

Interpretação de sequencias biológicas



Bruno Silvestre Lira
2023

Formato FASTA

>(nome/descrição)§
(Sequência)§

Evitar espaços e - no nome

Checar se não há caracteres na sequência (* = stop codon)

>Solyc01g088090.2.1

ATAAGAAAAAGCAATTCTAAAATCCCATATCACTTCTCCTAAGCTCTTGGAGATTTTGCTCATAATTATCTTAGAAAACTGATTAAAGGTATTCATTTT
AGATTTTAGAAAGCTGACTCTGGGATGGAATTTTGTTCTTTCTATTTCGACACCCTATTACTGCGTTGTAAATGTTAAACCGAATGTGATTCAAAAATG
CCTAATTTCAAATCAGACAATGTTTTCTAGACTAAAAGAAAGAAGATTGTGTTCTGGTTTAGATGTTTATAGCGTAGAGAAAAATAGGAAGCGGAG
ATCTTGCACTGTCGCTTCTGTTAAGGGGGTTGACAGTGTTGTTGACTCAAGTCTCTTGAATGAAAGCTATAACACAGATGTCGTGGATGGGAAAGT
CGGTACAGAAGGTGTGACTAGTAGAGGTAAATCAGTACCAAAGGTTATGATTCCTAGTTTACCAAGTGAAGCTAAAGGCGATGCTGTTGCTGTGAT
TGATAGCTGTTTGTGGGAATGGAAGCCAAAACCTGAATGTTTCATTATGAGAAATCTGGATGTCAAATGTAACTCCGCGCCAATACTTTTTCTTCCT

Cromossomo
**Versão/
Variante**

Solyc01g088090.2.1

Espécie
**Número
do gene**

Os05g28200.1

Sobic.009G104700.1

Glyma.07G152300.1

AT5G04410.1

Como é possível identificar o que está codificado por uma sequência nucleotídica?

>Solyc01g088090.2.1

```
ATAAGAAAAAGCAATTCTAAAATCCCATATCACTTCTCCTAAGCTCTTGGAGATTTTGCTCATAATTATCTTAGAAAACTGATTAAAGGTATTCATTTT
AGATTTTAGAAAGCTGACTCTGGGATGGAATTTTGTTCTTTCTATTTCGACACCCTATTACTGCGTTGTAAATGTTAAACCGAATGTGATTCAAAAATG
CCTAATTTCAAATCAGACAATGTTTTCTAGACTAAAAGAAAGAAGATTGTGTTCTGGTTTAGATGTTTATAGCGTAGAGAAAAATAGGAAGCGGAG
ATCTTGCACTGTCGCTTCTGTTAAGGGGGTGGACAGTGTTGTTGACTCAAGTCTCTTGAATGAAAGCTATAACACAGATGTCGTGGATGGGAAAGT
CGGTACAGAAGGTGTGACTAGTAGAGGTAAATCAGTACCAAAGGTTATGATTCCTAGTTTACCAAGTGAAGCTAAAGGCGATGCTGTTGCTGTGAT
TGATAGCTGTTTGTGGGAATGGAAGCCAAAACCTGAATGTTTCATTATGAGAAATCTGGATGTCAAATGTTAACCTCCGCGCCAATACTTTTTCTTCCT
```


Como é possível identificar o que está codificado por uma sequência nucleotídica?

Comparação com sequências já descritas

>Solyc01g088090.2.1

```
ATAAGAAAAAGCAATTCTAAAATCCCATATCACTTCTCCTAAGCTCTTGGAGATTTTGCTCATAATTATCTTAGAAAACTGATTAAAGGTATTCATTTT
AGATTTTAGAAAGCTGACTCTGGGATGGAATTTTGTTCTTTCTATTTCGACACCCTATTACTGCGTTGTAAATGTTAAACCGAATGTGATTCAAAAATG
CCTAATTTCAAATCAGACAATGTTTTCTAGACTAAAAGAAAGAAGATTGTGTTCTGGTTTAGATGTTTATAGCGTAGAGAAAAATAGGAAGCGGAG
ATCTTGCACTGTCGCTTCTGTTAAGGGGGTGGACAGTGTTGTTGACTCAAGTCTCTTGAATGAAAGCTATAACACAGATGTCGTGGATGGGAAAGT
CGGTACAGAAGGTGTGACTAGTAGAGGTAAATCAGTACCAAAGGTATGATTCCTAGTTTACCAAGTGAAGCTAAAGGCGATGCTGTTGCTGTGAT
TGATAGCTGTTTGTGGGAATGGAAGCCAAAACCTGAATGTTTCATTATGAGAAATCTGGATGTCAAATGTTAACCTCCGCGCCAATACTTTTTCTTCCT
```

Como é possível identificar o que está codificado por uma sequência nucleotídica?

Comparação com sequências já descritas

BLAST

Reconstrução filogenética

>Solyc01g088090.2.1

```
ATAAGAAAAAGCAATTCTAAAATCCCATATCACTTCTCCTAAGCTCTTGGAGATTTTGCTCATAATTATCTTAGAAAACTGATTAAAGGTATTCATTTT
AGATTTTAGAAAGCTGACTCTGGGATGGAATTTTGTTCTTTCTATTTCGACACCCTATTACTGCGTTGTAAATGTTAAACCGAATGTGATTCAAAAATG
CCTAATTTCAAATCAGACAATGTTTTCTAGACTAAAAGAAAGAAGATTGTGTTCTGGTTTAGATGTTTATAGCGTAGAGAAAAATAGGAAGCGGAG
ATCTTGCACTGTCGCTTCTGTTAAGGGGGGTTGACAGTGTTGTTGACTCAAGTCTCTTGAATGAAAGCTATAACACAGATGTCGTGGATGGGAAAGT
CGGTACAGAAGGTGTGACTAGTAGAGGTAAATCAGTACCAAAGGTTATGATTCCTAGTTTACCAAGTGAAGCTAAAGGCGATGCTGTTGCTGTGAT
TGATAGCTGTTTGTGGGAATGGAAGCCAAAACCTGAATGTTTCATTATGAGAAATCTGGATGTCAAATGTTAACCTCCGCGCCAATACTTTTTCTTCCT
```

BLAST

Basic
Local
Alignment
Search
Tool



Ferramenta básica de busca por alinhamento local

BLAST

Basic
Local
Alignment
Search
Tool



Ferramenta básica de busca por **alinhamento local**

Alinhamento Local x Alinhamento Global

Alinhamento Local – Alinhamento de regiões de alta similaridade.
Não importa as regiões adjacentes

Alinhamento Global – Alinhamento de toda a extensão das sequências

Alinhamento Local x Alinhamento Global

Seq1	A	C	T	A	C	T	T	C	A	T	T	T	C	T	T	A	C	G	G	A	T	C	A	G	C	T	A	C	T	T	T	A	G	A	G	G	C	T	T	G	C		
Seq2											T	A	C	T	C	A	C	T	G	A	T	C	A	G	G	T	A	C	T														

Alinhamento Local – Alinhamento de regiões de alta similaridade.
Não importa as regiões adjacentes

Alinhamento Local x Alinhamento Global

Alinhamento Global – Alinhamento de toda a extensão das sequencias

Seq1	A	C	T	A	C	T	T	C	A	T	T	T	C	T	T	A	C	G	G	A	T	C	A	G	C	T	A	C	T	T	T	A	G	A	G	G	C	T	T	G	C
Seq2	A	G	T	A	C	T	-	C	A	C	T	-	-	-	-	-	-	-	G	A	T	C	A	G	G	T	A	C	T	-	-	-	-	-	G	G	-	T	A	G	C

Alinhamento Local x Alinhamento Global

Alinhamento Global – Alinhamento de toda a extensão das sequencias

Seq1	A	C	T	A	C	T	T	C	A	T	T	T	C	T	T	A	C	G	G	A	T	C	A	G	C	T	A	C	T	T	T	A	G	A	G	G	C	T	T	G	C
Seq2	A	G	T	A	C	T	-	C	A	C	T	-	-	-	-	-	-	-	G	A	T	C	A	G	G	T	A	C	T	-	-	-	-	-	G	G	-	T	A	G	C

Seq1	A	C	T	A	C	T	T	C	A	T	T	T	C	T	T	A	C	G	G	A	T	C	A	G	C	T	A	C	T	T	T	A	G	A	G	G	C	T	T	G	C
Seq2											T	A	C	T	C	A	C	T	G	A	T	C	A	G	G	T	A	C	T												

Alinhamento Local – Alinhamento de regiões de alta similaridade.
Não importa as regiões adjacentes

Alinhamento Local x Alinhamento Global

Análise entre duas sequências

Análise em um contexto de grupo de sequências/família gênica

Alinhamento Local x Alinhamento Global

Análise entre duas sequências

Regiões conservadas

Busca por similaridade
em bancos de dados

Análise em um contexto de
grupo de sequências/família
gênica

Reconstrução
filogenética

Regiões de
conservação/variabilidade

BLAST

Basic
Local
Alignment
Search
Tool

BLASTn (nt x nt)
BLASTp (aa x aa)



Direta

BLASTx (nt x aa)
tBLASTn (aa x nt)
tBLASTx (nt x nt)



Nt Traduzidos

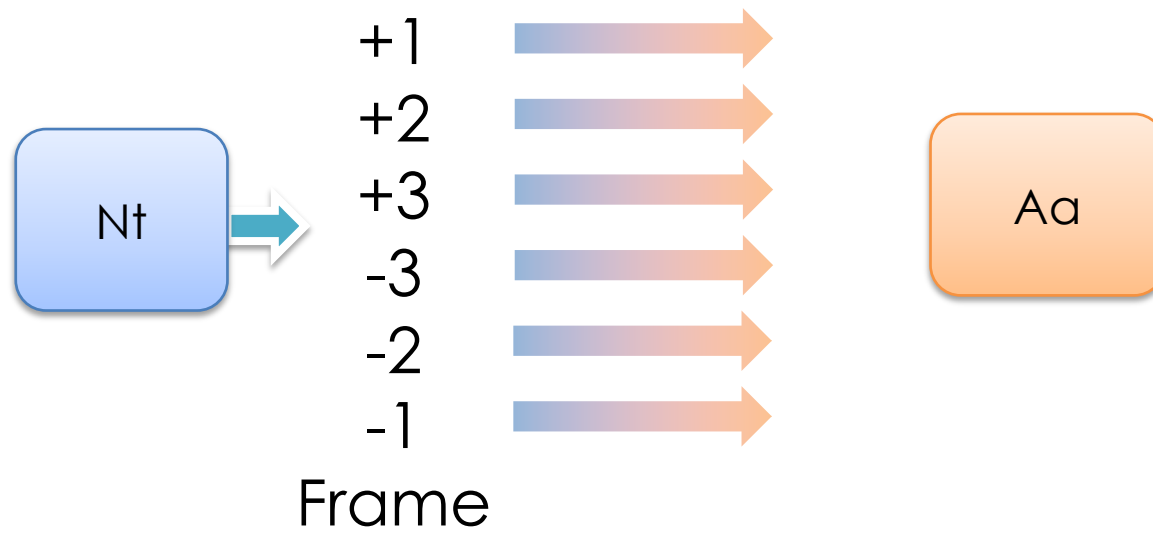
BLAST

Basic
Local
Alignment
Search
Tool

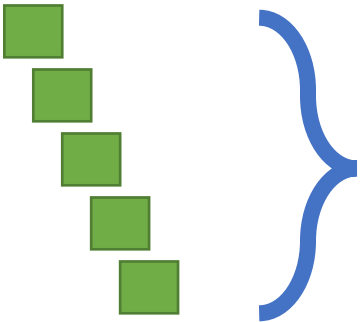
BLASTx (nt x aa)
tBLASTn (aa x nt)
tBLASTx (nt x nt)



Nt Traduzidos



>Soly01g088090.2.1 (1919 nt)



Dividir em fragmentos (Words) = 28 nt/6 aa (default)



Compara com o banco de dados

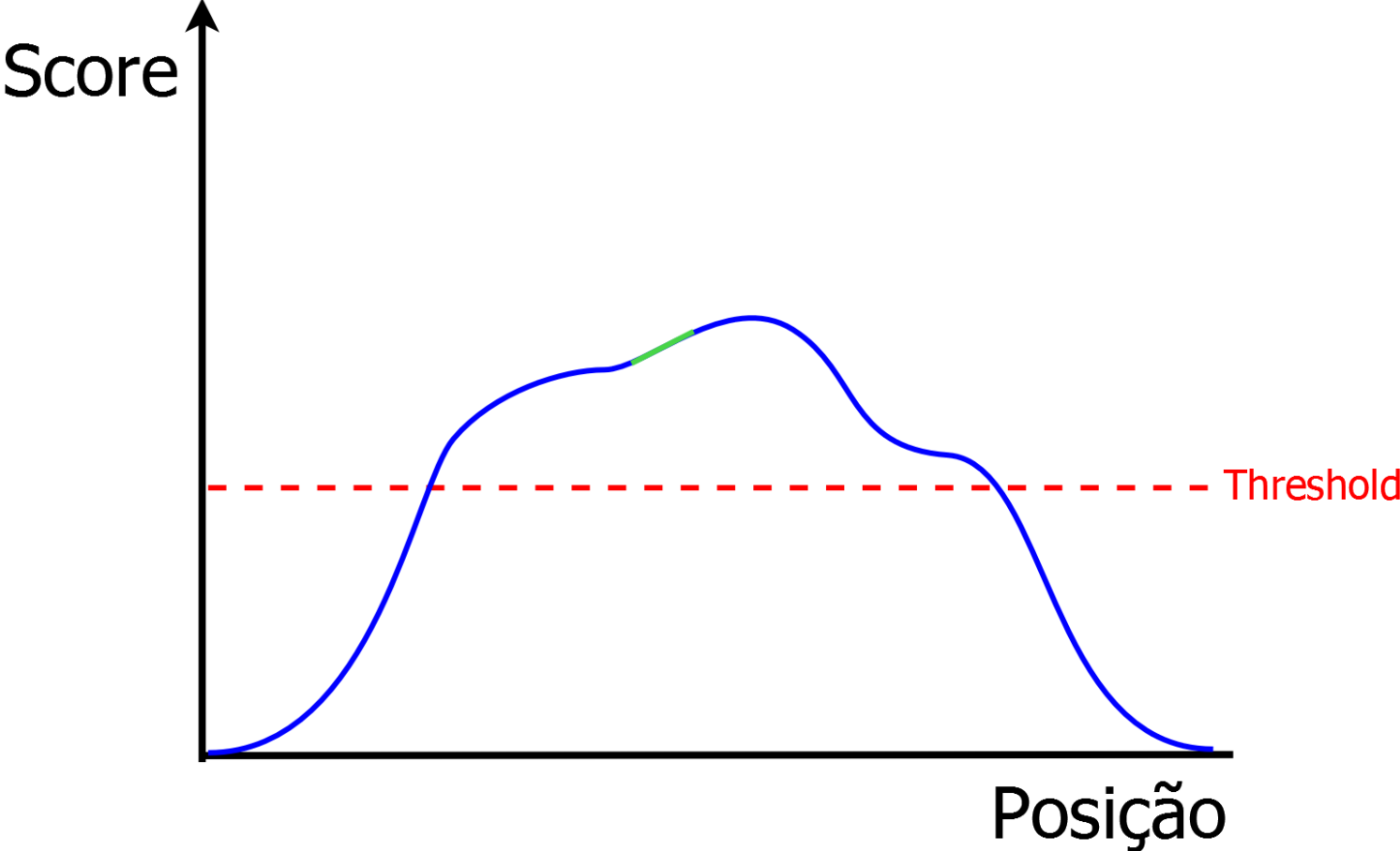
Banco de dados



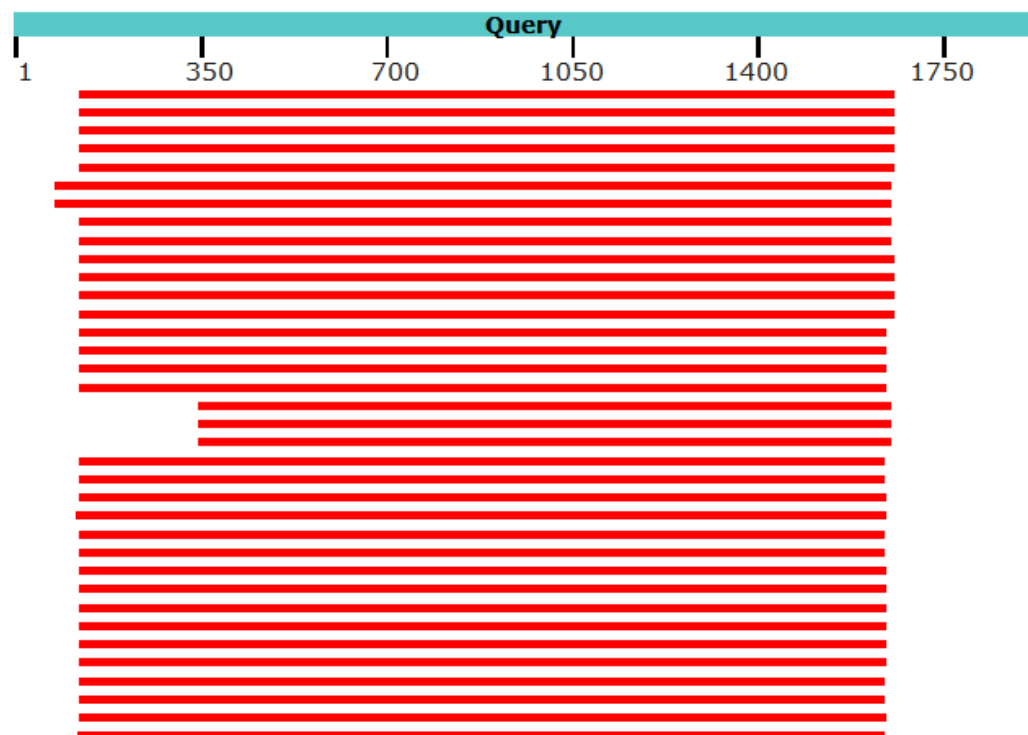
	A	G	C	T
A	10	-1	-3	-4
G	-1	7	-5	-3
C	-3	-5	9	0
T	-4	-3	0	8

	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W	
C	9																				C
S	-1	4																			S
T	-1	1	5																		T
P	-3	-1	-1	7																	P
A	0	1	0	-1	4																A
G	-3	0	-2	-2	0	6															G
N	-3	1	0	-2	-2	0	6														N
D	-3	0	-1	-1	-2	-1	1	6													D
E	-4	0	-1	-1	-1	-2	0	2	5												E
Q	-3	0	-1	-1	-1	-2	0	0	2	5											Q
H	-3	-1	-2	-2	-2	-2	1	-1	0	0	8										H
R	-3	-1	-1	-2	-1	-2	0	-2	0	1	0	5									R
K	-3	0	-1	-1	-1	-2	0	-1	1	1	-1	2	5								K
M	-1	-1	-1	-2	-1	-3	-2	-3	-2	0	-2	-1	-1	5							M
I	-1	-2	-1	-3	-1	-4	-3	-3	-3	-3	-3	-3	-3	1	4						I
L	-1	-2	-1	-3	-1	-4	-3	-4	-3	-2	-3	-2	-2	2	2	4					L
V	-1	-2	0	-2	0	-3	-3	-3	-2	-2	-3	-3	-2	1	3	1	4				V
F	-2	-2	-2	-4	-2	-3	-3	-3	-3	-3	-1	-3	-3	0	0	0	-1	6			F
Y	-2	-2	-2	-3	-2	-3	-2	-3	-2	-1	2	-2	-2	-1	-1	-1	-1	3	7		Y
W	-2	-3	-2	-4	-3	-2	-4	-4	-3	-2	-2	-3	-3	-1	-3	-2	-3	1	2	11	W
	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W	

Banco de dados



Distribution of the top 100 Blast Hits on 100 subject sequences



Sequences producing significant alignments:

Select: [All](#) [None](#) Selected:0

	Description
<input type="checkbox"/>	pheophytinase, chloroplastic [Solanum lycopersicum]
<input type="checkbox"/>	pheophytinase, chloroplastic [Solanum pennellii]
<input type="checkbox"/>	hypothetical protein EJD97_021004 [Solanum chilense]
<input type="checkbox"/>	PREDICTED: pheophytinase, chloroplastic-like [Solanum tuberosum]
<input type="checkbox"/>	Pheophytinase, chloroplastic [Capsicum annuum]
<input type="checkbox"/>	Pheophytinase, chloroplastic [Capsicum baccatum]
<input type="checkbox"/>	Pheophytinase, chloroplastic [Capsicum chinense]
<input type="checkbox"/>	PREDICTED: pheophytinase, chloroplastic [Capsicum annuum]
<input type="checkbox"/>	PREDICTED: pheophytinase, chloroplastic-like [Nicotiana sylvestris]
<input type="checkbox"/>	PREDICTED: pheophytinase, chloroplastic-like [Nicotiana attenuata]
<input type="checkbox"/>	PREDICTED: pheophytinase, chloroplastic-like [Nicotiana tomentosiformis]
<input type="checkbox"/>	pheophytinase, chloroplastic-like [Nicotiana tabacum]
<input type="checkbox"/>	PREDICTED: pheophytinase, chloroplastic-like isoform X1 [Nicotiana tomentosiformis]
<input type="checkbox"/>	PREDICTED: pheophytinase, chloroplastic-like [Nicotiana sylvestris]
<input type="checkbox"/>	PREDICTED: pheophytinase, chloroplastic-like [Nicotiana attenuata]
<input type="checkbox"/>	pheophytinase, chloroplastic-like [Olea europaea var. sylvestris]
<input type="checkbox"/>	pheophytinase, chloroplastic [Coffea eugenioides]
<input type="checkbox"/>	pheophytinase, chloroplastic-like [Coffea arabica]
<input type="checkbox"/>	Pheophytinase [Actinidia chinensis var. chinensis]

pheophytinase, chloroplastic [Solanum lycopersicum]

Sequence ID: [XP_004229681.1](#) Length: 512 Number of Matches: 1

Range 1: 1 to 512 [GenPept](#) [Graphics](#)

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

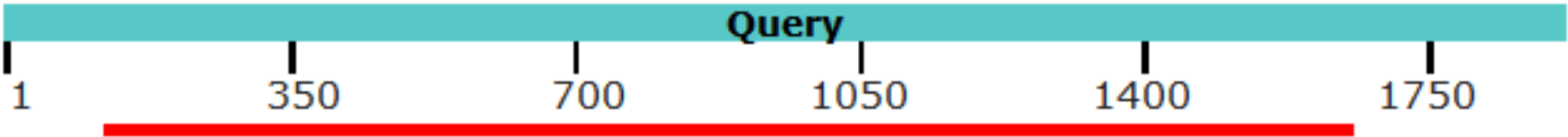
Score	Expect	Method	Identities	Positives	Gaps	Frame
1007 bits(2603)	0.0	Compositional matrix adjust.	512/512(100%)	512/512(100%)	0/512(0%)	+1
Query 124	MEFCSFYSTPYVCVNVKPNVIQKCLISNQTMFSRLKERRLCGLDVYSVEKNRKRRSCT				303	
Sbjct 1	MEFCSFYSTPYVCVNVKPNVIQKCLISNQTMFSRLKERRLCGLDVYSVEKNRKRRSCT				60	
Query 304	vasvkgvdsvvdsLLNESYNTDVVDGKVGTEGVTSRGKSVPKVMIPSLPSEAKGDAVAV				483	
Sbjct 61	VASVKGVDSVVDSSLLNESYNTDVVDGKVGTEGVTSRGKSVPKVMIPSLPSEAKGDAVAV				120	
Query 484	IDSCLEWKPKLNVHYEKSGCQNVNSAPILFLPGFGVGSFHYEKQLKDLGRDNRIWALDF				663	
Sbjct 121	IDSCLEWKPKLNVHYEKSGCQNVNSAPILFLPGFGVGSFHYEKQLKDLGRDNRIWALDF				180	
Query 664	IGQGKSLPSEDPTSRSKRLNESEGDGNNVLWGLGDEVEPWAKELVYSIDLWREQVRYFIE				843	
Sbjct 181	IGQGKSLPSEDPTSRSKRLNESEGDGNNVLWGLGDEVEPWAKELVYSIDLWREQVRYFIE				240	
Query 844	EVIKEPVYIVGNSLGGYVALYFAAYYPQLVKGVTLNATPFWGFLPNPVRSPRLSRLFPW				1023	
Sbjct 241	EVIKEPVYIVGNSLGGYVALYFAAYYPQLVKGVTLNATPFWGFLPNPVRSPRLSRLFPW				300	
Query 1024	AGTFPLPSNIRKLTENVWQKISAPESIAEVLKQVYADHTTKVDKVFSSILEVTEHPaaaa				1203	
Sbjct 301	AGTFPLPSNIRKLTENVWQKISAPESIAEVLKQVYADHTTKVDKVFSSILEVTEHPAAAA				360	
Query 1204	slasIMFAPQGQLNFKEALTGCRMNNVPVCLIIYGKEDPWVTPIWGLQVKRQFPEAPYYQI				1383	
Sbjct 361	SLASIMFAPQGQLNFKEALTGCRMNNVPVCLIIYGKEDPWVTPIWGLQVKRQFPEAPYYQI				420	
Query 1384	SPAGHCPHDEVPEIVNFLLRGWIRNVESDSSAALPLLDYPESVEYDVVKELEFVRQGVKK				1563	
Sbjct 421	SPAGHCPHDEVPEIVNFLLRGWIRNVESDSSAALPLLDYPESVEYDVVKELEFVRQGVKK				480	
Query 1564	SAKVQFYGSMTSQWERLSMFLKSRFQDGVYSP				1659	
Sbjct 481	SAKVQFYGSMTSQWERLSMFLKSRFQDGVYSP				512	

>Solyc01g088090.2.1 (1919 nt)
ATAAGAAAAAGCAATTCTAAAATCCCATATCACTTCTCCTAAGCTCTTGGAGATTTTGCTCATAATTATCTTAGAAAACTGATTAAAGGTATTCATTTT
AGATTTTAGAAAGCTGACTCTGGGATGGAATTTTGTTCCTTTCTATTTCGACACCCTATTACTGCGTTGTAAATGTTAAACCGAATGTGATTCAAAAATG

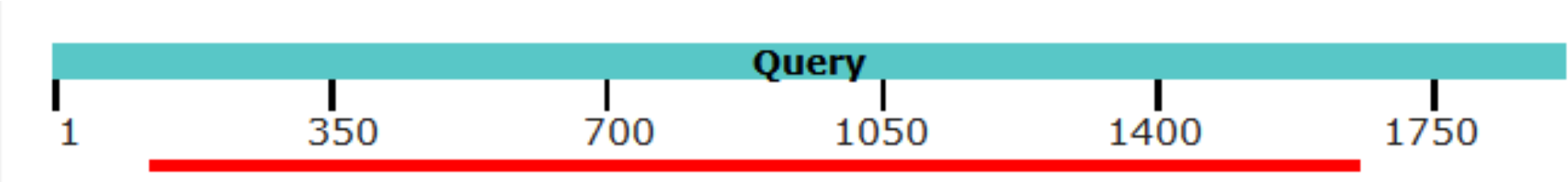


	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input type="checkbox"/>	pheophytinase, chloroplastic [Solanum lycopersicum]	1007	1007	80%	0.0	100.00%	XP_004229681.1

Distribution of the top 100 Blast Hits on 100 subject sequences



>Solyc01g088090.2.1



?

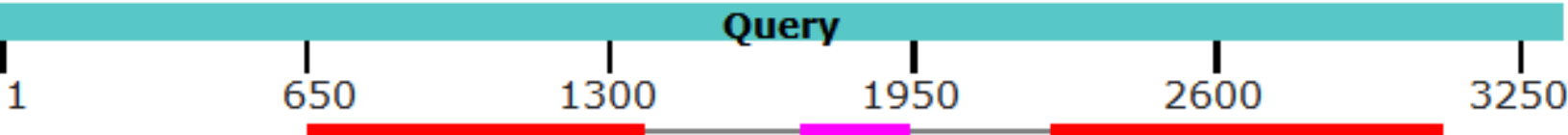
>Solyc01g088090.2.1 (3427 nt)
gDNA

Distribution of the top 298 Blast Hits on 100 subject sequences



>Solyc01g088090.2.1 (3427 nt)
gDNA

Distribution of the top 298 Blast Hits on 100 subject sequences



	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	pheophytinase, chloroplastic [Solanum lycopersicum]	Solanum lycopersicum	469	974	53%	5e-150	100.00%	512	XP_004229681.1

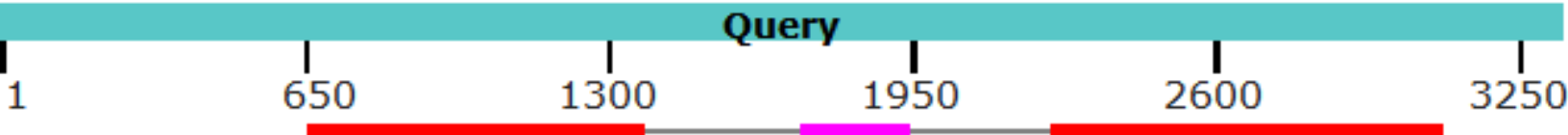
Score	Expect	Method	Identities	Positives	Gaps	Frame
469 bits(1207)	4e-150	Compositional matrix adjust.	242/242(100%)	242/242(100%)	0/242(0%)	+2

Score	Expect	Method	Identities	Positives	Gaps	Frame
155 bits(392)	8e-36	Compositional matrix adjust.	74/78(95%)	76/78(97%)	0/78(0%)	+3

Score	Expect	Method	Identities	Positives	Gaps	Frame
349 bits(896)	1e-104	Compositional matrix adjust.	195/279(70%)	195/279(69%)	84/279(30%)	+2

>Solyc01g088090.2.1 (3427 nt)
gDNA

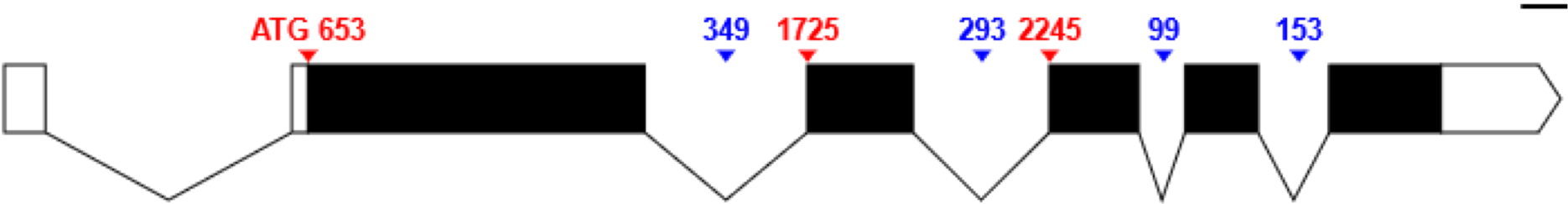
Distribution of the top 298 Blast Hits on 100 subject sequences



Range 2: 318 to 512 [GenPept](#) [Graphics](#) [▼ Next Match](#) [▲ Previous Match](#) [▲ First Match](#)

	Score	Expect	Method	Identities	Positives	Gaps	Frame
	349 bits(896)	1e-104	Compositional matrix adjust.	195/279(70%)	195/279(69%)	84/279(30%)	+2
Query	2246	WQKISAPESIAEVLKQVYADHTTKVDKVFSSILEVTEHPaaaaslasIMFAPQGQLNFKE					2425
Sbjct	318	WQKISAPESIAEVLKQVYADHTTKVDKVFSSILEVTEHPAAAAASLASIMFAPQGQLNFKE					377
Query	2426	ALTG*VKQVVLHSRVIWLAYVFLYSNKAEIFHLAISRCRMNNVPVCLIYGKEDPWVTPIW					2605
Sbjct	378	ALTG-----CRMNNVPVCLIYGKEDPWVTPIW					404
Query	2606	GLQVKRQFPEAPYYQISPAGHCPHDEVPEVPRLS*LYICF*HSILSVKKGLNIT*YKMMI					2785
Sbjct	405	GLQVKRQFPEAPYYQISPAGHCPHDEVPE-----					433
Query	2786	MLCHTH*L*ILVSPLNTNVQIVNFLLRGWIRNVESDSSAALPLLDYPESVEYDVVKELEF					2965
Sbjct	434	-----IVNFLLRGWIRNVESDSSAALPLLDYPESVEYDVVKELEF					473
Query	2966	VRQGVKKSAAKVQFYGSMTSQWERLSMFLKSRFQDGVYSP			3082		
Sbjct	474	VRQGVKKSAAKVQFYGSMTSQWERLSMFLKSRFQDGVYSP			512		

>Solyc01g088090.2.1 (3427 nt)
gDNA



Score	Expect	Method	Identities	Positives	Gaps	Frame
469 bits(1207)	4e-150	Compositional matrix adjust.	242/242(100%)	242/242(100%)	0/242(0%)	+2

Score	Expect	Method	Identities	Positives	Gaps	Frame
155 bits(392)	8e-36	Compositional matrix adjust.	74/78(95%)	76/78(97%)	0/78(0%)	+3



Score	Expect	Method	Identities	Positives	Gaps	Frame
349 bits(896)	1e-104	Compositional matrix adjust.	195/279(70%)	195/279(69%)	84/279(30%)	+2

Descriptions

Graphic Summary

Alignments

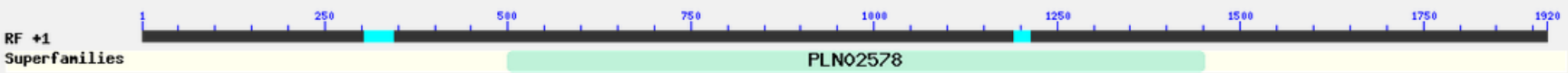
Taxonomy

 *hover to see the title*  *click to show alignments* ☒ *Show Conserved Domains*

Alignment Scores  < 40  40 - 50  50 - 80  80 - 200  >= 200 

16 sequences selected 

Putative conserved domains have been detected, click on the image below for detailed results.



Distribution of the top 16 Blast Hits on 16 subject sequences



pheophytinase, chloroplastic-like [Physcomitrium patens]

Sequence ID: [XP_024365762.1](#) Length: 540 Number of Matches: 1

[See 1 more title\(s\)](#) [See all Identical Proteins\(IPG\)](#)

Range 1: 138 to 531 [GenPept](#) [Graphics](#)

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

Score	Expect	Method	Identities	Positives	Gaps	Frame
369 bits(947)	1e-120	Compositional matrix adjust.	189/397(48%)	251/397(63%)	9/397(2%)	+1
Query	478	AVIDSCLWEWKPKLNVHYEKGSGCQNVNSAPILFLPGFGVGSFHYEKQLKDLGRDNRIWAL				657
		A I S +WEW+ + N+HYE +G +N N+ +L LPGFGVGSFHY +QL+DLG++ R+WA+				
Sbjct	138	ATIASHIWEWRHRWNIHYECAGMENTNAPAMLLLPFGFGVGSFHYHQQLRDLGQEYRVWAI				197
Query	658	DFIGQGKSLPSEDPTSRSKRLNESEGDGNNV-----LWGLGDEVEPWAKELVYSIDLWR				819
		DF+GQGKS PS DP E V W LG EPWA+ LVYS+D WR				
Sbjct	198	DFLGQGKSWPSHDPAPEEAEEVVEEVLEGQVNSKKIRHWSLGKNPEPWAEGLVYSVDTWR				257
Query	820	EQVRYFIEEVIKEPVYIVGNSLGGYVALYFAAYYPQLVKGVTLNATPFWGFLPNPVRSP				999
		+QV FIE+VI PVYIVGNSLGGYV YFAA P+LVKGVTLNATPFW F PN R P				
Sbjct	258	DQVHAFIEKVIIGGPVYIVGNSLGGYVGSYFAATNPVELVKGVTLNATPFWAFTPNRRYP				317
Query	1000	RLSRLFPWAGTFPLPSNIRKLTELWVQKISAPESIAEVLKQVYADHTTKVDKVFSSILEV				1179
		LS+L PW G P+P + + W + P +I +L VYA+ + K+ + I+E				
Sbjct	318	LLSKLTPWGGLLPVPIFAKAIIRFWDDLRLNPSTIRNMLGAVYANRSAINKKLITQIIIEA				377
Query	1180	TEHPaaaaslasIMFAPQGQLNFKEALTGCRMNNVPVCLIIYGKEDPWVTPIWGLQVKRQF				1359
		T+HPAA A+ ASI+FAP+ +F E L + +P+C+IYGKEDPWV P WG + K++				
Sbjct	378	TDHPAAFAAFASIVFAPRAHTDFGENLISLKERRMPMCMIIYGKEDPWVVPFWGQRAKQRN				437
Query	1360	PEAPYYQISPAGHCPHDEVPEIVNFLLRGWIRNVESDSSAALPLLDYPESVEYDVVKELE				1539
		P+A YY++SPAGHCPH E PE+VN LLR W+ +V++ A+ PLL+ + + V E E				
Sbjct	438	PDAIYYELSPAGHCPHHEAPEVVNTLLRKWVESVQT-LGASSPLLESRVNQLHWFVENE				496
Query	1540	FVRQGVKKSAAKVQFYGSMTSQWERLSMFLKSRLFQDGV		1650		
		R V+ F+ ++ + + FLK Q V				
Sbjct	497	GDRVSVRVKGTQNFQNLIDRIEI--FLKKSLQSQV		531		

Alinhamento Local x Alinhamento Global

Análise entre duas sequências

Regiões conservadas

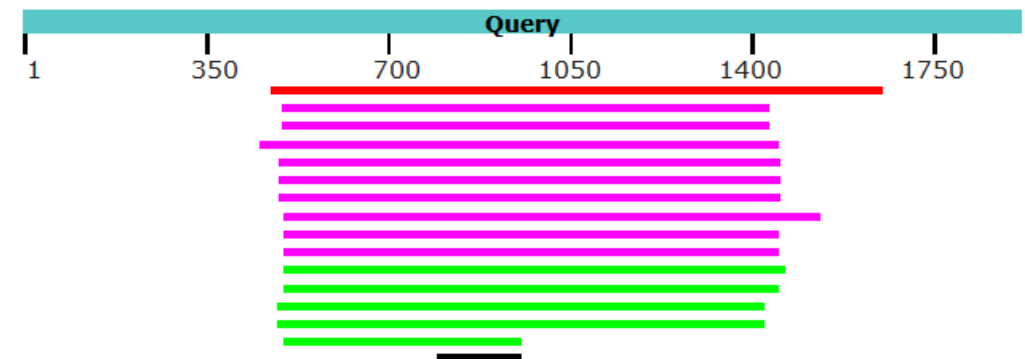
Busca por similaridade
em bancos de dados

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected:0

	Description
<input type="checkbox"/>	pheophytinase, chloroplastic [Solanum lycopersicum]
<input type="checkbox"/>	pheophytinase, chloroplastic [Solanum pennellii]
<input type="checkbox"/>	hypothetical protein EJD97_021004 [Solanum chilense]
<input type="checkbox"/>	PREDICTED: pheophytinase, chloroplastic-like [Solanum tuberosum]
<input type="checkbox"/>	Pheophytinase, chloroplastic [Capsicum annuum]

Distribution of the top 16 Blast Hits on 16 subject sequences

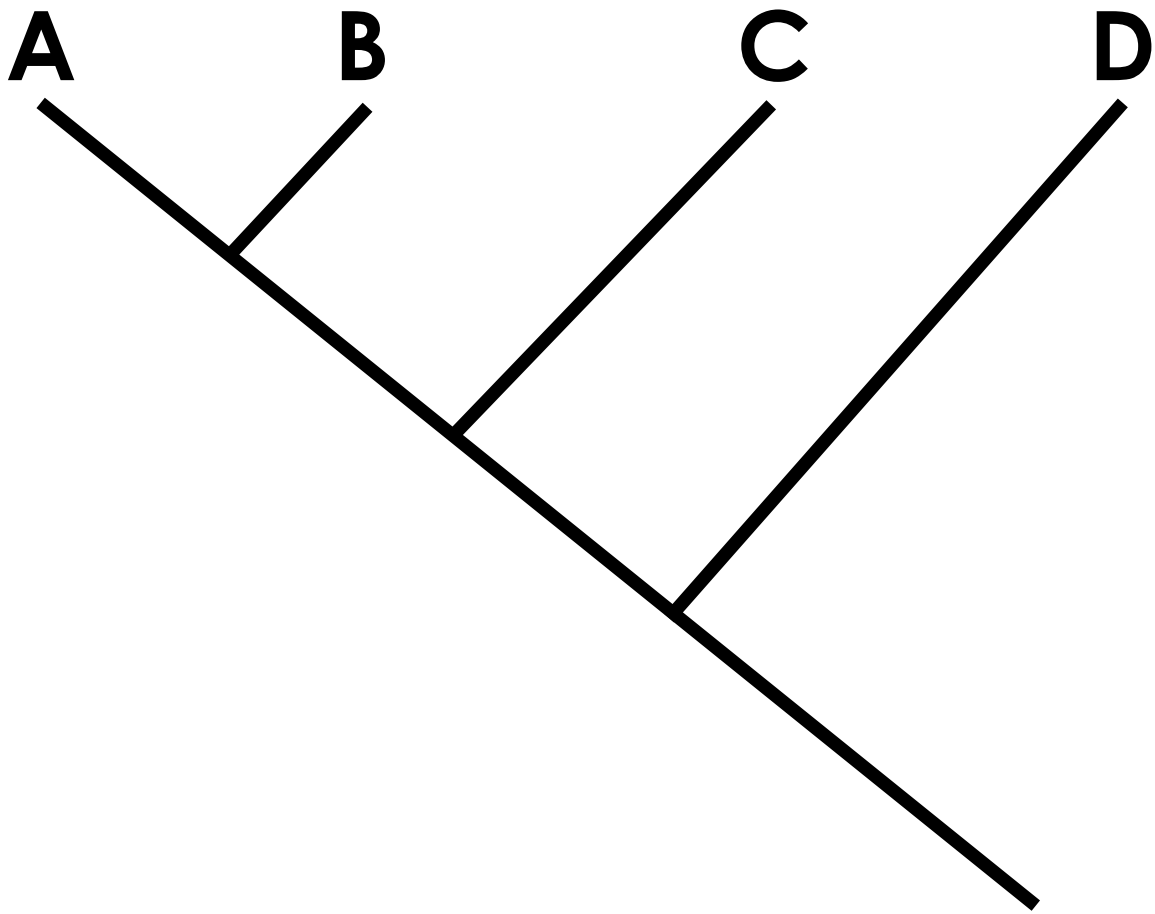


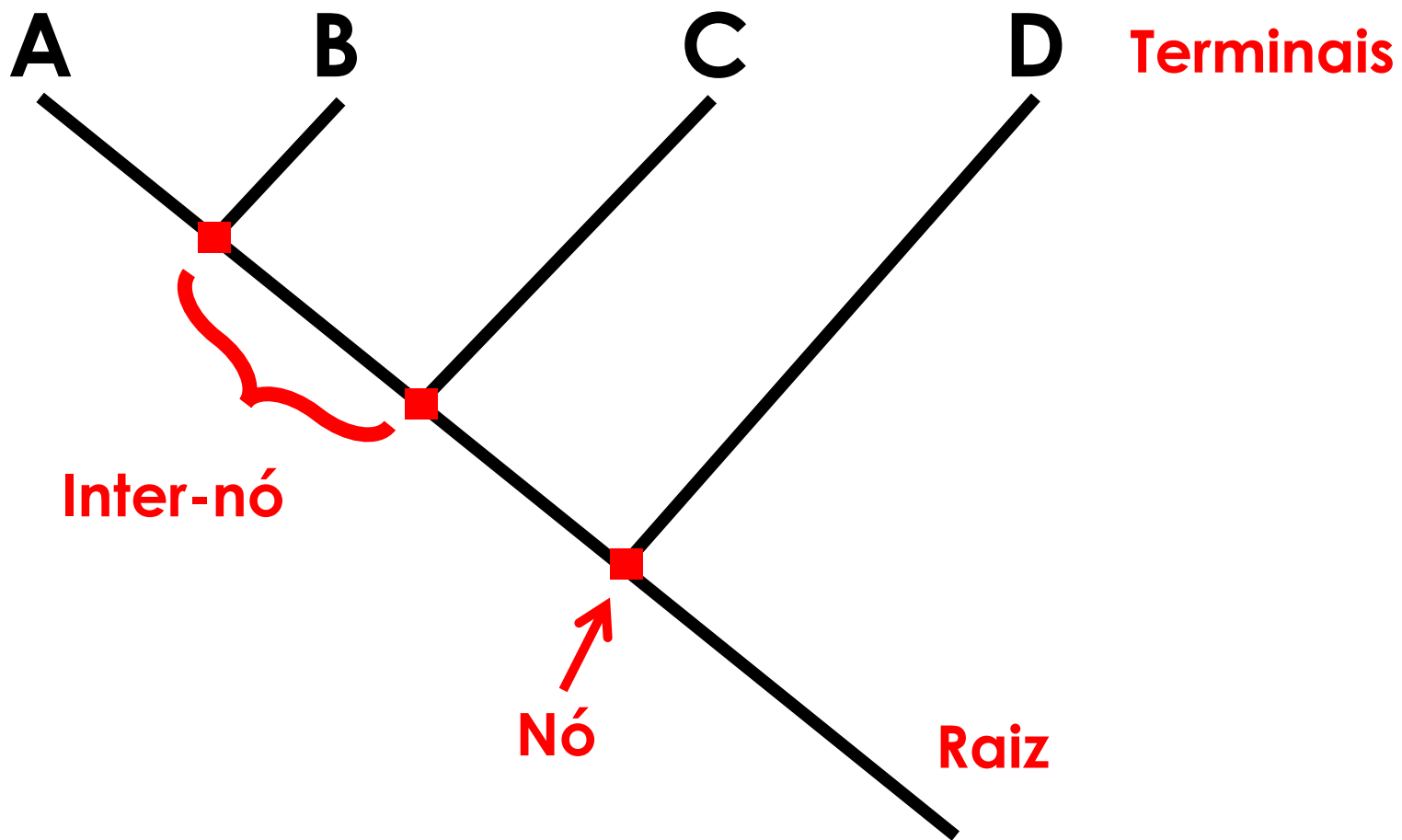
Alinhamento Local x Alinhamento Global

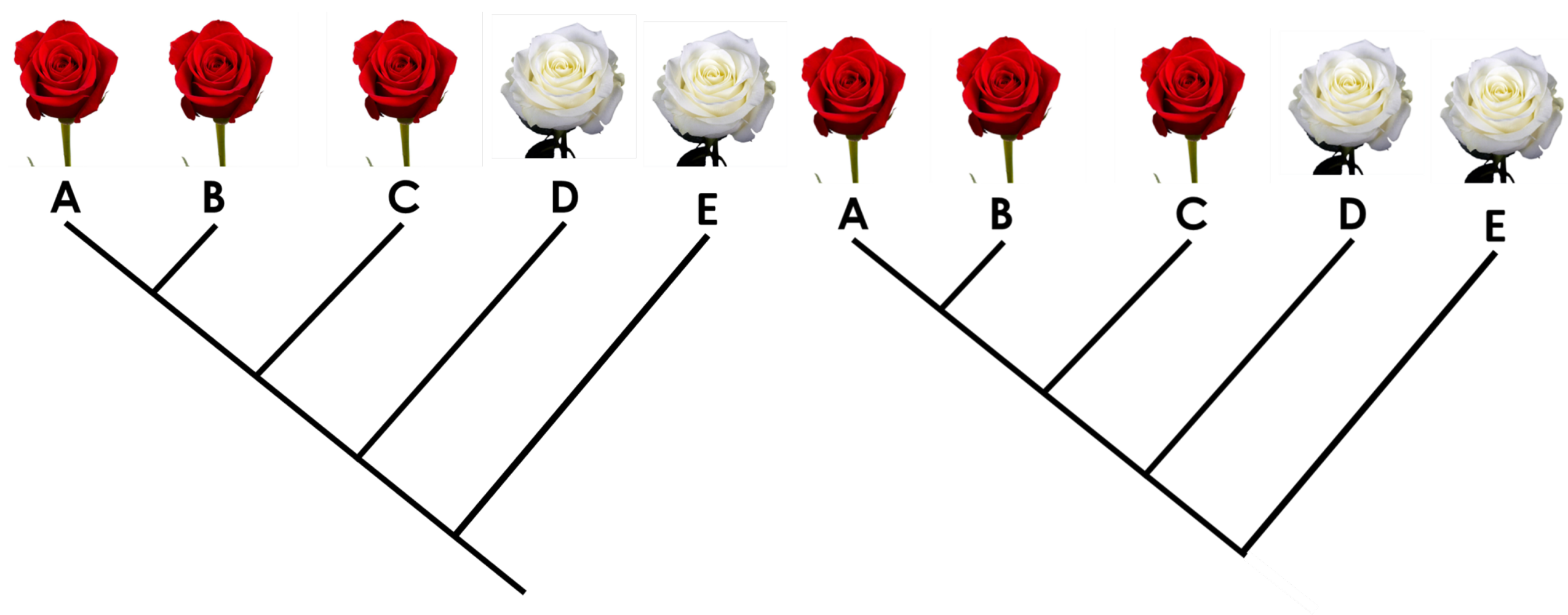
Análise em um contexto de grupo de sequências/família gênica

Reconstrução filogenética

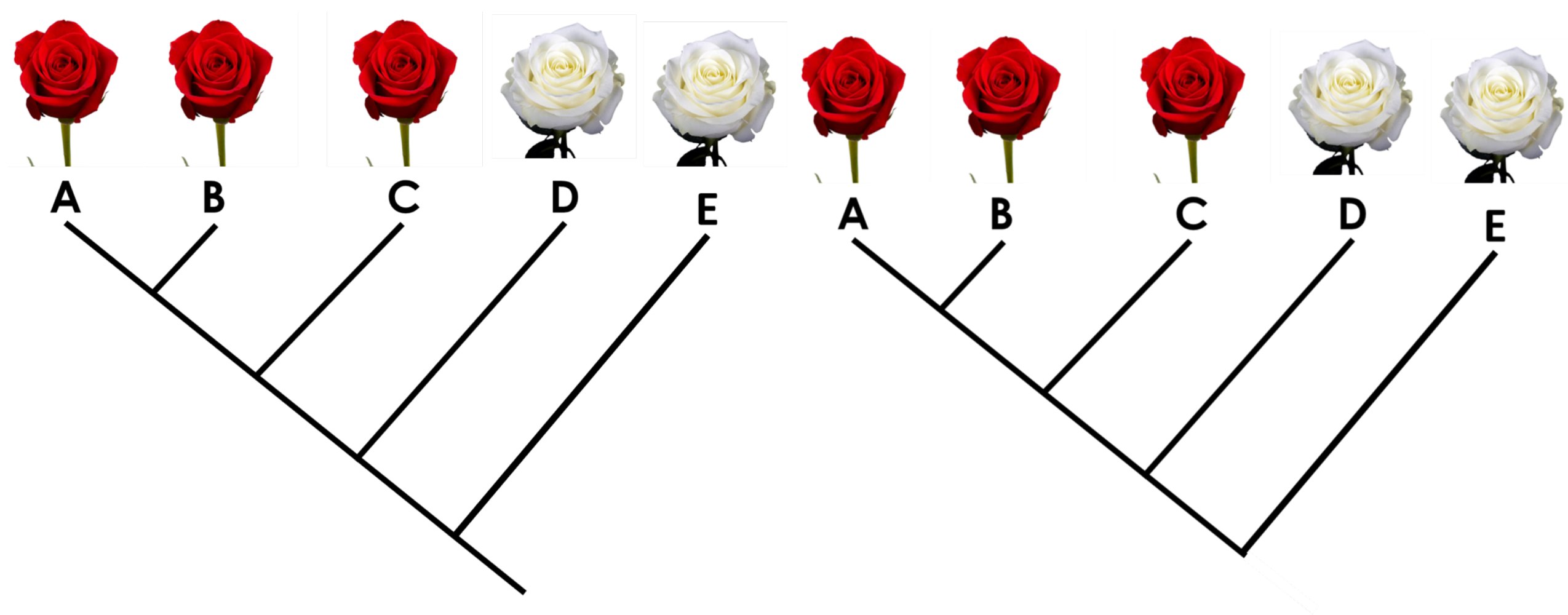
Regiões de conservação/variabilidade



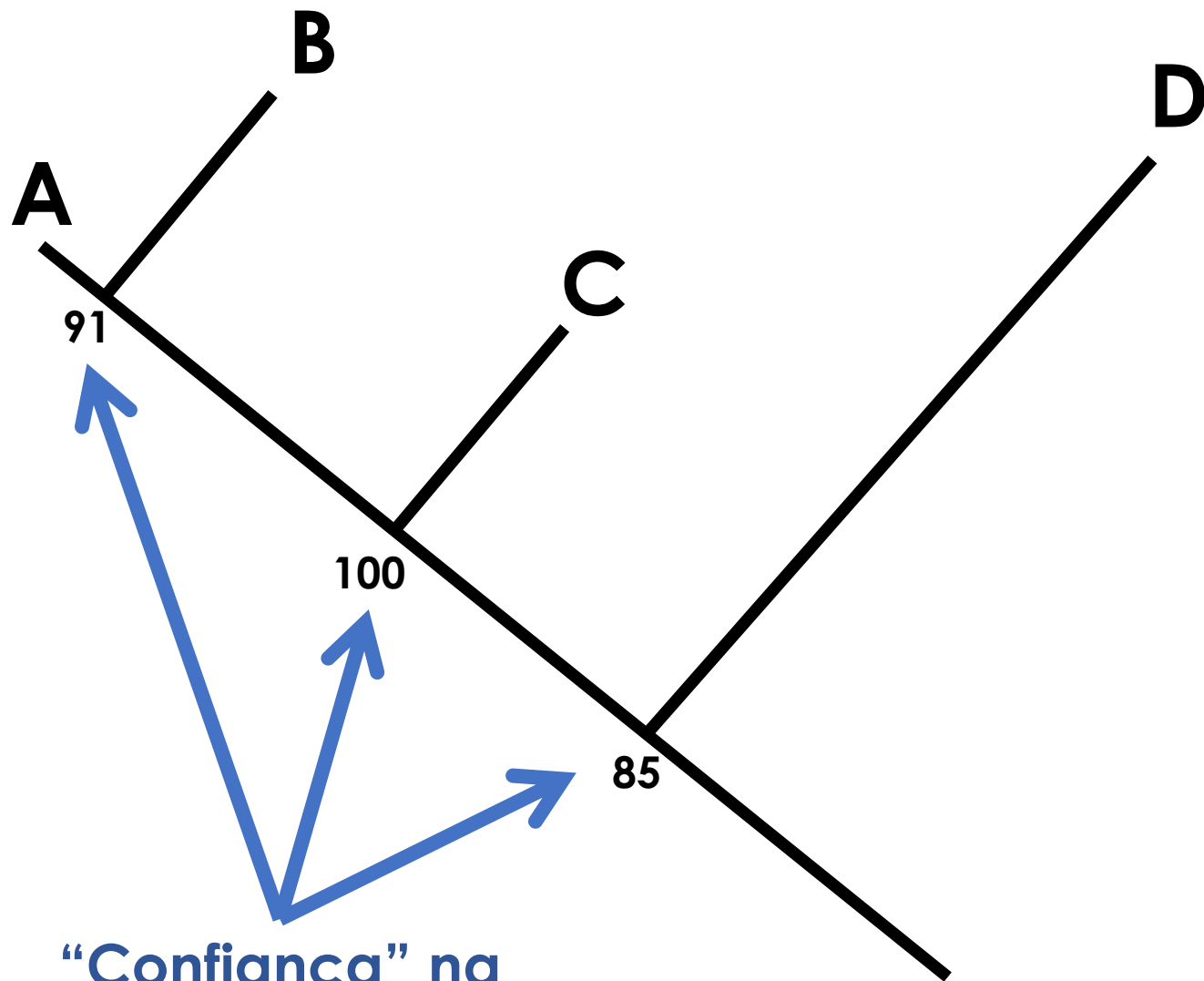




A informação que estas duas árvores traz é a mesma?



Raiz/grupo externo – polarização (sentido temporal)



“Confiança” na
sustentação do nó

Tamanho dos ramos =
acúmulo de
divergências

Sequência 1 (nt/aa)

Sequência 2 (nt/aa)

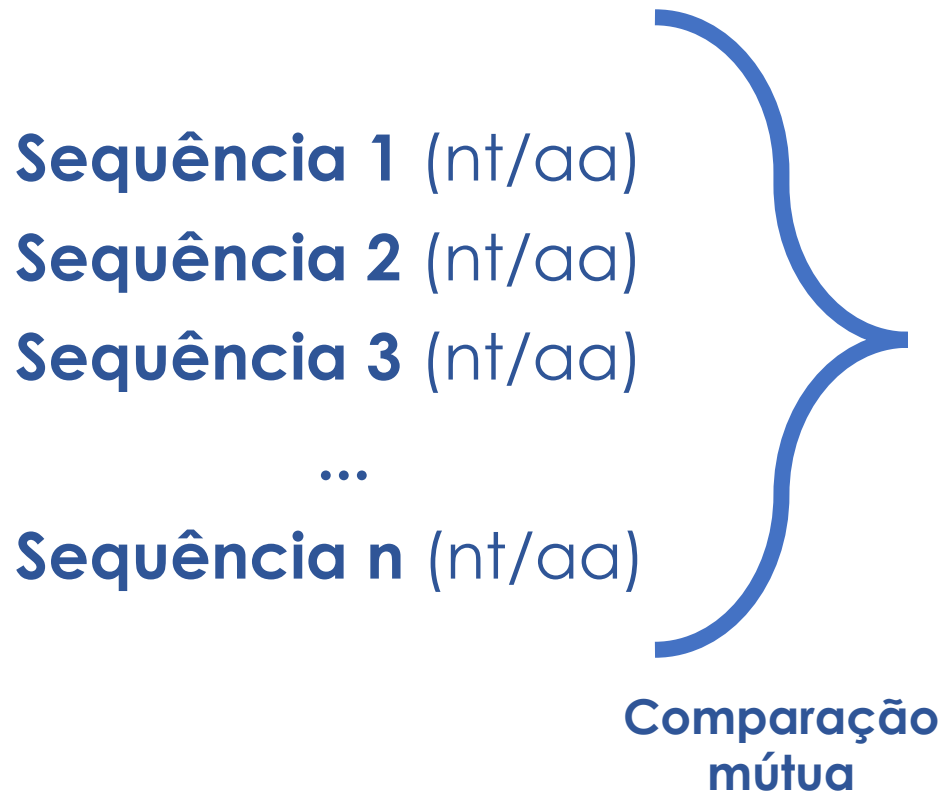
Sequência 3 (nt/aa)

...

Sequência n (nt/aa)

**Comparação
mútua**

**Estabelecimento
de relação**

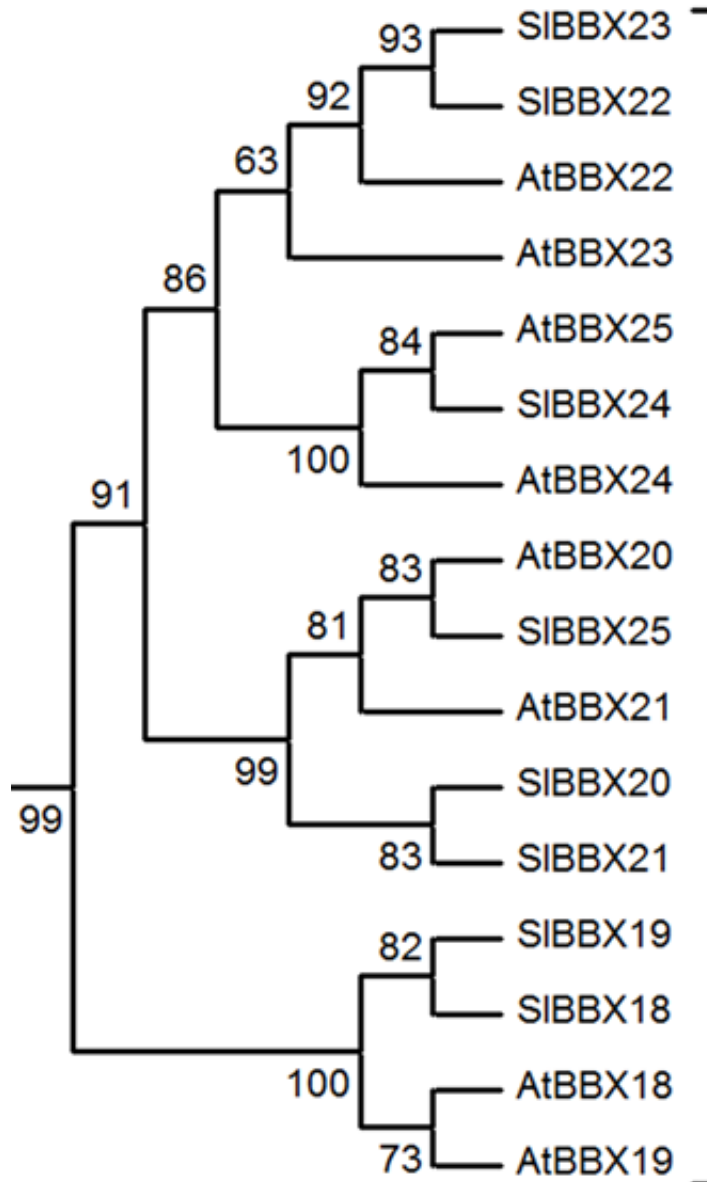


Estabelecimento de relação

Homologia – Entidades biológicas com origem evolutiva em comum

Paralogia – Originada por duplicação gênica (genes em uma mesma espécie)

Ortologia – Originada pela especiação (genes em espécies diferentes)

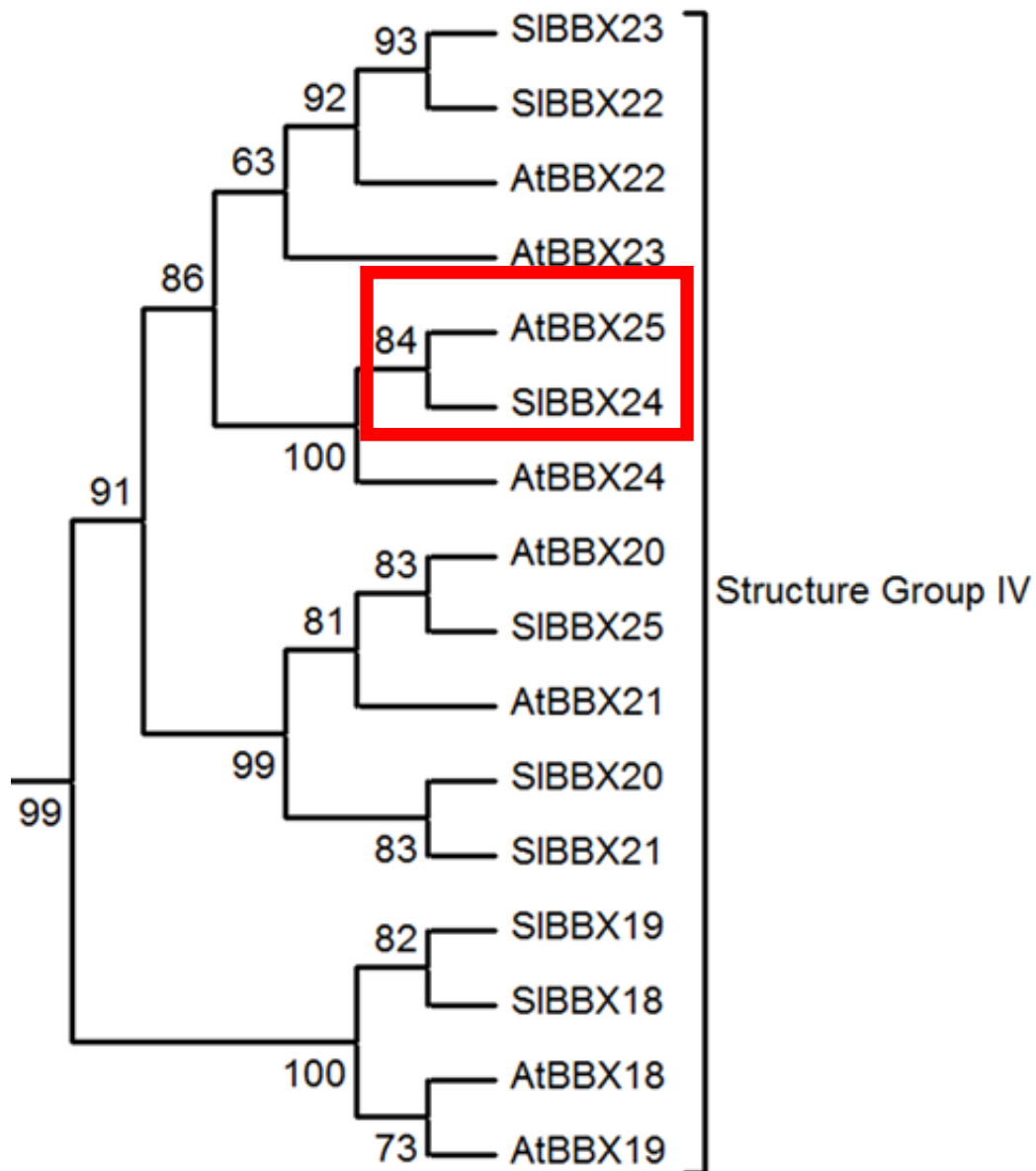


Structure Group IV



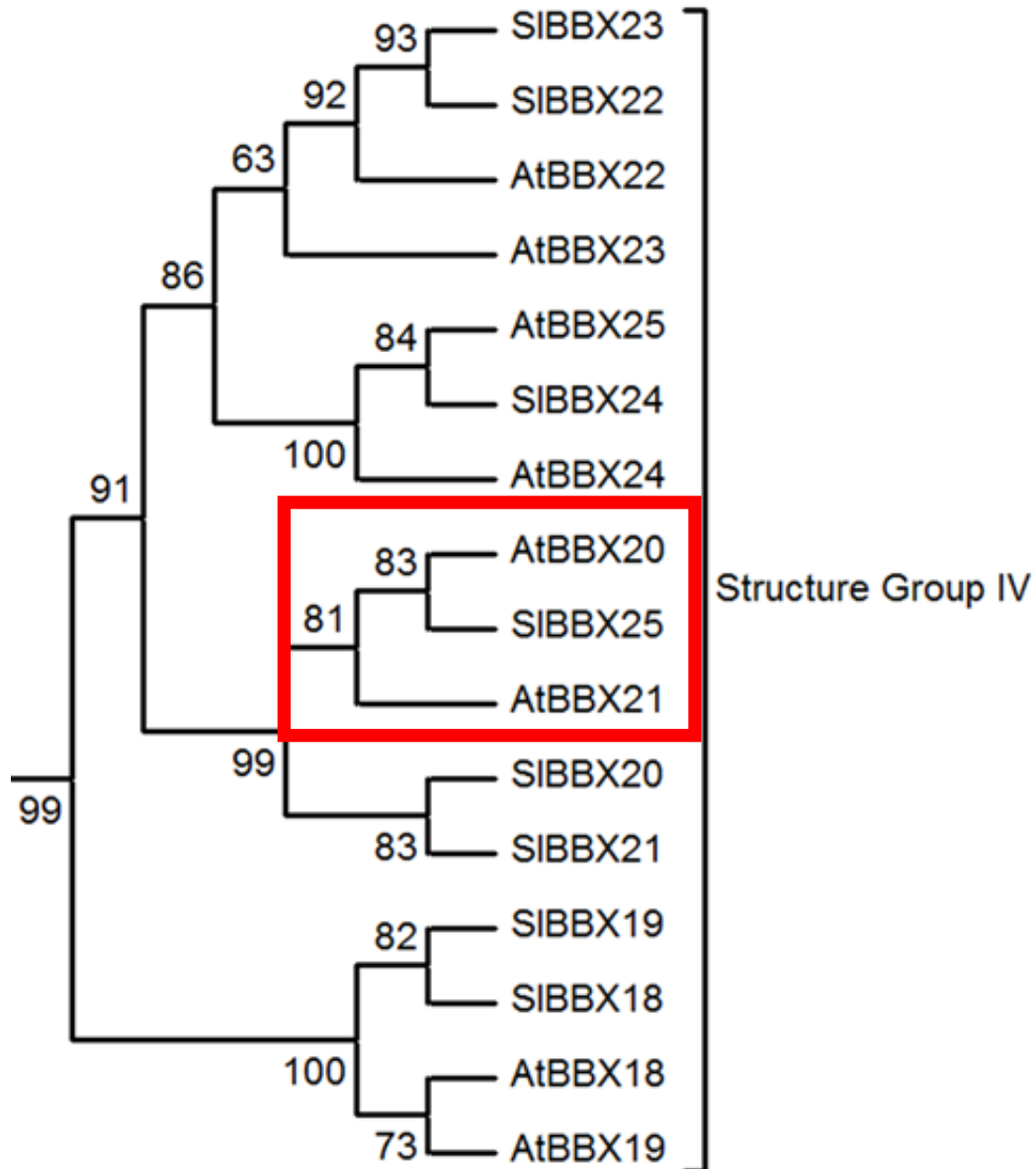
Ortologia – Originada pela
especiação (genes em espécies
diferentes)

Possuem a mesma função



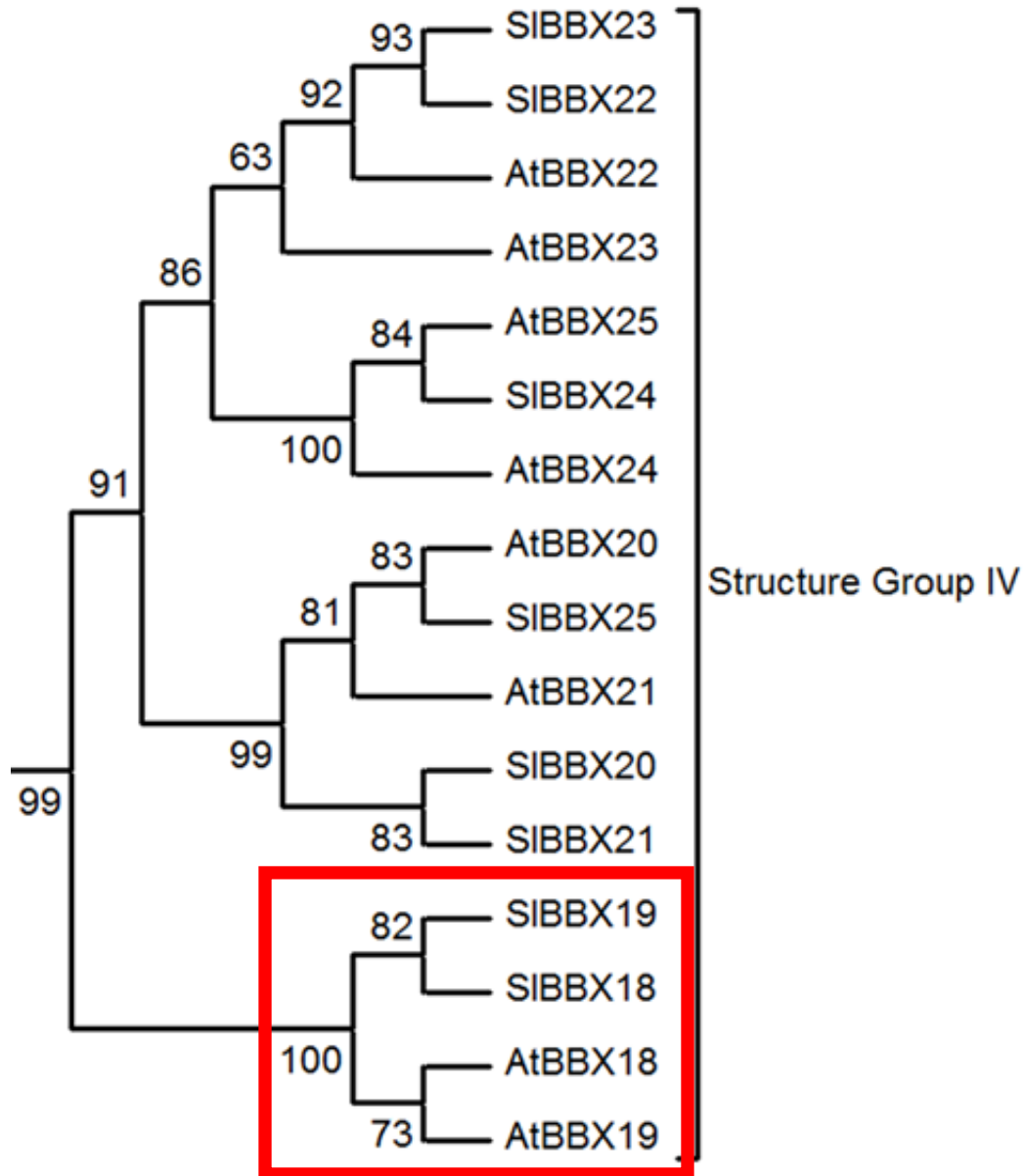
Ortologia – Originada pela
especiação (genes em espécies
diferentes)

Possuem a mesma função



Ortologia – Originada pela
especiação (genes em espécies
diferentes)

Possuem a mesma função



Ortologia – Originada pela
especiação (genes em espécies
diferentes)

Possuem a mesma função

Passos para se obter uma reconstrução filogenética

1. Obter sequencias
2. Alinhar as sequencias
3. Reconstruir a filogenia

Passos para se obter uma reconstrução filogenética

1. Obter sequencias
2. Alinhar as sequencias
3. Reconstruir a filogenia

Passos para se obter uma reconstrução filogenética

1. Obter sequencias

Sequencias próprias

Sequencias de bancos de dados

Passos para se obter uma reconstrução filogenética

1. Obter sequencias
2. Alinhar as sequencias
3. Reconstruir a filogenia

Passos para se obter uma reconstrução filogenética

2. Alinhar as sequencias

Alinhamentos globais

Diversos métodos (ClustalW, MUSCLE, TCOFFEE)

- 1 – Comparações par a par
- 2 – Comparação global

2.

Protein Sequences

Species/Abbrv	Sequence
1. 29615.t000016__Org	K-K-ELCKY-PAKTYCEFD-EASLCWGCDAKVHGANFLVARHTRTLCLCRSCQSLTPWKAS-GSRLGCVIVSVCDVCAANDANR
2. 29933.t000080__Org	K-K-ELCNS-LAKMYCESDQASLCWDCCARVHAANFLVAKHSTRTLLCHLCSFTPWTA-SGPRLRP-TVSICDNC-
3. 30027.t000047__Org	K-GCELGG-AAARMYCSDQASLCWSCDEKVHSANFLVAKHCRNLLCQVCQSPTPWKA-SGPKLGP-TVSICDSCFSLHNSSNN
4. 49712__Org_Smoell	M-EKS-R-DCELQV-RAAVYCCA-DEAYLCWKCDSKVHGANFIVARHLRSILCGRCHSP
5. Aco008479__Org_A	K-KGK-GCELCPG-TARMYCSEDEAMLCWECDARVHGAFNLVARHTRALLCRTCQSPTPWRA-SGSRLGPG-SVSLCHRCSPHSHSQSPAADDLRRGGGEAGGF
6. Aco009107__Org_A	K-K-ELCDG-VARMYCSDAHLCGGCDAAKVHGANFLVARHTRALLCRTCQSPTPWRA-AGARLGP-AVSVCERCLIRGRGGGG
7. AHYO_004215_O	KR-K-ELCSN-LATIYCDSEAILCYECDSKVHGANFLVAKHSTRILLCHVCQSPTLCSG-SGPKFGP-TFSVCPCTCTLGKEGV
8. AHYO_012866_O	K-G-ELCGL-AATTYCSDKASLCWHCDGKVHKANFLVIKHIRTLLCHLCSQNPTPWTA-SGPTLPP-SLSFCLACASSQSTHLM
9. AHYO_016308_O	LKE-S-ELCDM-PASTYCSDRARLCWECDTKVHKANFLVTKHTRILLCHICHNPWPWSA-SGPTLPP-SFSFCLSCHAANSND
10. AL7G25510__Org	K-KCDLCSG-VARMYCSDQASLCWDCDGKVHGANFLVAKHTRCLLCSACSQSLTPWKASGLRLLGP-TFSVCDSCVALKNAGGG
11. AL8G29820__Org	KK-K-ELCCG-VARMYCSDQASLCWDCDGKVHGANFLVAKHTRCLLCSACSQSHTSWKA-SGLRLLGP-TVSICESCLARKKNHSS
12. Aqcoe1G384600	K-E-ELCNK-PARMYCSDQASLCWECDFKVHSANFLVAKHSTRNLLCHVCQSITPWKA-SGSKLGP-TITACERCANRCGFQEN
13. Aqcoe7G312300	KK-E-ELCSL-PAKTYCESDEASLCWNCDWKIHSAANFLVAKHESRLLCHICSFPQTQWKS-SGSKLGS-TLSICEKCVKNNCKRRS
14. Araha.11735s0007	KK-K-ELCCG-VARIYCESDQASLCWDCDGKVHGANFLVAKHTRCLLCSACSQSHTSWKA-SGLRLLGP-TVSICDSCCLARKKNSS
15. Araha.18479s0007	K-KCDLCSG-VARMYCSDQASLCWDCDGKVHGANFLVAKHTRCLLCSACSQSLTPWKA-SGLRLLGP-TFSVCDSCVALKNAGGG
16. AT4G27310__Org	K-KCDLNG-VARMYCSDQASLCWDCDGKVHGANFLVAKHTRCLLCSACSQSLTPWKA-TGLRLLGP-TFSVCDSCVALKNAGGG
17. AT5G54470__Org	KK-K-ELCCG-VARMYCSDQASLCWDCDGKVHGANFLVAKHMRCLLCSACSQSHTPWKA-SGLNLLGP-TVSICESCLARKKNNS
18. Bol021177__Org_B	-MYCESDQASLCWNCDGKVHGANFLVAKHTRCLLCTSQCQSLTPWKA-TGLRLLGP-TFSVCDSCVALKSAAG
19. Bol042329__Org_B	K-KCDLCEG-VARMYCSDQASLCWDCDGKVHGANFLVAKHTRCLLCSACSQSLTPWKA-AGLRLLGP-TFSVCDSCVALKTAGGG
20. Bostr.26833s0826	GKK-K-ELCCG-VARMYCSDQASLCWDCDAKVHGANFLVAKHTRCLLCSACSQSHTPWKA-SGLRLLGP-TVSICESCLARKKSSVA
21. Bostr.7867s0551	K-KCDLCSG-VARMYCSDQASLCWDCDGKVHGANFLVAKHTRCLLCSACSQSLTPWKA-TGLRLLGP-TFSVCDSCVAIKHAAAA
22. Bradi3g15490__Org	DG-GKGSGKG-G-ELCGA-AARVYCCADDATLCWGCDAQVHGANFLVARHARALLCRGCARPPTWRA-AGPRLGP-TASLCERCVRGRG-GGGDG
23. Brara.A01701_Org	K-KCDLCEG-VARMYCSDQASLCWNCDAAKVHGANFLVAKHTRCLLCSACSQSLTPWKA-TGLRLLGP-TFSVCDSCVALKSAAGG
24. Brara.B01378_Org	KM-K-ELCEG-VARMFCSDDQASLCWDCDGNVHGANFLVAKHARCCLLCSACSQSLTPWKA-SGLRLCP-TVSICESCLARKNNSGA
25. Brara.H01521__Org	KK-KCDLCDG-VARMYCSDQASLCWNCDGKVHGANFLVAKHTRCLLCTSQCQSHTPWKA-TGLRLLGP-TFSVCYSVALKSAAG
26. Brara.J00918__Org	KK-N-ELCDG-VARMFCSDDQASLCWNCDGKVHGANFLVAKHTRCLLCSVCQSPTPWKA-SGLRFGP-TVSVCESCLARKNKTLS
27. Brara.K00095__Org	K-KCDLCEG-VARTYCSDQASLCWDCDGKVHGANFLVAKHTRCLLCSACSQSLTPWKA-AGLRLLGP-TFSVCDSCVALKSAAGG
28. Brast03G030200	DG-GGKAGKG-A-ELCGA-AARVYCCADDATLCWGCDAQVHGANFLVARHARALLCRGCARPPTWRA-AGPRLGP-TASLCERCVRGRGGGEE
29. Cagra.0570s0039	K-KCDLCSG-VARMYCSDQASLCWDCDGKVHGANFLVAKHTRCLLCSACSQSLTPWKA-TGLRLLGP-TFSVCDSCVALKNAPGA
30. Caora.1629s0023	KK-K-ELCSG-VARMYCSDDASLCWDCDGKVHGANFLVAKHTRCLLCSACSQSLTPWKA-SGLRLRP-TVSICESCCLSRKKKKS

Site # 1 with w/o gaps

Passos para se obter uma reconstrução filogenética

1. Obter sequencias
2. Alinhar as sequencias
3. Reconstruir a filogenia

Passos para se obter uma reconstrução filogenética

3. Reconstruir a filogenia

Interpretação do alinhamento
algoritmos (NJ, ML)
matrizes de substituição

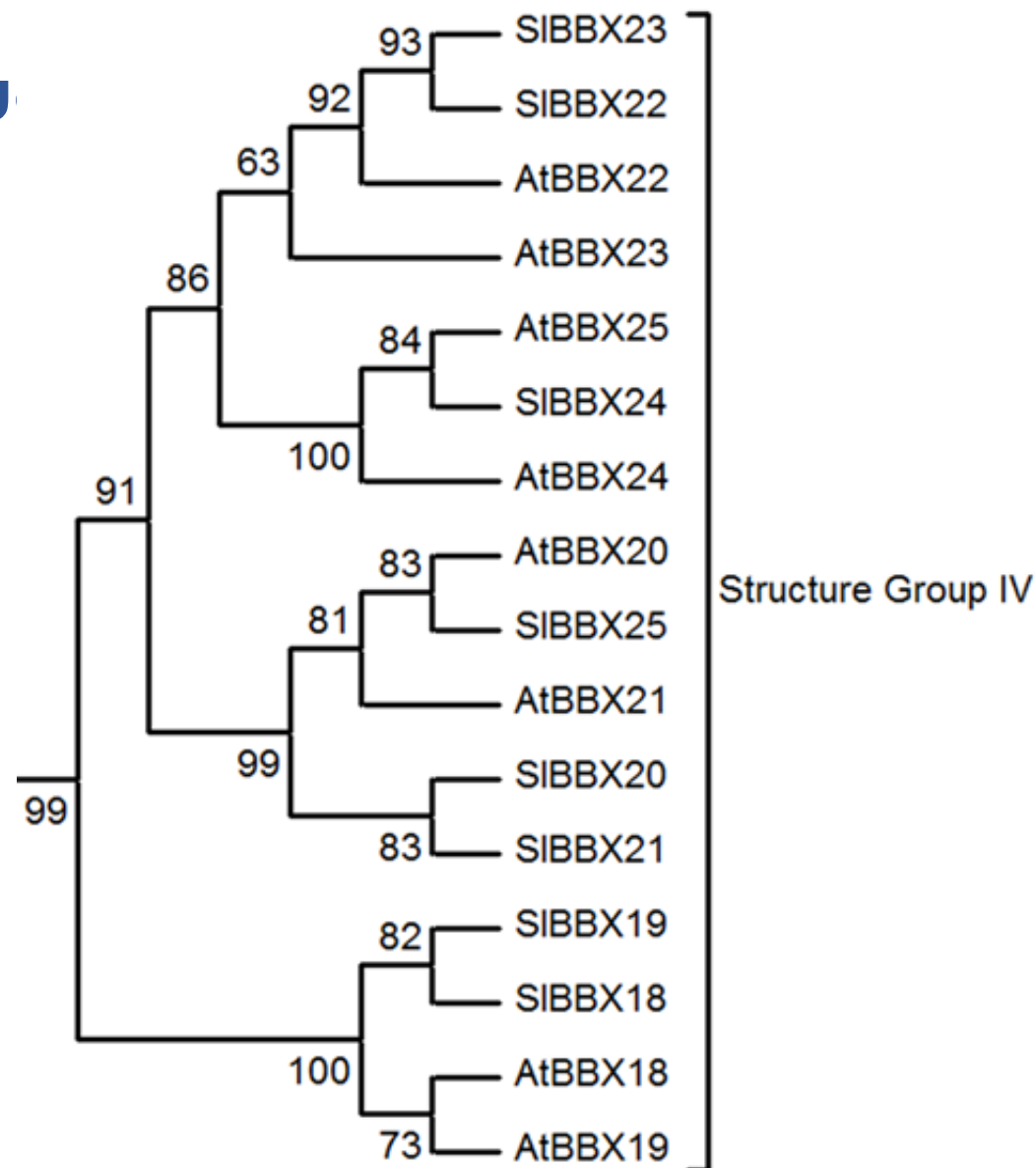
Validação
Réplicas (bootstrap)
Inferência estatística

Passos para se obter uma reconstru

3. Reconstruir a filogenia

Interpretação do alinhamento
algoritmos (NJ, ML)
matrizes de substituição

Validação
Réplicas (bootstrap)
Inferência estatística



Ferramentas

Bancos de dados

NCBI - <https://www.ncbi.nlm.nih.gov/>

Arabidopsis (TAIR) - <https://www.arabidopsis.org/>

Phytozome - <https://phytozome.jgi.doe.gov/>

TF db - <http://planttfdb.gao-lab.org/>

UniProt - <https://www.uniprot.org/>

Alinhamento e reconstrução filogenética

<https://tcoffee.crg.eu/apps/tcoffee/index.html>

<https://www.hiv.lanl.gov/content/sequence/PHYML/interface.html>

<http://www.atgc-montpellier.fr/phyml/usersguide.php>

<https://ngphylogeny.fr/>

Ferramentas

Elementos nas sequências

Motivos em promotores - <http://plantpan.itps.ncku.edu.tw/>

Domínios em proteínas:

Interpro - <https://www.ebi.ac.uk/interpro/search/sequence/>

Prosite - <https://prosite.expasy.org/scanprosite/>

Peptídeo sinal -

<https://services.healthtech.dtu.dk/services/SignalP-5.0/>


Interação PxP - <https://www.expasy.org/resources/string>

Ferramentas







Portal de recursos – Expasy

Expasy

Swiss Bioinformatics Resource Portal



e.g. [BLAST](#), [UniProt](#), [MSH6](#), [Albumin](#)...

<input type="checkbox"/>  Genes & Genomes <ul style="list-style-type: none"><input type="checkbox"/> Genomics<input type="checkbox"/> Metagenomics<input type="checkbox"/> Transcriptomics	<input type="checkbox"/>  Structural Biology <ul style="list-style-type: none"><input type="checkbox"/> Drug design<input type="checkbox"/> Medicinal chemistry<input type="checkbox"/> Structural analysis
<input type="checkbox"/>  Proteins & Proteomes	<input type="checkbox"/>  Systems Biology <ul style="list-style-type: none"><input type="checkbox"/> Glycomics<input type="checkbox"/> Lipidomics<input type="checkbox"/> Metabolomics
<input type="checkbox"/>  Evolution & Phylogeny <ul style="list-style-type: none"><input type="checkbox"/> Evolution biology<input type="checkbox"/> Population genetics	<input type="checkbox"/>  Text mining & Machine learning

Ferramentas

Desenho de estruturas

Gene - <http://wormweb.org/exonintron>

Gene/Proteína (IBS) - <https://mybiosoftware.com/ibs-illustrator-of-biological-sequences.html>

Proteína - <https://prosite.expasy.org/mydomains/>