

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

Atividade 4

Wellingson assunção Araújo

Usando o banco de dados do UniProt analisar a família de proteínas conhecida como "Squamosa promoter binding protein-like (**SBP** or **SPL**) family of transcription factors". Estudar a função dessas proteínas em *Arabidopsis thaliana* alinhando todos os membros.

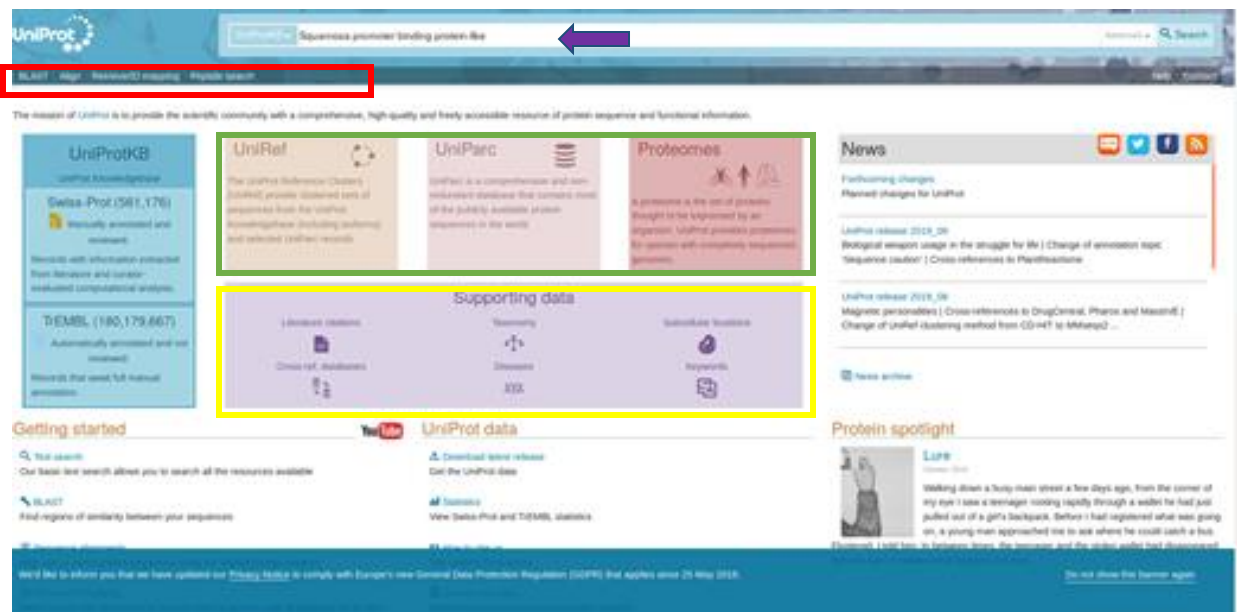
Artigo: Molecular Characterization of *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL)* Gene Family in *Betula luminifera*
Xiu-Yun Li, ErPei Lin, Hua-Hong Huang, Ming-Yue Niu, Zai-Kang Tong and Jun-Hong Zhang

PASSO 1 – Conhecendo o Uniprot

Aqui se buscou entender um pouco da interface do banco e como é organizado

O **UniProt** é um grande banco de dados que tem como objetivo fornecer à comunidade científica um recurso abrangente de sequências de proteínas e informações funcionais de alta qualidade e acesso livre.

Neste banco além de sequencias de proteínas é possível ter acesso a proteomas de diferentes espécies, clusters de proteínas, citações de literatura, referencias cruzadas dentre outras ferramentas.



➡ Campo de Busca

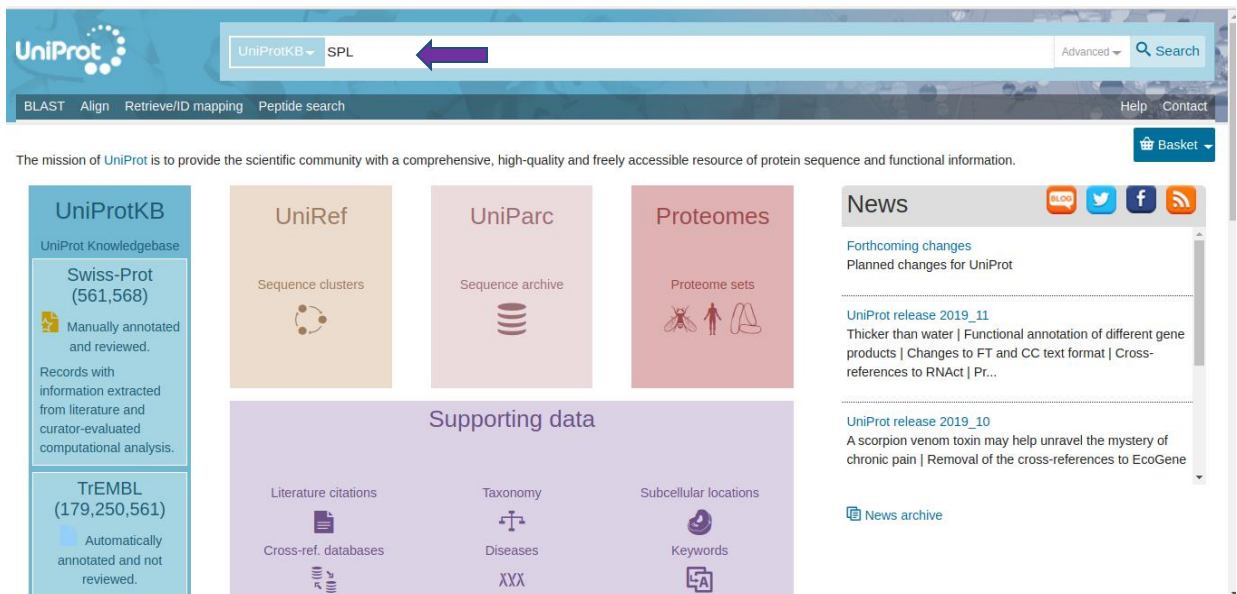
▭ Ferramentas principais

▭ Composição do banco de dados

▭ Alguns dados auxiliares fornecidos pelo banco

PASSO 2 – Realizar a pesquisa

Nesse momento foi informado no campo de busca o nome da família de proteínas



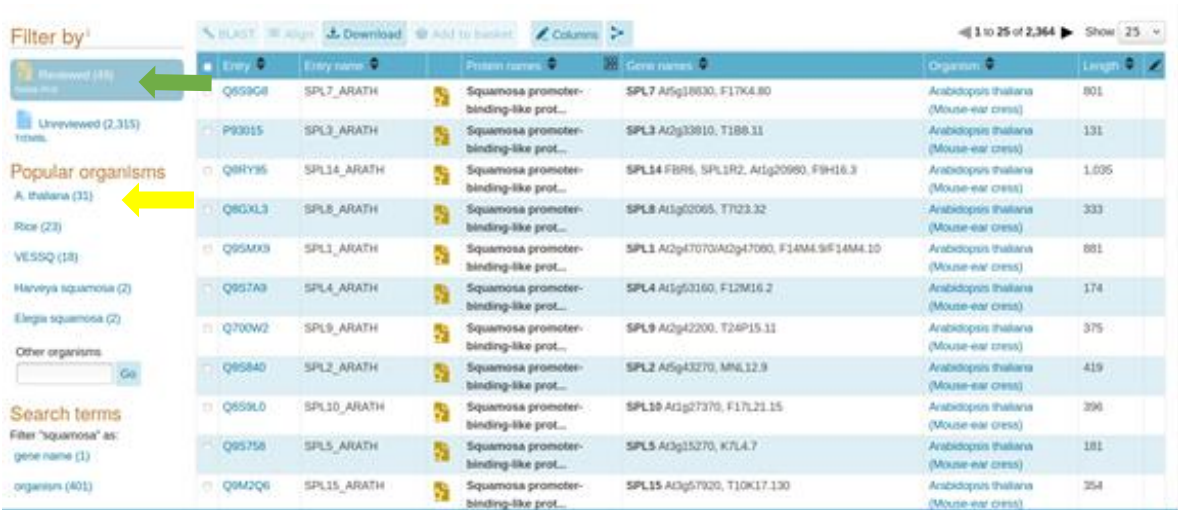
➡ Busca pela família de proteínas SPL

PASSO 3 – Refinar os resultados

Nesse momento foram mostrados todos os resultados possíveis, entretanto foi feito um refino selecionando a espécie (*Arabidopsis thaliana*) e também apenas proteínas anotadas e revisas manualmente (Swiss-Prot)

O UniProt fornece dois tipos de anotações e é possível trabalhar com as duas ou se escolher uma das categorias. São elas

- **Swiss-Prot - Revisado**
Registros anotados manualmente com informações extraídas da literatura e análise computacional avaliada pelo curador.
- **TrEMBL - Não revisado**
Registros analisados computacionalmente que aguardam anotação manual completa.



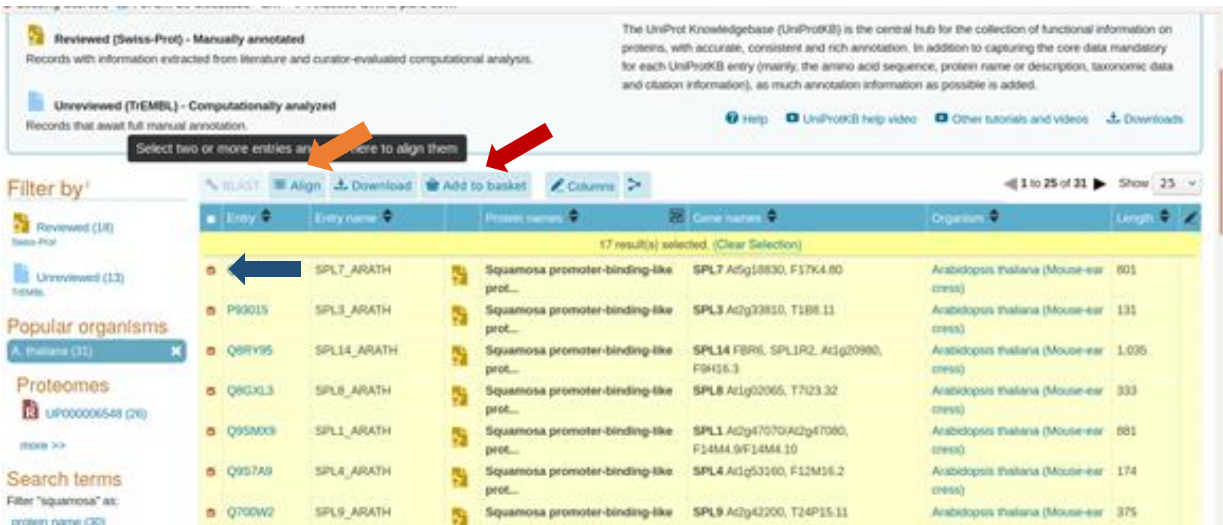
The screenshot shows the UniProt search results page. On the left, the 'Filter by' sidebar has 'Reviewed (48)' selected under 'Data type' and 'A. thaliana (31)' selected under 'Popular organisms'. The main table displays protein entries for *Squamosa promoter-binding-like protein* in *Arabidopsis thaliana*. The table has columns for Entry, Entry name, Protein names, Gene names, Organism, and Length. The first row is Q559G8, SPL7_ARATH, with a length of 801. A green arrow points to the 'Reviewed (48)' filter, and a yellow arrow points to the 'A. thaliana (31)' filter.

Entry	Entry name	Protein names	Gene names	Organism	Length
Q559G8	SPL7_ARATH	Squamosa promoter-binding-like prot...	SPL7 At5g18830, F17K4.80	Arabidopsis thaliana (Mouse-ear cress)	801
P90015	SPL3_ARATH	Squamosa promoter-binding-like prot...	SPL3 At2g33810, T188.11	Arabidopsis thaliana (Mouse-ear cress)	131
Q8RY95	SPL14_ARATH	Squamosa promoter-binding-like prot...	SPL14 FBR6, SPL1R2, At1g20980, F9H16.3	Arabidopsis thaliana (Mouse-ear cress)	1,035
Q8GXJ3	SPL8_ARATH	Squamosa promoter-binding-like prot...	SPL8 At1g02065, T723.32	Arabidopsis thaliana (Mouse-ear cress)	333
Q95MX9	SPL1_ARATH	Squamosa promoter-binding-like prot...	SPL1 At2g47070/At2g47080, F14M4.9/F14M4.10	Arabidopsis thaliana (Mouse-ear cress)	881
Q957A9	SPL4_ARATH	Squamosa promoter-binding-like prot...	SPL4 At1g53160, F12M16.2	Arabidopsis thaliana (Mouse-ear cress)	174
Q700W2	SPL9_ARATH	Squamosa promoter-binding-like prot...	SPL9 At2g42200, T24P15.11	Arabidopsis thaliana (Mouse-ear cress)	375
Q95M40	SPL2_ARATH	Squamosa promoter-binding-like prot...	SPL2 At5g43270, M6.12.9	Arabidopsis thaliana (Mouse-ear cress)	419
Q559L0	SPL10_ARATH	Squamosa promoter-binding-like prot...	SPL10 At1g27370, F17L21.15	Arabidopsis thaliana (Mouse-ear cress)	296
Q95758	SPL5_ARATH	Squamosa promoter-binding-like prot...	SPL5 At3g15270, K7L4.7	Arabidopsis thaliana (Mouse-ear cress)	181
Q9M2Q6	SPL15_ARATH	Squamosa promoter-binding-like prot...	SPL15 At3g57920, T10K17.130	Arabidopsis thaliana (Mouse-ear cress)	254

- ➡ Filtragem para resultados Swiss-Prot
- ➡ Filtragem para resultados em *Arabidopsis thaliana*

PASSO 4 – Seleção das sequencias para o alinhamento

Nesse momento foram selecionados os resultados de interesse e estes foram adicionados a cesta que é um recurso lúdico que guarda todas as sequencias permitindo que se continue pesquisa e se guarde todas de interesse em um só lugar, depois foi realizado o alinhamento múltiplo



The screenshot shows the UniProt search results page with the 'Align' button highlighted. The table displays protein entries for *Squamosa promoter-binding-like protein* in *Arabidopsis thaliana*. The first row is Q559G8, SPL7_ARATH, with a length of 801. An orange arrow points to the 'Align' button, a blue arrow points to the 'A. thaliana (31)' filter, and a red arrow points to the 'Add to basket' button.

Entry	Entry name	Protein names	Gene names	Organism	Length
Q559G8	SPL7_ARATH	Squamosa promoter-binding-like prot...	SPL7 At5g18830, F17K4.80	Arabidopsis thaliana (Mouse-ear cress)	801
P90015	SPL3_ARATH	Squamosa promoter-binding-like prot...	SPL3 At2g33810, T188.11	Arabidopsis thaliana (Mouse-ear cress)	131
Q8RY95	SPL14_ARATH	Squamosa promoter-binding-like prot...	SPL14 FBR6, SPL1R2, At1g20980, F9H16.3	Arabidopsis thaliana (Mouse-ear cress)	1,035
Q8GXJ3	SPL8_ARATH	Squamosa promoter-binding-like prot...	SPL8 At1g02065, T723.32	Arabidopsis thaliana (Mouse-ear cress)	333
Q95MX9	SPL1_ARATH	Squamosa promoter-binding-like prot...	SPL1 At2g47070/At2g47080, F14M4.9/F14M4.10	Arabidopsis thaliana (Mouse-ear cress)	881
Q957A9	SPL4_ARATH	Squamosa promoter-binding-like prot...	SPL4 At1g53160, F12M16.2	Arabidopsis thaliana (Mouse-ear cress)	174
Q700W2	SPL9_ARATH	Squamosa promoter-binding-like prot...	SPL9 At2g42200, T24P15.11	Arabidopsis thaliana (Mouse-ear cress)	375

- ➡ Alinhamento múltiplo
- ➡ Seleção dos resultados em *Arabidopsis thaliana*
- ➡ Cesta de sequencias

PASSO 5 – Análise das sequencias

o tempo de análise varia conforme a quantidade de sequencias, mas é bem rápido.
Abaixo é mostrado a tela indicadora das análises



PASSO 6 – interpretação das saídas

A ferramenta de Alinhamento Múltiplo proporciona ao usuário diferentes abordagens a depender do objetivo da pesquisa
Foi selecionado alguns parâmetros para demonstrar algumas dessas abordagens

Inicialmente optamos por estudar o motivo conservado entre as proteínas o qual tem relação com a funcionalidade da proteína e conseguimos perceber a existência deste em todas as sequencias analisadas

How to print an alignment in color

Job status: COMPLETED

Highlight

Annotation

☒ Beta strand

☐ Compositional bias

☐ Region

☐ Metal binding

☐ Alternative sequence

☐ Helix

☐ Repeat

☐ Sequence conflict

☐ Chain

☐ Mutagenesis

☐ Zinc finger

☐ Turn

Amino acid properties

☐ Similarity

☐ Hydrophobic

☐ Negative

☐ Positive

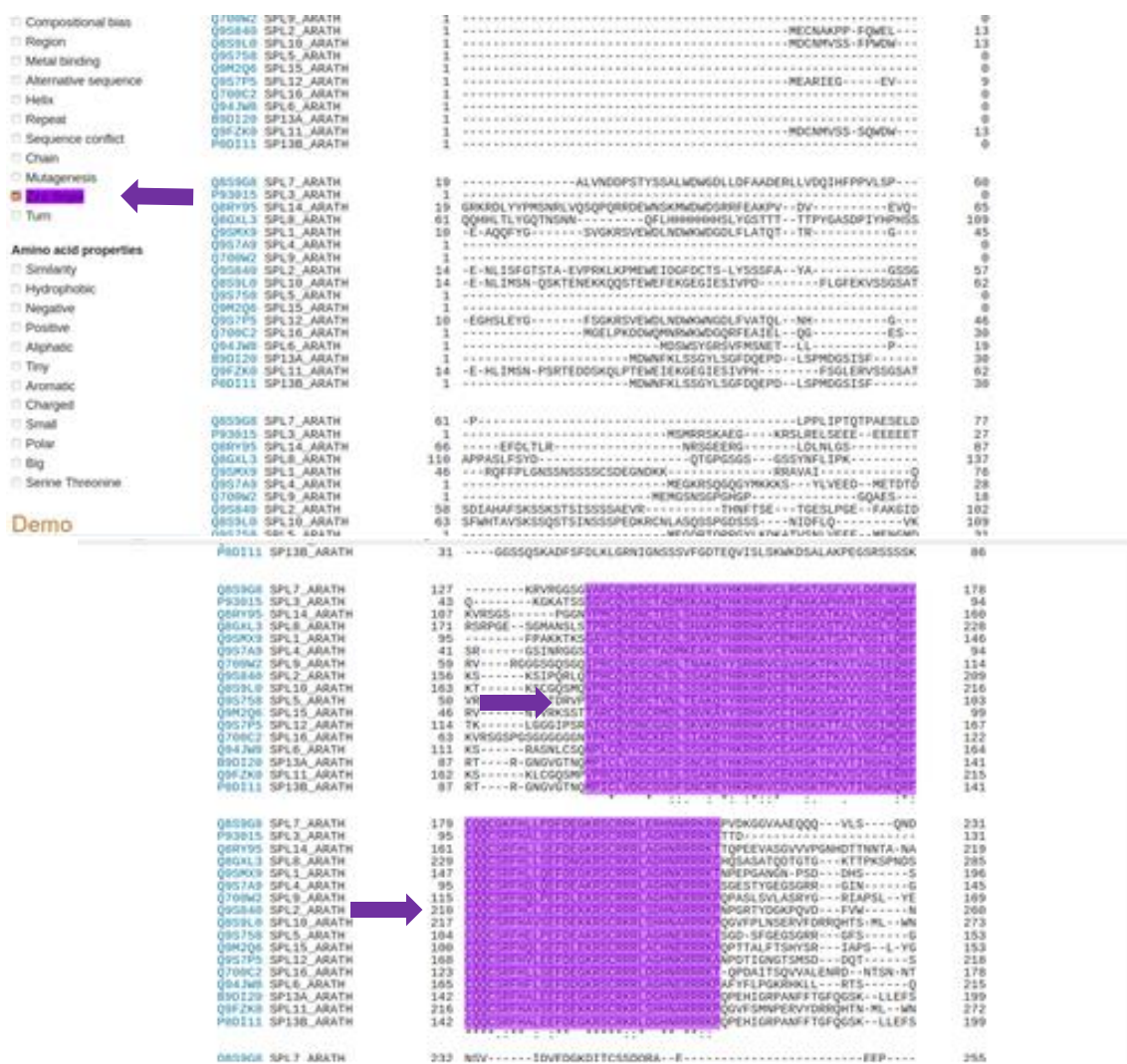
☐ Aliphatic

☐ Tiny

Q05908	SPL7_ARATH	1NSSLSQSPF.....	18
P93015	SPL3_ARATH	1	0
Q0RY35	SPL14_ARATH	1HDEVSQAQVAAPRPHQS...	18
Q05X13	SPL8_ARATH	1	MLDYEMWNPSSIVLSGDERNPPOSOTRSSSFDPDPISHYNDHRHITISPPLLSSFSVQQ	60
Q05MX9	SPL1_ARATH	1HEARIDE.....GG...	9
Q05TA9	SPL4_ARATH	1	0
Q700W2	SPL9_ARATH	1	0
Q05E40	SPL2_ARATH	1MECAKFP...FQEL...	13
Q05910	SPL10_ARATH	1MDCNVSS...FQWQ...	13
Q05T58	SPL5_ARATH	1	0
Q0M2Q6	SPL15_ARATH	1	0
Q05TP5	SPL12_ARATH	1HEARIEG.....EV...	9
Q700C2	SPL16_ARATH	1	0
Q942W0	SPL6_ARATH	1	0
E00120	SP13A_ARATH	1	0
Q9F2K0	SPL11_ARATH	1MDCNVSS...SQWQ...	13
P00111	SP13B_ARATH	1	0
Q05908	SPL7_ARATH	19ALVNDPSTYSSALNDWGDLLDFAADERLLVDQIHFPVLSP...	60
P93015	SPL3_ARATH	19	0
Q0RY35	SPL14_ARATH	19	GRKRLLVYPHSMRLVGGQPGQDHWASKHWDSRRFEAKPV...EVQ...	65
Q05X13	SPL8_ARATH	61	Q05HLLTLVGGTNSM...QLHSMHMSLVSSTT...TTFYASDPZYHMS	164
Q05MX9	SPL1_ARATH	10	...E...AQGFYS...SVGKRSVVDLNDWMDGLFLATQT...TR...	45
Q05TA9	SPL4_ARATH	1	0
Q700W2	SPL9_ARATH	1	0
Q05E40	SPL2_ARATH	14	...E...NLISFGTSTA...EVPKLPKMEWIDGFOCTS...LYSSFA...YA...	57
Q05910	SPL10_ARATH	14	...E...NLIRSN...QSKTENKKQSTEWFEKGGEGIESIVP...	62
Q05T58	SPL5_ARATH	1	0
Q0M2Q6	SPL15_ARATH	1	0
Q05TP5	SPL12_ARATH	10	...E...EGHSLEYE...FSGKRSVVDLNDWMDGLFLVATQL...NH...	46
Q700C2	SPL16_ARATH	1	...MSELPKDQWQMRNWDGQRFEAIEL...QG...	30
Q942W0	SPL6_ARATH	1	0
E00120	SP13A_ARATH	1	0
Q9F2K0	SPL11_ARATH	1	0
P00111	SP13B_ARATH	1	0
Q700W2	SPL9_ARATH	59	RV...RGGGSGGGQEPKQVEGSCMDLTNAKGYYSRHHVGVNSKTPKYTVAGIEGRF	114
Q05E40	SPL2_ARATH	156	KS...KSIQPRLOTPHQVEGCNLDLSSAKDYHRRHICENHSKFFKVVVSGVERRF	209
Q05910	SPL10_ARATH	103	KT...KSCGGSMVPRQIDGCELDLSSKDYHRRHVCETHSKCPWVYVSGLERRF	216
Q05T58	SPL5_ARATH	59	VR...GPSTQVPSRLQVDNCTVNLTEAKDYHRRHVCETHAKASAATVAGVQGRF	103
Q05MX9	SPL1_ARATH	46	RV...NTVRKSSTTARQVEGCRMDLNVKAVYYSRHKVCCIHSSKSKVIVSGLHGRF	99
Q05TP5	SPL12_ARATH	114	TK...LGGGIPRAICQVDNCGADLSKVDYHRRHVCETHSKATLALVGGIHGRF	167
Q700C2	SPL16_ARATH	61	KVRSGPSGGGGGAPYKGVNCKEQLSIKDYHRRHVCETHSKATLALVGGIHGRF	122
Q942W0	SPL6_ARATH	111	KS...RASNLCQNPLCQVYGCSKDLSSKDYHRRHVCETHSKATLALVGGIHGRF	164
E00120	SP13A_ARATH	07	RT...R...GNVSTNQPICLVDGCSDFSNVREYHRRHVCETHSKATLALVGGIHGRF	141
Q9F2K0	SPL11_ARATH	162	KS...KLCQGSMPVPRQIDGCELDLSSAKDYHRRHVCETHSKATLALVGGIHGRF	215
P00111	SP13B_ARATH	07	RT...R...GNVSTNQPICLVDGCSDFSNVREYHRRHVCETHSKATLALVGGIHGRF	141
Q05908	SPL7_ARATH	219	CQCKPHLLPQDEK...SCRRHLEHNRK...KPVKGGVAAEQQ...VLS...QND	231
P93015	SPL3_ARATH	95	CQCSRPHLLSEFDEK...SRHLEHNRK...STTD	131
Q0RY35	SPL14_ARATH	161	CQCSRPHLLSEFDEK...SRHLEHNRK...TTQPEVAGGVVPGHDTTNNTA...NA	219
Q05X13	SPL8_ARATH	229	CQCSRPHLLSEFDEK...SRHLEHNRK...CHQSASATQGTGTG...KTFKSPMDL	285
Q05MX9	SPL1_ARATH	147	CQCSRPHLLSEFDEK...SRHLEHNRK...TNPEPGANGN...PSD...DHS...S	198
Q05TA9	SPL4_ARATH	95	CQCSRPHLLSEFDEK...SRHLEHNRK...AGASTQDSSGSESTYGESSRR...GIN...	145
Q700W2	SPL9_ARATH	115	CQCSRPHLLSEFDEK...SRHLEHNRK...QOPASLVLASRYD...RIAPSL...YE	169
Q05E40	SPL2_ARATH	218	CQCSRPHLLSEFDEK...SRHLEHNRK...PWGRTYDGKQVQD...FVW...N	268
Q05910	SPL10_ARATH	217	CQCSRPHLLSEFDEK...SRHLEHNRK...QGVYPLNSEVPDRQNTS...HL...N	273
Q05T58	SPL5_ARATH	104	CQCSRPHLLSEFDEK...SRHLEHNRK...ISGQ...SFGSGGAK...GPS...S	153
Q0M2Q6	SPL15_ARATH	100	CQCSRPHLLSEFDEK...SRHLEHNRK...QPTTALFTSHYSR...IAPS...L...YS	153
Q05TP5	SPL12_ARATH	168	CQCSRPHLLSEFDEK...SRHLEHNRK...ANPOTISNGTSMRD...DQT...S	218
Q700C2	SPL16_ARATH	123	CQCSRPHLLSEFDEK...SRHLEHNRK...QPDWITSDVVALENRD...NTSN...N	178
Q942W0	SPL6_ARATH	165	CQCSRPHLLSEFDEK...SRHLEHNRK...PAFYFLPKGRHKL...RTS...Q	215
E00120	SP13A_ARATH	142	CQCSRPHLLSEFDEK...SRHLEHNRK...TGASRGGG...QPEHIGRPAF...TGFQSSK...LLEFS	199
Q9F2K0	SPL11_ARATH	216	CQCSRPHLLSEFDEK...SRHLEHNRK...QGVVSNRERYRERQNTN...HL...N	272
P00111	SP13B_ARATH	142	CQCSRPHLLSEFDEK...SRHLEHNRK...QPEHIGRPAF...TGFQSSK...LLEFS	199
Q05908	SPL7_ARATH	232	NSV...IDVEDGKDITCSSDQRL...E...	255
P93015	SPL3_ARATH	132	...	131
Q0RY35	SPL14_ARATH	220	NMD...LMALLALACAGGNVKKPPVGPSP...	250
Q05X13	SPL8_ARATH	206	GVKSSSPSSAPPTISLECFRQRFQTASSST...	322
Q05MX9	SPL1_ARATH	197	NYL...LITLLKIL...SAP...H...N...H...	217
Q05TA9	SPL4_ARATH	146	QVY...M...QNG...E...SSRVEH...	162
Q700W2	SPL9_ARATH	170	NGD...AGNSSF...SAG...EIGMPSSK...	195
Q05E40	SPL2_ARATH	261	RFA...LINPRS...EEK...F...IMPSSK...	283
Q05910	SPL10_ARATH	274	GLS...LNTWSEKY...TWGT...T...	299
Q05T58	SPL5_ARATH	154	QLI...QTO...E...SRNVDS...	169

➡ Motivo

Os dedos de zinco são domínios proteicos que têm a propriedade de se ligarem ao DNA. Este é composto por duas folhas-beta antiparalelas e uma alfa-hélice. O íon de zinco é fundamental para garantir a estabilidade deste tipo de domínio.



➡ Zinc finger protein

Por fim também foi obtida um dendrograma em forma de árvore que indica a ligação filogenética das proteínas. É mostrado de forma clara a grande relação que as proteínas possuem

