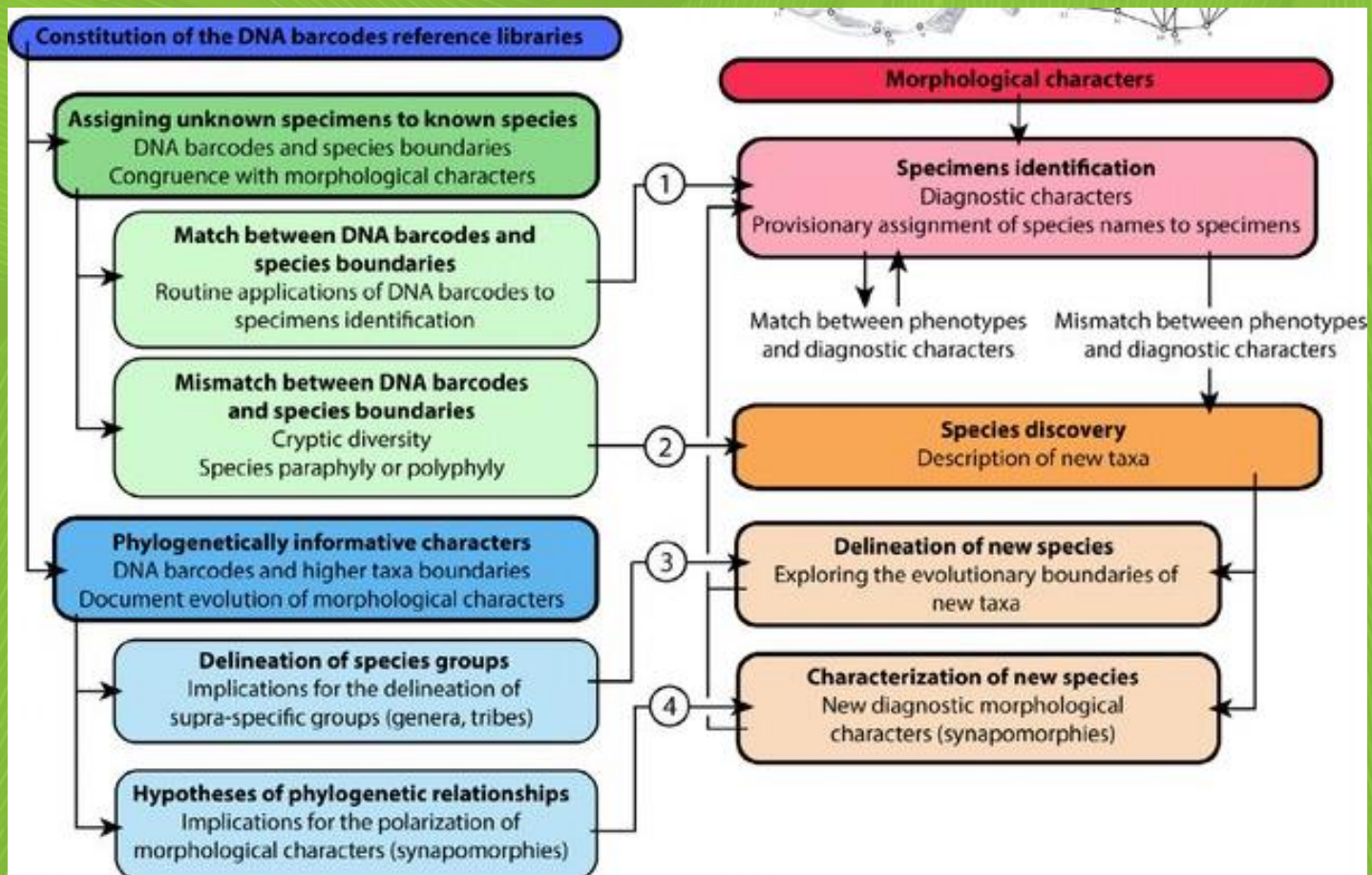
2



4



6



COMPARAÇÃO DA NOSSA SEQUÊNCIA DNA COM ABASE DE DADOS



🌿 **BLAST** (**B**asic **L**ocal **A**lignment **S**earch **T**ool) é uma ferramenta online do NCBI (National Center for Biotechnology Information) que permite encontrar regiões de similaridade entre sequências biológicas (nucleótidos ou proteínas).

U.S. National Library of Medicine | NCBI | Sign in to NCBI

BLAST®

Home Recent Results Saved Strategies Help

Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. [Learn more](#)

NEW
IgBLAST 1.8.0 released
A new version of IgBLAST is now available.
Wed, 15 Nov 2017 16:00:00 EST | [More BLAST news...](#)

Web BLAST

Nucleotide BLAST
nucleotide → nucleotide

blastx
translated nucleotide → protein

tblastn
protein → translated nucleotide

Protein BLAST
protein → protein

BLAST® - blastn suite | Home | Recent Results | Saved Strategies | Help

Standard Nucleotide BLAST

BLASTn programs search nucleotide databases using a nucleotide query.

Enter Query Sequence

Enter accession number(s), url(s), or FASTA sequence(s) [View](#)

Query subrange: From To

or upload file: Choose File | No file chosen

Job Title: COLFF759-13

Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

Database: **Standard databases (nr, etc.)** | rRNA/ITS databases | Genomic + transcript databases | Detacoronavirus

Nucleotide collection (nr/nt)

Organism: Exclude: [Add organisms](#)

Exclude: ☐ Models (CDX/P) ☐ Unpublished/non-referenced sample sequences

Limit to: ☐ Sequences from type material

Enter Query: [View](#) [Create custom database](#)

Enter an EMBL query to find search

Program Selection

Optimize for: **Highly similar sequences (megablast)** | More dissimilar sequences (discontinuous megablast) | Somewhat similar sequences (blastn) | Choose a BLAST algorithm

BLAST Search database Nucleotide collection (nr/nt) using Megablast (Optimize for highly similar sequences)

☐ Show results in a new window

Após a comparação

BLAST[®] » blastn suite » results for RID-RYR0M8CC016

Home Recent Results Saved Strategies Help

[Edit Search](#) [Save Search](#) [Search Summary](#)

Job Title: sample_denv2

RID: RYR0M8CC016 Search expires on 20-10-2023 02:36 am [Download All](#)

Program: BLASTN [Citation](#)

Database: nt [See details](#)

Query ID: lcl|Query_52797

Description: sample_denv2

Molecule type: dna

Query Length: 10703

Other reports: [Distance tree of results](#) [MSA viewer](#)

Filter Results

Organism: only top 20 will appear ☐ exclude

Type common name, binomial, taxid or group name

[Add organism](#)

Percent Identity: to E value: to Query Coverage: to

[Filter](#) [Reset](#)

[Descriptions](#) [Graphic Summary](#) [Alignments](#) [Taxonomy](#)

Sequences producing significant alignments

[Download](#) [Manage Columns](#) [Show](#) [Feedback](#)

Max score: A pontuação mais alta é calculada a partir de correspondências de alinhamento e incompatibilidades.

Total score: a soma das pontuações do alinhamento todos segmentos da sequência.

Query coverage: a % do comprimento do contig que se alinha com a sequência na base de dados (NCBI).

Descriptions						
Sequences producing significant alignments						
Download Manage Columns Show 100						
select all 100 sequences selected						
		GenBank	Graphics	Distance tree of results		
Description		Max Score	Total Score	Query Cover	E value	Per. Ident
☒ Dengue virus 2 isolate CHRF_DenV002... partial genome		4566	15967	81%	0.0	98.58%
☒ Dengue virus 2 B17-1489 RNA... nearly complete genome		4540	15849	81%	0.0	98.43%
☒ Dengue virus 2 B17-1556 RNA... nearly complete genome		4495	15780	81%	0.0	98.11%
☒ Dengue virus 2 B17-1357 RNA... nearly complete genome		4494	15740	81%	0.0	98.11%
☒ Dengue virus 2 B17-1571 RNA... nearly complete genome		4488	15729	81%	0.0	98.07%
☒ Dengue virus 2 strain TM38 polyprotein gene... complete cds		4479	16098	81%	0.0	97.96%
☒ Dengue virus 2 B17-1634 RNA... nearly complete genome		4477	15706	81%	0.0	97.99%
☒ Dengue virus 2 strain TM132 polyprotein gene... complete cds		4468	15896	81%	0.0	97.88%
☒ Dengue virus 2 strain TM39 polyprotein gene... complete cds		4468	15847	81%	0.0	97.88%
☒ Dengue virus 2 isolate SGIEHI0252862Y15... complete genome		4462	15655	81%	0.0	97.84%
☒ Dengue virus 2 isolate SGIEHI0259687Y15... complete genome		4462	15655	81%	0.0	97.84%

- **E value:** o número de resultados esperados na base de dados. Quanto mais próximo de 0, melhor.
- **Percent identity:** a % de bases idênticas ao genoma de referência.
- **Accession [number]:** um identificador único atribuído aos registos nas bases de dados NCBI.

Descriptions							Graphic Summary	Alignments	Taxonomy
Sequences producing significant alignments							Download	Manage Columns	Show 100
select all 100 sequences selected							GenBank	Graphics	Distance tree of results
	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession		
☒	Denque virus 2 isolate CHRF_DenV002_narrat genome	4566	15967	81%	0.0	98.58%	MG320061.1		
☒	Denque virus 2 B17-1489.RNA_nearly complete genome	4540	15849	81%	0.0	98.43%	LC436672.1		
☒	Denque virus 2 B17-1556.RNA_nearly complete genome	4495	15780	81%	0.0	98.11%	LC436673.1		
☒	Denque virus 2 B17-1367.RNA_nearly complete genome	4494	15740	81%	0.0	98.11%	LC436669.1		
☒	Denque virus 2 B17-1571.RNA_nearly complete genome	4488	15729	81%	0.0	98.07%	LC436674.1		
☒	Denque virus 2 strain TM38 poliovirulen pene complete cds	4479	16098	81%	0.0	97.96%	KX452017.1		
☒	Denque virus 2 B17-1634.RNA_nearly complete genome	4477	15706	81%	0.0	97.99%	LC436675.1		
☒	Denque virus 2 strain TM132 poliovirulen pene complete cds	4468	15896	81%	0.0	97.88%	KX452030.1		
☒	Denque virus 2 strain TM39 poliovirulen pene complete cds	4468	15847	81%	0.0	97.88%	KX452018.1		
☒	Denque virus 2 isolate SGIEHI020862Y15 complete genome	4462	15655	81%	0.0	97.84%	MG513444.1		
☒	Denque virus 2 isolate SGIEHI020862Y15 complete genome	4462	15655	81%	0.0	97.84%	KY921905.1		
☒	Denque virus 2 strain GML16 complete genome	4457	15638	81%	0.0	97.84%	KX452019.1		

Download		GenBank	Graphics
Fusarium tricinctum isolate KGS17 clone DR-102 translation elongation factor 1-alpha			
Sequence ID: MT462120.1		Length: 699	Number of Matches: 1
Range 1: 1 to 699		GenBank	Graphics
Score	1291 bits(699)	Expect	0.0
		Identities	699/699(100%)
		Gaps	0/699(0%)
		Strand	Plus/Plus
Query 1:	ATGGGTAAAGGAGGACAAAGATCACTTAAAGTCGTGTCATCGGACACGTCGACTCTGGC	60	
Subject 1	ATGGGTAAAGGAGGACAAAGATCACTTAAAGTCGTGTCATCGGACACGTCGACTCTGGC	60	
Query 61	AAGTCGACCACTGTAAAGTAAACAAACAGCGGGTTGCTTATCTGCACTCGGAATCCGCCA	120	
Subject 61	AAGTCGACCACTGTAAAGTAAACAAACAGCGGGTTGCTTATCTGCACTCGGAATCCGCCA	120	
Query 121	AACCTGGCAGGGTATCAACAAACATCTTGCTAACTTTTGAAGACCGGTCACCTTGATCT	180	
Subject 121	AACCTGGCAGGGTATCAACAAACATCTTGCTAACTTTTGAAGACCGGTCACCTTGATCT	180	



Anacardium occidentale isolate 92 maturase K (matK) gene, partial cds; chloroplast

GenBank: AY594459.1

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

Go to:

LOCUS AY594459 1945 bp DNA linear PLN 26-JUL-2016
DEFINITION Anacardium occidentale isolate 92 maturase K (matK) gene, partial cds; chloroplast.
ACCESSION AY594459
VERSION AY594459.1
KEYWORDS .
SOURCE chloroplast Anacardium occidentale (cashew)
ORGANISM [Anacardium occidentale](#)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; eudicotyledons; Gunneridae; Pentapetales; rosids; malvids; Sapindales; Anacardiaceae; Anacardium.
REFERENCE 1 (bases 1 to 1945)
AUTHORS Pell,S.K.
TITLE Molecular Phylogeny of Anacardiaceae: Intrafamilial Classification and Evolutionary Relationships of Noted Genera
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1945)
AUTHORS Pell,S.K.
TITLE Direct Submission
JOURNAL Submitted (07-APR-2004) Cullman Program for Molecular Systematics Studies, New York Botanical Garden, 200th St. and Kazimiroff Blvd., Bronx, NY 10458, USA
FEATURES
source Location/Qualifiers
1..1945
/organism="Anacardium occidentale"
/organelle="plastid:chloroplast"
/mol_type="genomic DNA"
/isolate="92"
/specimen_voucher="Mori 24142"
/db_xref="taxon:171929"
/country="French Guiana"
/note="authority: Anacardium occidentale L."
399..>1945
gene

Customize view

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

Related information

Protein

Taxonomy

LinkOut to external resources

Order MATK cDNA clone/Protein/Antibody/RNAi
[OriGene]

Recent activity

Turn Off Clear

Anacardium occidentale isolate 92 maturase K (matK) gene, partial cds; chloroplast Nucleotide

anacardium occidentale matK (3) Nucleotide

Anacardium occidentale voucher Zhang L sn(KUN) internal transcribed spacer Nucleotide

anacardium occidentale Internal (461) Nucleotide

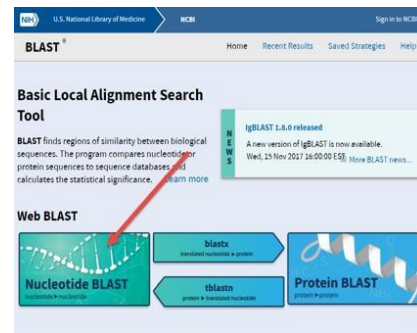
anacardium occidentale ITS (1591) Nucleotide

EXERCÍCIO

Siga o link "https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&BLAST_SPEC=GeoBlast&PAGE_TYPE=BlastSearch" na página principal do NCBI Blast.

- Selecione "Core nucleotide database (nt)" como base de dados de dados de pesquisa. NR é a base de dados "Não redundante", que contém todas as sequências não redundantes (não idênticas) do GenBank e dos bancos de dados completos do genoma.

- Clique no botão BLAST para iniciar a pesquisa.



Choose Search Set

Database ☒ Standard databases (nr etc.): ☐ rRNA/ITS databases ☐ Genomic + transcript databases ☐ Betacoronavirus ☐ Experimental databases

Core nucleotide database (core nt)

Organism Optional ☐ exclude

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

Limit to Optional ☐ Sequences from type material

Entrez Query Optional

Program Selection

Optimize for ☒ Highly similar sequences (megablast) ☐ More dissimilar sequences (discontiguous megablast) ☐ Somewhat similar sequences (blastn)

Choose a BLAST algorithm ?

Search database core_nt using Megablast (Optimize for highly similar sequences) ☐ Show results in a new window

RESULTADO DO BLAST

Após a conclusão da pesquisa, analisar a página resultante do BLAST, observando-se três partes principais

BLAST * » blastn suite » results for RID-M3JFFUAU010

Home Recent Results Saved Strat

[Edit Search](#) [Save Search](#) [Search Summary](#) [How to read this report?](#) [BLAST Help Videos](#) [Back to Traditional](#)

Job Title NM_001275:Homo sapiens chromogranin A (CHGA),...
RID M3JFFUAU010 [Search expires on 08-02 01:03 am](#) [Download All](#)
Program BLASTN [Citation](#)
Database refseq_rna [See details](#)
Query ID NM_001275.4
Description Homo sapiens chromogranin A (CHGA), transcript variant 1, mRNA
Molecule type nucleic acid
Query Length 1985
Other reports [Distance tree of results](#) [MSA viewer](#)

Filter Results
Organism only top 20 will appear
Type common name, binomial, taxid or group name
[Add organism](#)
Percent Identity to **E value** to **Filter**

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments [Download](#) [Manage Columns](#) [Show](#) 10

☒ select all 7 sequences selected

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	A
☒	Macaca mulatta chromogranin A (CHGA). mRNA	3081	3081	99%	0.0	94.77%	NM_001275.4
☒	Macaca fascicularis chromogranin A (CHGA). mRNA	3075	3075	99%	0.0	94.72%	NM_001275.4
☒	Equus caballus chromogranin A (CHGA). mRNA	1628	1628	93%	0.0	83.08%	NM_001275.4
☒	Bos taurus chromogranin A (CHGA). mRNA	1548	1548	95%	0.0	81.91%	NM_001275.4
☒	Sus scrofa chromogranin A (CHGA). mRNA	1415	1415	80%	0.0	83.08%	NM_001275.4
☒	Rattus norvegicus chromogranin A (Chga). mRNA	278	278	20%	4e-72	80.34%	NM_001275.4
☒	Mus musculus chromogranin A (Chga). mRNA	176	176	11%	2e-41	81.61%	NM_001275.4



1. Descrições

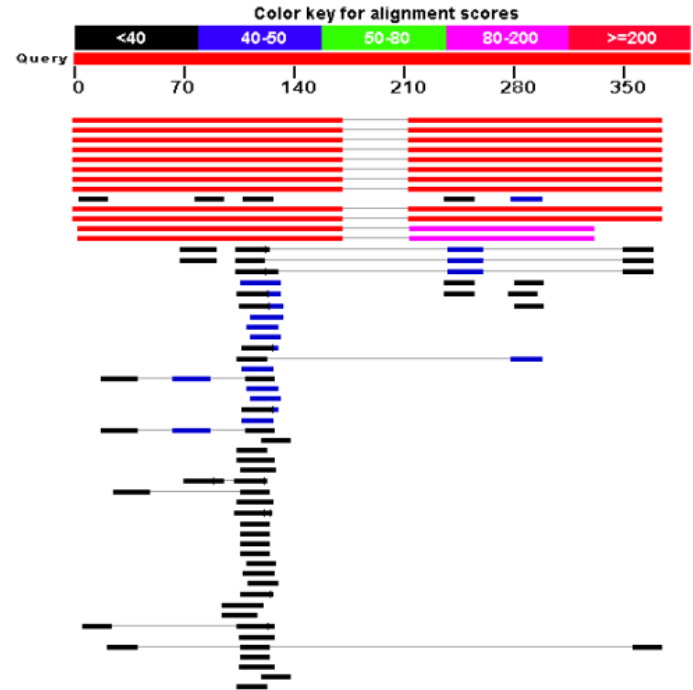
uma tabela com uma descrição de uma linha de cada acerto com algumas estatísticas de alinhamento.

RESULTADO DO BLAST



2. Gráfico Resumo

cada resultado é representado por uma linha que mostra qual a parte da sequência alinha com a sequência submetida à base de dados, através de linhas coloridas de acordo com a pontuação de alinhamento.



RESULTADO DO BLAST



3. Alinhamentos

os alinhamentos obtidos através da consulta e os resultados da base de dados

Staphylococcus aureus subsp. aureus RN4220 plasmid pRM27, complete sequence

Sequence ID: [KT780704.1](#) Length: 64127 Number of Matches: 1

Range 1: 22081 to 22559 [GenBank](#) [Graphics](#)

[Next Match](#) [Previous Match](#)

Score	Expect	Identities	Gaps	Strand	
856 bits(463)	0.0	477/483(99%)	4/483(0%)	Plus/Plus	
Query	1	TGAGCGAATTGATTAAATAAGCTAATAGATTATACATTAGAAAAATGAAAGGGGA	60		
Sbjct	22081	TGAGCGAATTGATTAAATAAGCTAATAGATTATACATTAGAAAAATGAAAGGGGA	22136		
CDS: Putative 1	1	V R M L Q S I P A L P V G D I K K			
Query	61	TTTTATTCGTGAATGTTACAGTCTATCCCGGCATTGCCAGTCGGGGATATTA AAAAGA	120		
Sbjct	22137	TTTTATTCGTGAATGTTACAGTCTATCCCGGCATTGCCAGTCGGGGATATTA AAAAGA	22196		
CDS: B1e0 (plasmid) [	1	M L Q S I P A L P V G D I K K			
CDS: Putative 1	18	S I G F Y C D K L G F T L V H H E D G F			
Query	121	GTATAGGTTTTATTGCGATAAACTAGGTTTCACTTTGGTTCAACATGAAGATGGATTCTG	180		
Sbjct	22197	GTATAGGTTTTATTGCGATAAACTAGGTTTCACTTTGGTTCAACATGAAGATGGATTCTG	22256		
CDS: B1e0 (plasmid) [	18	S I G F Y C D K L G F T L V H H E D G F			

