

Universidade de São Paulo - USP
Escola Superior de Agricultura Luiz de Queiroz - ESALQ
Banco de Dados e Ferramentas Computacionais Aplicados à
Genômica

Atividade 3

Wellingson assunção Araújo

- 1 - Ler e interpretar o artigo abaixo sobre a caracterização da proteína codificada pelo gene Brevis Radix.
- 2 - Utilizando as ferramentas disponíveis no NCBI reavaliar o alinhamento apresentado
- 3 - Apresentar de maneira didática a metodologia para chegar aos resultados.
 - a) Quanto a conservação,
 - b) Número de introns,
 - c) Posicionamento no genoma completo de Arabidopsis thaliana.
- 3 - Reescrever o parágrafo com os resultados reavaliados.

Artigo: Characterization of the Plant-Specific BREVIS RADIX Gene Family Reveals Limited Genetic Redundancy Despite High Sequence Conservation¹
Autores: Georgette C. Briggs, Céline F. Mouchel, and Christian S. Hardtke

3 - Apresentar de maneira didática a metodologia para chegar aos resultados.:

• PASSO 1

- Para o alinhamento múltiplo das sequencias buscando a similaridade foi utilizado a ferramenta COBALT do NCBI
- COBALT = **Constraint-based Multiple Alignment Tool**
- O COBALT é uma ferramenta de alinhamento de múltiplas seqüências que encontra uma coleção de restrições em pares derivadas do banco de dados de domínio conservado, banco de dados de motivos de proteínas e similaridade de sequência, usando RPS-BLAST, BLASTP e PHI-BLAST. Restrições aos pares são então incorporadas em um alinhamento múltiplo progressivo.
- **LINK DE ACESSO:**
<https://www.ncbi.nlm.nih.gov/tools/cobalt/cobalt.cgi?CMD=Web>

Interface do COBALT

The screenshot shows the COBALT web interface. At the top, there's a navigation bar with 'NIH' and 'NCBI' logos, and links for 'Home', 'Recent Results', and 'Help'. Below this, the title 'COBALT Constraint-based Multiple Alignment Tool' is displayed. The main section is titled 'Enter Query Sequences' and contains a large text input field for 'Enter at least 2 protein accessions, gis, or FASTA sequences'. Below this field is a button labeled 'Align'. To the right of the 'Align' button is a checkbox labeled 'Show results in a new window'. Below the 'Align' button is a link for 'Advanced parameters'. At the bottom of the page, there is a footer with copyright information and links for 'Copyright', 'Disclaimer', 'Privacy', 'Accessibility', 'Contact', and 'Send feedback'.

Detalhes Importantes

This annotated screenshot provides detailed instructions for using the COBALT interface. Red arrows point from text boxes to specific UI elements:

- A box on the right states: "Aqui devem ser colocadas as queries na forma fasta uma em seguida da outra" (Here, the queries must be placed in FASTA format, one after the other), with an arrow pointing to the main text input field.
- A box below it states: "É possível criar um documento com todas as sequencias (FASTA) e carrega-lo neste campo" (It is possible to create a document with all the sequences (FASTA) and load it into this field), with an arrow pointing to the 'Escolher arquivo' (Choose file) button.
- A box at the bottom right states: "Manter esse campo selecionada faz com o seu resultado seja mostrado em uma nova guia" (Keeping this field selected makes your result be shown in a new tab), with an arrow pointing to the 'Show results in a new window' checkbox.
- A box at the bottom left states: "Após as sequencias FASTA carregadas esse botão da origem ao trabalho" (After the FASTA sequences are loaded, this button returns you to the work area), with an arrow pointing to the 'Align' button.
- A box at the bottom center states: "Clicando nesse item é possível mudar as configurações padrões do alinhamento" (Clicking on this item it is possible to change the default alignment settings), with an arrow pointing to the 'Advanced parameters' link.

PASSO 2

Depois de escolher a ferramenta é preciso preparar as entradas. Isso se deu conforme as seguintes etapas:

- Aqui será mostrado como foi feita a busca de uma sequencias e as demais seguiram o mesmo processo

Table 1. BRXL genes from different plants

Table of BRXL genes identified in Arabidopsis (At), poplar (Pt), and rice (Os), with respective gene identification (ID) numbers and existence of ESTs indicated.

Gene Name	Gene ID	ESTs Exist?
AtBRX	At1g31880	Yes
AtBRXL1	At2g35600	Yes
AtBRXL2	At3g14000	Yes
AtBRXL3	Fuse At1g54180 + At1g54190	Yes
AtBRXL4	At5g20540	Yes
PtBRXL1	C-scaff-290001	Yes
PtBRXL2	LG_III000279	No
PtBRXL3	LGIII1601	Yes
PtBRXL4	LGV1000394	No
PtBRXL5	LGXVIII000005	No
OsBRXL1	Os08g36020	Yes
OsBRXL2	Os02g47230	Yes
OsBRXL3	Os04g51170	Yes
OsBRXL4	Os03g63650	Yes
OsBRXL5	Os12g09080	No

Etapa 1 – Como tínhamos em mãos o artigo podemos consultar os nomes dos genes e também os seus IDs

NCBI
National Center for
Biotechnology Information

NCBI Home

Resource List (A-Z)

All Resources

Chemicals & Bioassays

Data & Software

DNA & RNA

Domains & Structures

Genes & Expression

Genetics & Medicine

Genomes & Maps

Homology

Literature

Proteins

Sequence Analysis

Taxonomy

Gene

All Databases

Assembly

Biocollections

BioProject

BioSample

BioSystems

Books

ClinVar

Conserved Domains

dbGaP

dbVar

Gene

Genome

GEO DataSets

GEO Profiles

GTR

HomoloGene

Identical Protein Groups

MedGen

MeSH

At1g31880

Search

NCBI

enter for Biotechnology Information advances sci...

ical and genomic information.

| Mission | Organization | NCBI News & Blog

nit

to NCBI

Transfer t...

your com...

Bookshelf

PubMed Central

BLAST

leotide

nome

ne

Protein

PubChem

NCBI News & Blog

August-October 2019 RefSe...

annotations: mouse, firefly a

Develop

Use NCBI APIs and code

Analyze

Identify an NCBI tool for

Research

Explore NCBI research

Etapa 2 – Realizar a pesquisa com o ID do gene

Obs.: Para facilitar a busca seleciono GENE

AT2G21030 BREVIS RADIX-like protein [Arabidopsis thaliana (thale cress)]

Gene ID: 816637, updated on 22-Feb-2019

Summary

Gene symbol

AT2G21030

Gene description

BREVIS RADIX-like protein

Primary source

[Araport:AT2G21030](#)

Locus tag

AT2G21030

Gene type

protein coding

RNA name

BREVIS RADIX-like protein

RefSeq status

REVIEWED

Organism

[Arabidopsis thaliana \(ecotype: Columbia\)](#)

Lineage

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Gunneridae; Pentapetalae; rosids; malvids; Brassicales; Brassicaceae; Camelineae; Arabidopsis

Also known as

F26H11.21

Podemos ver aqui o nome do gene e também o organismo

Genomic context

Location:

chromosome: 2

See AT2G21030 in [Genome Data Viewer](#)

Exon count:

4

Sequence:

Chromosome: 2; NC_003071.7 (9024536..9025764, complement)

Descendo mais um pouco a pagina

mRNA and Protein(s)

1. [NM_127672.5](#) → [NP_179699.4](#) BREVIS RADIX-like protein [Arabidopsis thaliana]

[See identical proteins and their annotated locations for NP_179699.4](#)

Status: REVIEWED

UniProtKB/TrEMBL

[F4IFN0](#)

Conserved Domains (1)

[summary](#)

[pfam08381](#)

Location:13 – 55

BRX; Transcription factor regulating root

Clicando nesse link terenos as informações de proteína

Related sequences

Nucleotide	Protein
Heading	Accession and Version
genomic	CP002685.1 AEC07113.1

Additional links

BREVIS RADIX-like protein [Arabidopsis thaliana]

NCBI Reference Sequence: NP_179699.4

[Identical Proteins](#) [FASTA](#) [Graphics](#)

Go to:

LOCUS

NP_179699

158 aa

Linear

DEFINITION

BREVIS RADIX-like protein [Arabidopsis thaliana].

ACCESSION

NP_179699

VERSION

NP_179699.4

DBLINK

BioProject: [PRJNA116](#)

BioSample: [SAMN03081427](#)

DBSOURCE

REFSEQ: accession [NM_127672.5](#)

KEYWORDS

RefSeq.

SOURCE

Arabidopsis thaliana (thale cress)

ORGANISM

[Arabidopsis thaliana](#)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Gunneridae; Pentapetalae; rosids; malvids; Brassicales; Brassicaceae; Camelineae; Arabidopsis.

REFERENCE

1 (residues 1 to 158)

AUTHORS

Lin,X., Kaul,S., Rounsley,S., Shea,T.P., Benito,M.I., Town,C.D., Fujii,C.Y., Mason,T., Bowman,C.L., Barnstead,M., Feldblyum,T.V., Buell,C.R., Ketchum,K.A., Lee,J., Ronning,C.M., Koo,H.L., Moffat,K.S., Cronin,L.A., Shen,M., Pai,G., Van Aken,S., Umayam,L.,

Clicando aqui temos acesso ao formato FASTA

NCBI Resources How To

Protein

Protein

Advanced

FASTA

BREVIS RADIX-like protein [Arabidopsis thaliana]

NCBI Reference Sequence: NP_179699.4

[GenPept](#) [Identical Proteins](#) [Graphics](#)


>NP_179699.4 BREVIS RADIX-like protein [Arabidopsis thaliana]
MVKLTCTIILENSHITFASLPTGGNDLKRIKRFREMYNKWQAQRWGENYDKIVELYNVQRFNRQALQTPA
RSDDOQSQRDSTYSKMDSARESKDWTPRHFRPPGVSPPHHFYGGSSNYGPGSYHGGPPMDAARTTTSSRDD
PPSMNSASEMQAEWIEED

Uma segunda opção é copiar o número de referência da sequência, pois ela também pode ser usada pela ferramenta para alinhar as sequencias dos códigos informados

Agora é só copiar a sequência para montar o arquivo de entrada no COBALT

Após a preparação de todas as entradas preparadas podemos rodar as análises clicando em Align

 U.S. National Library of Medicine

 National Center for Biotechnology Information

Sign in to NCBI

COBALT

Constraint-based Multiple Alignment Tool

[Home](#) [Recent Results](#) [Help](#)

Enter Query Sequences

COBALT computes a multiple protein sequence alignment using conserved domain and local sequence similarity information. [Reset page](#)

Enter at least 2 protein accessions, gis, or FASTA sequences [Clear](#)

NP_179699.4

NP_565809.1

NP_174471.2

NP_188016.2

NP_974309.1

NP_001154422.1

Or, upload FASTA file

Escolher arquivo

Nenhum arquivo selecionado

Job Title

Align

☐ Show results in a new window

[Advanced parameters](#)

Aqui foi usado o codigo da proteina

BLAST is a registered trademark of the National Library of Medicine.
[Copyright](#) | [Disclaimer](#) | [Privacy](#) | [Accessibility](#) | [Contact](#) | [Send feedback](#)

[NCBI](#) | [NLM](#) | [NIH](#) | [DHHS](#)

NIH U.S. National Library of Medicine NCBI National Center for Biotechnology Information Sign in to NCBI

COBALT

Constraint-based Multiple Alignment Tool

Home Recent Results Help

COBALT computes a multiple protein sequence alignment using conserved domain and local sequence similarity information. [Reset page](#)

Enter Query Sequences

Enter at least 2 protein accessions, gis, or FASTA sequences [Clear](#)

```

XP_025878614.1
>p|084765|BRXL4_ORYS| Protein Brevis radix-like 4 OS=Oryza sativa subsp. japonica
OQ=39947 GN=BRXL4 PE=3 SV=1
MIQLKDMVMKLSGTSRHHGQORRGSPPPRGRTTSVYRSGYRPGMVDDMAVPPATYLG
GGGTSMSASSSTPAMDFAAPGAEGAREWVAQVEPGVQITFVSLAGGGNDLKRIRFSREM
YDKWQAQKWGWGNNERIMELVYVRRFSQVLPTPPRSSDGGERESFYSGVSGTRGSPAATP
CDABITFDDVLTGKACALVDRKACDNDNNECDRIICDGGCCCEIICDCEPAAWDDNDNDGACV
  
```

Or, upload FASTA file

Nenhum arquivo selecionado

Job Title

☐ Show results in a new window

[Advanced parameters](#)

Aqui a sequência FASTA

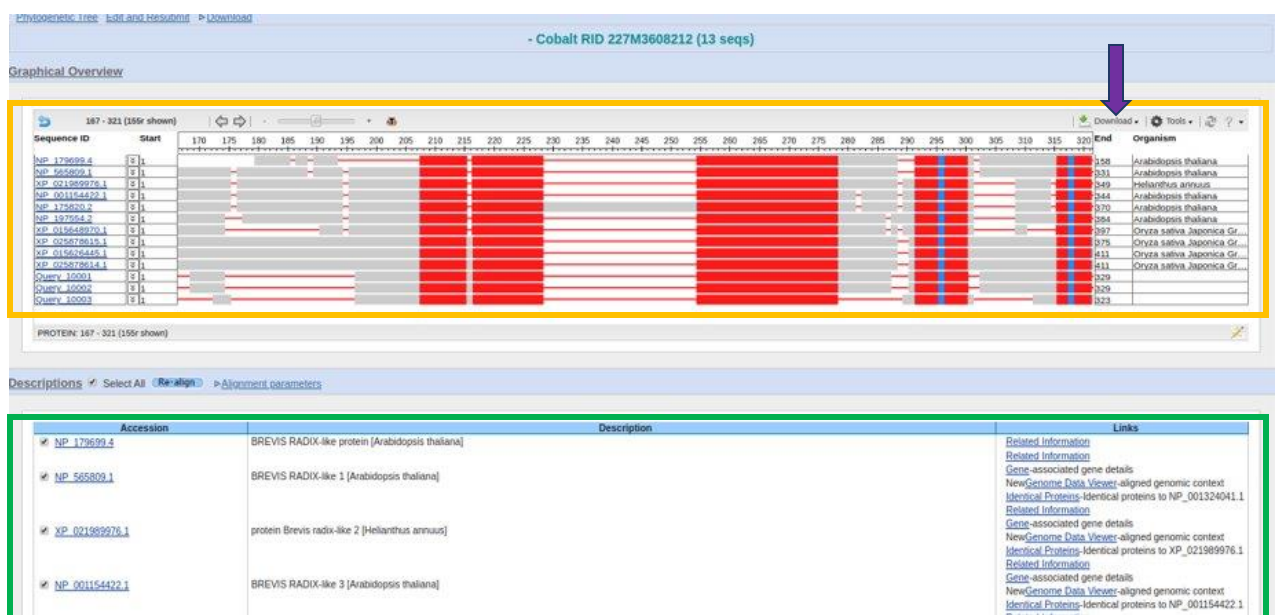
BLAST is a registered trademark of the National Library of Medicine.
 Copyright | Disclaimer | Privacy | Accessibility | Contact | Send feedback

NCBI | NLM | NIH | DHHS

Obs: Aqui foi usada a sequência fasta devido ela ter sido retirada de outro banco de dados pois as sequências dos genes *Oryza sativa* OsBRXL3, *Oryza sativa* OsBRXL4 e *Oryza sativa* OsBRXL5 não estavam contidas no NCBI

PASSO 4 – RESULTADOS

As saídas obtidas



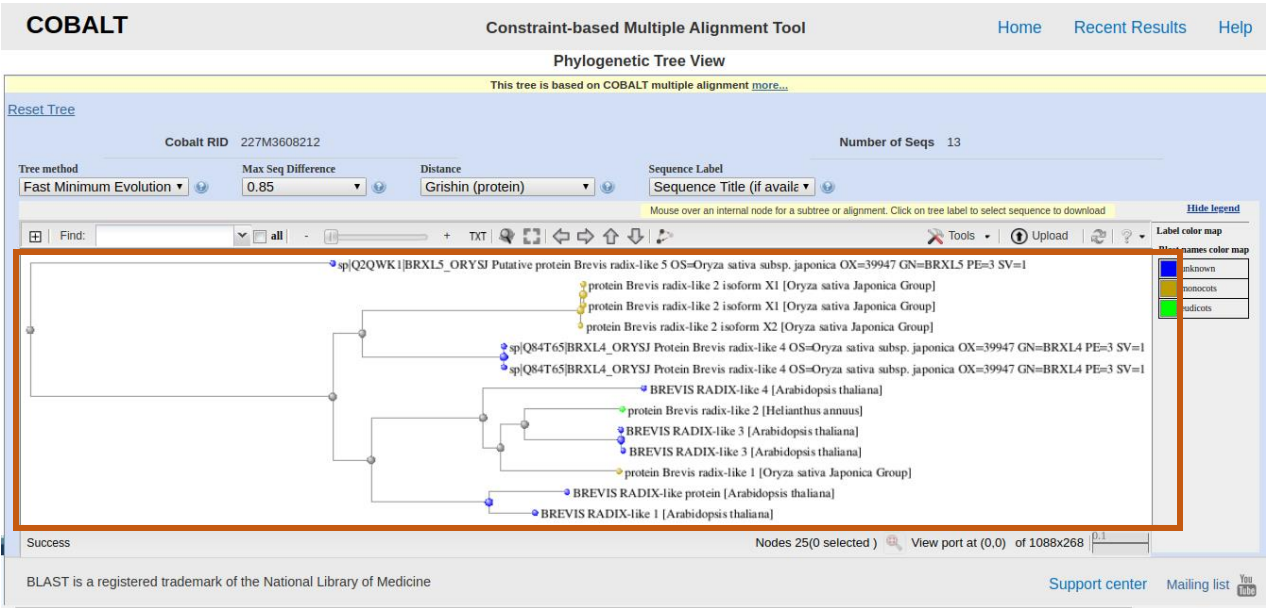
➡ É possível fazer o download do arquivo fasta do alinhamento mutiplo

☐ Mostra o alinhamento das sequencias

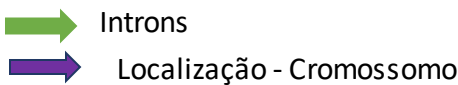
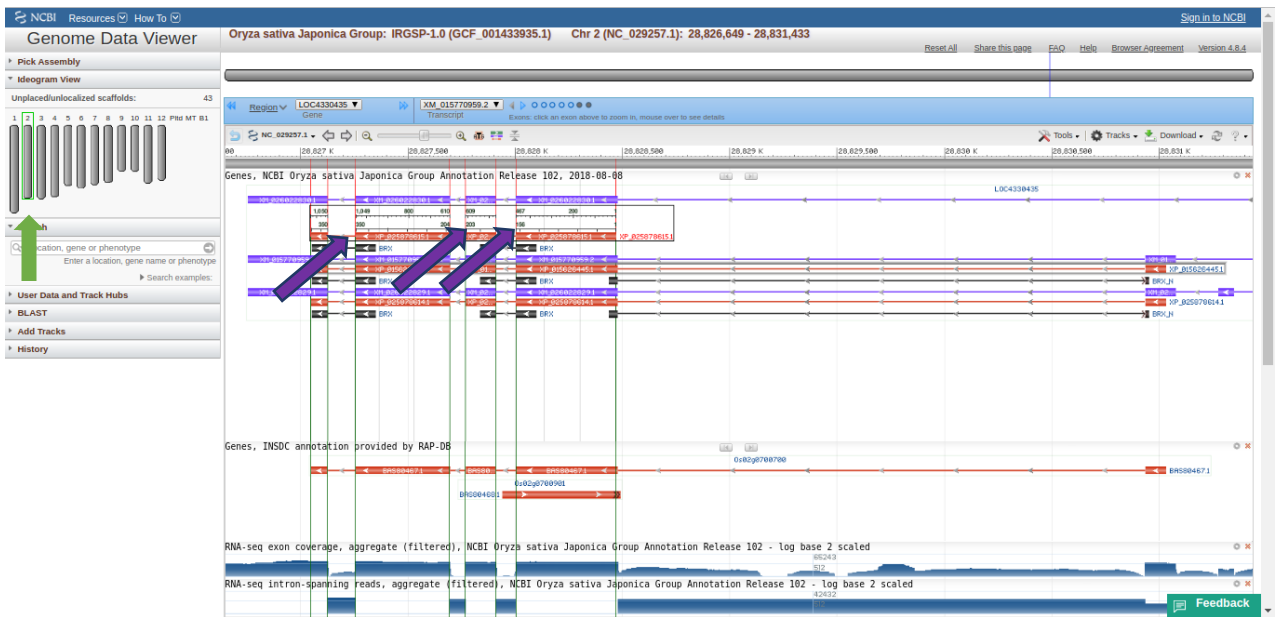
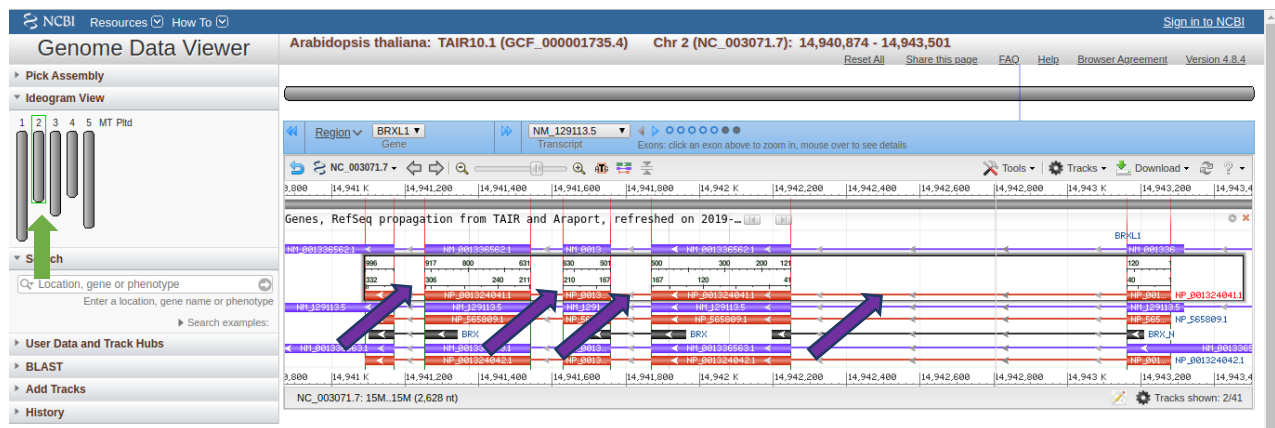
A região conservada das sequencias analisadas

☐ Mostra as principais informações das sequencias como ID, Nome e links relacionados a estas

NP_179699.4	13	-----HITFASLP-TGNDLKRIRFS	REYFKWGAQRMWGENYDKIVELYVQRFNRDAL---	QTFAASD	73
NP_565809.1	137	KENVAQVEPGLITFVSLP-SGGDLKRIRFS	REYFKWGAQRMWGENYDKIVELYVQRFNRDAL---	QTFAASE	208
XP_021989976.1	135	KENVAQVEPGLITFVSLP-QGGDLKRIRFS	REYFKWGAQRMWGENYDKIVELYVQRFNRDAL---	QVPAAPKAE	208
NP_001154422.1	140	KENVAQVEPGLITFVSLP-QGGDLKRIRFS	REYFKWGAQRMWGENYDKIVELYVQRFNRDAL---	QVPAAPKAE	212
NP_175829.2	140	KENVAQVEPGLITFVSLP-QGGDLKRIRFS	REYFKWGAQRMWGENYDKIVELYVQRFNRDAL---	QVPAAPKAE	238
NP_197554.2	150	KENVAQVEPGLITFVSLP-QGGDLKRIRFS	REYFKWGAQRMWGENYDKIVELYVQRFNRDAL---	QVPAAPKAE	223
XP_015648976.1	150	KENVAQVEPGLITFVSLP-QGGDLKRIRFS	REYFKWGAQRMWGENYDKIVELYVQRFNRDAL---	QVPAAPKAE	222
XP_025878615.1	125	RENTAQVEPGLITFVSLP-QGGDLKRIRFS	REYFKWGAQRMWGENYDKIVELYVQRFNRDAL---	QVPAAPKAE	198
XP_015648976.1	161	RENTAQVEPGLITFVSLP-QGGDLKRIRFS	REYFKWGAQRMWGENYDKIVELYVQRFNRDAL---	QVPAAPKAE	234
XP_025878614.1	161	RENTAQVEPGLITFVSLP-QGGDLKRIRFS	REYFKWGAQRMWGENYDKIVELYVQRFNRDAL---	QVPAAPKAE	234
Query_10001	86	RENTAQVEPGLITFVSLP-QGGDLKRIRFS	REYFKWGAQRMWGENYDKIVELYVQRFNRDAL---	QVPAAPKAE	158
Query_10002	86	RENTAQVEPGLITFVSLP-QGGDLKRIRFS	REYFKWGAQRMWGENYDKIVELYVQRFNRDAL---	QVPAAPKAE	158
Query_10003	109	RENTAQVEPGLITFVSLP-QGGDLKRIRFS	REYFKWGAQRMWGENYDKIVELYVQRFNRDAL---	QVPAAPKAE	173



A ferramenta Genome Data Wiewer indica o cromossomo que a sequência pertence e também possível fazer um estudo do número de introns



PASSO 6 – RESULTADOS

Tabela 1 – Resultados obtidos pela ferramenta Genome Data Viewer

ESPÉCIE / GENE	ID	INTRONS	POSIÇÃO
Arabidopsis thaliana AtBRX	AT1G31880,	4	Cromossomo 1
Arabidopsis thaliana AtBRXL1	AT2G35600	4	Cromossomo 2
Arabidopsis thaliana AtBRXL2	AT3G14000	4	Cromossomo 3
Arabidopsis thaliana AtBRXL3	At1g54190	4	Cromossomo 1
Arabidopsis thaliana AtBRXL4	At5g20540	4	Cromossomo 5
Oryza sativa OsBRXL1	Os08g36020	4	Cromossomo 8
Oryza sativa OsBRXL2	Os02g47230	4	Cromossomo 2

OBS: Como as sequências dos genes *Oryza sativa* OsBRXL3, *Oryza sativa* OsBRXL4 e *Oryza sativa* OsBRXL5 não estavam contidas no NCBI não foi possível usar a ferramenta Genome Data Wiewer para essas sequências.

PASSO 7 - Considerações finais

Diante dos resultados é possível inferir que família de genes BRX em *Arabidopsis thaliana* e *Oryza sativa* possuem regiões comuns bem conservadas e também se mostram agrupados diante da filogenética o que pode ser visto no dendrograma. Os diferentes genes são encontrados em diferentes cromossomos das duas espécies e também possuem número de introns entre 3 e 5. O gene OsBRXL5 (*Oryza sativa*) demonstrou uma maior dissimilaridade quando comparado aos demais o mesmo ficou alocado em um grupo diferente dos demais no dendrograma.