



XI Jornadas de Usuarios de R , Madrid 16 noviembre 2019

geneHummus : automatización computacional para el estudio de familias génicas

Jose V. Die Ramón
Dept Genética, ETSIAM, Universidad Córdoba

 @jdieramon



ABOUT ME



- 2012 - 2017 U.S. Department of Agriculture
- 2018 Scientific Visitor at NCBI
- 2017 - Dep. Genética, ETSIAM, UCO
- Usuario R ~ 2015
- Intereses : bioinformática , mejora vegetal , R , comunidad
-  @jdieramon



To develop new information technologies to aid in the understanding of fundamental molecular and genetic processes that control health and disease.

Fuente : <https://www.ncbi.nlm.nih.gov/home/about/mission/>

NCBI Resources How To

All Databases Search

jvdie My NCBI Sign Out

NCBI
National Center for Biotechnology Information

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

[About the NCBI](#) | [Mission](#) | [Organization](#) | [NCBI News & Blog](#)

Submit
Deposit data or manuscripts into NCBI databases


Download
Transfer NCBI data to your computer


Learn
Find help documents, attend a class or watch a tutorial


Develop
Use NCBI APIs and code libraries to build applications


Analyze
Identify an NCBI tool for your data analysis task


Research
Explore NCBI research and collaborative projects


Popular Resources

- PubMed
- Bookshelf
- PubMed Central
- BLAST
- Nucleotide
- Genome
- SNP
- Gene
- Protein
- PubChem

NCBI News & Blog

November 13 NCBI Minute: Resources for next-gen sequence analysis
06 Nov 2019
On Wednesday, November 13, 2019 at 12 PM, NCBI staff will present a webinar

NIH Biomedical Data Science Codeathon in Pittsburgh, Jan 8-10
30 Oct 2019
NCBI is pleased to announce a Biomedical Data Science Codeathon in

GenBank release 234 is available
28 Oct 2019
GenBank release 234.0 (10/14/2019) is now available on the NCBI FTP site. This release has 6.69 trillion bases and 1.68

[More...](#)

You are here: NCBI > National Center for Biotechnology Information

Support Center

GETTING STARTED

- NCBI Education
- NCBI Help Manual
- NCBI Handbook
- Training & Tutorials
- Submit Data

RESOURCES

- Chemicals & Bioassays
- Data & Software
- DNA & RNA
- Domains & Structures
- Genes & Expression
- Genetics & Medicine
- Genomes & Maps
- Homology
- Literature
- Proteins
- Sequence Analysis
- Taxonomy
- Variation

POPULAR

- PubMed
- Bookshelf
- PubMed Central
- BLAST
- Nucleotide
- Genome
- SNP
- Gene
- Protein
- PubChem

FEATURED

- Genetic Testing Registry
- GenBank
- Reference Sequences
- Gene Expression Omnibus
- Genome Data Viewer
- Human Genome
- Mouse Genome
- Influenza Virus
- Primer-BLAST
- Sequence Read Archive

NCBI INFORMATION

- About NCBI
- Research at NCBI
- NCBI News & Blog
- NCBI FTP Site
- NCBI on Facebook
- NCBI on Twitter
- NCBI on YouTube
- Privacy Policy

National Center for Biotechnology Information, U.S. National Library of Medicine
8600 Rockville Pike, Bethesda MD, 20894 USA
[Policies and Guidelines](#) | [Contact](#)



20 years ago, Prof R. Serrano ([@UPV](#)) taught us how to run a BLAST. Now, I tweet from this building #NCBI.



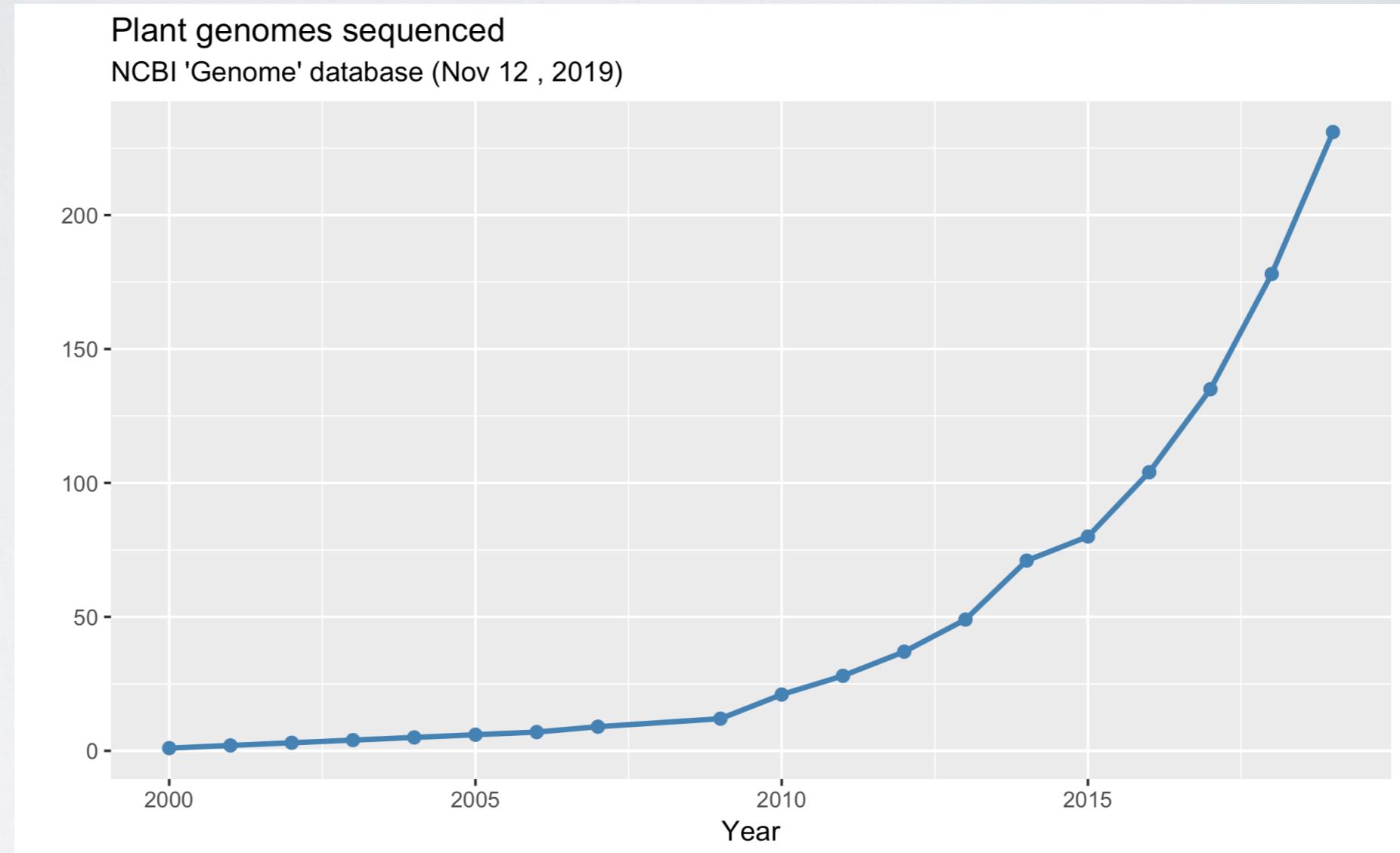
3:47 PM - 2 May 2018

Identificación de familias génicas a partir de proyectos secuenciación

- Estructura
- Evolución
- Función

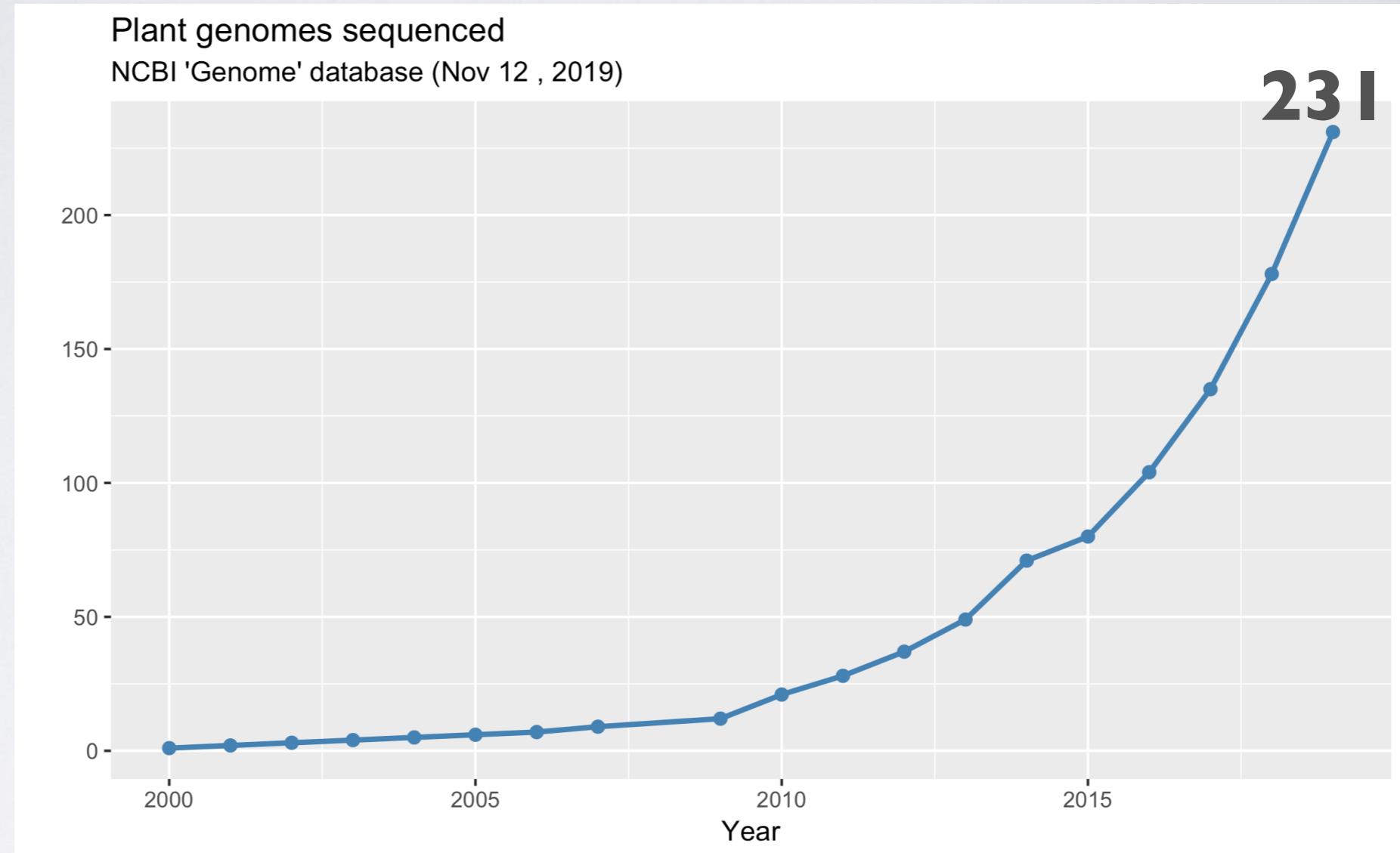
Identificación de familias génicas a partir de proyectos secuenciación

- Estructura
- Evolución
- Función



Identificación de familias génicas a partir de proyectos secuenciación

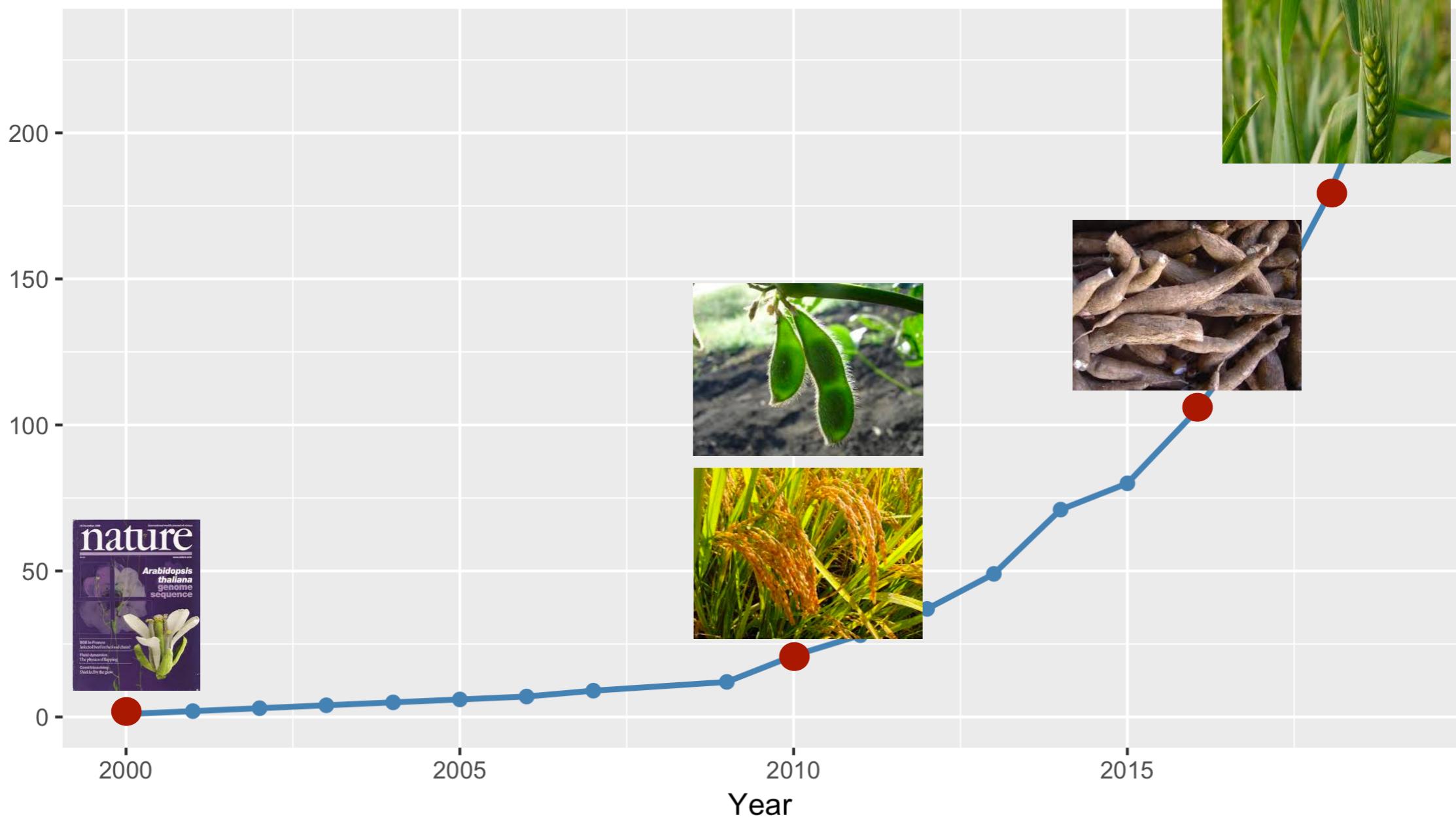
- Estructura
- Evolución
- Función



Identificación de familias génicas a partir de proyectos secuenciación

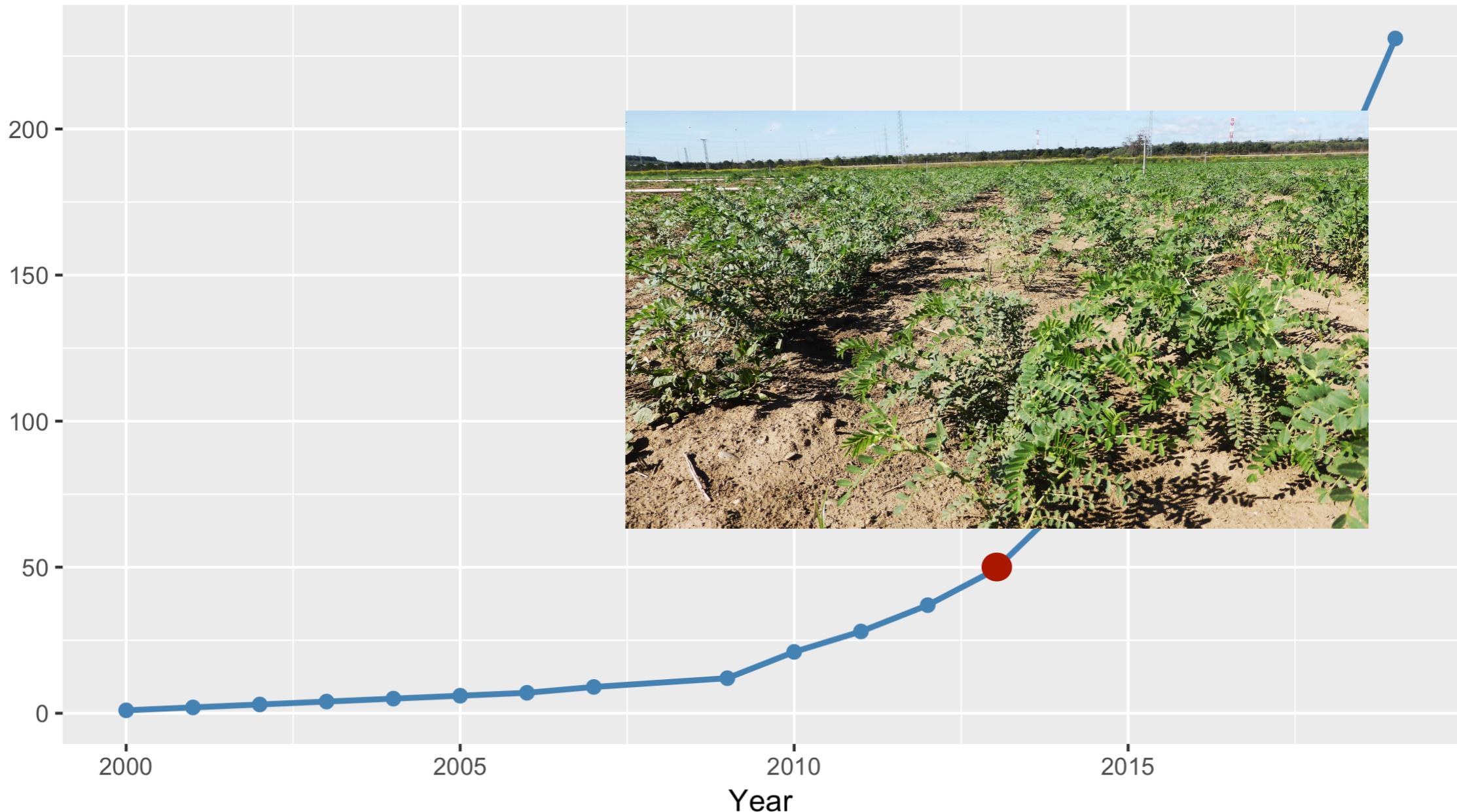
Plant genomes sequenced

NCBI 'Genome' database (Nov 12 , 2019)

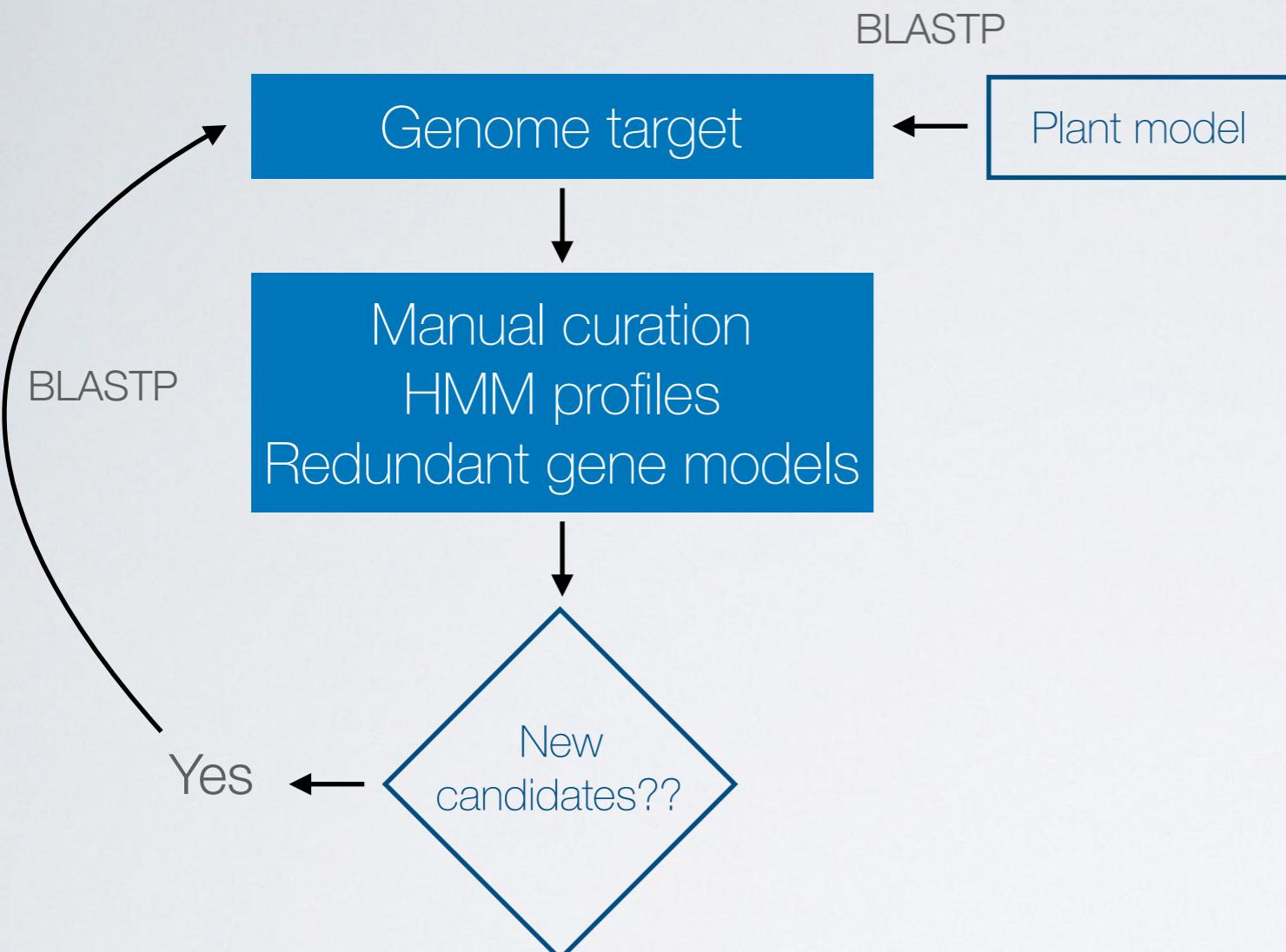


Plant genomes sequenced

NCBI 'Genome' database (Nov 12 , 2019)

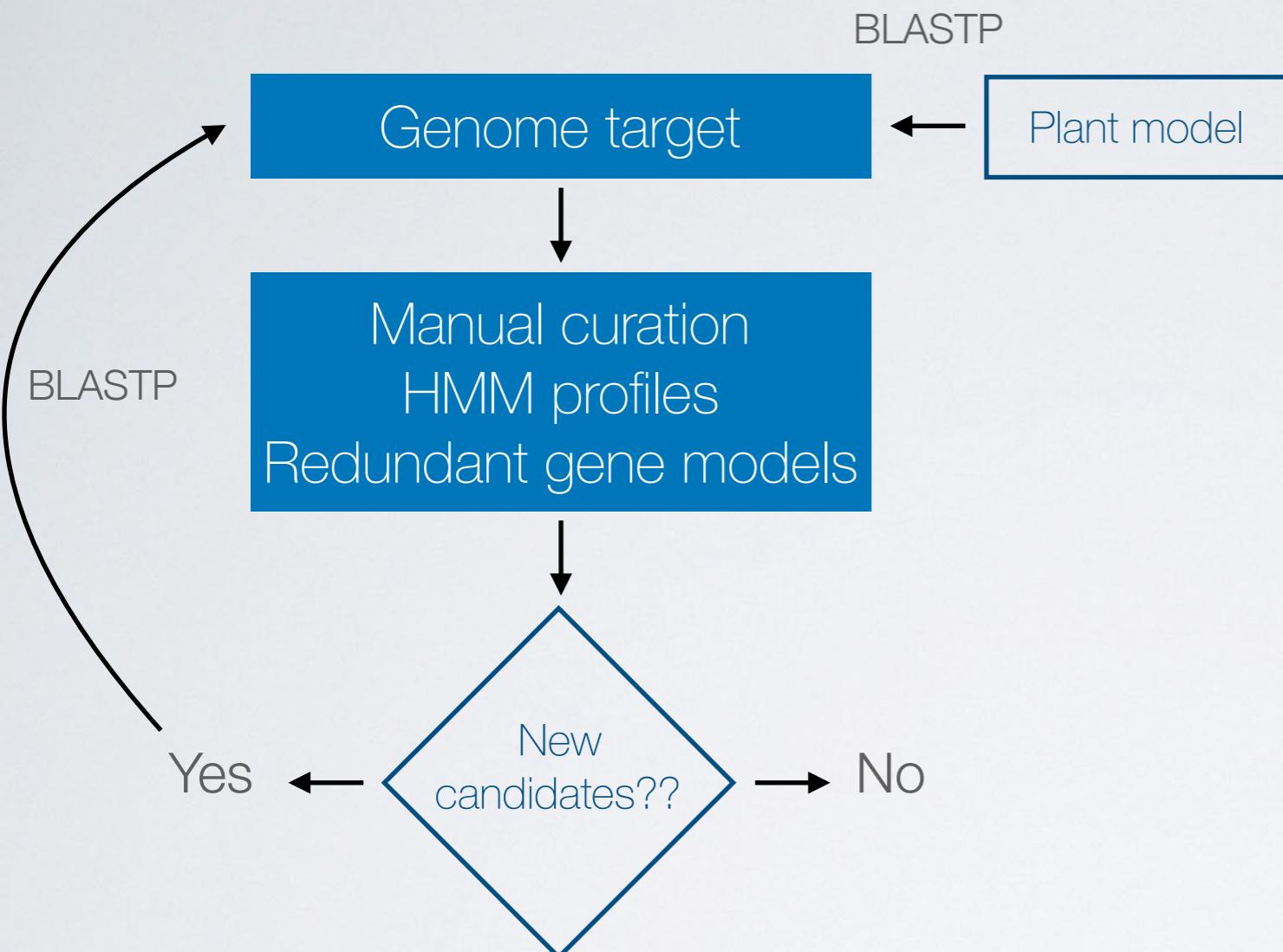


Aproximación para el estudio de familia génicas



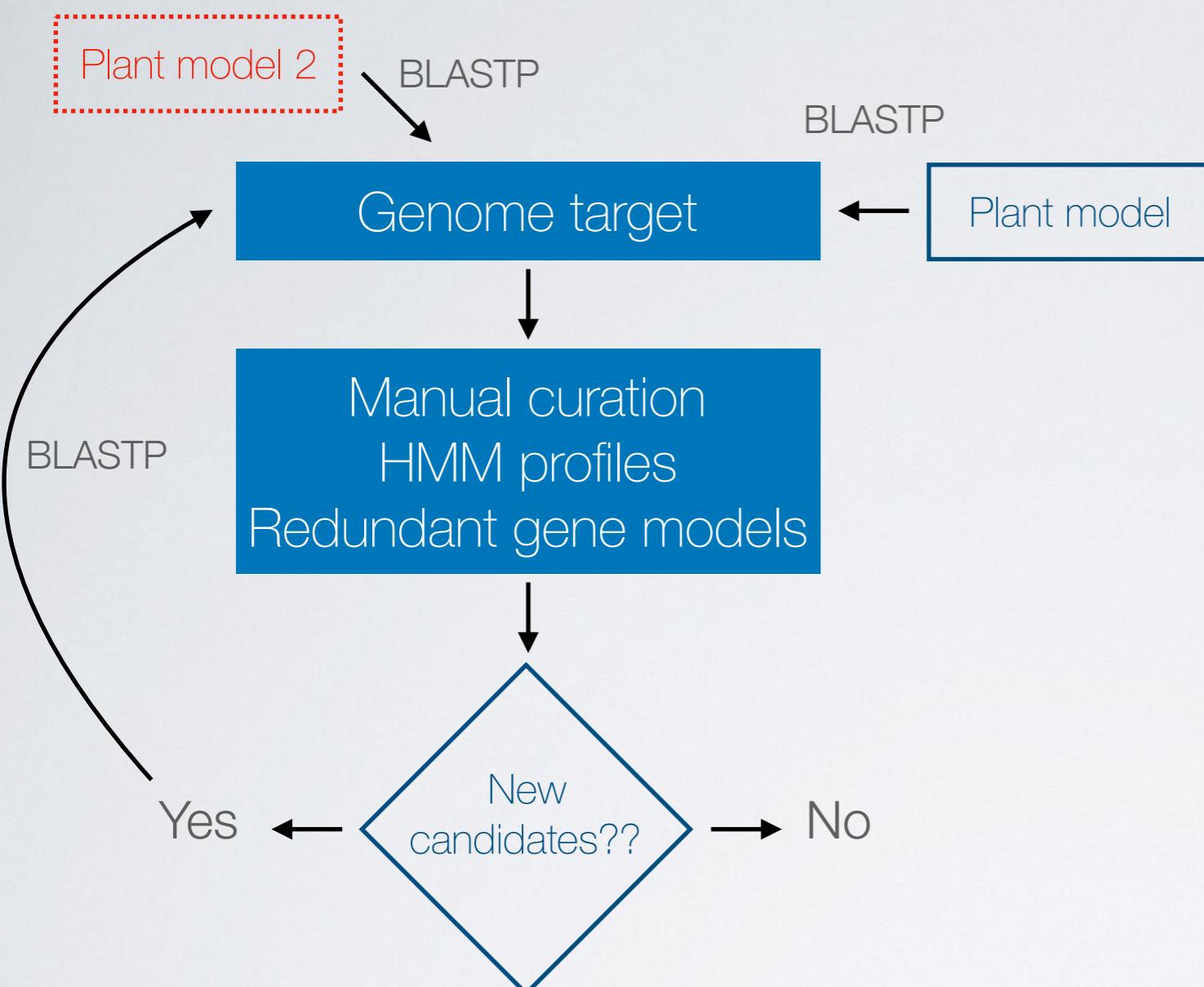
Esquema para búsqueda exhaustiva BLAST

Aproximación para el estudio de familia génicas



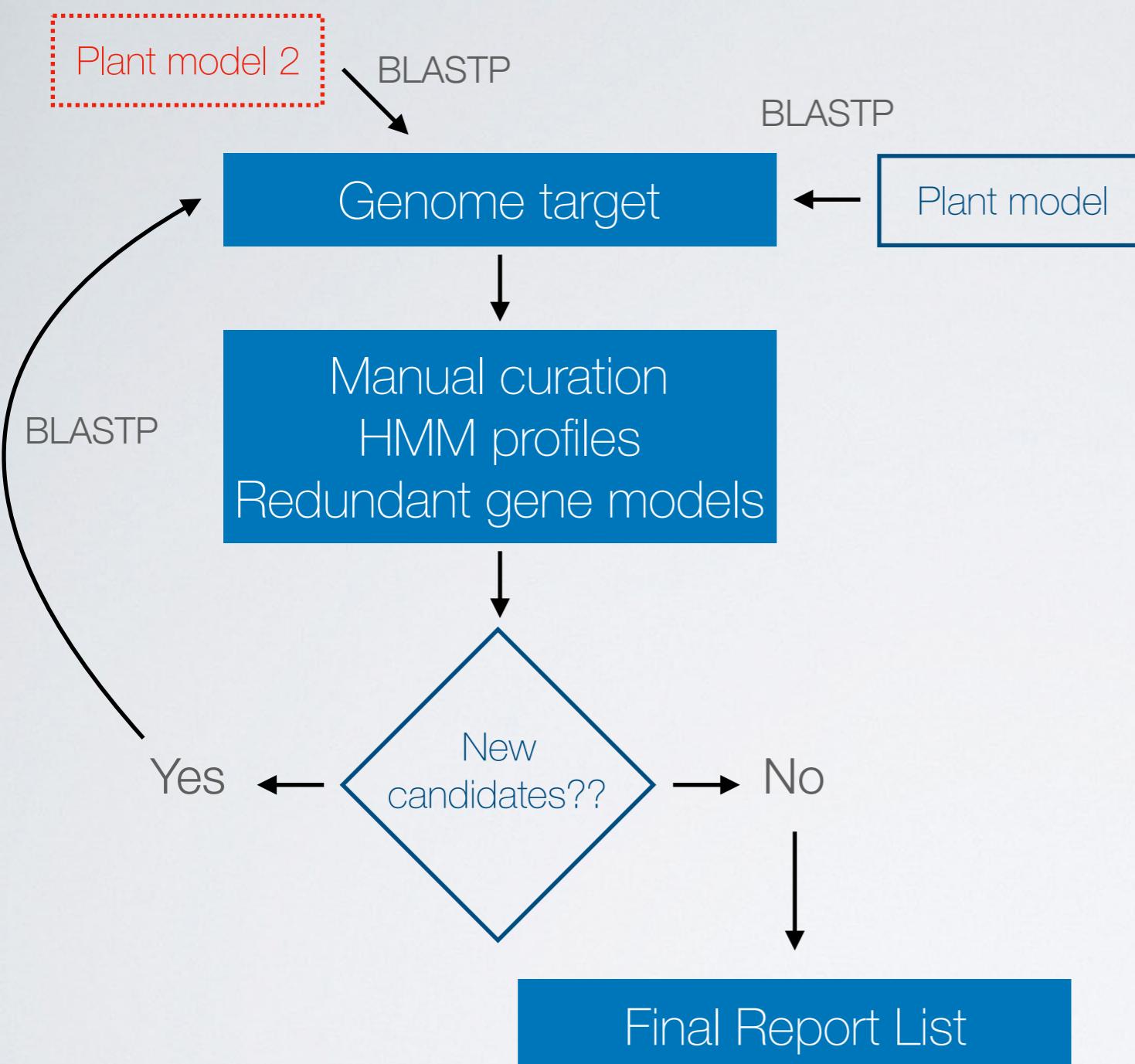
Esquema para búsqueda exhaustiva BLAST

Aproximación para el estudio de familia génicas



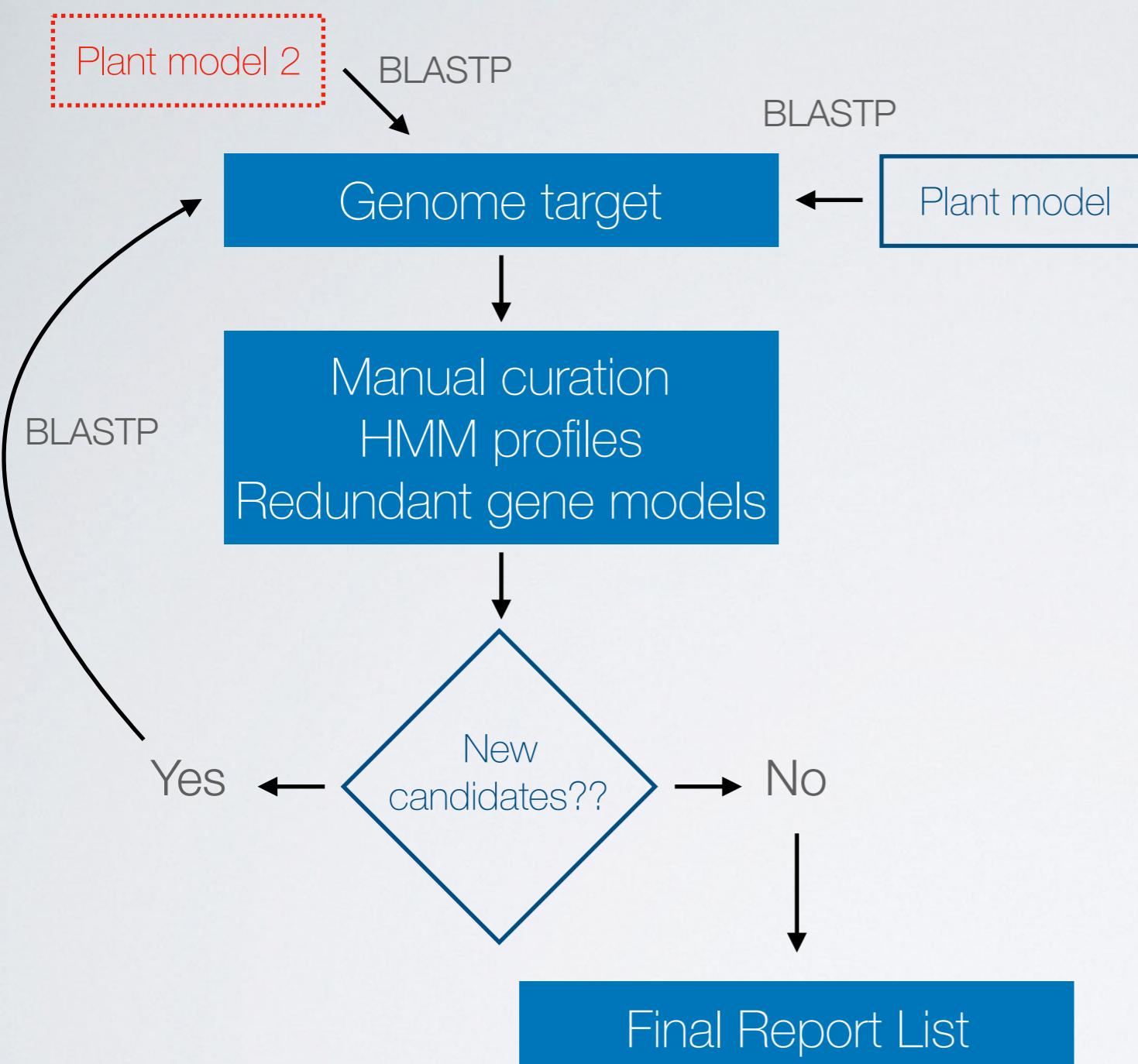
Esquema para búsqueda exhaustiva BLAST

Aproximación para el estudio de familia génicas



Esquema para búsqueda exhaustiva BLAST

Aproximación para el estudio de familia génicas



Die et al. BMC Genomics (2018) 19:301
<https://doi.org/10.1186/s12864-018-4695-9> BMC Genomics

RESEARCH ARTICLE Open Access

Genome-wide identification of the auxin response factor gene family in *Cicer arietinum*

Jose V. Die , Juan Gil and Teresa Millan

Abstract

Background: Auxin Response Factors act as critical components of the auxin-signaling pathway by regulating the transcription of auxin-responsive genes. The release of the chickpea reference genome provides an opportunity to identify and characterize the ARF gene family in this important legume by a data mining coupled by comparative genomics approaches.

Results: We performed a comprehensive characterization and analysis of 24 ARF genes in the chickpea reference genome. Comparative phylogenetic analysis of the ARF from chickpea, *Medicago* and *Arabidopsis* suggests that recent duplications have played a very limited role in the expansion of the ARF chickpea family. Gene structure analysis based on exon-intron organization provides additional evidence to support the evolutionary relationship among the ARF members. Conserved motif analysis shows that most of the proteins fit into the canonical ARF structure model, but 9 proteins lack or have a truncated dimerization domain. The mechanisms underlying the diversification of the ARF gene family are based on duplications, variations in domain organization and alternative splicing. Concerning duplications, segmental, but not tandem duplications, have contributed to the expansion of the gene family. Moreover, the duplicated pair genes have evolved mainly under the influence of purifying selection pressure with restricted functional divergence. Expression profiles responding to various environmental stimuli show a close relationship between tissue and expression patterns. Promoter sequence analysis reveals an enrichment of several *cis*-regulatory elements related to symbiosis, and modulation of plant gene expression during the interaction with microbes.

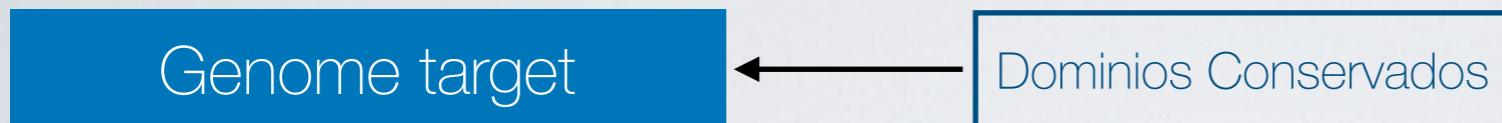
Conclusions: In conclusion, this study provides a comprehensive overview of the ARF gene family in chickpea. Globally, our data supports that auxin signaling pathway regulates a wide range of physiological processes and stress responses. Our findings could further provide new insights into the complexity of the regulation of ARF at the transcription level that may be useful to develop rational chickpea breeding strategies to improve development or stress responses. Our study also provides a foundation for comparative genomic analyses and a framework to trace the dynamic evolution of ARF genes on a large time-scale within the legume family.

Keywords: ARF, Bioinformatics, Chickpea, Gene duplication, Gene expression, RT-qPCR, Rstat

Esquema para búsqueda exhaustiva BLAST

Auxin Response Factor gene family
in the chickpea genome (2018)

Aproximación para el estudio de familia génicas



Die et al. *BMC Genomics* (2018) 19:301
<https://doi.org/10.1186/s12864-018-4695-9> BMC Genomics

RESEARCH ARTICLE Open Access

 CrossMark

Genome-wide identification of the auxin response factor gene family in *Cicer arietinum*

Jose V. Die , Juan Gil and Teresa Millan

Abstract

Background: Auxin Response Factors act as critical components of the auxin-signaling pathway by regulating the transcription of auxin-responsive genes. The release of the chickpea reference genome provides an opportunity to identify and characterize the ARF gene family in this important legume by a data mining coupled by comparative genomics approaches.

Results: We performed a comprehensive characterization and analysis of 24 ARF genes in the chickpea reference genome. Comparative phylogenetic analysis of the ARF from chickpea, *Medicago* and *Arabidopsis* suggests that recent duplications have played a very limited role in the expansion of the ARF chickpea family. Gene structure analysis based on exon-intron organization provides additional evidence to support the evolutionary relationship among the ARF members. Conserved motif analysis shows that most of the proteins fit into the canonical ARF structure model, but 9 proteins lack or have a truncated dimerization domain. The mechanisms underlying the diversification of the ARF gene family are based on duplications, variations in domain organization and alternative splicing. Concerning duplications, segmental, but not tandem duplications, have contributed to the expansion of the gene family. Moreover, the duplicated pair genes have evolved mainly under the influence of purifying selection pressure with restricted functional divergence. Expression profiles responding to various environmental stimuli show a close relationship between tissue and expression patterns. Promoter sequence analysis reveals an enrichment of several *cis*-regulatory elements related to symbiosis, and modulation of plant gene expression during the interaction with microbes.

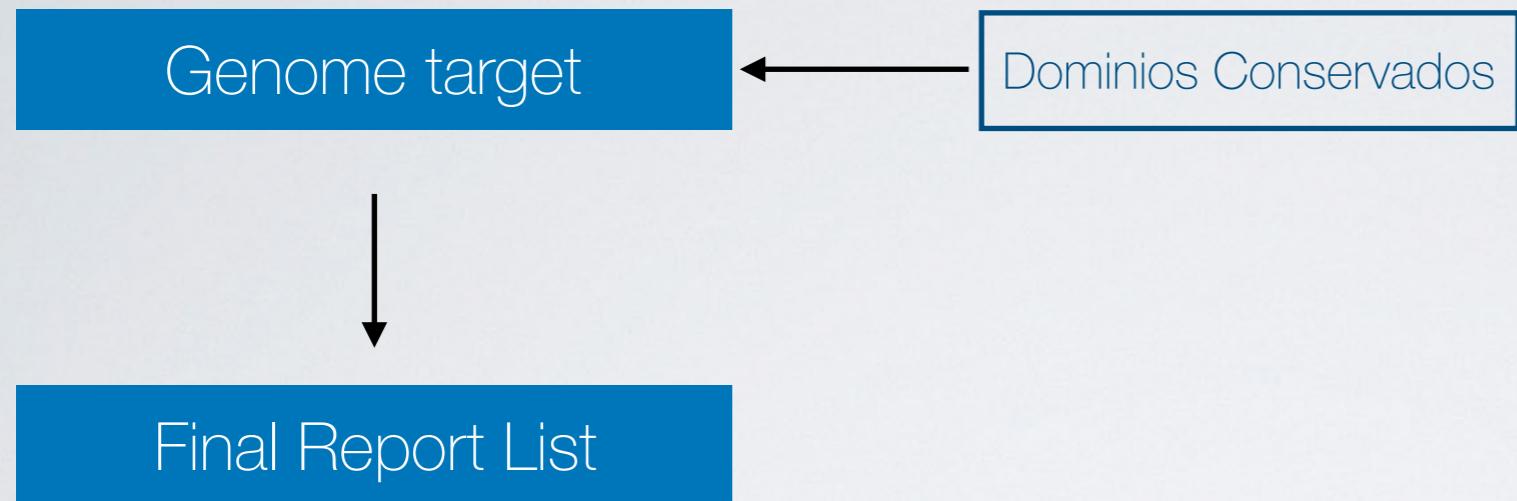
Conclusions: In conclusion, this study provides a comprehensive overview of the ARF gene family in chickpea. Globally, our data supports that auxin signaling pathway regulates a wide range of physiological processes and stress responses. Our findings could further provide new insights into the complexity of the regulation of ARF at the transcription level that may be useful to develop rational chickpea breeding strategies to improve development or stress responses. Our study also provides a foundation for comparative genomic analyses and a framework to trace the dynamic evolution of ARF genes on a large time-scale within the legume family.

Keywords: ARF, Bioinformatics, Chickpea, Gene duplication, Gene expression, RT-qPCR, Rstat

Esquema para búsqueda automática

Auxin Response Factor gene family
in the chickpea genome (2018)

Aproximación para el estudio de familia génicas



Die et al. *BMC Genomics* (2018) 19:301
<https://doi.org/10.1186/s12864-018-4695-9> BMC Genomics

RESEARCH ARTICLE Open Access

 CrossMark

Genome-wide identification of the auxin response factor gene family in *Cicer arietinum*

Jose V. Die , Juan Gil and Teresa Millan

Abstract

Background: Auxin Response Factors act as critical components of the auxin-signaling pathway by regulating the transcription of auxin-responsive genes. The release of the chickpea reference genome provides an opportunity to identify and characterize the ARF gene family in this important legume by a data mining coupled by comparative genomics approaches.

Results: We performed a comprehensive characterization and analysis of 24 ARF genes in the chickpea reference genome. Comparative phylogenetic analysis of the ARF from chickpea, *Medicago* and *Arabidopsis* suggests that recent duplications have played a very limited role in the expansion of the ARF chickpea family. Gene structure analysis based on exon-intron organization provides additional evidence to support the evolutionary relationship among the ARF members. Conserved motif analysis shows that most of the proteins fit into the canonical ARF structure model, but 9 proteins lack or have a truncated dimerization domain. The mechanisms underlying the diversification of the ARF gene family are based on duplications, variations in domain organization and alternative splicing. Concerning duplications, segmental, but not tandem duplications, have contributed to the expansion of the gene family. Moreover, the duplicated pair genes have evolved mainly under the influence of purifying selection pressure with restricted functional divergence. Expression profiles responding to various environmental stimuli show a close relationship between tissue and expression patterns. Promoter sequence analysis reveals an enrichment of several *cis*-regulatory elements related to symbiosis, and modulation of plant gene expression during the interaction with microbes.

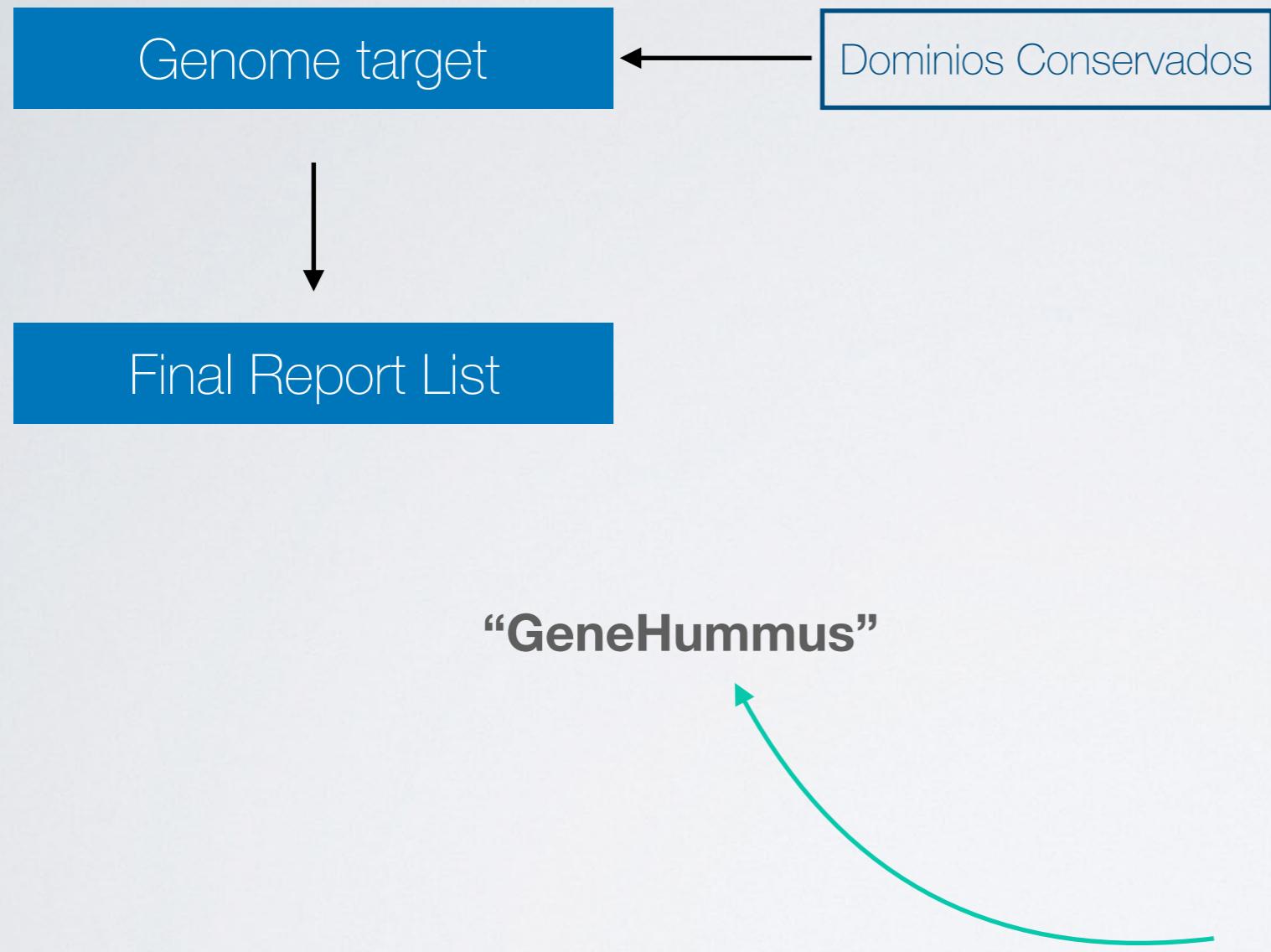
Conclusions: In conclusion, this study provides a comprehensive overview of the ARF gene family in chickpea. Globally, our data supports that auxin signaling pathway regulates a wide range of physiological processes and stress responses. Our findings could further provide new insights into the complexity of the regulation of ARF at the transcription level that may be useful to develop rational chickpea breeding strategies to improve development or stress responses. Our study also provides a foundation for comparative genomic analyses and a framework to trace the dynamic evolution of ARF genes on a large time-scale within the legume family.

Keywords: ARF, Bioinformatics, Chickpea, Gene duplication, Gene expression, RT-qPCR, Rstat

Esquema para búsqueda automática

Auxin Response Factor gene family
in the chickpea genome (2018)

Aproximación para el estudio de familia génicas



Esquema para búsqueda automática

Auxin Response Factor gene family
in the chickpea genome (2018)

Die et al. *BMC Genomics* (2018) 19:301
<https://doi.org/10.1186/s12864-018-4695-9>

BMC Genomics

RESEARCH ARTICLE **Open Access**

 CrossMark

Genome-wide identification of the auxin response factor gene family in *Cicer arietinum*

Jose V. Die , Juan Gil and Teresa Millan

Abstract

Background: Auxin Response Factors act as critical components of the auxin-signaling pathway by regulating the transcription of auxin-responsive genes. The release of the chickpea reference genome provides an opportunity to identify and characterize the ARF gene family in this important legume by a data mining coupled by comparative genomics approaches.

Results: We performed a comprehensive characterization and analysis of 24 ARF genes in the chickpea reference genome. Comparative phylogenetic analysis of the ARF from chickpea, *Medicago* and *Arabidopsis* suggests that recent duplications have played a very limited role in the expansion of the ARF chickpea family. Gene structure analysis based on exon-intron organization provides additional evidence to support the evolutionary relationship among the ARF members. Conserved motif analysis shows that most of the proteins fit into the canonical ARF structure model, but 9 proteins lack or have a truncated dimerization domain. The mechanisms underlying the diversification of the ARF gene family are based on duplications, variations in domain organization and alternative splicing. Concerning duplications, segmental, but not tandem duplications, have contributed to the expansion of the gene family. Moreover, the duplicated pair genes have evolved mainly under the influence of purifying selection pressure with restricted functional divergence. Expression profiles responding to various environmental stimuli show a close relationship between tissue and expression patterns. Promoter sequence analysis reveals an enrichment of several *cis*-regulatory elements related to symbiosis, and modulation of plant gene expression during the interaction with microbes.

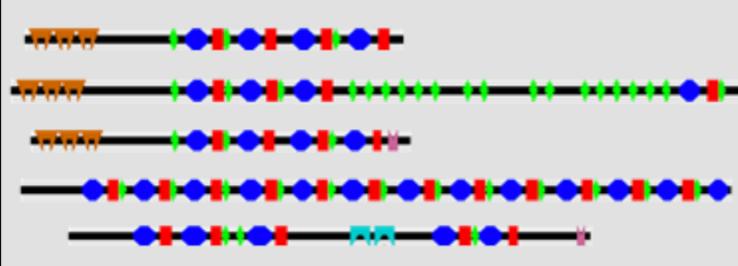
Conclusions: In conclusion, this study provides a comprehensive overview of the ARF gene family in chickpea. Globally, our data supports that auxin signaling pathway regulates a wide range of physiological processes and stress responses. Our findings could further provide new insights into the complexity of the regulation of ARF at the transcription level that may be useful to develop rational chickpea breeding strategies to improve development or stress responses. Our study also provides a foundation for comparative genomic analyses and a framework to trace the dynamic evolution of ARF genes on a large time-scale within the legume family.

Keywords: ARF, Bioinformatics, Chickpea, Gene duplication, Gene expression, RT-qPCR, Rstat

geneHummus : fundamento

NCBI Resources ▾ How To ▾ jvdie My NCBI Sign Out

SPARCLE Sparcle Advanced Search Help



SPARCLE

SPARCLE (Subfamily Protein Architecture Labeling Engine) is a resource for the functional characterization and labeling of protein sequences that have been grouped by their characteristic conserved domain architecture.

Using SPARCLE

- [About SPARCLE](#)
- [Input protein query sequence](#)
- [Help](#)
- [FTP](#)
- [News](#)
- [Publications](#)

Tools

- [Overview of CDD Resources](#)
- [Conserved Domain Database \(CDD\)](#)
- [CD-Search](#)
- [Batch CD-Search](#)
- [CDART \(Domain Architectures\)](#)
- [CDTree \(classification and research tool\)](#)
- [BLAST](#)

Other Resources

- [Structure Group Home Page](#)
- [Entrez Structure \(Molecular Modeling Database\)](#)
- [Entrez Gene](#)
- [Entrez Protein](#)
- [BioSystems](#)
- [FLink](#)



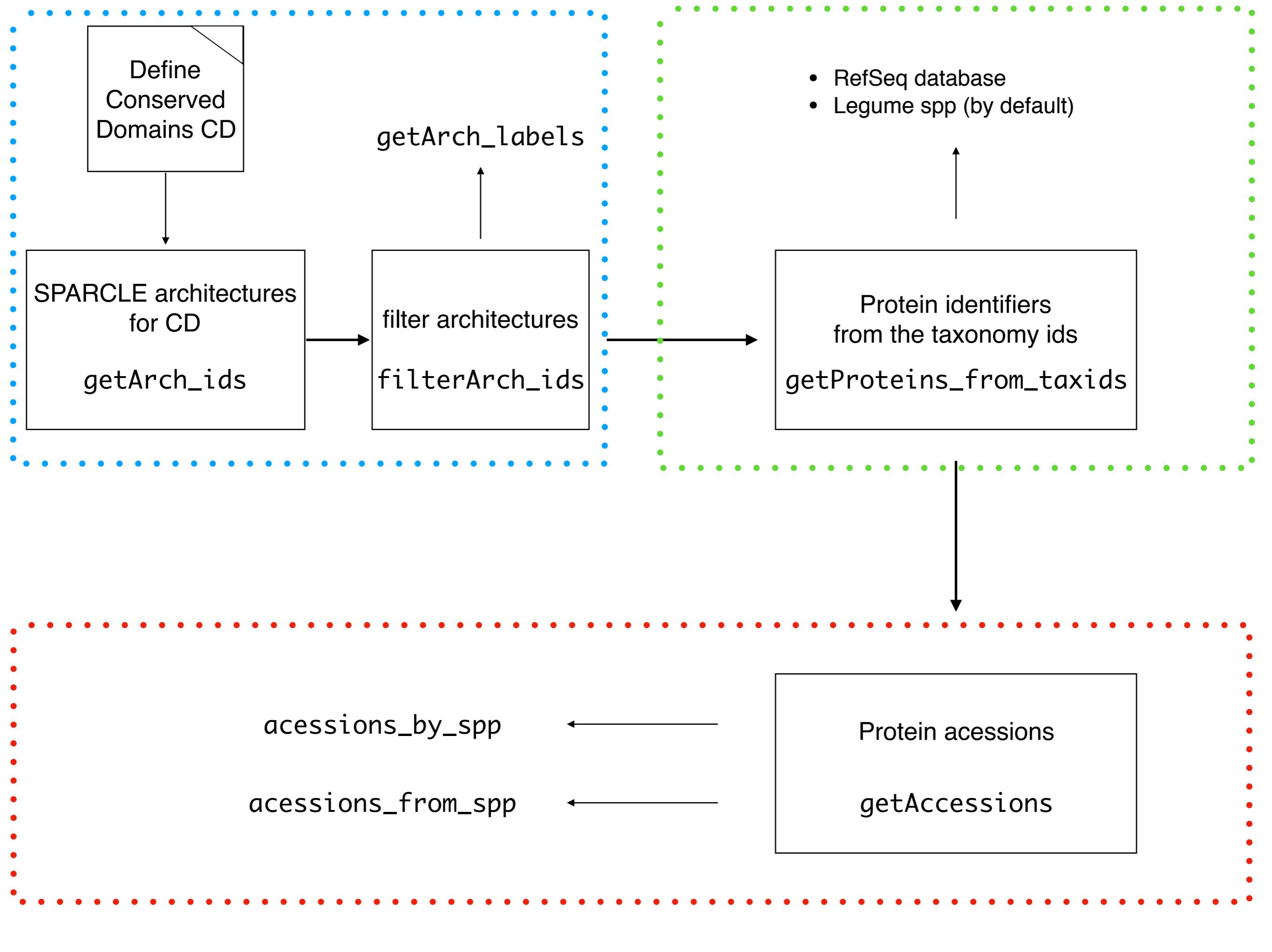
A. Marchler-Bauer



G. Y. Lewis

References:

 Marchler-Bauer A et al. (2017), "CDD/SPARCLE: functional classification of proteins via subfamily domain architectures.", **Nucleic Acids Res.** 45(D)200-3.



geneHummus : input

Auxin Response Factor: Estructura Genómica

[1]



[2]



[3]

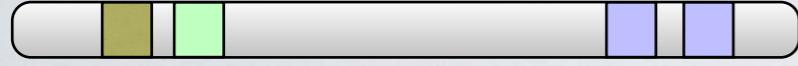


geneHummus : input

Auxin Response Factor: Estructura Genómica

```
arf <- c("pfam02362", "pfam06507", "pfam02309")
```

[1]



[2]



[3]



geneHummus en acción

Auxin Response Factor: Estructura Genómica

```
arf <- c("pfam02362", "pfam06507", "pfam02309")
```

[1]

[2]

[3]

```
# Obtener las arquitecturas  
# -----  
archids <- getArch_ids(arf)
```

getArch_ids.R

geneHummus en acción

Auxin Response Factor: Estructura Genómica

```
arf <- c("pfam02362", "pfam06507", "pfam02309")
```



```
# Obtener las arquitecturas
```

```
# -----  
archids <- getArch_ids(arf)
```

```
# Obtener los identificadores de proteínas
```

```
# -----  
arf_legumes = getProteins_from_tax_ids(archids, legumesIds)
```

getArch_ids.R

getProteins_from_tax_ids.R

```
> head(arf_legumes, 50)
[1] "1150156484" "1150156482" "950933327"  "593705262"  "1379669790"
[6] "357520645"  "1044541344" "357461711"  "1117466571" "1540557966"
[11] "1431657702" "828288607" "593699460"  "1117381062" "1540515334"
[16] "922356548"  "502082215"  "1021494095" "1431702815" "1379613255"
[21] "1540542505" "1021494093" "1044577571" "1431640353" "357443233"
[26] "1535895172" "1012179943" "922369540"  "1150138816" "1117565788"
[31] "1540505299"  "571493944"  "571498730"  "1021585233" "922369542"
[36] "1431696201"  "951074745"  "1117371114" "1117541684" "1117487898"
[41] "1431611783"  "502158399"  "1535875058" "502121960"  "1150090403"
[46] "1540492592"  "1117381058" "1117302354" "1044598473" "593268035"
>
```

geneHummus en acción

Auxin Response Factor: Estructura Genómica

```
arf <- c("pfam02362", "pfam06507", "pfam02309")
```



```
# Obtener las arquitecturas
```

```
# -----  
archids <- getArch_ids(arf)
```

```
# Obtener los identificadores de proteínas
```

```
# -----  
arf_legumes = getProteins_from_tax_ids(archids, legumesIds)
```

```
# Obtener los accessions de proteínas
```

```
# -----  
arf_accs = getAccessions(arf_legumes)
```

getArch_ids.R

getProteins_from_tax_ids.R

getAccession.R

```
> str(arf_accs)
'data.frame': 789 obs. of 2 variables:
 $ accession: chr "XP_016169706" "XP_003540131" "XP_015952369" "XP_025666965" ...
 $ organism : chr "Arachis ipaensis" "Glycine max" "Arachis duranensis" "Arachis h
ypogaea" ...
>
```

```
> head(arf_accs)
  accession      organism
1 XP_016169706 Arachis ipaensis
2 XP_003540131      Glycine max
3 XP_015952369 Arachis duranensis
4 XP_025666965     Arachis hypogaea
5 XP_020205510      Cajanus cajan
6 XP_020959846     Arachis ipaensis
>
```

geneHummus en acción

```
# Obtener las arquitecturas
# -----
archids <- getArch_ids(arf)

# Obtener los identificadores de proteínas
# -----
arf_legumes = getProteins_from_tax_ids(archids, legumesIds)

# Obtener los accessions de proteínas
# -----
arf_accs = getAccessions(arf_legumes)

# Resumir n.accesions / especie
accessions_by_spp(arf_accs)
```

getArch_ids.R

getProteins_from_tax_ids.R

getAccession.R

accessions_by_spp.R

```
> accessions_by_spp(arf_accs)
# A tibble: 12 x 2
  organism                N.seqs
  <chr>                  <int>
1 Abrus precatorius        94
2 Arachis duranensis       57
3 Arachis hypogaea      120
4 Arachis ipaensis        63
5 Cajanus cajan            56
6 Cicer arietinum         52
7 Glycine max              87
8 Lupinus angustifolius    92
9 Medicago truncatula      52
10 Phaseolus vulgaris      31
11 Vigna angularis          43
12 Vigna radiata var. radiata  42
> |
```

geneHummus en acción

```
# Obtener las arquitecturas
# -----
archids <- getArch_ids(arf)

# Obtener los identificadores de proteínas
# -----
arf_legumes = getProteins_from_tax_ids(archids, legumesIds)

# Obtener los accessions de proteínas
# -----
arf_accs = getAccessions(arf_legumes)

# Resumir n.accesions / especie
accessions_by_spp(arf_accs)

# Extraer accession de una especie concreta
accessions_from_spp(arf_accs, "Cicer arietinum")
```

getArch_ids.R

getProteins_from_tax_ids.R

getAccession.R

accessions_by_spp.R

accessions_from_spp.R

```
> head(accessions_from_spp(arf_accs, "Cicer arietinum"), 20)
[1] "XP_004506012" "XP_027191758" "XP_012573395" "XP_004505104"
[5] "XP_004485416" "XP_027190526" "XP_012569005" "XP_004488113"
[9] "XP_004505103" "XP_027191452" "XP_004497510" "XP_004508019"
[13] "XP_004485844" "XP_027191757" "XP_012570274" "XP_004487099"
[17] "XP_004505967" "XP_004510661" "XP_004504542" "XP_027191276"
>
```

GeneHummus

Visiting Bioinformaticians Program

29/8/2018

Find ARFs in your favorite legume species

GeneHummus is a pipeline for the identification, characterization and expression analysis of plant gene families. As a case study we focused on the auxin receptor factors (ARF) gene family in legumes. Here you'll see the output of GeneHummus for the ARF classification. If you like to use **GeneHummus**, we have written a tutorial in our GitHub repository '[NCBI-Hackathons/GeneHummus](#)'.

Enjoy your genehummus!

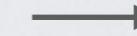
Auxin Response Factors (ARFs) in Legumes

Genomas leguminosas disponibles en NCBI

Nov. 2019

- *Abrus precatorius*
- *Arachis duranensis*
- *Arachis ipaensis*
- *Arachis hypogaea*
- *Cajanus cajan*
- *Cicer arietinum*
- *Glycine max*
- *Lupinus angustifolius*
- *Medicago truncatula*
- *Phaseolus vulgaris*
- *Vigna angularis*
- *Vigna radiata*

```
arf <- c("pfam02362", "pfam06507", "pfam02309")
```



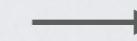
	> accessions_by_spp(arf_accs)	N.seqs
	# A tibble: 12 x 2	<int>
	organism <chr>	
1	<i>Abrus precatorius</i>	94
2	<i>Arachis duranensis</i>	57
3	<i>Arachis hypogaea</i>	120
4	<i>Arachis ipaensis</i>	63
5	<i>Cajanus cajan</i>	56
6	<i>Cicer arietinum</i>	52
7	<i>Glycine max</i>	87
8	<i>Lupinus angustifolius</i>	92
9	<i>Medicago truncatula</i>	52
10	<i>Phaseolus vulgaris</i>	31
11	<i>Vigna angularis</i>	43
12	<i>Vigna radiata</i> var. <i>radiata</i>	42

Genomas leguminosas disponibles en NCBI

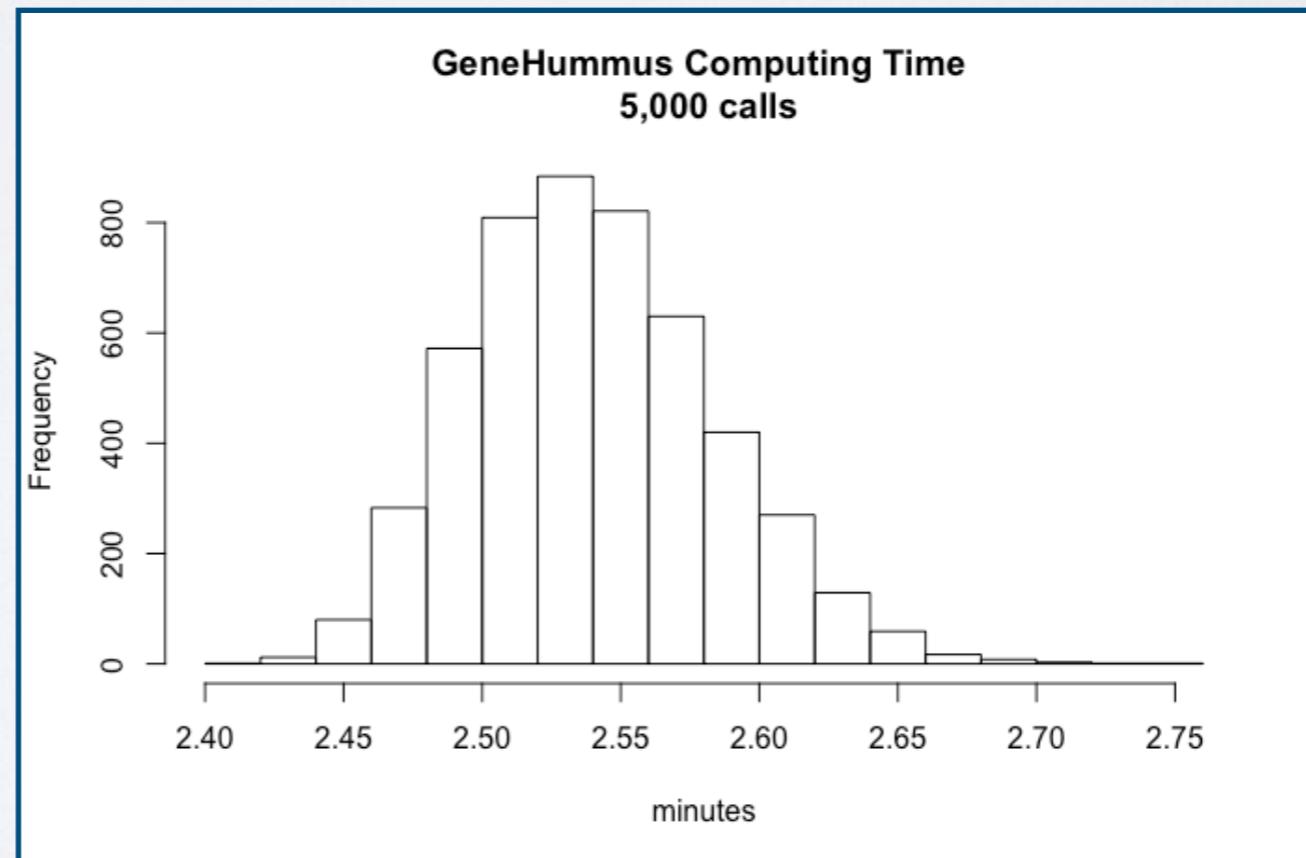
Nov. 2019

- *Abrus precatorius*
- *Arachis duranensis*
- *Arachis ipaënsis*
- *Arachis hypogaea*
- *Cajanus cajan*
- *Cicer arietinum*
- *Glycine max*
- *Lupinus angustifolius*
- *Medicago truncatula*
- *Phaseolus vulgaris*
- *Vigna angularis*
- *Vigna radiata*

```
arf <- c("pfam02362", "pfam06507", "pfam02309")
```



	> accessions_by_spp(arf_accs)	N.seqs
	# A tibble: 12 x 2	<int>
	organism <chr>	
1	<i>Abrus precatorius</i>	94
2	<i>Arachis duranensis</i>	57
3	<i>Arachis hypogaea</i>	120
4	<i>Arachis ipaënsis</i>	63
5	<i>Cajanus cajan</i>	56
6	<i>Cicer arietinum</i>	52
7	<i>Glycine max</i>	87
8	<i>Lupinus angustifolius</i>	92
9	<i>Medicago truncatula</i>	52
10	<i>Phaseolus vulgaris</i>	31
11	<i>Vigna angularis</i>	43
12	<i>Vigna radiata</i> var. <i>radiata</i>	42

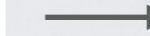


Genomas leguminosas disponibles en NCBI

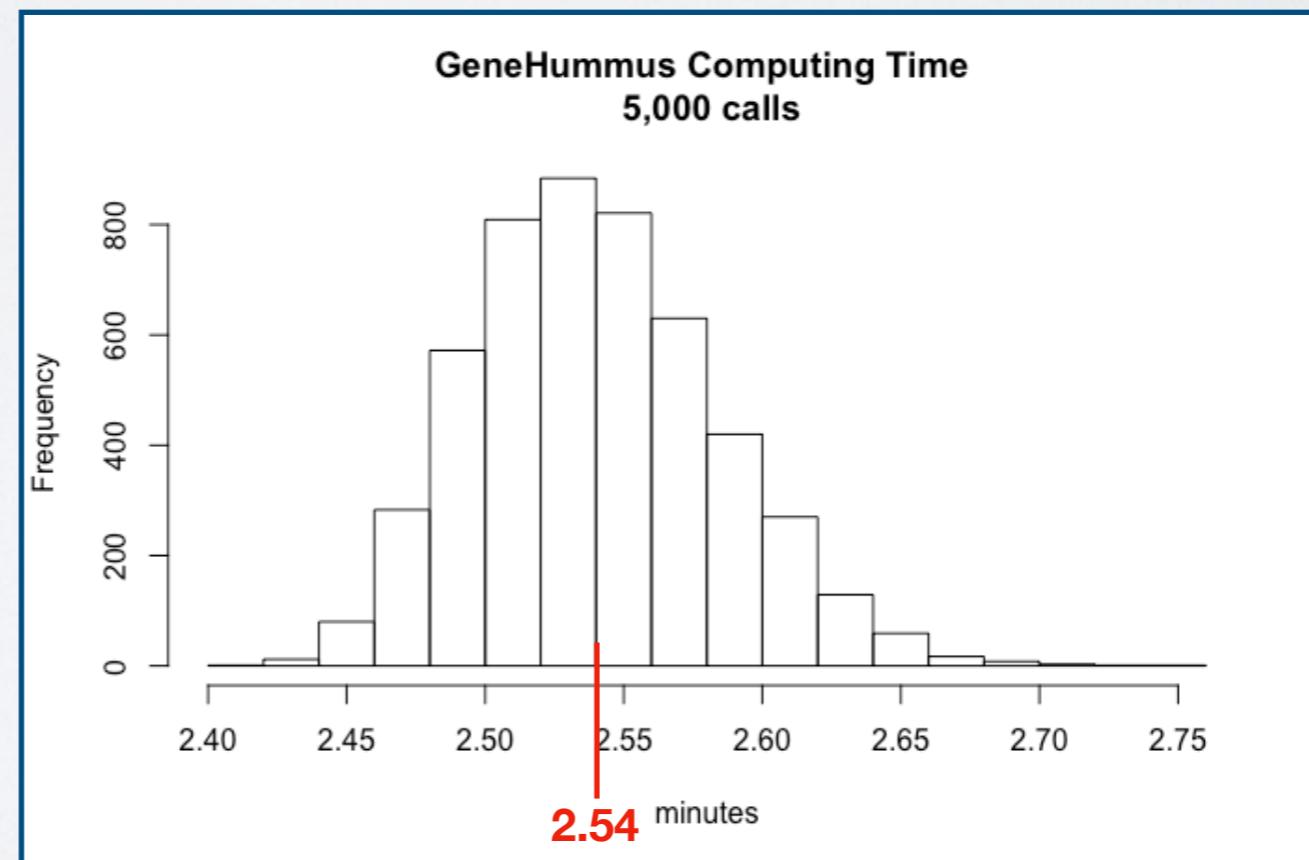
Nov. 2019

- *Abrus precatorius*
- *Arachis duranensis*
- *Arachis ipaensis*
- *Arachis hypogaea*
- *Cajanus cajan*
- *Cicer arietinum*
- *Glycine max*
- *Lupinus angustifolius*
- *Medicago truncatula*
- *Phaseolus vulgaris*
- *Vigna angularis*
- *Vigna radiata*

```
arf <- c("pfam02362", "pfam06507", "pfam02309")
```



	> accessions_by_spp(arf_accs)	N.seqs
	# A tibble: 12 x 2	<int>
	organism <chr>	
1	<i>Abrus precatorius</i>	94
2	<i>Arachis duranensis</i>	57
3	<i>Arachis hypogaea</i>	120
4	<i>Arachis ipaensis</i>	63
5	<i>Cajanus cajan</i>	56
6	<i>Cicer arietinum</i>	52
7	<i>Glycine max</i>	87
8	<i>Lupinus angustifolius</i>	92
9	<i>Medicago truncatula</i>	52
10	<i>Phaseolus vulgaris</i>	31
11	<i>Vigna angularis</i>	43
12	<i>Vigna radiata</i> var. <i>radiata</i>	42



<https://github.com/NCBI-Hackathons/GeneHummus>

NCBI-Hackathons / GeneHummus

Unwatch ▾ 6 Star 4 Fork 2

Code Issues 0 Pull requests 0 Projects 0 Wiki Security Insights Settings

An Automated Pipeline to Classify Gene Families based on Protein Domain Organization using Auxin Response Factors in Legumes as an Example Edit

Manage topics

184 commits 2 branches 2 releases 3 contributors View license

Branch: master New pull request Create new file Upload files Find file Clone or download ▾

jdieramon update publication BMC Latest commit 1635ffa on 20 Jul

R	wrap examples -- elapsed time	7 months ago
data	R pack setup	9 months ago
man	wrap examples -- elapsed time	7 months ago
vignettes	Update genHummus.md	6 months ago
DESCRIPTION	tryCatch warnings + check on platforms for CRAN	7 months ago
LICENSE	R pack setup	9 months ago
LICENSE.md	add license	9 months ago
NAMESPACE	tryCatch warnings + check on platforms for CRAN	7 months ago
README.md	update publication BMC	4 months ago

[Code](#)[Issues 0](#)[Pull requests 0](#)[Projects 0](#)[Wiki](#)[Security](#)[Insights](#)[Settings](#)

An Automated Pipeline to Classify Gene Families based on Protein Domain Organization using Auxin Response Factors in Legumes as an Example

[Edit](#)

Customizing GeneHummus

`geneHummus` can be customized to be suitable for other agronomically important taxonomic families beyond legumes. For that, we'll use the same functions as we did earlier and pass the corresponding taxonomic filter as argument. You can download from the NCBI other taxonomy ids and customize your search for your own species. When installing the `genehummus` package, you'll have access to several objects that contain the taxonomy ids for the families:

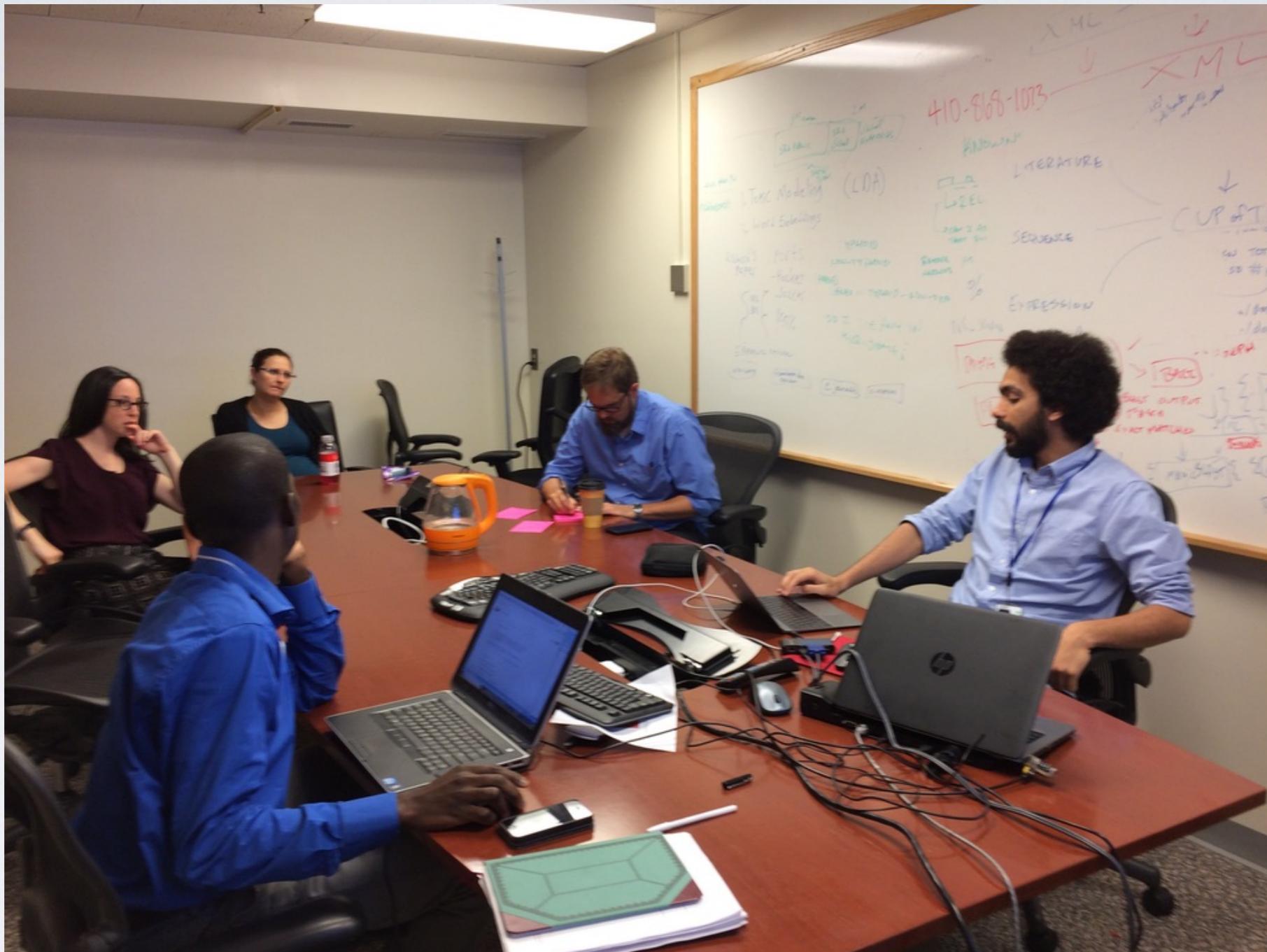
- Brassicaceae: `brassicaceaeIds`
- Cucurbitaceae: `cucurbitaceaeIds`
- Rosaceae: `rosaceaeIds`
- Solanaceae: `solanaceaeIds`

• <i>Brassica napus</i>	123	• <i>Cucurbita maxima</i>	71	• <i>Pyrus x bretschneideri</i>	51	• <i>Nicotiana tabacum</i>	103
• <i>Camelina sativa</i>	86	• <i>Cucurbita pepo</i>	71	• <i>Malus domestica</i>	48	• <i>Nicotiana tomentosiformis</i>	73
• <i>Raphanus sativus</i>	52	• <i>Cucurbita moschata</i>	67	• <i>Prunus avium</i>	33	• <i>Nicotiana sylvestris</i>	56
• <i>Brassica oleracea</i>	49	• <i>Cucumis sativus</i>	27	• <i>Rosa chinensis</i>	28	• <i>Nicotiana attenuata</i>	49
• <i>Brassica rapa</i>	48	• <i>Momordica charantia</i>	26	• <i>Prunus persica</i>	27	• <i>Capsicum annuum</i>	46
• <i>Arabidopsis thaliana</i>	44	• <i>Cucumis melo</i>	24	• <i>Prunus mume</i>	25	• <i>Solanum tuberosum</i>	43
• <i>Capsella rubella</i>	28			• <i>Fragaria vesca</i>	23	• <i>Solanum lycopersicum</i>	39
• <i>Eutrema salsugineum</i>	26					• <i>Solanum pennellii</i>	34
• <i>Arabidopsis lyrata</i>	24						



NCBI Visiting Bioinformaticians Group

Class 2018



NCBI 2018 Visiting Bioinformaticians Group



Moamen Elmassy

Department of Biological Sciences
Texas Tech University



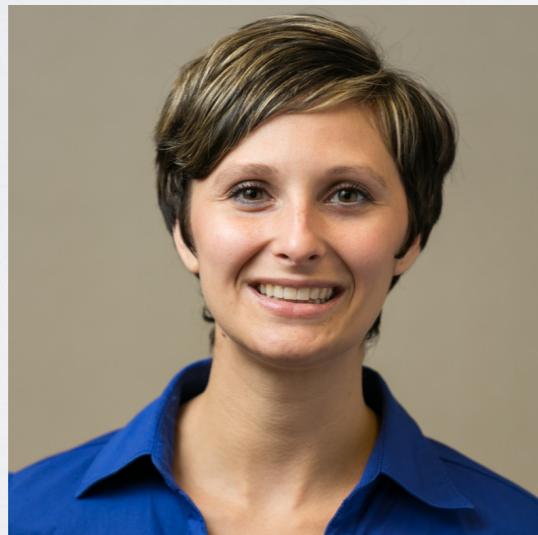
Kimberly LeBlanc

National Institute on Drug Abuse
NIH



Olaitan I. Awe

Department of Computer Science
University of Ibadan



Allissa Dillman

NCBI
US National Library of Medicine



Ben Busby

NCBI
US National Library of Medicine



XI Jornadas de Usuarios de R , Madrid 16 noviembre 2019

geneHummus : automatización computacional para el estudio de familias génicas

Jose V. Die Ramón
Dept Genética, ETSIAM, Universidad Córdoba

 @jdieramon

