

DETERMINAÇÃO DA ESTRUTURA ORGANIZACIONAL DAS VIAS
DE MAP KINASES EM SORGO, *Arabidopsis lyrata* E CANA-DE-
AÇÚCAR POR MEIO DE ANÁLISE DE BIOINFORMÁTICA

VERÔNICA AGUIAR DA SILVA

UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE
DARCY RIBEIRO - UENF

CAMPOS DOS GOYTACAZES - RJ
AGOSTO – 2010

DETERMINAÇÃO DA ESTRUTURA ORGANIZACIONAL DAS VIAS
DE MAP KINASES EM SORGO, *Arabidopsis lyrata*, E CANA-DE-
AÇÚCAR POR MEIO DE ANÁLISE DE BIOINFORMÁTICA

VERÔNICA AGUIAR DA SILVA

“Tese apresentada ao Centro de Ciências e
Tecnologia Agropecuária da Universidade
Estadual do Norte Fluminense Darcy Ribeiro,
como parte das exigências para obtenção do
título de Doutor em Genética e Melhoramento
de Plantas”.

Orientador: Prof. Gonçalo Apolinário de Souza Filho

CAMPOS DOS GOYTACAZES - RJ
AGOSTO – 2010

DETERMINAÇÃO DA ESTRUTURA ORGANIZACIONAL DAS VIAS
DE MAP KINASES EM SORGO, *Arabidopsis lyrata* E CANA-DE-
AÇÚCAR POR MEIO DE ANÁLISE DE BIOINFORMÁTICA

VERÔNICA AGUIAR DA SILVA

“Tese apresentada ao Centro de Ciências e Tecnologia Agropecuária da Universidade Estadual do Norte Fluminense Darcy Ribeiro, como parte das exigências para obtenção do título de Doutor em Genética e Melhoramento de Plantas”.

Aprovada em 23 de agosto de 2010.

Comissão Examinadora:

Prof. José Roberto da Silva (D.Sc. Biociências e Biotecnologia) - UFRJ

Prof. Victor M. Quintana Flores (D.Sc. em Biociências e Biotecnologia) - UENF

Prof. Jorge Hernandez Fernandez (D.Sc. Biologia Celular e Estrutural) - UENF

Prof. Gonçalo A. de Souza Filho (D.Sc. em Biociências e Biotecnologia) - UENF
(Orientador)

Dedicatória

Dedico esta tese a minha filha, Catarina Sayuri, que apesar dos seus 7 aninhos de vida tem sido a minha grande companheira durante todos estes anos de estudo e que, mesmo sendo privada de inúmeros momentos ao meu lado, ainda assim nunca deixou de demonstrar o seu amor incondicional e uma compreensão infinita. Te amo filha! Obrigada por tudo!

AGRADECIMENTOS

A minha mãe, Maria das Graças, e ao meu pai, Aguinaldo, pelo apoio incondicional durante todos estes anos.

Ao meu tio e padrinho, Evaldo, meu segundo pai e grande amigo, por estar sempre presente, sempre me orientando com suas sábias palavras nos momentos mais difíceis da minha vida.

Ao meu orientador, professor Dr. Gonçalo Apolinário pelo seu bom senso e notável inteligência.

Não poderia deixar de agradecer aos meus colegas, membros do grupo GONCA, pelas observações sinceras e amigas. Em especial gostaria de agradecer a Dr^a. Beatriz Ferreira.

Deixo aqui externado o meu agradecimento à professora e coordenadora do programa, Prof^a. Dra. Telma Nair Santana, e ao amigo Daniel, pelo apoio dado na reta final para a conclusão deste trabalho.

E, por fim, agradeço a todos os meus professores das diversas disciplinas e colegas de sala de aula pelo apoio e paciência nos momentos iniciais do programa.

SUMÁRIO

LISTA DE FIGURAS.....	v
ÍNDICE DE TABELAS.....	vii
ABREVIATURAS E SIMBOLOS.....	viii
RESUMO	ix
ABSTRACT	xii
I- INTRODUÇÃO	1
II- REVISÃO DE LITERATURA	5
2.1. As vias de sinalização em plantas.....	5
2.2. Vias MAP Kinases.....	6
2.3. Constituição das Vias de Map Kinases em Plantas.....	13
2.4. Base de dados públicos para o estudo de proteínas.....	24
2.5. Cana-de-açúcar, sorgo e <i>A. lyrata</i>	25
2.6. Objetivos.....	32
3. MATERIAL E MÉTODOS	33
4. RESULTADOS	45
5. DISCUSSÃO	68
6. CONCLUSÃO	74
7. REFERÊNCIAS BIBLIOGRÁFICAS	76
APÊNDICE.....	89

LISTA DE FIGURAS

Figura 1. Esq. ilustrativo da cascata de ativação das vias de MAPKs	7
Figura 2. Convergência da cascata de ativação das vias de MAP Kinases	9
Figura 3. Sítios que caracterizam as proteínas da família kinase	11
Figura 4. Assinaturas de consenso de MPK/MKK em animais e plantas	12
Figura 5. Genes homólogos, parálogos e ortólogos: esquema ilustrativo	14
Figura 6. Árvore filogenética de MPK de <i>A. thaliana</i> , <i>A. lyrata</i> e álamo	15
Figura 7. Árvore filogenética de MPK de <i>A. thaliana</i> , arroz, cana e sorgo	17
Figura 8. Interações de MKKs com MPKs em <i>A. thaliana</i>	21
Figura 9. Esquema de fosfatases que interagem com MPKs em <i>A. thaliana</i>	23
Figura 10. Esquema dos passos para obtenção dos ESTs	28
Figura 11. Anotação dos ESTs de cana-de-açúcar no SUCEST	28
Figura 12. Genes ortólogos em <i>A. thaliana</i> , sorgo, arroz e álamo.	30
Figura 13. Modelo de genes de <i>A. lyrata</i>	31
Figura 14. Esq. ilustrativo dos passos para obtenção dos ESTs em cana	37
Figura 15. Ferramentas utilizadas para obtenção das MPK/MKK de cana	39
Figura 16. Montagem dos transcritos de cana-de-açúcar	40
Figura 17. Esquema da etapa de sequenciamento dos clones	44
Figura 18. Alinhamento das sequências de MPK de sorgo, cana e <i>A. lyrata</i>	49
Figura 19. Árvores filogenética de MPKs de <i>A. thaliana</i> , <i>A. lyrata</i> e álamo	52
Figura 20. Árvore filogenética de MPKs de <i>A. thaliana</i> , arroz, sorgo e cana	53
Figura 21. Alinhamento das sequencias de ortólogos de MPK12	55
Figura 22. Árvore com as MPK12 de todas as espécies	56

Figura 23. Regiões de assinaturas conservadas de MKKs	58
Figura 24. Árvore de MKK de <i>A. thaliana</i> , <i>A. lyrata</i> e álamo	60
Figura 25. Árvore de MKK de <i>A. thaliana</i> , arroz, sorgo e cana-de-açúcar	61
Figura 26. Árvore de MKK das sequências de todas as espécies do clado D	62
Figura 27. Alinhamento de domínio NTF das sequências de MKK3	63
Figura 28. Visão conjunta dos ortólogos de MPKs de <i>A. thaliana</i>	65
Figura 29. Visão conjunta dos ortólogos de MKKs de <i>A. thaliana</i>	66
Figura 30. Componentes MPK/MKK predominantes em dicotiledôneas	66

ÍNDICE DE TABELAS

Tabela 1 - Relação dos grupos de tecidos de cana-de-açúcar obtidos a partir das bibliotecas de cDNA de cana-de-açúcar do projeto SUCEST.....	27
Tabela 2 - Nomenclatura para MPK e MPKK em <i>A. thaliana</i> (<i>At</i>), <i>A. lyrata</i> (<i>Al</i>) e <i>Populus trichocarpa</i> (<i>Pt</i>).....	47
Tabela 3 - Nomenclatura para MPK e MPKK em <i>Sorgo bicolor</i> (<i>Sb</i>), <i>Oryza sativa</i> (<i>Os</i>) e Cana-de-açúcar (<i>Sc</i>)	48

ABREVIATURAS E SÍMBOLOS

Al	<i>Arabidopsis lyrata</i>
At	<i>Arabidopsis thaliana</i>
ATP	Adenosina trifosfato
BAR	The Bio-Array Resource for Plant Biology
cDNA	Ácido desoxirribonucléico complementar
ESTs	Expressed sequence tags
MAPK ou MPK	Proteína Map Kinase
MAPKK ou MKK	Proteína Map Kinase Kinase
MAPKKK ou MKKK	Proteína Map Kinase Kinase Kinase
NTF	Fator de Transcrição Nuclear
Os	<i>Oryza Sativa</i>
PCR	Reação em cadeia de polimerase
PK	Proteína Kinase
Pt	<i>Populus Trichocarpa</i>
Sb	<i>Sorghum Bicolor</i>
Sc	Cana-de-açúcar

RESUMO

SILVA, Verônica Aguiar da; D.Sc.; Universidade Estadual do Norte Fluminense Darcy Ribeiro; Agosto 2010; Determinação da estrutura organizacional das vias MAP KINASES em sorgo, *Arabidopsis lyrata* e cana-de-açúcar por meio de análise de Bioinformática; Orientador: Gonçalo Apolinário de Souza Filho; Conselheiros: Victor M. Quintana Flores e Jorge Hernandez Fernandez.

A habilidade das plantas em se defender de inúmeros estresses bióticos e abióticos e coordenar processos de desenvolvimento está relacionada a uma subfamília das proteínas kinases (PK): MAP Kinases. As MAP Kinases são módulos universais de uma notável rede de sinais de transdução em eucariotos e tem sido relacionada na literatura a uma bem sucedida história de conservação evolucionária de genes. Estudos com MAP Kinases, em *A. thaliana*, revelaram que componentes destas famílias estão distribuídos em 4 clados (A, B, C e D). Em plantas, papéis biológicos têm sido associados a módulos MAP Kinases que pertencem a um mesmo clado. Entretanto, questões relacionadas ao processo de expansão, duplicação e caracterização destas vias em plantas ainda não estão claras. No presente trabalho, teve-se por principal objetivo, identificar os componentes MAP Kinases (módulos MKKs e MPKs) em sorgo, cana-de-açúcar e *A. lyrata* e verificar se a estrutura organizacional destas vias segue o mesmo padrão previamente estabelecido para *A. thaliana*, arroz e álamo. Nossas análises identificaram componentes MPKs e MKKs a partir dos bancos de dados públicos,

onde foram obtidas 16, 13 e 20 MPKs e 13, 5 e 10 MKKs em sorgo, cana-de-açúcar e *A. lyrata*, respectivamente. Eventos de expansão, observados no clado D de MPKs nestas três espécies reforçam evidências de duplicações de genes MPKs que ocorreram antes da divergência entre monocotiledôneas e dicotiledôneas, enquanto outras, aparentemente, ocorreram posteriormente, como pode ser observado no grupo de genes MPK21. A análise filogenética e o alinhamento conjunto das sequências MPK12 permitiram observar uma particularidade para este grupo de genes em todas as espécies analisadas. Estas proteínas apresentam a assinatura PK, -DLR-, atípica para estas famílias de genes. Com relação aos módulos MKKs, nossas análises também permitiram observar que, em sorgo, cana-de-açúcar e *A. lyrata*, módulos MKK3 também são caracterizados pelo fato de possuírem um domínio NTF (“Fator de Transferência Nuclear”) e são os únicos constituintes do clado B de MKKs. Além disso, foi possível observar que proteínas MPKs/MKKs de *A. thaliana* e *A. lyrata* não apresentam divergências evolutivas significativas. Por meio desta abordagem, também foi possível identificar componentes MPK/MKKs que são exclusivos de plantas dicotiledôneas, permanecendo monofiléticos ao longo do processo evolutivo. Em conjunto, os dados obtidos permitiram concluir que a organização das vias de MPKs/MKKs nas três espécies analisadas segue o modelo proposto inicialmente em *A. thaliana*, arroz e álamo e que estas famílias de genes continuam se expandindo, dando origem ao grupo de genes MPK21 do clado D. Entretanto, os dados sugerem uma futura confirmação *in vitro* e acredita-se que os resultados derivados desta abordagem devem ajudar a nortear os atuais trabalhos desenvolvidos nestas espécies em respostas a inúmeros estresses ambientais

ABSTRACT

SILVA, Verônica Aguiar da; D.Sc.; Universidade Estadual do Norte Fluminense Darcy Ribeiro; August 2010; Determination of the organizational structure of the routes of MAP KINASES in Sorghum Bicolor, Arabidopsis lyrata and sugarcane using analysis of Bioinformatics; Advisor: Gonçalo Apolinário de Souza Filho; Committee members: Victor M. Quintana Flores e Jorge Hernandez Fernandez

The ability of plants to defend themselves from innumerable biotic and abiotic stresses, and coordinate development processes, is related to a subfamily of kinases proteins (PK): MAP Kinases. The MAP Kinases are universal modules of a remarkable signal transduction network in eukaryotes and have been related on the literature to a successful history evolutionary conservation of genes. Studies with MAP Kinases, in *A. thaliana*, revealed that components of these families are distributed across 4 clades (A, B, C e D). In plants, biological roles have been associated to MAP Kinases modules that belong to a same clad. However, issues related to the process of expansion, duplication and characterization of these routes in plants are not yet clear. The present work had as main goal, to identify the components MAP Kinases (modules MKKs and MPKs) in sorghum bicolor, sugarcane and *A. lyrata* and to verify if the organizational structure of these routes follows the same pattern previously established for *A. thaliana*, rice and poplar. Our Analysis identified the components MPKs and MKKs, from public databases, where 16, 13 and 20 MPKs and 13, 5 and 10 MKKs in sorghum bicolor, sugarcane and *A. lyrata*, were obtained respectively. Expansion events, observed in clad D of MPKs in these 3 species,

reinforce evidence of MPKs genes duplications that occurred previous to the divergence between monocotyledons and dicotyledons, while others, apparently, occurred later, as can be observed in MPK21 genes group. The phylogeny analysis and the conjunct alignment of MPK12 sequences allowed observing a particularity to this group of genes in all analyzed species. These proteins show the PK signature, -DL R-, atypical for those gene families. In relation to the MKKs modules, our analysis also allowed to observe that, in sorghum bicolor, sugarcane and *A. lyrata*, MKK3 modules are also characterized by the fact of having a domain NTF ("Nuclear Transfer Factor") and are the only constituents of clad B of MKKs. Furthermore, it was possible to observe that MPKs/MKKs proteins of *A. thaliana* and *A. lyrata* do not present signifying evolutionary divergence. Through this approach, it has been also possible to identify MPK/MKKs components that are exclusive of dicotyledons plants, remaining monophyletic along the evolutionary process. Together, the data obtained allowed to conclude that the organization of MPKs/MKKs routes in the three analyzed species, follow the model originally proposed in *A. thaliana*, rice and poplar and that those genes families are still expanding, giving origin to the group of genes MPK21 of clad D. However, the data suggest a future confirmation in vitro and it is believed that the results derived from this approach, should help to guide the current works developed in these species in response to innumerable environmental stresses.

1. INTRODUÇÃO

A habilidade das plantas em se defender de inúmeros estresses bióticos e abióticos tem sido relacionada a uma notável rede de sinalização de proteínas, que atuam num processo de fosforização em cascata. Dentre as famílias de proteínas envolvidas neste complexo mecanismo de defesa, destacam-se as proteínas da família kinase ativadas por mitogênese (MAP Kinases). As MAP Kinases em plantas podem ser do tipo MAP Kinase Kinase Kinase (MKKK), que fosforilam proteínas kinases do tipo MAP Kinase Kinase (MKK), que por sua vez ativam as MAPKs (MPK). Pesquisas com MAP Kinases têm sido relacionadas a várias respostas a estresses bióticos e abióticos, respostas a hormônios, proliferação celular, diferenciação e a várias respostas programadas e adaptativas em eucariotos (Rodriguez et al., 2010; Nakagami et al., 2005).

Pesquisas têm revelado que MAP Kinases são módulos universais de uma notável rede de sinais de transdução em eucariotos e tem sido relacionada na literatura a uma bem sucedida história de conservação evolucionária de uma família de genes e seus componentes são codificados por componentes altamente conservados (Hamel et al., 2006). Estudos realizados com o pequeno genoma da *A. thaliana* revelaram 60, 10 e 20 genes MKKKs, MKKs e MPKs, respectivamente, e definiram uma nomenclatura sistemática para esta família de genes (MAPK Group, 2002). Posteriormente, esta nomenclatura foi considerada suficientemente robusta para ser estendida a 21 sequências MPKs de álamo e 15 de arroz (Hamel et al., 2006). Embora tenham sido observadas variações neste

padrão em outras espécies, a adoção de uma nomenclatura única parece ser apropriada para estabelecer um modelo de anotação para estas proteínas que compõe a arquitetura dos módulos MPKs/MKKs em eucariotos. A análise filogenética feita com MAP Kinases de *A. thaliana* e seus ortólogos em arroz e álamo situou os membros destas famílias em 4 clados (MAPK Group, 2002; Hamel et al., 2006). Pesquisas têm associado módulos MPKs/MKKs situados em clados específicos a inúmeros processos fisiológicos em plantas (Bardwell et al., 2009; Wang et al., 2008; Qiu et al., 2008; Seo et al., 2007; Brodersen et al., 2006).

A cana-de-açúcar e o sorgo são espécies próximas na escala evolutiva. Pesquisas têm se dedicado a um melhor entendimento da fisiologia, metabolismo e genética dessas duas espécies de plantas e têm demonstrado que o genoma de uma espécie pode ser estudado por meio do entendimento do genoma de espécies próximas. O alto grau de conservação do conjunto gênico (sintenia) entre espécies de Gramíneas (co-linearidade) tem sugerido que características identificadas em espécies com genomas grandes e complexos (neste caso, a cana-de-açúcar) poderiam ser dissecadas pelo isolamento dos mesmos genes em espécies relacionadas com genoma pequeno (neste caso, o sorgo) (Albino et al., 2006; KUITTINEN, H. et al., 2004). Dessa forma, a inclusão de dados provenientes do genoma do sorgo é uma estratégia para um melhor entendimento da organização estrutural dos componentes MAP Kinases que foram identificados em cana-de-açúcar a partir do banco de transcrito do SUCEST (<http://www.sucest-fun.org>), constituindo, portanto, o principal objetivo neste trabalho.

Como espécie modelo de Gramíneas, o pequeno genoma do sorgo (*Sorghum bicolor* L., 760 Mpb) é também o complemento lógico mais próximo do genoma de arroz (*Oryza sativa*, 440 Mpb), que encontra-se, atualmente, totalmente sequenciado (Kresovich et al., 2005). A mesma estratégia foi utilizada neste trabalho para a identificação das MAP Kinases em *A. lyrata* (genoma grande), uma vez que o genoma de *A. thaliana* (genoma pequeno) tem sido bastante estudado.

Bancos de expressão têm auxiliado diversas pesquisas na elucidação do papel biológico desempenhado pelos componentes MPKs e MKKs dentro do contexto de plantas superiores (Zhou, C et al., 2009; Guo et al., 2004).

Recentemente foram observados diversos componentes predominantes de MAP Kinases em *A. thaliana* e pesquisas sugerem a necessidade de se focar na identificação de substratos de MAP Kinases e compreender como esta especificidade é alcançada entre as rotas da sinalização de MAPKs (Rodriguez et al., 2010; Erik et al., 2009).

No presente trabalho, almejou-se ampliar a visão acerca do grau de conservação e diversidade estrutural em módulos MPKs/MKKs de plantas. Dentro deste contexto, uma análise conjunta desses componentes identificados em sorgo, cana-de-açúcar e *A. lyrata* e a comparação com os módulos MPKs/MKKs já identificados em *A. thaliana*, álamo e arroz ajudam a descrever um cenário comparativo que permite discutir algumas questões que, até o momento, não estavam claras. Como objetivo geral, espera-se definir se a organização destas vias seguem os mesmos padrões já identificados em outras espécies. Dessa forma, pretende-se com este estudo, identificar, por meio de análise *in silico*, os componentes MPK/MKK a partir dos bancos de dados genômicos públicos de sorgo (JGI, 2010) e *A. lyrata* (JGI, 2010), e do banco de transcritos da cana-de-açúcar de açúcar – SUCEST (SUCEST, 2010). Além disso, almeja-se discutir melhor o grau de conservação estrutural dos membros para estas famílias de genes e gerar a árvore filogenética para MPKs e MKKs de cana-de-açúcar, sorgo e *A. lyrata* em conjunto com *A. thaliana*, arroz e álamo, como o objetivo de averiguar se, para o primeiro grupo de plantas, seguem ou não os mesmos padrões já identificados para os segundos. Deseja-se também promover uma melhor discussão a respeito da expansão observada no clado D do conjunto de módulos MAP Kinases de cana-de-açúcar, sorgo e *A. lyrata*, também observados no genoma de álamo, arroz e *A. thaliana*, onde algumas questões permaneceram obscuras. Além disso, pretende-se investigar o conjunto de componente MAP Kinases que permaneceram monofiléticos e, portanto, estão presentes apenas no genoma de plantas dicotiledôneas e sua relação com os diversos processos biológicos por meio de investigação em banco de dados de expressão. Dessa forma, deseja-se estabelecer o grau de divergência/similaridade identificados em módulos MPKs/MKKs nas espécies monocotiledôneas e dicotiledôneas analisadas neste trabalho. Também objetiva-se neste trabalho identificar se cana-de-açúcar, sorgo e *A. lyrata* possuem o componente MKK3, que preserva o fator de transcrição NTF identificado em ortólogos de *A. thaliana*, arroz e álamo

(Hamel, 2006) e confirmar as suspeitas de que módulos MKK3 constituem os únicos membros que formam o clado B das MKKs em plantas. Além disso, neste estudo também objetiva-se verificar se *Arabidopsis (thaliana e lyrata)* apresentam peculiaridades em relação aos seus componentes MAP Kinases.

A determinação da estrutura organizacional das vias de sinalização de proteínas kinases ativadas por mitogênese em cana-de-açúcar, sorgo e *A. lyrata* por meio de análise “*in silico*” e os resultados decorrentes desta pesquisa sugerem uma futura confirmação “*in vitro*”. Acredita-se que os resultados derivados desta abordagem devem ajudar a nortear os atuais trabalhos desenvolvidos nestas espécies em respostas a inúmeros estresses ambientais, além de servir de estímulo para a caracterização “*in silico*” desta importante via de sinalização em plantas superiores, minimizando significativamente os custos de pesquisa.

2. REVISÃO DE LITERATURA

2.1. As vias de sinalização em plantas

A comunicação célula-ambiente e célula-célula (vias de transdução de sinais) é crucial para o desenvolvimento de organismos eucariotos. Plantas respondem a estímulos endógenos e exógenos, ativando uma rede de sinalização celular que compreende: sensores e sinalizadores de cálcio, sinalizadores de açúcar, respostas hormonais, cascatas de fosforilação e proteínas, que desempenham papéis específicos na coordenação de respostas celulares na fisiologia da planta, no seu desenvolvimento e processos de defesa e diferenciação (Rodrigues, 2010). Embora recentes pesquisas tenham permitido a identificação de muitas rotas de sinalização, pouco ainda se conhece sobre a integração dos processos de sinalização em plantas (Erik et al., 2009).

Proteínas da família kinases desempenham um importante papel dentro do processo de resposta utilizado por plantas e animais. Muitas kinases foram descobertas desde o início dos anos 80 e estima-se que o genoma humano apresente aproximadamente 2000 kinases (Bridges, 2002). As proteínas kinases são a maior família de proteínas em eucariotos (Manning, 2002) e são a chave central de comunicação no controle intracelular, regulação e transdução de sinais. O processo de fosforilação e desfosforilação de proteínas é catalisado por kinases e fosfatases, respectivamente, e muitos processos de transdução de sinais dependem de tal mecanismo reversível (Mehta et al., 2007). Dentro deste contexto, destacam-se as proteínas da grande família das kinases: as MAP

Kinases. Em todos os eucariotos, vias de sinalização composta por MAP Kinases (Proteínas Kinases Ativadas por Mitogênese) funcionam como centros regulatórios do crescimento, morte, diferenciação, proliferação celular e na fisiologia de hormônios, bem como na sinalização de estresses bióticos e abióticos (Erik et al., 2009; Nakagami, 2005). A organização hierárquica dessas três classes de kinases funcionalmente relacionadas (MKKKs, MKKs e as MPKs) permite que estas proteínas operem como transmissores de sinais capazes de eficientemente amplificar, integrar e canalizar informações entre o ambiente celular e os centros metabólicos e de transcrição das células animais e vegetais (Rodrigues, 2010). Os mecanismos regulatórios incluem vários fenômenos que vão desde alterações químicas e estruturais da proteína até o controle do processo de transcrição. Por este motivo, um detalhado entendimento das vias de sinalização que envolve as proteínas kinases, com especial interesse na subfamília MAP kinases, é foco de interesse de muitas pesquisas. Entretanto, até o presente momento, apenas alguns agentes envolvidos no processo de desativação de MPKs são conhecidos (Erik et al., 2009).

2.2. Vias MAP Kinases

As MAPKs foram primeiramente descobertas em 1986 a partir de experiências com células animais por Sturgill e Ray (Sturgill, 1986). A primeira proteína foi inicialmente denominada *microvolt* e, posteriormente, associada à proteína-2 kinase (MAP-2 kinase). Mais tarde, estas kinases foram relacionadas a processos de mitogênese, envolvendo a fosforização de resíduos de tirosina e, a partir de então, renomeadas *mitogen-activated protein kinase* (MAP kinase) (Rossomando, 1987).

O sequenciamento de genomas, tais como o de *Arabidopsis thaliana*, revelaram três classes de kinases ativadas por mitogênese que operam como sinalizadores em cascata: as **MKKKs** (ou MKKKs ou MEKKs – acrônimo de MAP kinase kinase kinase), que fosforilam **MKKs** (ou MKKs ou MEKs – acrônimo de MAP kinase kinase), que fosforilam **MPKs** (ou MPKs - acrônimo de MAP kinase), que fosforilam outros substratos das vias de sinalização (Erik et al., 2009; Nakagami, 2005). A cascata MAPKs consiste no mínimo nestes módulos de

componentes (MAPKKK → MAPKK → MAPK) que formam uma combinação de conexões entre eles.

Análises bioquímicas e genéticas apontam para uma rede complexa, na qual kinases de um nível podem ser ativadas por mais de uma kinases do nível acima, inclusive em espécies distantes filogeneticamente (Rodrigues, 2010). Há também evidências de que atividade de MAP kinases pode ser influenciada por proteínas acessórias (Mapes, 2004; Levchenko, 2000; Whitmarsh, 1998) e proteínas fosfatases (Stephen, 2000). Estas proteínas, por sua vez, podem ser ativadas por kinases, o que torna o sistema ainda mais complexo (Ver Figura 1).

As diferentes possibilidades de resposta nas vias de transdução são atribuídas não só à duração e intensidade de estímulo, mas também à interação com as vias paralelas de sinalização. Além disso, a existência de múltiplos genes codificando cada uma das três kinases, bem como as proteínas que, através da interação com estas kinases, modulam a propagação do sinal, confere uma complexidade adicional à via em organismos superiores. (Bardwell et al., 2009; Peyssonnaud e Eychène, 2001).

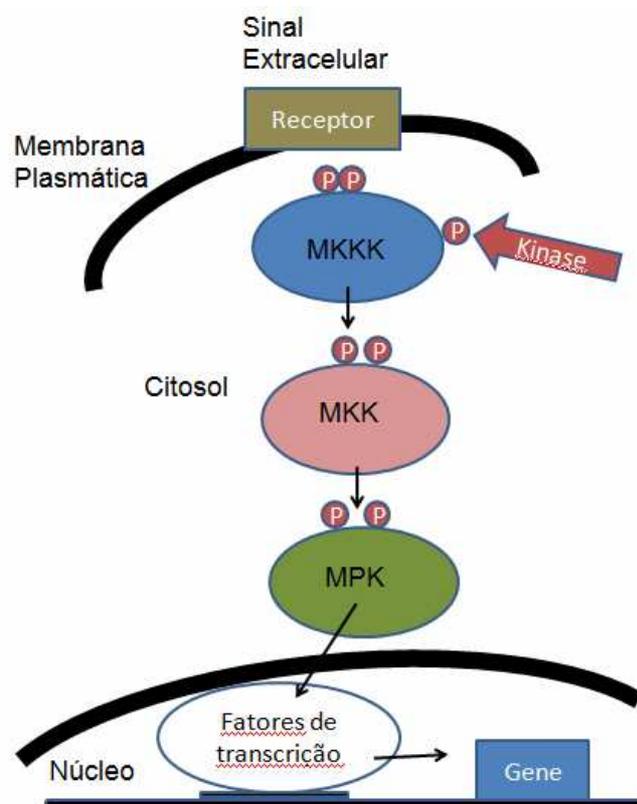


Figura 1. Esquema ilustrativo da cascata de ativação de vias MAP Kinases adaptado de (Tim et al., 2002).

2.2.1. Caracterização dos componentes MAP Kinases

As cascatas de MAP Kinases consistem dos módulos: MKKK, que ativam MKKs, que por sua vez vão ativar as MPKs, formando uma combinação de conexões entre eles. Os membros destas famílias de proteínas possuem uma espécie de “assinatura”, que faz com que módulos individuais sejam agrupados dentro de uma família específica e podendo ser, dessa forma, caracterizados. Um aspecto interessante é que muitos domínios específicos, que caracterizam estas famílias de genes se mantiveram altamente conservados durante o processo evolutivo, numa ampla variedade de espécies, inclusive na espécie humana (Widmann, 1999). Toda proteína pertencente à grande família kinase possui uma assinatura PK (DxK), descrita até o momento na literatura, que a caracteriza (Hamel et al., 2006; MAP Group, 2002). A assinatura (DxK) é comum às proteínas da grande família das kinases, onde x representa um dos aminoácidos leucina, isoleucina ou valina (L, I ou V), respectivamente posicionados numa região específica (MAP Group, 2002). As proteínas da família kinase ativadas por mitogênese (MAP Kinases) constituem uma subfamília das proteínas kinases e possuem, além do “*motivo*” DxK, outras assinaturas específicas, que permitem com que sejam identificadas como componentes específicos MKKK, MKK ou MPK, conforme descrito a seguir.

a) Proteínas do Módulo MPKs

Os componentes MAPKs diferem no comprimento de sua alça T (alça de ativação) e na identidade do aminoácido que separa os resíduos ativadores (Thr - treonina e Tyr- tirosina) que são: Glutamato (E), Glicina (G), Prolina (P) e Ácido Aspártico (D). Sendo assim, a maioria das MAPKs de plantas, incluindo os genes de *Arabidopsis*, contém o *motivo* TEY (Treonina - ácido glutâmico -Tirosina) ou TDY (Treonina- ácido aspártico- Tirosina) no seu domínio de ativação, semelhante ao observado em mamíferos e leveduras (HIRT, 2000). Porém, em mamíferos e leveduras foram identificados domínios de ativação TGY (Treonina - Glicina - Tirosina) e TPY (Treonina - Prolina -Tirosina), ainda não identificados em plantas até o momento (Rodrigues, 2010; Ichimura, 2002).

As MPKs em plantas se organizam em quatro diferentes grupos (A,B,C e D), muito embora haja algumas redundâncias funcionais entre eles (Ichimura,

2000; Jonak, 2002). A inspeção na estrutura primária das MPKs em plantas revela que seus domínios catalíticos são altamente conservados, porém compartilham pouca identidade na região N-terminal. Esta característica indica que MPKs diferentes reconhecem substratos distintos. As MPKs ativadas, por sua vez, podem fosforilar alvos citoplasmáticos ou fatores de transcrição no núcleo da célula (Zhou C, 2008; Kolch, 2000).

As proteínas MPKs são o último componente ativado pela cascata de proteínas de fosforilação da família de proteínas MAP kinases.

b) Proteínas do Módulo MKKs

Os componentes MKKs são proteínas que ativam as proteínas MPKs. As MKKs fosforilam MPKs nos resíduos de treonina e tirosina (sequências Thr-X-Tyr) por meio de dupla fosforilação. Estes resíduos estão localizados na alça de ativação da MPK (também chamada alça T; Elsen e Davis, 2001).

Existem 10 MKKs putativas codificadas no genoma de arábido, isto é, o número de MKK é metade de MPKs em arábido. Provavelmente, nesta espécie, as MKKs ativam múltiplas MPKs e o cruzamento de várias rotas de sinais de transdução se concentra neste nível da cascata de MAP Kinases em plantas. (Rodrigues, 2010; Mishra, 2006). Ver Figura 2.

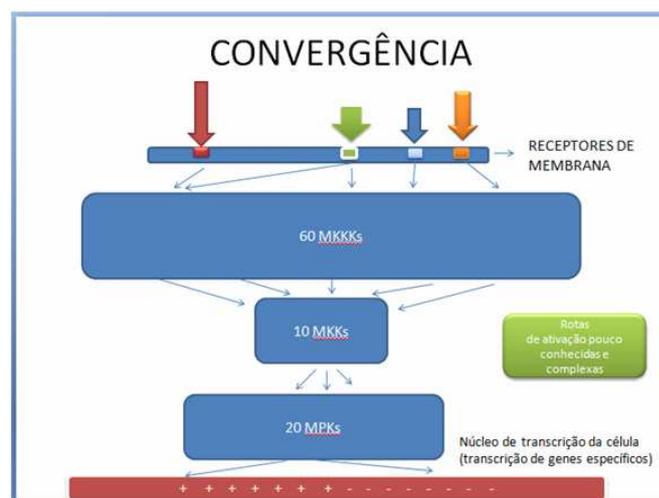


Figura 2. Convergência da cascata de ativação: A camada intermediária da cascata (nível MKK) concentra a maior parte dos sinais

Assim como as MPKs, componentes MKKs em plantas se organizam em quatro diferentes clados (A,B,C e D). A sequência do sítio de fosforilação de MKKs em plantas é diferente das kinases de mamíferos. As MKKs em plantas têm sequências consenso S/TxxxxS/T, enquanto enzimas de mamíferos têm S/TxxxS/T, sendo T o aminoácido treonina, S representando o aminoácido serina e x que pode ser qualquer aminoácido. As MKKs têm sido observadas como sendo as menos numerosas da família das kinases ativadas por mitogênese em plantas (Rodrigues, 2010).

c) Proteínas do Módulo MKKKs

As vias MKKKs são caracterizadas por serem enzimas que possuem o “motivo” (assinatura) *serina/treonina*. Tais regiões são ativadas em resposta a estímulo extracelular, após fosforilação por kinases, ou através de interação com proteínas ligadoras de GTP monoméricas (proteínas G) da família Ras ou Rho (Elsen e Davis, 2001).

As MKKKs de plantas apresentam estruturas primárias diferentes e, ainda, pouco se conhece sobre a estrutura de seus domínios conservados que permitam uma efetiva caracterização “*in silico*”. Evidências apontam que nem todas MKKKs ativam as MKKs por fosforilação nos resíduos de serina/treonina na sequência conservada (*motivo*) S/T x3ou5 S/T (Rodrigues, 2010; Hirt, 2009). Embora as famílias das MKKKs formem um grande e heterogêneo grupo da rota de componentes MAP Kinases, poucos trabalhos têm demonstrado a efetiva fosforilação de MKKs específicas por MKKKs. Em *arabidopsis thaliana* já foram identificadas cerca de 60-80 genes que, possivelmente, funcionam como MKKKs. Sessenta destas kinases, em *arabidopsis*, já foram relacionadas e podem ser divididas em dois grandes subgrupos: - as MKKKs do tipo MEKK, para as quais existem evidências de que elas funcionam como MKKKs em plantas; e as do tipo RAF, para as quais evidências de que elas funcionam como MKKKs ainda são poucas (Rodrigues, 2010). Além disso, uma MKKK (AtMKKK1) pode interagir diretamente com a AtMPK4 *in vitro*, sugerindo que esta MPKKK pode ser multifuncional, ou seja, pode operar na cascata AtMKKK1→MKK1/2→AtMPK4 ou diretamente na cascata AtMKKK1→AtMPK4 (Miao et al., 2007; Ichimura et al., 1998).

As figuras 3 e figuras 4a e 4b ilustram as principais regiões (motivos) identificadas para proteínas da família de MAP Kinases, situando os módulos MAP Kinases (MPK/MKK e MKKK) como uma subfamília das proteínas kinases (PK). Como pode ser observado, ainda não existe um consenso sobre as regiões conservadas para módulos MKKKs.

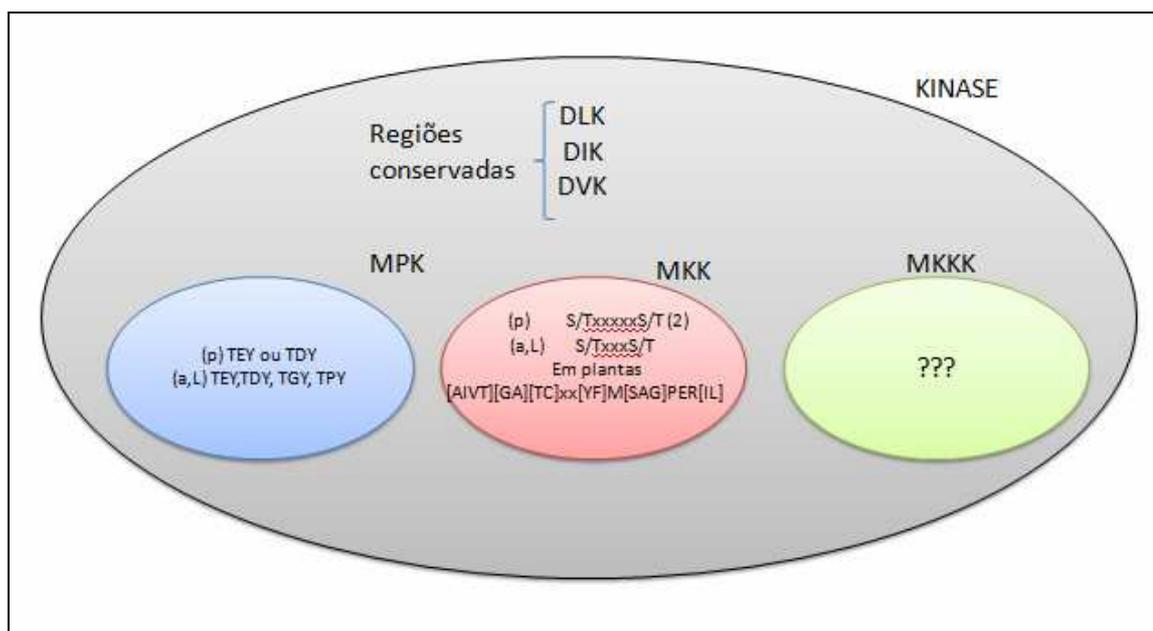


Figura 3. Sítios específicos que caracterizam as proteínas da grande família cinase e das subfamílias das cinases: MAPK e MKKs. Onde as letras (p) se referem às plantas, (a) animais e (L) leveduras.

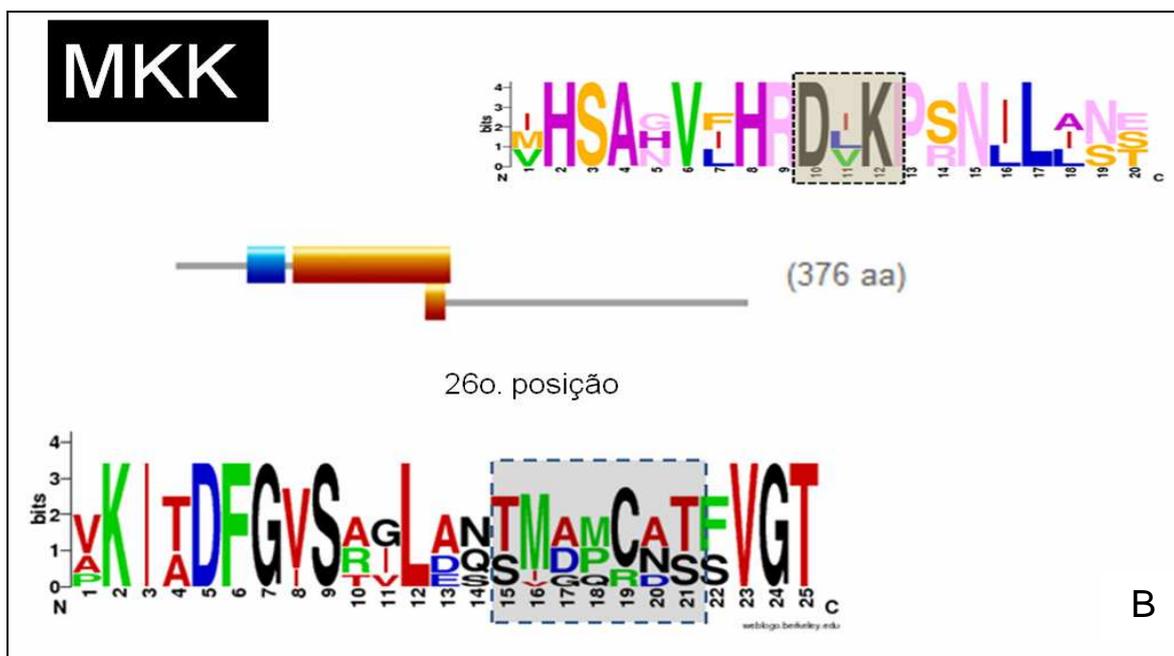
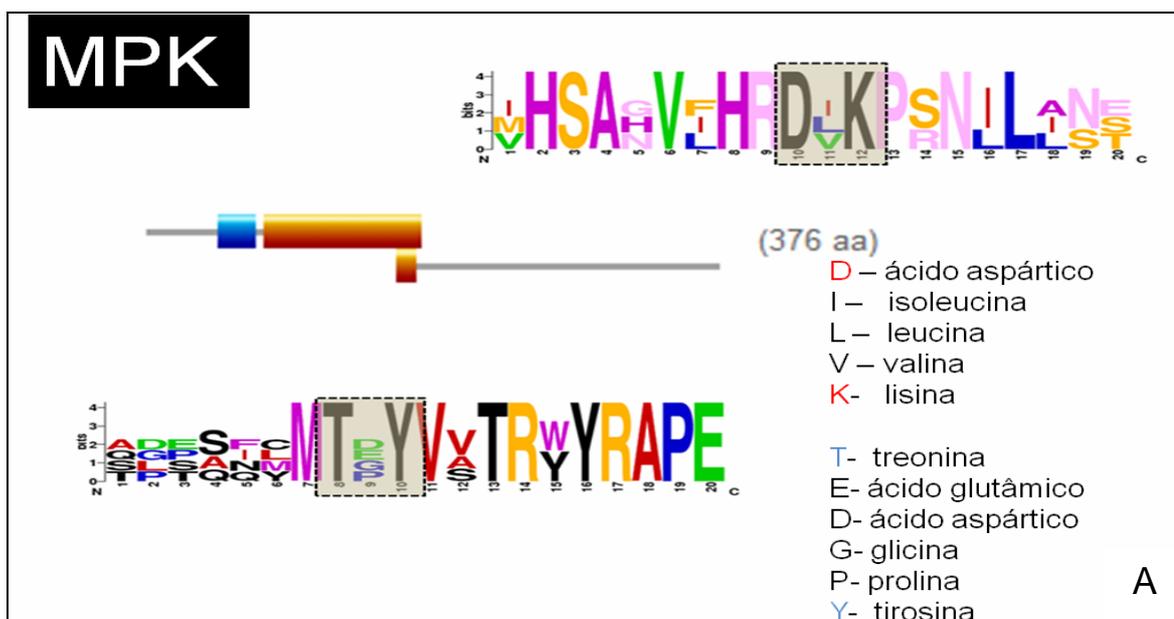


Figura 4. WebLogo (Crooks et al., 2004) ilustra as assinaturas de consenso conservadas nos domínios catalíticos de plantas da família kinase. O alinhamento foi feito com o ClustalW2, onde são ilustrados (A) Motivos MPKs: domínio kinase DxK + domínio TxY, específico de MPKs; (B) Motivo MKKs (domínio kinase DxK + domínio MKK S/TxxxxxS/T específico de MKK em plantas. A região em destaque representa a região conservada e os números abaixo representam a frequência relativa de cada resíduo.

2.3. Constituição das vias MAP Kinases em Plantas

Para um melhor entendimento deste tópico, faz-se necessário abordar as definições referentes a genes homólogos, ortólogos e parálogos, também crucial para o entendimento deste projeto de pesquisa.

2.3.1. Definição de Ortologia e Paralogia

Após a conclusão do sequenciamento de vários genomas, a pesquisa genômica tem se preocupado com a comparação entre genes de diferentes espécies, numa tentativa de estabelecer uma relação funcional e evolutiva entre eles. A história da evolução dos genes baseia-se em alguns eventos elementares, a saber: descendência vertical (especiação), duplicação gênica, perda de genes, transferência horizontal de genes, fusão, fissão e outros rearranjos genéticos (Yu et al., 2005; Gu et al., 2005).

A especiação e a duplicação gênica, contudo, são os mais relevantes e servem de base para a classificação de dois importantes conceitos em genética: a ortologia e a paralogia. Esses dois são casos particulares de homologia entre genes. Dois genes são considerados homólogos se têm a mesma origem evolutiva. Dois genes homólogos, originados por um evento de especiação são ditos ortólogos, ao passo que dois genes originados por duplicação dentro do mesmo genoma são denominados parálogos. Tal similaridade deriva de um processo de duplicação gênico ocorrido antes da existência de cada uma dessas espécies, ou seja, no ancestral comum a todas elas (Jouannic et al., 1999; Jouannic et al., 2001). Na Figura 5 ilustra-se melhor esta definição, que é de suma importância para as discussões propostas neste trabalho.

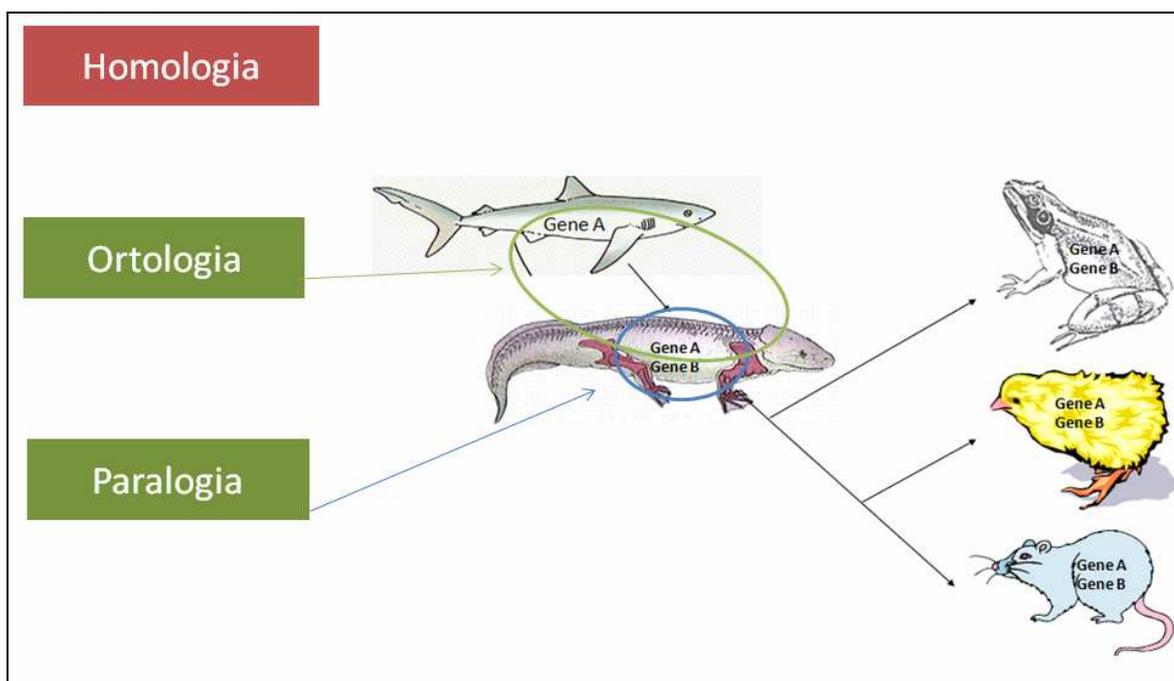


Figura 5. Ilustração de homologia, ortologia e paralogia de genes. Os genes A são ortólogos e estão em espécies diferentes. O gene B é derivado do gene A por evento de duplicação no mesmo genoma, portanto, representam casos de paralogia de genes. Tanto a ortologia quanto a paralogia são casos especiais de homologia de genes.

2.3.2. Determinação da estrutura organizacional dos módulos MPKs/MKKs em álamo e arroz por meio de análise *in silico*

Estudos das famílias MAPKs e MKKs têm sido amplamente realizados em *Arabidopsis*, motivando, inclusive, a adoção de uma nomenclatura única para os componentes individuais destas famílias de genes (MAPK GROUP, 2002). Outros grupos têm sugerido a extensão da nomenclatura proposta para *Arabidopsis*, como suficientemente robusta para ser estendida para caracterização e padronização dos genes dessas famílias também para outras espécies (Hamel et al., 2006). No estudo desenvolvido por Hamel, feito com proteínas MAPKs e MKKs em Arroz (*Oryza sativa*), Álamo (*Populus trichocarpa*) e *Arabidopsis thaliana*, confirma que, nas três espécies analisadas, domínios específicos se mantiveram altamente conservados nestas famílias de genes. Em *Arabidopsis*, por exemplo, já foram identificados 20 (vinte) genes que codificam 20 MAPKs e 10 (dez) que codificam MKKs comparadas com 6 (seis) MAPKs e 6 (seis) MKKs

em leveduras e 10 (dez) MAPKs e 7 (sete) MKKs em humanos (Hamel et al., 2006). Estudos de filogenia têm situado MPKs e MKKs dentro de 4 clados (A-D) (Hamel et al., 2006) . A figura 6 e 7 ilustram o resultado dos estudos de filogenia desenvolvidos por Hamel, feito para estas espécies. Um estudo similar foi proposto neste trabalho para cana-de-açúcar, sorgo e *A. lyrata*.

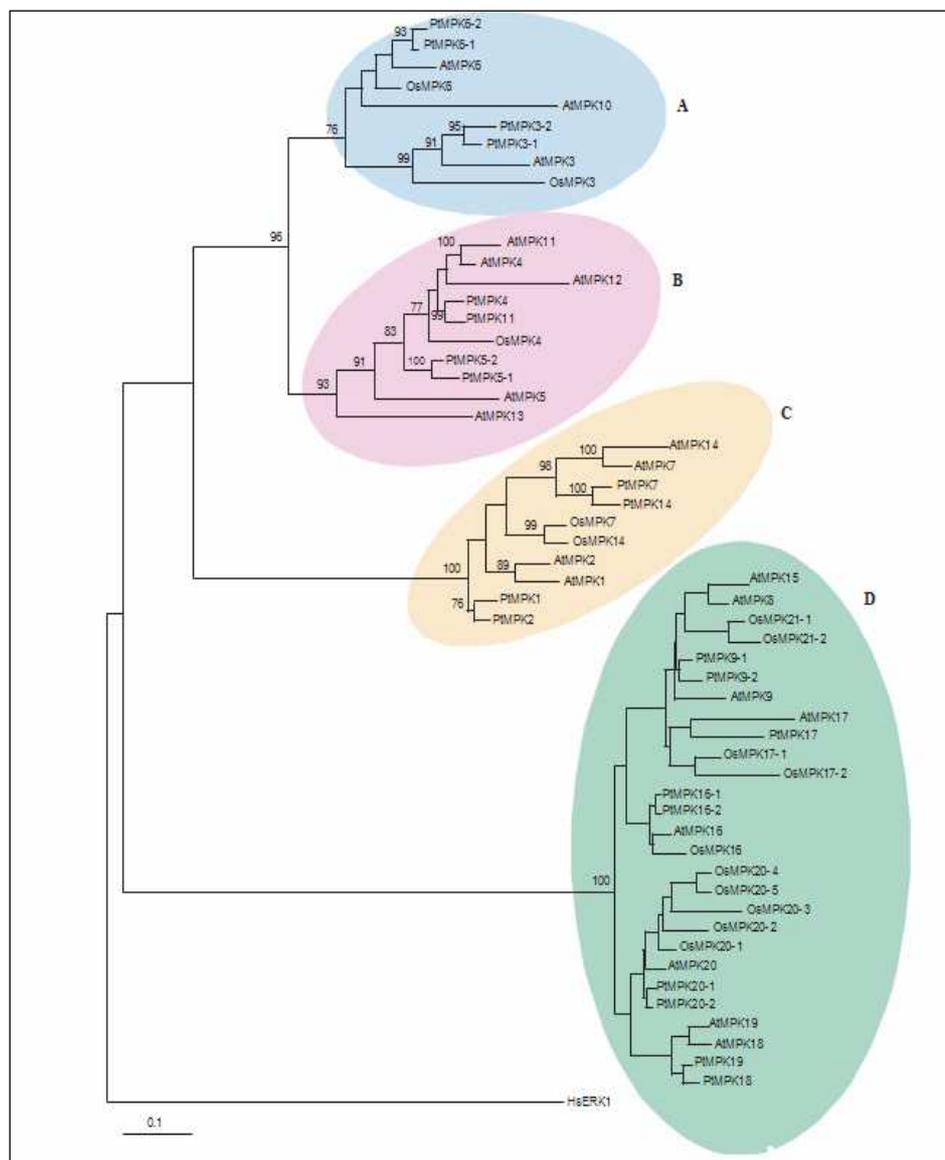


Figura 6. Árvore filogenética das vias MPKs em arroz, álamo e arabidopsis (Hamel et al., 2006).

A análise funcional e consultas a bancos de expressão disponíveis tem revelado que, apesar do alto nível de conservação estrutural destas famílias de genes, frequentemente membros individuais (ortólogos em outras espécies)

desempenham papéis distintos, embora algum nível de redundância funcional seja também observado (Rodrigues, 2010; Erik et al., 2009; Hamel et al., 2006; Feilner, 2005; Sterck, 2005; Jonak, 2002).

Uma análise filogenética inicial das famílias de genes que codificam as três classes de kinases, no genoma de *Arabidopsis*, sugere que algumas destas famílias podem ser maiores ou menores comparadas a outras espécies (Mapk Group, 2002; Hamel et al., 2006).

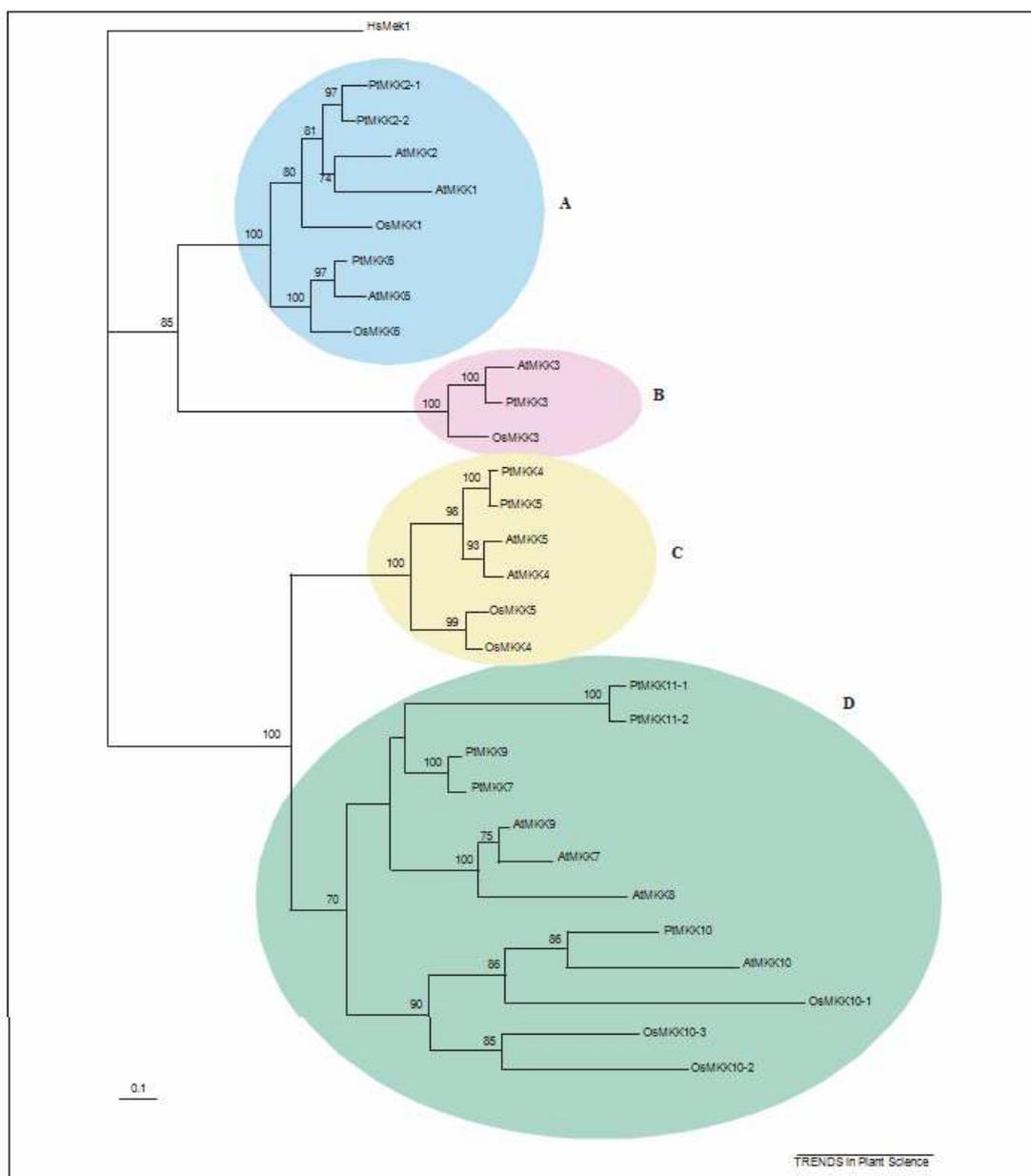


Figura 7. Árvore filogenética das vias MKKs em arroz, álamo e arábida (Hamel et al., 2006).

2.3.3. Padrões de ativação de MAP Kinases em plantas

O entendimento de como funcionam módulos MAP Kinases em plantas superiores muito se deve a pesquisas realizadas com *A. thaliana*, cujos componentes MPKs, MKKs e MKKKs já foram praticamente todos identificados. Entretanto, até o momento, o conhecimento sobre a cascata de MAP Kinases é obtido, principalmente, em espécies dicotiledôneas, com ênfase em pesquisas

realizadas em *A. thaliana*, havendo ainda poucos estudos neste campo com monocotiledôneas (Rodrigues et al., 2010). Boa parte dos trabalhos têm se concentrado no entendimento de como estas vias funcionam e como são ativadas e, a maior parte das informações disponíveis, atualmente, se concentram em interações MKKs→MPKs. (Erik et al., 2009; Rodrigues, 2010).

O grupo A das MKKs em *A. thaliana* inclui a MKK1, MKK2 e MKK6, que interagem e fosforilam múltiplas MPKs em arabidopsis (Rodrigues et al., 2010; Erik et al., 2009). Vários estudos têm mostrado a importância da MKK1/2 na sinalização de respostas à patógenos (Brader et al., 2007; Gao et al., 2008; Qiu et al., 2008) e a fatores abióticos, tais como frio e salinidade (Teige et al., 2004; Xing et al., 2004). Entretanto, estes estudos tem sido focados na cascata MEKK1→MKK1/2→MPK4 e/ou MPK6. O terceiro membro do grupo A de MKKs, MKK6, tem sido relacionado em plantas, à divisão celular. Pesquisas têm demonstrado que ortólogos do MKK6 em *Nicotiana tabacum* (tabaco) e *Medicago sativa* (alfava), o NtMEK1 (NtNQK1) e o MPK13/NRK1, respectivamente, desempenham papel central na citocinese (Nishihama, R, et al., 2002; Bogre, L. 1999). Outras interações *in vitro* também foram registradas, onde a MKK1/2 utiliza outras MPKs como substrato. A Figura 9 mostra estas interações na parte superior (nas setas em azul). As MPKs (MPK1, MPK2, MPK3, MPK4 e MPK6) têm sido estudadas extensivamente e pesquisas têm demonstrado que estes componentes estão envolvidos numa gama de estresses bióticos e abióticos (Mishra et al., 2006; Teige et al., 2004; Asai et al., 2002; Yuasa et al., 2001; Ichimura et al., 2000). Outros estudos têm demonstrado que em *A. thaliana* os genes MPK1 e MPK2 são ativados por JA, ABA e H₂O₂. A duplicação de genes tem sido registrada para *A. thaliana* (Ermolaeva et al., 2003; Sterck et al., 2005; Yu et al., 2005). Em *A. thaliana*, estudos relatam o envolvimento do gene MPK4 na sinalização de SA (Ácido Salicílico) durante respostas imunes (Brodersen et al., 2006; Petersen et al., 2000). Pesquisas em leveduras têm revelado que os componentes MPK1, MPK2, MPK7 e MPK14 são substratos da MKK3 na cascata MKK/MPK (Doczi et al., 2007). Em *A. thaliana*, tanto o gene MKK1 quanto o MKK2 têm sido relacionados como mediadores de resposta imune e associam AtMKK1 à resposta a estresse oxidativo, frio e alta salinidade (Qiu et al., 2008-1; Meszaros et al., 2006; Teige et al., 2004). Ainda em *A. thaliana*, estudos têm demonstrado que a MKK1 atua na cascata de ativação da MPK4 (Qiu et al., 2008-

2; Ichimura et al., 1998) e que os módulos MKK1→MPK6 regulam o metabolismo H_2O_2 através da enzima CAT1 (Xing et al., 2008). Um mesmo módulo MKK é utilizado em várias rotas, ativando MPKs diferentes, o que tem sugerido que uma mesma MKK pode funcionar em vários módulos MPKs diferentes (Rodriguez et al., 2010). Em *A. thaliana*, o gene MKK4 está associado a respostas a patógenos mediadas por SA e no desenvolvimento e diferenciação de estômatos; e o gene MKK5 à abscisão do órgão floral (Wang et al., 2007). Ainda em *A. thaliana*, os genes MKK7 e MKK9, que pertencem ao clado D, possuem resíduos de ácido aspártico em um dos sítios de fosforilação conservados, o que sugere que estas kinases são autoativadas (Cardinale et al., 2002). A proteína MKK9 associada à biossíntese do etileno (Zhou et al., 2009) e à MKK7 tem sido relacionada ao transporte polar de auxinas e à resistência basal e sistêmica (Zhang et al., 2007). A Figura 8 ilustra um panorama geral da cascata de fosforilação *in vitro* e *in planta* de módulos MKK→MPKs.

O componente MKK3 representa o grupo B das MKKs e é caracterizado por possuir uma estrutura única para membros MKKs em plantas superiores. Esta família de genes é distinta por possuir uma extensão C-terminal relacionada, em eucariotos, a fatores de transferência nuclear. A proteína MKK3 está envolvida na sinalização ativada por patógenos e por hormônios (Doczi, R. et al., 2007; Takahashi, F. et al., 2007). Estes papéis são provavelmente mediados pela habilidade da MKK3 de interagir com um gama de MPKs (ver Figura 8) (Lee et al., 2008; Doczi, R. et al., 2007)

O grupo C das MKKs em *A. thaliana* é composto pelos componentes MKK4 e MKK5 que são parálogos e que interagem com as MPKs parálogas (MPK3 e MPK6) em *A. thaliana*. Intrigantemente, este mesmo módulo desempenha um papel essencial durante o processo de desenvolvimento, na determinação da diferenciação estomática (Lampard, G. et al., 2008; Lampard, G. et al., in press) e dentro do controle da abscisão do órgão floral em *Arabidopsis* (Cho, S. K. et al., 2008). O modelo de uma relação funcional da MKK4/5 com a MPK6 foi suportado pelos dados de sistemas duplo-híbrido (“two hybrid-system”) em leveduras (Lee et al., 2008) mas, ao contrário da MPK6, a MPK3 nestes mesmos ensaios interagiu somente com MKK4. Entretanto, ainda não está claro qual o papel desempenhado pela MKK4 e MKK5 dentro da cascata de sinalização de MAP Kinases. Tanto a MKK4 quanto a MKK5 são ativamente expressas em

vários tecidos de *Arabidopsis* em todos os estágios do desenvolvimento da planta (Genevestigator, 2010), sugerindo que estas MKKs se expressam constitutivamente dentro da célula (Bush, S. M., 2007) (Samuel M. et al., 2008).

O grupo D de MKKs em *Arabidopsis thaliana* forma um clado com 4 membros (AtMKK7, AtMKK8, AtMKK9 e AtMKK10) (Ichimura, K. et al., 2002) e ortólogos putativos já foram encontrados em outras espécies, como arroz e álamo (Hammel, et al., 2006). Entretanto, ainda não está claro qual o papel biológico desempenhado por estes módulos. Em ensaios com leveduras, a MKK8 falhou ao interagir com qualquer uma das 20 MPKs em ensaios de sistema duplo-híbrido (Lee, et al., 2008) Todavia, a AtMKK10 e seu parálogo AtMKK9 interagiram com MPK17 (Menges, M. et al., 2008). Outros ensaios ainda mostraram que a MKK10 fosforila as MPKs (MPK1, MPK2, MPK5 e MPK10) *in vitro* (Popescu, 2009).

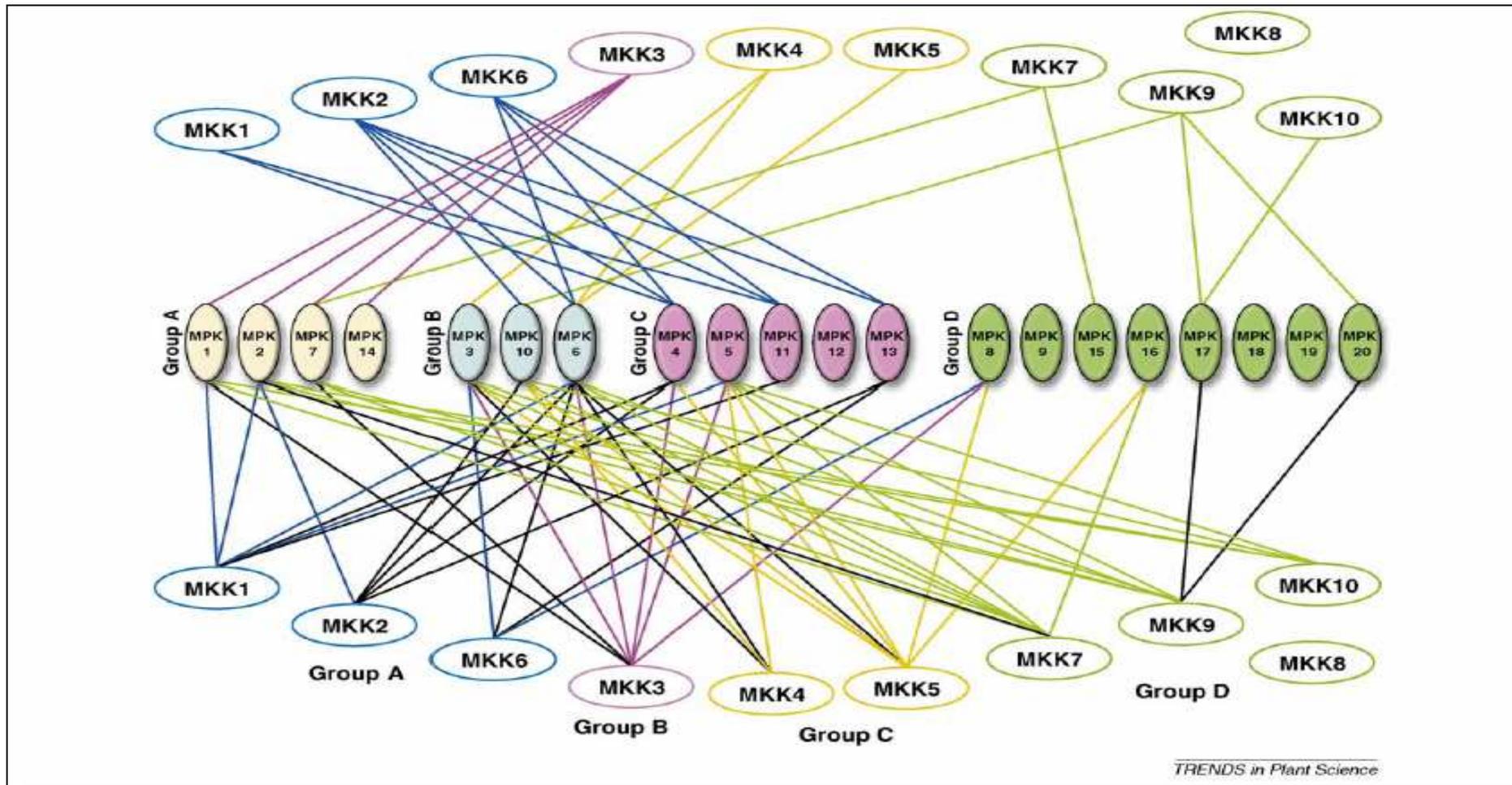


Figura 8. Interações de MKKs com MPKs relacionadas a pesquisas com *Arabidopsis thaliana*. Os componentes MPK/MKKs foram coloridos de acordo com o grupo filogenético à que pertencem. As conexões da parte superior ilustram interações positivas, enquanto as conexões na parte inferior da figura representam interações MKK→MPK onde houve fosforilação in vitro e/ou in planta. As setas em preto, na parte de baixo, representam tanto interação proteína-proteína MKK→MPK quanto de fosforilação. (Erik et al., 2009)

Embora estudos acerca da relação entre MPKs e seus ativadores forneceram um esboço inicial dos padrões de evolução da sinalização por MAPK em plantas, muitas questões ainda precisam ser exploradas. Muitos dados derivados de interações *in vitro* ainda precisam ser transpostos para análise dentro de um contexto espacial e temporal. A Figura 9, a seguir, resume o resultado das últimas pesquisas obtidas em *A. thaliana* sobre interações MKK →MPKs *in vitro* e *in planta*.

A figura 9 (a) ilustra algumas dessas fosfatases. As MKPs em amarelo derivam de pesquisas em *A. thaliana*. As linhas em vermelho indicam que estes dados derivam de experiências *in vitro*, as linhas em azul demonstram interações positivas proteína-proteína. As linhas em preto indicam que as pesquisas foram confirmadas tanto *in vitro* quanto em interação proteína-proteína. Entretanto, sabe-se também que inúmeros substratos MPK podem interagir na cascata de sinalização e um esboço geral do que se conhece sobre o relacionamento de MPKs e seus substratos específicos encontra-se ilustrado na figura 9 (b).

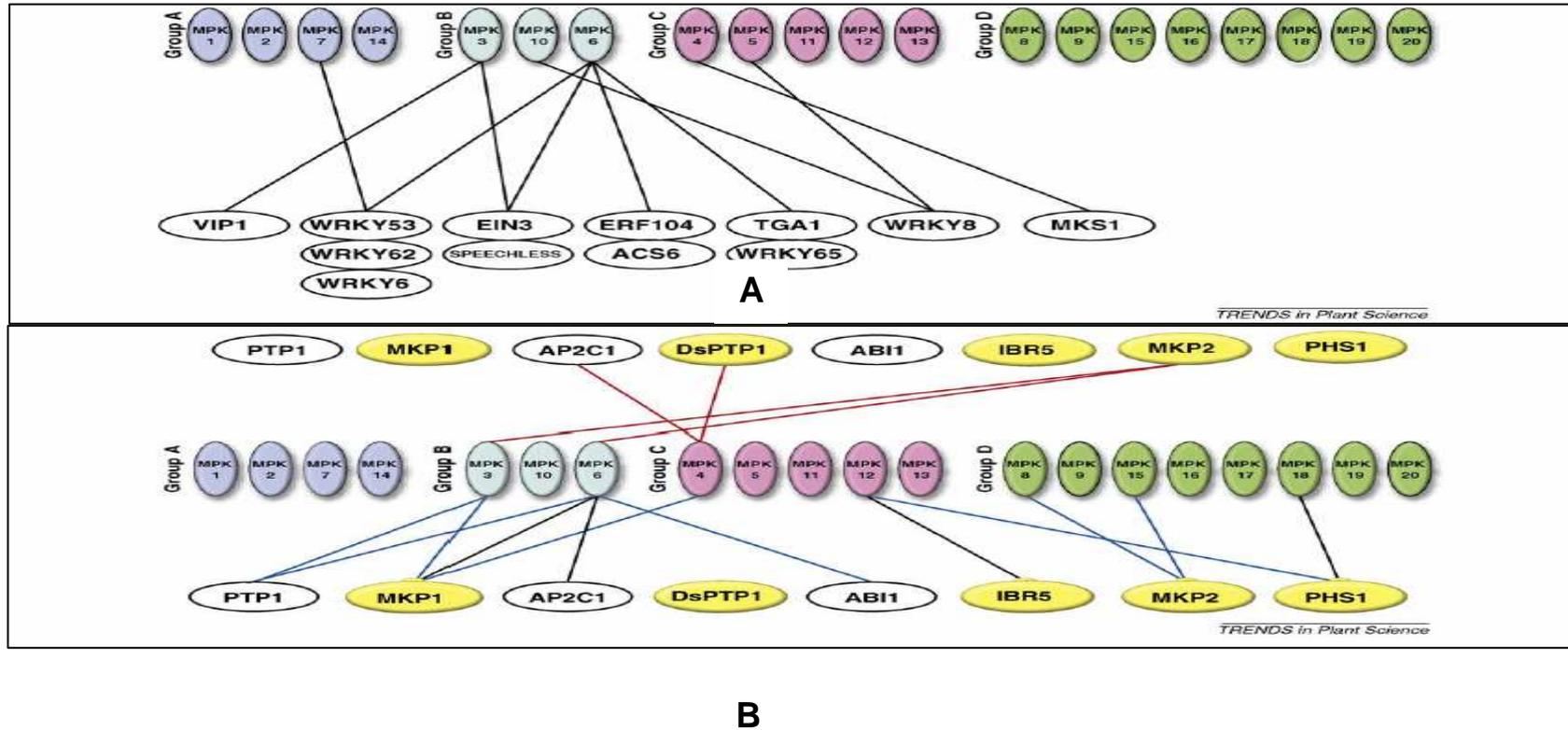


Figura 9. (A) Esquema que reflete a situação de proteínas fosfatases (B) Substratos identificados que interagem com MPKs de *A. thaliana* (Erik et al., 2009).

2.4. Bases de Dados públicas para o estudo de proteínas

Atualmente, com a disponibilidade de sequência de uma variedade de genomas armazenados em bancos de dados públicos, tanto de espécies de plantas monocotiledôneas quanto dicotiledôneas, tais como: *Oryza sativa* (arroz), *Populus trichocarpa* (álamo), *Zea Mays* (milho), *Sorghum bicolor* (sorgo), dentre outros, tem sido possível um exame mais detalhado sobre a estrutura dos genes que codificam proteínas e famílias de proteínas destas espécies.

Alguns bancos de dados públicos de proteínas contêm um alto nível de anotações que contemplam informações, tais como: a descrição e a função da proteína, estrutura dos seus domínios, modificações pós-tradução etc. Além disso, possuem uma estrutura que facilita o acesso computacional a diferentes campos de informações por meio de uma interface gráfica. Bancos de dados de proteínas possuem a seguinte classificação: banco de dados de sequências primárias, banco de dados de sequências secundárias e bancos de estruturas. Nos bancos de dados de estruturas, tem-se por objetivo compreender melhor a relação sequência-estrutura e os processos evolutivos subjacentes que dão origem ao desdobramento de diferentes famílias. Os bancos de dados de sequências primárias contêm anotações, tais como: a descrição e a função da proteína, estrutura dos seus domínios, modificações pós-tradução, além de ter uma estrutura que facilita o acesso computacional a diferentes campos de informações. Os bancos de dados secundários contêm informações de como as famílias de proteínas podem ser caracterizadas por meio de uma única assinatura, geralmente a mais conservada, obtida pelo alinhamento múltiplo de sequências já conhecidas.

Alguns desses bancos são descritos a seguir:

- Interpro: Ambiente que integra vários bancos anteriores. É uma fonte integrada de informações sobre famílias de proteínas, domínios e sítios funcionais. Facilita a complementação entre informações obtidas por enfoques distintos. (www.ebi.ac.uk/interpro/)
- TAIR- Os dados disponíveis de TAIR incluem a sequência completa do genoma junto com a estrutura do gene, a informação do produto gênico, metabolismo, a expressão do gene, sequência de DNA, os mapas do genoma,

marcadores genéticos, publicações e informação sobre a comunidade de pesquisa de *Arabidopsis*. (<http://www.arabidopsis.org/>);

- PHYTOZOME v5.0 - Phytozome é um projeto integrado com outras bases de dados públicas, que permite estudos genômicos comparativos entre plantas, por meio de arranjos filogenéticos (<http://www.phytozome.net/>)
- PlantsP - É uma base de dados de genômica funcional que estuda os genes ligados ao processo de fosforilação de proteínas em plantas. Nele, o domínio CD (domínio catalítico) de todas as proteínas da família kinase ativadas por mitogênese (MAPK, MAPKK e MAPKKK) de *Arabidopisthalianathaliana*, por exemplo, está identificado. (<http://plantsp.genomics.purdue.edu/html/>, obtido em março de 2009).

Ao longo dos anos, métodos estatísticos capazes de analisar grandes quantidades de dados biológicos, na tentativa de identificar e prever as funções dos genes e proteínas, por eles codificados, têm sido utilizados (WAGENINGEN, 2007). Dentre estas ferramentas, podemos citar todo o suporte de bioinformática disponibilizado pelo NCBI (<http://www.ncbi.nlm.nih.gov/>, acessado em julho de 2010).

2.5. Cana-de-açúcar, sorgo e *Arabidopsis lyrata*

2.5.1. A cana-de-açúcar

A Cana-de-açúcar é uma das espécies de plantas mais importantes das regiões tropicais e subtropicais e é cultivada em cerca de 20 milhões de hectares em mais de 90 países. Atualmente, o principal produto de interesse comercial derivado da cana-de-açúcar é a sacarose que, uma vez extraída, pode ser purificada em fábricas especializadas e utilizada como matéria-prima nas indústrias alimentícias e no processo de produção de etanol. No mundo, o etanol representa cerca de 3% do combustível utilizado em motores de automóveis. (Globalbioenergy, 2009).

Um grande impulso ao conhecimento da genética e biologia molecular da cana-de-açúcar foi obtido com a criação de um projeto de sequenciamento de ESTs (“Expressed Sequence Tags”) desta espécie. Tal projeto foi denominado SUCEST (“Sugarcane EST Genome Project”) e dedicou-se ao sequenciamento e

análise de clones de cDNA obtidos a partir de bibliotecas de diversos tecidos da planta sob diferentes condições fisiológicas. O projeto SUCEST foi fundado no estado de São Paulo, sob coordenação da FAPESP, e teve como objetivo identificar 50.000 genes de cana-de-açúcar ou sequenciar cerca de 300.000 ESTs (Expressed Sequence TAGs). Maiores informações podem ser obtidas em: (<http://www.sucest-fun.org/>, acessado em maio de 2010).

O projeto SUCEST gerou uma gama de informações acerca do transcrito da cana-de-açúcar. As sequências foram obtidas a partir de clones de 37 bibliotecas de cDNA construídas de diferentes tecidos ou condições fisiológicas da planta (Ver tabela 1). Para tanto, os mRNAs transcritos em um tecido/condição específica foram submetidos à transcrição reversa para a obtenção de moléculas de cDNA. Os cDNAs obtidos foram clonados em vetores plasmidiais. A partir desse ponto seguiu-se o sequenciamento. As sequências obtidas foram denominadas ESTs (Expressed Sequence TAGs). Ver Figura 10.

Todos os ESTs obtidos foram analisados através de programas computacionais (bioinformática). Cada clone de cDNA sequenciado é então denominado “*read*”. Os “*reads*” são agrupados de acordo com o seu grau de similaridade em arquivos denominados “*clusters*”. Uma vez obtidas as sequências, estas foram anotadas no SUCEST, seguindo as definições da Figura 11.

O SUCEST contém cerca de 43 mil CLUSTERS (ou seja, 43 mil transcritos) obtidos a partir da mineração e análise de 260 mil READS.

Uma vez identificados os genes de interesse em cana-de-açúcar, pode-se analisar a ocorrência dos mesmos nas 37 diferentes bibliotecas de cDNA do projeto SUCEST, específicas para determinado tecido ou condição. Essa abordagem é conhecida como análise da expressão gênica “*in silico*” ou “*Electronic Northern*” e baseia-se na contabilização do número de vezes que cada EST aparece nas diferentes bibliotecas de cDNA (RAFALSKI *et al.*, 1998).

Tabela 1 - Relação dos grupos correspondentes aos tecidos de cana-de-açúcar obtidos a partir da reunião das bibliotecas de cDNA de cana-de-açúcar do projeto SUCEST.

Tecido	Descrição
AD	Clones das bibliotecas de cDNA de AD1
AM	União dos clones das bibliotecas de cDNA AM1 e AM2
CL	Clones das bibliotecas de cDNA CL6
FL	União dos clones das bibliotecas de cDNA FL1 a FL8
HR	Clones das bibliotecas de cDNA HR1
LB	União dos clones das bibliotecas de cDNA LB1 e LB2
LR	Clone da biblioteca de cDNA LR1
LV	Clones da bibliotecas de cDNA LV1
RT	União dos clones das bibliotecas de cDNA RT1 a RT3
RZ	União dos clones das bibliotecas de cDNA RZ2 e RZ3
SB	Clones da biblioteca de cDNA SB1
SD	União dos clones das bibliotecas de cDNA SD1 e SD2
ST	União dos clones das bibliotecas de cDNA ST1 e ST2

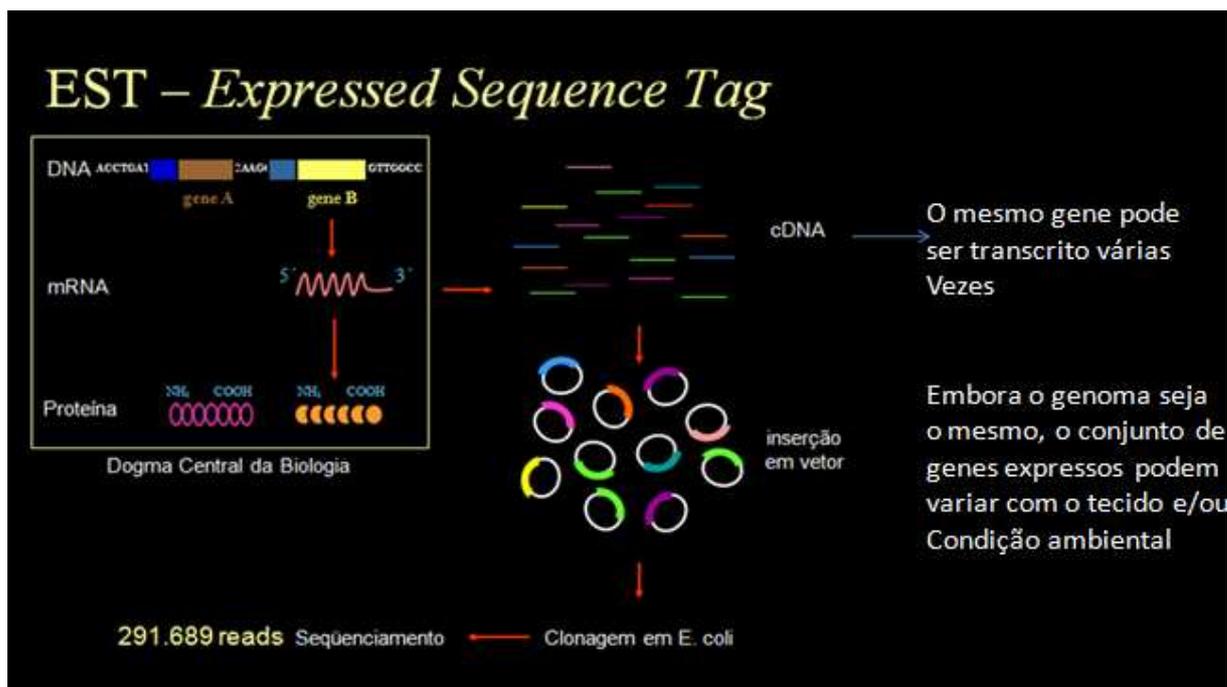


Figura 10. Esquema ilustrativo dos passos para a obtenção dos ESTs em cana-de-açúcar.

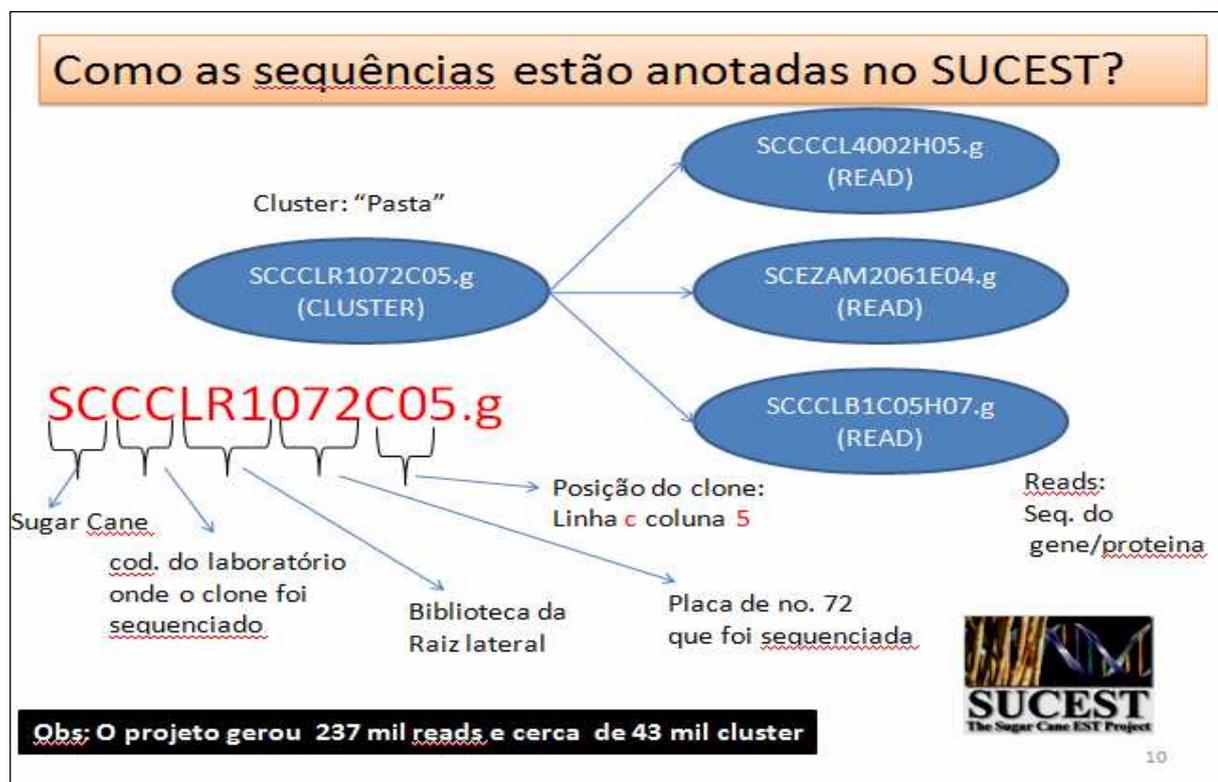


Figura 11. Anotação dos EST de cana-de-açúcar no SUCEST

2.5.2. *Sorghum bicolor* (sorgo)

O *Sorghum bicolor* (sorgo) é originário da África e é uma espécie geneticamente próxima da cana-de-açúcar, arroz e do milho. Conhecido por sua resistência à seca e alta produtividade, o sorgo é atualmente a segunda planta mais utilizada para produção de biocombustíveis nos Estados Unidos, perdendo apenas para o milho. Os grãos de sorgo produzem a mesma quantidade de etanol que o milho com apenas dois terços da água (Paterson et al., 2009).

O genoma do sorgo é 75% maior que o do arroz e muito semelhante ao da cana-de-açúcar, exceto pelo fato de que o sorgo é uma espécie diplóide, enquanto a cana-de-açúcar de açúcar é poliplóide (variando de octaploide à undecaploide), e o da cana-de-açúcar tem oito a dez cópias, obtido em (FAPESP, 2010)

Cientistas do *Joint Genome Institute* (JGI), ligado ao Departamento de Energia dos Estados Unidos (DOE), em parceria com diversas instituições, publicaram o sequenciamento e a análise do genoma completo do sorgo (JGI, 2010). Uma análise inicial do genoma do *Sorghum bicolor*, que possui cerca de 730 milhões de pares de bases, mapeou cerca de 98% do seu genoma. A técnica utilizada para a montagem da base de dados genômica foi a *shotgun*.

A figura 12 ilustra a família de genes ortólogos de *A. thaliana* identificadas em sorgo (*sorghum*), arroz e álamo (*poplar*). Nesta figura, é possível observar que cerca de 10.000 genes são compartilhados por todas as espécies citadas e que a maior parte dos genes encontrados em arroz possui ortólogos em sorgo.

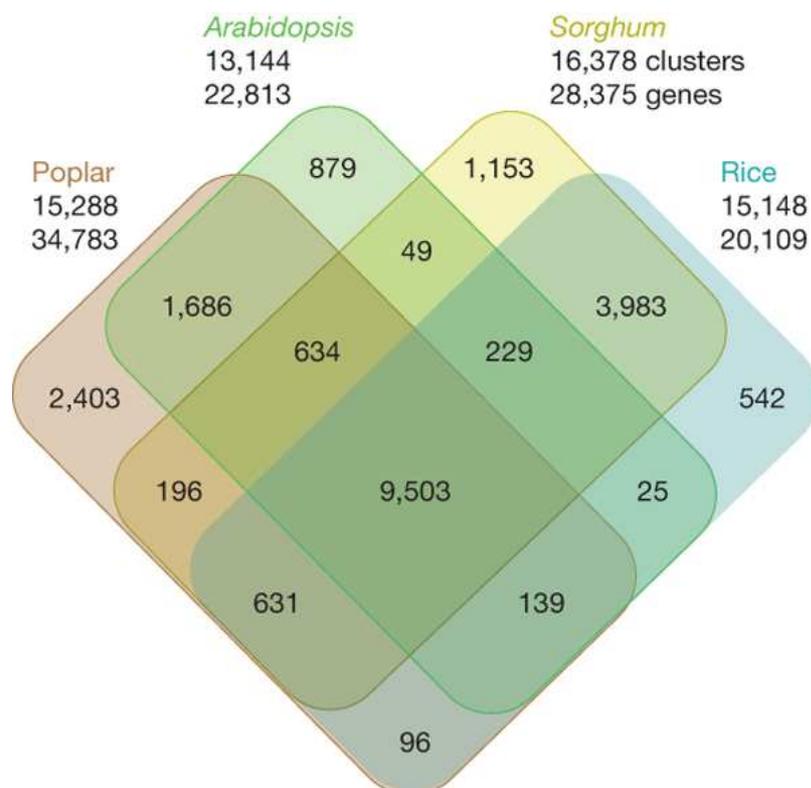


Figura 12. Ilustração do compartilhamento de genes ortólogos de *A. thaliana* em sorgo, arroz e álamo. Figura extraída de (Paterson et al., 2009).

2.5.3. *Arabidopsis lyrata*

Arabidopsis lyrata é um parente próximo, porém incompatível, da *A. thaliana*. Trata-se de uma espécie diplóide que possui um número de cromossomas igual a 8 ($n=8$) e diverge de *A. thaliana*, cujo genoma é menor ($n=5$).

A. thaliana tem se tornado um modelo para a biologia molecular de plantas e, em parte, isto se deve ao seu pequeno genoma. A comparação entre *A. lyrata* e *A. thaliana* tem sido utilizada em pesquisas para o entendimento e evolução de genomas e tem sido uma oportunidade para se entender como uma espécie pode ganhar ou perder cerca de um terço do seu genoma em apenas 5 milhões de anos.

O interesse de se estudar *A. lyrata* deriva, em parte, tanto de sua estreita relação com *A. thaliana* quanto de suas propriedades biológicas originais. Além

disso, estudos comparativos com estas duas espécies têm auxiliado no entendimento de processos de seleção e mutação em plantas.

O sequenciamento do genoma de *A. lyrata* v1.0 foi um programa derivado do *US Department of Energy's Office of Science, Biological and Environmental Research the University of California, Lawrence Berkeley National Laboratory*. A técnica utilizada na montagem do genoma para a criação da base de dados foi a *shotgun* e as sequências podem ser obtidas publicamente em (JGI, 2010). A figura 13 representa um resumo do modelo de genes obtido do genoma da *A. lyrata*.

Modelo de Genes V1.0	
Comprimento (pares de base) de:	faixa
gene	2,080
transcrito	1,192
Exon	223
Intron	207
DESCRIÇÃO:	
Comprimento da proteína(aa):	361
Exon por gene	5.3
Genes por Mb	158,5
# do modelo de genes	32,670

Figura 13: Modelo de genes obtido do sequenciamento do genoma *A. lyrata* adaptado do site <http://genome.jgi-psf.org/>

2.6. Objetivos neste trabalho

2.6.1. Objetivo Geral

O principal objetivo deste trabalho é avaliar a estrutura organizacional das vias de MAP kinases em cana-de-açúcar, sorgo e *A. lyrata* e analisar se tais vias seguem os padrões previamente estabelecidos para *A. thaliana*, arroz e álamo.

2.6.2. Objetivos específicos

1. Identificar grupos de genes MPK/MKKs ortólogos/parálogos de *A. thaliana* em cana-de-açúcar, sorgo e *A. lyrata*.
2. Geração da árvore filogenética envolvendo genes já caracterizados em arroz, *A. thaliana* e álamo, em conjunto com os componentes MAP Kinases identificados por este trabalho em cana-de-açúcar, sorgo e *A. lyrata*, estabelecendo o padrão de organização das vias MAPKs nestas espécies.
3. Verificar, a partir de uma análise conjunta, se processos de expansão/duplicação de genes, pertencentes ao clado D, observados em álamo e arroz, ocorreram também em sorgo, cana-de-açúcar e *A. lyrata*.
4. Verificar se existem divergências nos padrões estruturais observados em módulos MPK/MKKs descritos até a presente data na literatura, bem como novas regiões conservadas de genes, a partir do conjunto de módulos MPKs/MKKs das espécies analisadas por este trabalho.
5. Averiguar o grau de divergência/similaridade em relação aos módulos MPKs/MKKs de *A. thaliana* e *A. lyrata*.

3. MATERIAL E MÉTODOS

3.1. Levantamento e caracterização de MAP Kinases identificadas na literatura

Foram recuperadas, em artigos publicados e no GenBank (<http://www.ncbi.nlm.nih.gov/>), sequências de MAP Kinases (MPKs/MKKs) identificadas em várias espécies, além daquelas de *Arabidopsis thaliana*, arroz e álamo, utilizadas para análise de filogenia feita neste trabalho e descritas na tabela 2 e 3). Tais sequências, tanto de monocotiledôneas quanto de dicotiledôneas utilizadas como entradas (Blast) em consultas nos banco de dados de sorgo, *A. lyrata* e cana-de-açúcar, encontram-se descritas no ANEXO A. O principal objetivo foi recuperar um conjunto inicial de genes com alto grau de similaridade com as MAP Kinases identificadas na literatura, dando início aos trabalhos de caracterização dos componentes MPK/MKKs nestas espécies.

3.2. Identificação das sequências de MPK/MKKs de sorgo e *A. lyrata*

Ambos os conjuntos de sequências de sorgo e *A. lyrata* foram obtidos por meio de consultas diretas ao site (JGI, 2010), utilizando como parâmetros de entradas no BLAST, principalmente, as 20 sequências de aminoácidos de MPK de *A. thaliana* e as 10 MKKs desta mesma espécie. As sequências foram obtidas com base no perfil HMM – *Hidden Markov Model* (HMMER <http://hmmer.wustl.edu/>), e só foram aceitas aquelas que possuíam um

posicionamento adequado da assinatura MPK/MKK no *loop* de ativação da proteína. Os genes redundantes foram identificados por meio de inspeções manuais.

A partir das sequências recuperadas com as consultas diretas, o processo de identificação de MPKs e MKKs de sorgo e *A. lyrata* foi dividido em etapas e analisadas de forma separada. Entretanto, as tarefas desempenhadas foram comuns para a identificação de módulos MAP Kinases em ambas as espécies e são descritas a seguir:

a) Identificação das MPKs de sorgo e *A. lyrata*

Passos:

1- A partir das sequências candidatas obtidas na etapa de alinhamento (*Blast*), foi feita uma seleção inicial de todos os transcritos (sorgo/*A. lyrata*) que possuíam as sequências TEY ou TDY, independente da região onde estavam localizadas na proteína;

2- Em seguida, foi feito um alinhamento utilizando o programa ClustalW dos transcritos recuperados no item 1 de sorgo e cana-de-açúcar, com as sequências das MPKs já caracterizadas na literatura. O objetivo foi identificar todos os componentes que possuem a assinatura (TxY) na região VII e VIII da proteína ou possíveis variações nestes domínios; além da assinatura DxK de proteínas kinases;

Obs: O ClustalW (<http://www.ebi.ac.uk/Tools/clustalw/index.html>) permite o alinhamento de múltiplas sequências e gera os arquivos para estudo de filogenia no formato Phylip e MEGA. As sequências foram fornecidas no formato FASTA. Como todas as sequências a serem alinhadas estavam completas (inteiras), pode-se utilizar os parâmetros *gap open* e *gap extension default* do Clustal.

3- O próximo passo foi investigar cada uma das sequências obtidas no item 2 com a ferramenta *BlastO/Expasy* para detectar similaridades com as sequências anotadas no GenkBank e identificar domínios kinases, respectivamente;

BlastO – fornece, a partir do Genbank (NCBI), ortólogos da proteína fornecida como parâmetro da consulta, caso existam. (<http://oxytricha.princeton.edu/BlastO/>).

EXPASY – identifica assinaturas na proteína alvo, que são especificadas de determinadas famílias de proteínas, auxiliando também na predição de qual família a proteína, objeto da consulta, pertence. (<http://www.expasy.ch/prosite/>).

4- Uma vez feita uma pré-validação no item 3, o próximo passo foi agrupar todas as sequências validadas neste item com as sequências de MPK caracterizadas na literatura e gerar o arquivo no formato *PHYLIP* com a ferramenta ClustalW (onde os parâmetros *gap open* e *gap extension iniciais da ferramenta foram mantidos***).

5- De posse do arquivo no formato PHYLIP obtido no item 4, foi gerada a árvore filogenética com a ferramenta PhyML e feita a validação final dos componentes MPKs identificados.

b) Levantamento de MKKs de sorgo e *A. lyrata*

Passos:

1- O primeiro passo foi alinhar as sequências candidatas a MKK, com MKKs identificadas na literatura, onde foi feita uma pré-validação das sequências (sorgo/*A. lyrata*), verificando-se quais destas continham a assinatura DxK posicionadas adequadamente e também a assinatura *S/TxxxxxS/T*, situada a partir do 26º aminoácido, contado a partir da assinatura DxK. Entretanto, nenhuma sequência que possuía variações na região desses domínios foi descartada nesta etapa.

2- O passo seguinte foi investigar as sequências candidatas (sorgo/*A. lyrata*) obtidas no item 1, como a ferramenta BlastO/Expasy, onde foi possível agrupar todas as que deram alta similaridade com ortólogos de MKKs do NCBI.

3- Uma vez feita uma pré-validação das sequências do item 2, as mesmas foram utilizadas para gerar o arquivo no formato PHYLIP.

4- Uma vez obtido o arquivo do item 3, o mesmo foi utilizado para gerar a árvore filogenética com a ferramenta PhyML para validar as sequências MKKs de sorgo/*A. lyrata*.

** Quando se faz o alinhamento, onde o produto de todos os genes representam sequências inteiras, então pode-se utilizar os parâmetros *default* do CLUSTALW para alinhamento das sequências e obtenção do arquivo no formato PHYLIP para geração da árvore filogenética.

3.3. Identificação dos transcritos MPK/MKKs em cana-de-açúcar

As sequências recuperadas no item 3.1 (literatura) e as identificadas no item 3.2 (de sorgo) foram utilizadas como entrada no Blast do SUCEST para recuperação de um conjunto inicial de transcritos candidatos para análise em cana-de-açúcar. O objetivo foi o de gerar uma base de dados local, uma vez que poderia haver sequências de MAP Kinases (MPK/MKK) codificadas parcialmente no SUCEST (sequências incompletas), onde somente a análise de domínio não seria suficiente para identificá-las.

A identificação dos transcritos de cana-de-açúcar, portanto, constituiu a etapa mais elaborada do processo de identificação em relação às demais espécies e foram realizadas em 3 etapas:

Etapa 1: Geração de uma base de dados local para início dos trabalhos;

Etapa 2: Identificação dos transcritos que correspondem a sequências MPK/MKKs ou que representam “pedaços” da sequência da proteína;

Etapa 3: Compra e sequenciamento dos clones que representam MPK/MKK, mas que não continham todas as assinaturas MPKs/MKKs por não estarem completos o SUCEST.

A seguir, cada uma das etapas é descrita:

3.3.1. Geração da Base de dados local

Uma vez obtido um conjunto de ESTs no SUCEST (SUCEST, 2005), que deram similaridade com MPK/MKKs da literatura e que representavam as sequências de consenso, provenientes de consultas BLAST, o processo de mineração dos transcritos candidatos à MPK/MKKs deram início. O banco de dados de partida para as análises conteve um conjunto inicial de 640 transcritos ou “clusters” (ESTs). As sequências dos “*clusters*” foram recuperadas uma a uma no SUCEST, manualmente, e armazenadas em arquivos texto. Os passos realizados nesta fase estão ilustrados na Figura 14.

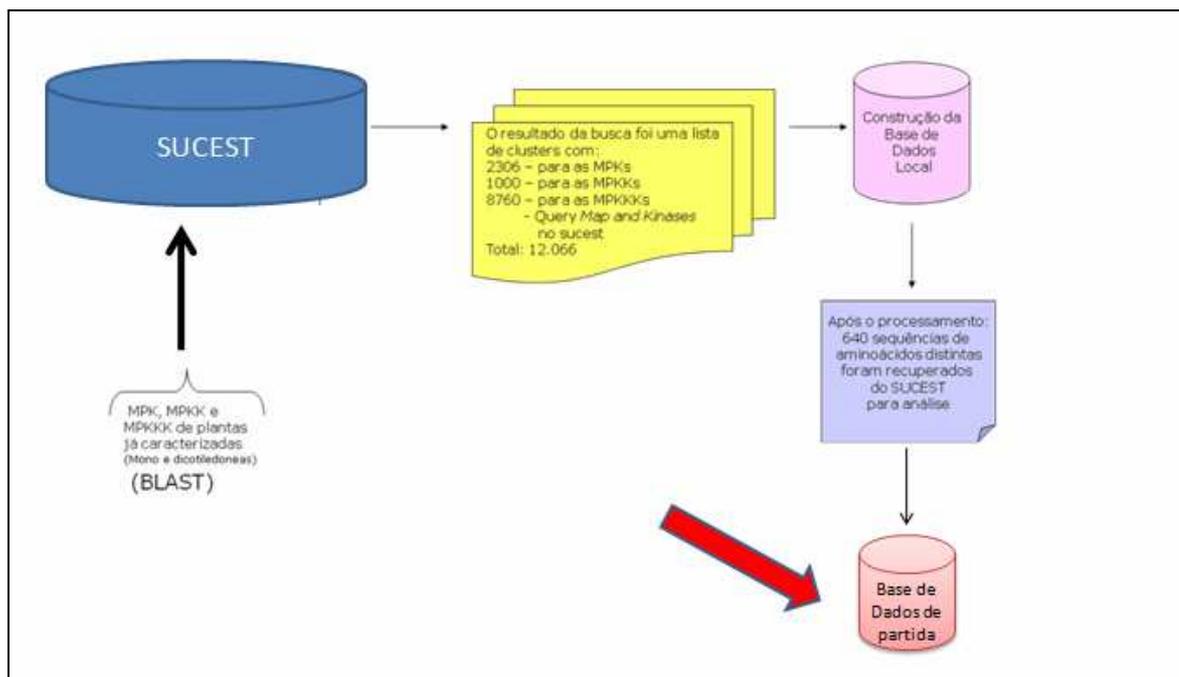


Figura 14: Esquema ilustrativo da Base de dados de partida para a obtenção dos transcritos de cana-de-açúcar

3.3.2. Identificação dos transcritos MPK/MKKs em cana-de-açúcar no SUCEST

O processo de identificação MPKs e MKKs em cana-de-açúcar foi dividido nas etapas descritas a seguir:

a) Levantamento das MPKs de cana-de-açúcar

Passos:

1- A partir das 640 sequências obtidas no SUCEST, o passo seguinte foi inspecionar cada uma das sequências candidatas com a ferramenta BlastO, com o objetivo de detectar transcritos (clusters) candidatos à MPKs em cana-de-açúcar. Nesta etapa, todos os ESTs que deram similaridade com MPKs do NCBI foram agrupados, independente de possuírem ou todos os domínios que caracterizavam a proteína.

2- Uma vez obtidas as sequências candidatas à MPKs em cana-de-açúcar, no item 1, foi feito um o alinhamento com o ClustalW2, em conjunto com as sequências das MPKs já caracterizadas na literatura, com o objetivo de fazer uma pré-validação das mesmas e identificar os ESTs que estavam parcialmente sequenciados. Nestas análises, os parâmetros “gap open” e “gap extension” do clustal tiveram que ser ajustados (ver Anexo B).

Obs: ClustalW (<http://www.ebi.ac.uk/Tools/clustalw/index.html>) – permite alinhamento de múltiplas sequências e gera os arquivos para estudo de filogenia no formato Phylip e MEGA. As sequências devem ser fornecidas no formato FASTA, onde devem ser utilizados os parâmetros *gap open= 50 (ou 100)* e *gap extension = 0.5*. Ver anexo B.

3- Em seguida, as sequências pré-validadas no item 2, em conjunto com as MPKs obtidas em arroz, sorgo e *A. thaliana*, foram agrupadas para gerar o arquivo no formato PHYLIP no ClustalW. Os parâmetros utilizados no ClustalW foram “gap open”= 50 (ou 100) e “gap extension” = 0.5.

4- Uma vez obtido o arquivo do item 3, a árvore filogenética foi gerada com a ferramenta PhyML, completando a etapa de validação das MPK de cana-de-açúcar. Ver anexo C.

A árvore filogenética gerada no item 4, com dos dados de sorgo e cana, foi crucial para uma melhor identificação dos ESTs correspondentes a ortólogos/parálogos nesta espécie e auxiliaram na identificação daqueles que representavam sequências parcialmente codificadas de uma mesma proteína.

b) Levantamento de MKKs de cana-de-açúcar

Passos:

1- No passo 1, do item referente à “identificação de MPKs”, foi feita uma verificação com a ferramenta BlastO das 640 sequências candidatas à (MPKs/MKKs). Nesta etapa, não só ESTs candidatos à MPKs foram identificadas, como também àqueles que deram similaridade com MKKs. Estas sequências foram da mesma forma agrupadas, independente de possuírem ou não a assinatura S/TxxxxxS/T;

2- O passo seguinte foi alinhar as sequências candidatas à MKKs em cana-de-açúcar pré-validadas no item 1 com as MKKs identificadas em sorgo, arroz e *A. thaliana* e verificar quais possuíam, além da assinatura DxK, uma sequência S/TxxxxS/T situada a partir do 26º aminoácido, contado a partir da assinatura DxK. Porém, os transcritos identificados nesta fase não possuíam todas as assinaturas típicas para membros destas famílias que não foram descartados;

3- Uma vez obtidas as sequências completas ou incompletas de transcritos candidatos à MKKs em cana-de-açúcar no passo 2, estas sequências foram utilizadas para gerar o arquivo no formato PHYLIP no ClustalW.

4- O arquivo obtido no item 3 foi utilizado para gerar a árvore filogenética com a ferramenta PhyML e validar os ESTs candidatos à MKKs em cana-de-açúcar.

Um resumo desta etapa e as ferramentas de bioinfo utilizadas podem ser ilustrados por meio da Figura 15.

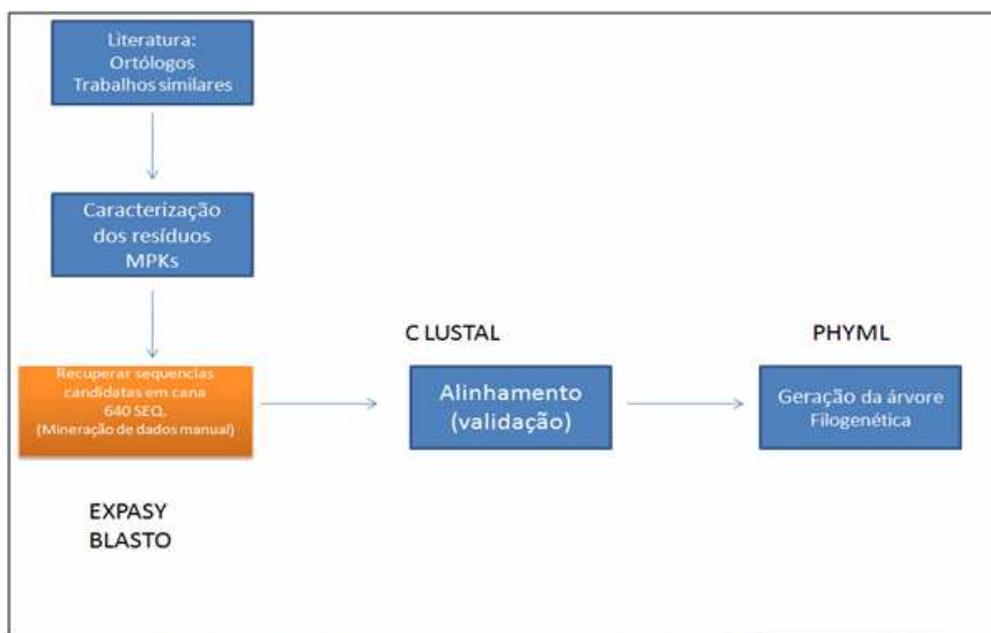


Figura 15. Ferramentas específicas de bioinfo utilizadas em cada fase do processo de identificação dos transcritos de cana-de-açúcar correspondentes a MPK/MKKs

c) Montagem dos ESTs de cana-de-açúcar

Dentro do processo de identificação dos ESTs de cana-de-açúcar que corresponderiam a MPK/MKKs, foram identificadas sequências consenso que representavam partes isoladas de uma mesma proteína MAP Kinase. A árvore filogenética das sequências candidatas de cana-de-açúcar, em conjunto com as sequências de sorgo, forneceram tais evidências, uma vez que sendo sorgo uma espécie próxima geneticamente da cana-de-açúcar, era esperado um número equitativo de genes homólogos para módulos MPK/MKKs para ambas as espécies.

Para a obtenção da sequência ScMKK4, foram utilizados os transcritos SCQSRT2036C06.g e SCEZAM2059E11.g (SUCEST, 2005) e mais a sequência SbMKK4 de sorgo. O alinhamento das sequências foi feito utilizando o ClustalW2, num pareamento 2 a 2 sorgo/cana-de-açúcar. O mesmo procedimento foi feito para a obtenção das sequências do gene ScMKK3 e ScMPK7e seus respectivos ortólogos em sorgo. O esquema pode ser observado na Figura 16.

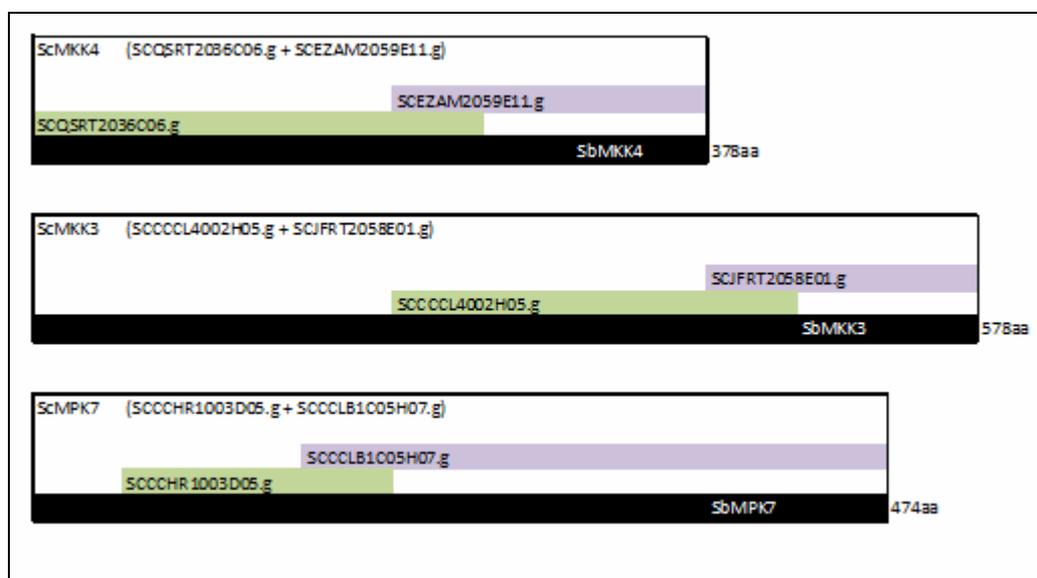


Figura 16. Montagem dos transcritos que correspondem a uma única sequência MPK/MKK em cana-de-açúcar-de-açúcar, a partir das sequências de ortólogos em sorgo.

3.3.3. Sequenciamento dos clones na UENF

Os clones cujas sequências completas não se encontram disponíveis no banco de dados do SUCEST foram adquiridos a partir do “Brazilian Clone Collection Center (BCCC)” FCAV/UNESP/Jaboticabal-SP-Brasil. Foram adquiridos os clones SCACLR2029C10.g (ScMKK1), SCQSRT2036C06.g (ScMKK4), SCUTRZ3070G12.g (ScMKK10) e SCVPRT2084F07.g (ScMPK14). A descrição dos procedimentos realizados para multiplicação e sequenciamento de tais clones está descrita a seguir:

- Crescimento de bactérias para extração de DNA plasmidial.

As bactérias contendo os clones de cDNA do SUCEST, previamente adquiridas, foram crescidas em tubos de 50 mL contendo 5 mL do meio de cultura CircleGrown (Q-Biogene) suplementado com 60 µg/mL de ampicilina. As bactérias foram incubadas a 37 °C sob agitação constante durante 16 h. Em seguida, as células foram centrifugadas a 2500 Xg, durante 20 min a 8°C. O sedimento foi, então, utilizado para extração de DNA plasmidial.

- Extração de DNA Plasmidial contendo os clones de cDNA do SUCEST.

O sedimento resultante da cultura de bactéria foi ressuspeso em 50 µL do próprio meio de cultura. Foram acrescentados 300 µL de TENS (0,5% de SDS, 100 mM de NaOH, 10 mM de Tris-HCl pH 8,0 e 1 mM de EDTA) ao tubo e a amostra agitada intensamente durante 3 min. Imediatamente após, foram acrescentados 150 µL de acetato de sódio 3 M, pH 5,2 e a amostra foi novamente agitada durante 30 s e centrifugada a 13.000 Xg por 3 min. A fase superior foi transferida sem a espuma para um tubo de 1,5 mL e o DNA precipitado com 1 mL de etanol absoluto gelado. A amostra foi misturada por inversão e centrifugada a 13.000 Xg durante 2 min. O sobrenadante foi descartado e o precipitado foi lavado com 1 mL de etanol 70% gelado. O tubo foi agitado lentamente por inversão e, em seguida, centrifugado a 13.000 Xg por 2 min. O sobrenadante foi eliminado e o precipitado foi colocado para secar a 37°C. Após a secagem, o precipitado foi ressuspeso em 40 µL de TE RNase (10 mM de Tris-HCl pH 8,0, 1 mM de EDTA pH 8,0 e 20 µg/mL de RNase) e incubado em banho-maria a 37 °C

por 1 h. Em seguida, para eliminar proteínas, a amostra foi purificada com igual volume de fenol:clorofórmio:alcool isoamílico (25:24:1). Para separar as fases, o tubo foi centrifugado a 12.000 Xg durante 5 min. Em seguida, a fase aquosa foi recuperada, o excesso de fenol eliminado e a amostra foi tratada com igual volume de clorofórmio. Para separar as fases, o tubo foi centrifugado nas mesmas condições e o sobrenadante coletado. Posteriormente, a amostra foi armazenada a -20 °C.

- PCR

Para a PCR foram utilizados 0,5 µl de cDNA, 0,2 mM dNTPs, 1,5 mM MgCl₂, 1X Taq buffer (MBI Fermentas, Lituânia), 10 pmol de iniciadores (tabela do fabricante), 2 U de Taq DNA polimerase I (MBI Fermentas, Lithuania) em um programa (94 °C por 45 s, T_m °C [tabela do fabricante] por 45 s, 75 °C por 1 min e 30 s). As condições de PCR para cada gene foram estão listadas na Tabela do fabricante.

As amostras da PCR foram separadas em gel de agarose 1,2% e coradas em SYBR Gold Nucleic Acid Stain (Invitrogen), segundo recomendações do fabricante. A intensidade das bandas foi determinado através do software TotalLab-TL100 (Nonlinear Dynamics). A intensidade relativa de cada banda foi normalizada pelos valores de intensidade de banda da actina (controle interno), gerando-se, assim, uma razão numérica corrigida.

O DNA do clone 2(SCQSRT2036C06.g) foi amplificado por PCR com o par de oligonucleotídeos iniciadores MapCL2For (5'-TCCTGCCACATCATTTCCTG -3') e MapCL2Rev (5'-:TGAAACTCCTACGCCATCTCC -3') e o DNA do clone 3 (SCVPRT2084F07.g) foi amplificado por PCR com o par de oligonucleotídeos iniciadores MapCL3For (5'- CTACGGACCTTGCGGGAGC -3') e MapCL3Rev (5'-TTTTTACCCAAGCGGAA -3'). A reação de PCR de ambos os clones foi feita em um volume final de 50 µl contendo 1µl (20ng) de DNA total, 20 pmol de cada primer, 200µM de dNTP, 2,5µl de tampão de reação da Taq DNA polimerase 10X, 2mM de MgCl₂ e 1U de Taq DNA polimerase e água ultrapura para completar o volume.

A PCR foi realizada em termociclador com o seguinte ciclo: 5 min. de desnaturação a 95°C, seguido de 40 ciclos com desnaturação a 95°C por 1 min., anelamento a 50°C por 45 seg., extensão a 72°C por 1 min. e extensão final a 72°C por 5 min. Uma alíquota da reação será aplicada em gel de agarose (1%) para confirmar a amplificação.

Em seguida, os produtos amplificados foram precipitados com isopropanol 100% e etanol 70%.

A amplificação de sequenciamento dos fragmentos foi feita em um volume final de 10 µL, onde foram utilizados 1 µL dos fragmentos de interesse obtidos na PCR; 3,2 pmoles de oligonucleotídeos iniciadores; 2,0 µL de tampão 2,5 X; 0,5 µL de Big Dye Terminator Cycle Sequencing v.3 (Applied Biosystems, São Paulo, Brasil). Os oligonucleotídeos iniciadores utilizados foram MapCL2For e MapCL2Rev para o clone 2 e MapCL3For e MapCL3Rev para o clone 3. As condições de amplificação foram: 1 min. e 30 seg. de desnaturação a 95°C; seguidos de 25 ciclos com desnaturação a 95°C, por 25 segundos, anelamento a 50°C por 20 segundos e extensão a 60°C por 4 minutos.

Em seguida, os produtos reamplificados foram purificados utilizando isopropanol 100% e etanol 60% e sequenciados em um sequenciador capilar ABI PRISM 3130 Genetic Analyzer (Applied Biosystems).

As sequências geradas foram editadas através dos programas Phred/Phrap/Consed em sistema operacional Linux (EWING & GREEN, 1998; EWING et al., 1998; GORDON et al., 1998). As sequências foram analisadas pelo BLASTn contra a base de dados do NCBI (National Center for Biotechnology Information, website <http://www.ncbi.nlm.nih.gov>).

3.4. Geração das árvores MPK/MKK com cana-de-açúcar, sorgo e *A. lyrata*

Uma vez identificados todos os módulos correspondentes as sequências MPK/MKK de sorgo, cana-de-açúcar e *A. lyrata*, foi feito um alinhamento conjunto dos componentes MPK/MKK destas espécies em dois grupos de MPK (cana-de-açúcar, sorgo, arroz e *A. thaliana*) e (*A. thaliana*, *A. lyrata* e álamo), separando o grupo das monocotiledôneas e o das dicotiledôneas. Porém, sempre mantendo os genes MPK/MKK de *A. thaliana* para ambos os grupos, uma vez que boa parte do entendimento destas vias se deve a pesquisas feitas nesta espécie. O alinhamento foi feito utilizando a ferramenta ClustalW, marcando a opção (formato

PHYLIP), utilizando os parâmetros (gap open=50 e gap extension=0.5). Uma vez obtido o arquivo (que possui extensão .aln), o mesmo é entregue para análise de filogenia para o programa PhyML on line (<http://www.atgc-montpellier.fr/phyml/>). A árvore resultante é enviada para o e-mail do solicitante. Um resumo do processo de sequenciamento pode ser observado na figura 17.

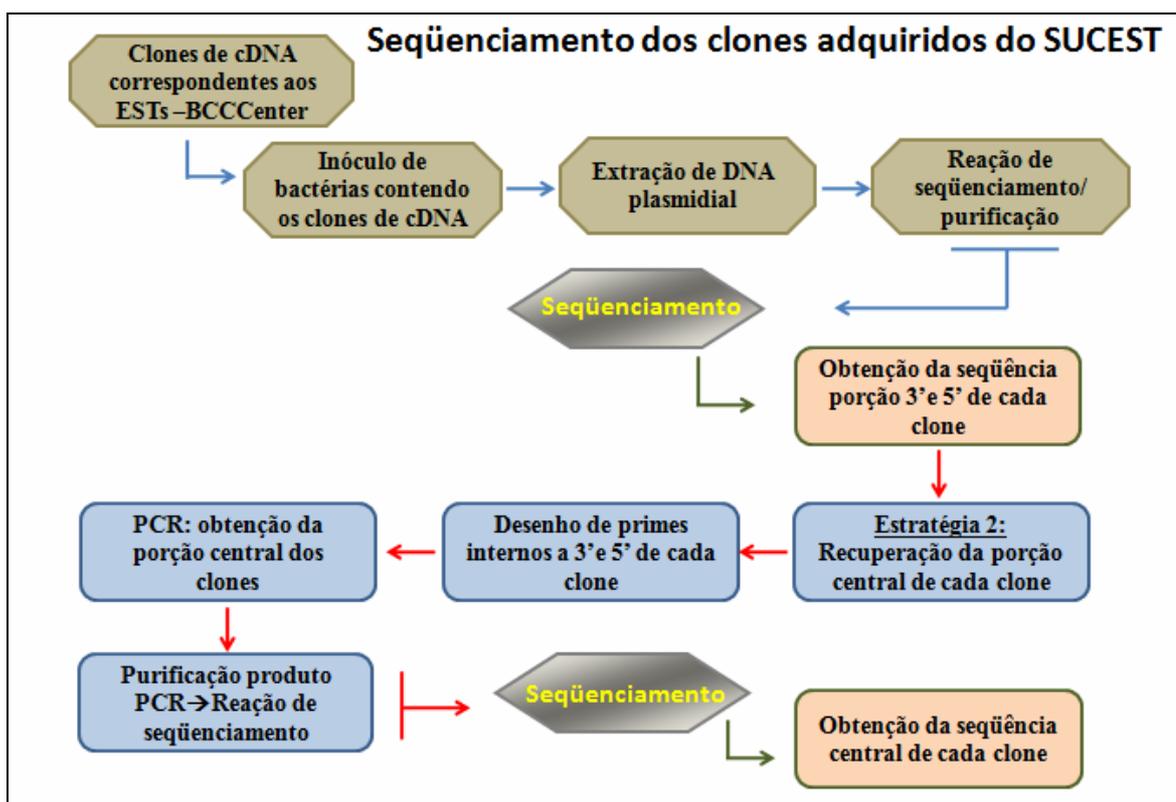


Figura 17. Resumo da etapa de sequenciamento dos clones adquiridos do SUCEST

4. RESULTADOS

4.1. Componentes MPKs

4.1.1. Identificação de componentes MPKs de Sorgo, Cana-de-açúcar e *A. lyrata*

Pesquisas *in silico* em banco de dados genômicos de sorgo e *A. lyrata* e no banco de transcrito da cana-de-açúcar (SUCEST - <http://www.sucest-fun.org>) identificaram, respectivamente, 16, 20 e 13 sequências MPKs. As MPKs de *A. lyrata*, sorgo e cana-de-açúcar foram denominadas seguindo a nomenclatura estabelecida inicialmente para *A. thaliana* (Tabela 2 e Tabela 3).

Todas as sequências MPKs identificadas, com exceção do componente MPK20-4 em cana-de-açúcar, possuem assinaturas específicas para membros destas famílias. Um espectro que caracteriza proteínas MPKs é a presença da assinatura -TxY- situada entre os subdomínios VII e VIII (Figura 18), onde x pode ser o aminoácido (E ou D). Tal domínio é crucial para a ativação da proteína, sendo o alvo de fosforilação pelas proteínas MKKs.

Além disso, toda proteína pertencente a grande família de proteínas kinases (PK), possui uma assinatura típica (-DxK-), onde x pode ser um dos aminoácidos L,V ou K. O alinhamento dos componentes MPKs permitiu uma comparação na estrutura desses genes e revelou que, com exceção das sequências SbMPK12 e AIMP12, que possuem uma variação (-DLR-, onde R

representa o aminoácido arginina), todas as demais MPKs apresentam a assinatura kinase (PK) –DLK-.

Tabela 2. Nomenclatura para MPK e MPKK em *A. thaliana* (At), *A. lyrata* (Al) e *Populus trichocarpa* (Pt).

	<i>Arabidopsis thaliana</i>		<i>Arabidopsis lyrata</i>		<i>Populus trichocarpa</i>	
	At gene (clado)	At gene code	Al gene (clado)	Al Protein ID	Pt gene (clado)	Pt gene Mode
MPK	AtMPK3 (A)	At3g45640	AlMPK3 (A)	484948	PtMPK3-1 (A)	estExt_fgenes4_pm.C_LG_IX0462
	AtMPK6 (A)	At2g43790	AlMPK6 (A)	903782	PtMPK3-2 (A)	fgenes4_pm.C_LG_I000779
	AtMPK10 (A)	At3g59790	AlMPK10 (A)	867420	PtMPK6-1 (A)	estExt_fgenes4_pm.C_LG_VII0025
	AtMPK4 (B)	At4g01370	AlMPK4 (B)	912096	PtMPK6-2 (A)	estExt_Genewise1_v1.C_LG_XVII0005
	AtMPK5 (B)	At4g11330	AlMPK5 (B)	911474	PtMPK4 (B)	estExt_fgenes4_pg.C_LG_XIV0319
	AtMPK11 (B)	At1g01560	AlMPK11 (B)	887292	PtMPK5-1 (B)	fgenes4_pm.C_LG_I000354
	AtMPK12 (B)	At2g46070	AlMPK12 (B)	322261	PtMPK5-2 (B)	fgenes4_pm.C_LG_III000431
	AtMPK13 (B)	At1g07880	AlMPK13 (B)	878970	PtMPK11 (B)	estExt_fgenes4_pg.C_LG_II1481
	AtMPK1 (C)	At1g10210	AlMPK1 (C)	471135	PtMPK1 (C)	grail3.0003026801
	AtMPK2 (C)	At1g59580	AlMPK2 (C)	876031	PtMPK2 (C)	fgenes4_pm.C_LG_V000601
	AtMPK7 (C)	At2g18170	AlMPK7 (C)	319769	PtMPK7 (C)	fgenes4_pm.C_scaffold_57000068
	AtMPK14 (C)	At4g36450	AlMPK14 (C)	353194	PtMPK14 (C)	grail3.0019023301
	AtMPK8 (D)	At1g18150	AlMPK8 (D)	880114	PtMPK9-1 (D)	estExt_fgenes4_pm.C_LG_XII0156
	AtMPK9 (D)	At3g18040	AlMPK9 (D)	898157	PtMPK9-2 (D)	estExt_fgenes4_pm.C_LG_XV0119
	AtMPK15 (D)	At1g73670	AlMPK15 (D)	476507	PtMPK16-1 (D)	eugene3.00081905
	AtMPK16 (D)	At5g19010	AlMPK16 (D)	488785	PtMPK16-2 (D)	estExt_Genewise1_v1.C_LG_X4145
	AtMPK17 (D)	At2g01450	AlMPK17 (D)	864560	PtMPK17 (D)	estExt_Genewise1_v1.C_LG_X6513
	AtMPK18 (D)	At1g53510	AlMPK18 (D)	474479	PtMPK18 (D)	estExt_fgenes4_pg.C_1680031
	AtMPK19 (D)	At3g14720	AlMPK19 (D)	478884	PtMPK19 (D)	estExt_fgenes4_pg.C_LG_I3001
	AtMPK20 (D)	At2g42880	AlMPK20 (D)	483386	PtMPK20-1 (D)	gw1.V.609.1
MPKK	AtMKK1 (A)	At4g26070	AlMKK1 (A)	329070	PtMKK2-1 (A)	eugene3.00180195
	AtMKK2 (A)	At4g29810	AlMKK2 (A)	491777	PtMKK2-2 (A)	estExt_fgenes4_pm.C_LG_VI0403
	AtMKK6 (A)	At5g56580	AlMKK6 (A)	495747	PtMKK6 (A)	estExt_Genewise1_v1.C_1450125
	AtMKK3 (B)	At5g40440	AlMKK3 (B)	493944	PtMKK3 (B)	estExt_fgenes4_pg.C_LG_I2149
	AtMKK4 (C)	At1g51660	AlMKK4 (C)	474280	PtMKK4 (C)	gw1.X.3215.1
	AtMKK5 (C)	At3g21220	AlMKK5 (C)	479642	PtMKK5 (C)	eugene3.00080074
	AtMKK7 (D)	At1g18350	AlMKK7 (D)	472055	PtMKK7 (D)	gw1.122.164.1
	AtMKK8 (D)	At3g06230	AlMKK8 (D)	477923	PtMKK9 (D)	gw1.XII.274.1
	AtMKK9 (D)	At1g73500	AlMKK9 (D)	895208	PtMKK10 (D)	eugene3.00290225
	AtMKK10 (D)	At1g32320	AlMKK10 (D)	890800	PtMKK11-1 (D)	fgenes4_pg.C_LG_VIII001629

Tabela 3. Nomenclatura para MPK e MPKs em *Oryza sativa* (Os), *Sorghum bicolor* (Sb) e Cana-de-açúcar (Sc).

	<i>Oryza sativa</i>		<i>Sorghum bicolor</i>		Sugarcane	
	Os gene	Os gene code	Sb gene	Sb Protein ID	Sc gene (clado)	Sc code SUCEST
MPK	OsMPK3 (A)	Os03g17700	SbMPK3 (A)	4998321/5049657	ScMPK3 (A)	SCCCRZ2C04G10.g/SCSBAM1084E01.g
	OsMPK6 (A)	Os06g06090	SbMPK6 (A)	5061319	ScMPK6 (A)	SCEQRT1030A09.g
	OsMPK4 (B)	Os10g38950	SbMPK4 (B)	5049158	ScMPK4 (B)	SCVPCL6042B11.g
	OsMPK12 (B)	Os08g06060	SbMPK12 (B)	4890234 / 5058423	ScMPK7 (C)	SCCCHR1003D05.g + SCCCLB1C05H07.g/ SCEQRZ3022F11.g
	OsMPK7 (C)	Os06g48590	SbMPK7 (C)	4898568 / 5062195		
	OsMPK14 (C)	Os02g05480	SbMPK14 (C)	148488 / 5054673	ScMPK14 (C)	SCCCRZ1003E02.g/SCCCLB1C01H11.g/ /SCVPRT2084F07.g
	OsMPK16 (D)	Os11g17080	SbMPK16-1(D)	4886131 / 5056541		
	OsMPK17-1 (D)	Os06g49430	SbMPK16-2(D)	146217/4893907/ 5060086	ScMPK16-1(D)	SCQSAM2099A11.g
	OsMPK17-2 (D)	Os02g04230	SbMPK17-1(D)	5062230	ScMPK16-2(D)	SCEZRZ1014F04.g
	OsMPK20-1 (D)	Os01g43910	SbMPK17-2(D)	5054632	ScMPK17-1(D)	SCCCRZ2C04F05.g/SCEZST3149C01.g
	OsMPK20-2 (D)	Os05g50560	SbMPK20-1(D)	125771 / 5053387	ScMPK17-2(D)	SCCCLB1002F04.g
	OsMPK20-3 (D)	Os06g26340	SbMPK20-2(D)	4769120 / 5061034	ScMPK20-1(D)	SCCCLR1067H01.g
	OsMPK20-4 (D)	Os01g47530	SbMPK20-4(D)	143446 / 5053474	ScMPK20-2(D)	SCSGRT2064G11.g/SCCCLR1076B01.g
	OsMPK20-5 (D)	Os05g49140	SbMPK20-5(D)	4895827 / 5060977	ScMPK20-4(D)	SCEPLR1030H10.g
	OsMPK21-1 (D)	Os01g43910	SbMPK21-1(D)	130250 / 5061016	ScMPK20-5(D)	SCSGLR1084B06.g
	OsMPK21-2 (D)	Os01g45620	SbMPK21-2(D)	5036319		
MPKK	OsMKK1 (A)	Os06g05520	SbMKK1 (A)	4896497	ScMKK1 (A)	SCACLR2029C10.g / SCRLAM2050C06.g
	OsMKK6 (A)	Os01g32660	SbMKK6 (A)	4753151 / 5053627	ScMKK6 (A)	SCCCLR1072C05.g/SCEZAM2061E04.g
	OsMKK3 (B)	Os06g27890	SbMKK3 (B)	5007712	ScMKK3 (B)	SCCCL4002H05.g + SCJFRT2058E01.g
	OsMKK4 (C)	Os02g54600	SbMKK4 (C)	5056011 / 144825	ScMKK4 (C)	SCQSRT2036C06.g + SCEZAM2059E11.g
	OsMKK5 (C)	Os06g09180	SbMKK5-1 (C)	5040047 / 4824793	ScMKK10 (D)	SCUTRZ3070G12.g
	OsMKK10-1 (D)	Os02g46760	SbMKK5-2 (C)	4001264		
	OsMKK10-2 (D)	Os03g12390	SbMKK10-1(D)	4809718 / 5038903		
	OsMKK10-3 (D)	Os03g50550	SbMKK10-2(D)	4745203 / 5049890		
			SbMKK10-3(D)	4964479 / 5028418		
			SbMKK10-4(D)	4772587 / 4775909 / 5028419		
			SbMKK10-5(D)	4824504 / 5039573 4807042 / 5036220		
		SbMKK10-6(D)	4970552 / 5033641			
		SbMKK10-7(D)	4804526 / 5037047			

	VII	VIII
AlMPK3	KYIHSANIILHRDLKPSNILLNANCDLKI	CD FGLAR PTSEND----FMTEYVWTRWYRAPE
SbMPK3	KYIHSANVILHRDLKPSNILLNANCDLKI	CD FGLAR PSSES D----MTEYVWTRWYRAPE
ScMPK3	KYIHSANVILHRDLKPSNILLNANCDLKI	CD FGLAR PSSES D----MTEYVWTRWYRAPE
AlMPK6	KYIHSANVILHRDLKPSNILLNANCDLKI	CD FGLAR VTSES D----FMTEYVWTRWYRAPE
SbMPK6	KYIHSANVILHRDLKPSNILLNANCDLKI	CD FGLAR TTSET D----FMTEYVWTRWYRAPE
ScMPK6	KYIHSANVILHRDLKPSNILLNANCDLKI	CD FGLAR TTSET D----FMTEYVWTRWYRAPE
AlMPK10	KYIHSANVILHRDLKPSNILLTTQCDLKI	CD FGLARATPESN----LMTEYVWTRWYRAPE
AlMPK4	KYVHSANVILHRDLKPSNILLNANCDLKI	GD FGLAR TKSET D----FMTEYVWTRWYRAPE
SbMPK4	KYVHSANILHRDLKPSNLLFNANCDLKI	AD FGLAR TTSET D----LMTEYVWTRWYRAPE
ScMPK4	KYVHSANILHRDLKPSNLLFNANCDLKI	AD FGLAR TTSET D----LMTEYVWTRWYRAPE
AlMPK11	KYVHSANVILHRDLKPSNILLNANCDLKI	GD FGLAR TKSET D----FMTEYVWTRWYRAPE
AlMPK5	KYIHSANVILHRDLKPSNILLNSNCDLKI	TD FGLAR TTSET E----FMTEYVWTRWYRAPE
SbMPK12	KYVHSANVILHRDLKPSNILLNANCDLKI	GD FGLAR TTTET D----FMTEYVWTRWYRAPE
AlMPK12	KYVHSANILHRDLKPSNILLNSPQELKI	GD FGLAR TTSDT D----FMTEYVWTRWYRAPE
AlMPK13	KYIHSANVILHRDLKPSNVLNTNCDLKI	CD FGLAR TSNET E----DMTEYVWTRWYRAPE
AlMPK1	KYIHSANILHRDLKPSNILLVNTNCDLKI	CD FGLARAS---NTWQFMTEYVWTRWYRAPE
AlMPK2	KYIHSANILHRDLKPSNILLVNTNCDLKI	CD FGLAR TS---NTWQFMTEYVWTRWYRAPE
AlMPK7	KYLHSANILHRDLKPSNILLVNTNCDLKI	CD FGLAR TS---QGNQFMTEYVWTRWYRAPE
SbMPK7	KYLHSAGILHRDLKPSNILLVNTNCDLKI	CD FGLAR TN---NTWQFMTEYVWTRWYRAPE
ScMPK7	KYLHSAGILHRDLKPSNILLVNTNCDLKI	CD FGLAR TN---NTWQFMTEYVWTRWYRAPE
SbMPK14	KYLHSAEILHRDLKPSNILLVNTNCDLKI	CD FGLAR TN---SSWQFMTEYVWTRWYRAPE
ScMPK14	KYLHSAEILHRDLKPSNILLVNTNCDLKI	CD FGLAR TN---SSWQFMTEYVWTRWYRAPE
AlMPK14	KYLHSANILHRDLKPSNILLVNTNCDLKI	CD FGLAR TS---RGNQFMTEYVWTRWYRAPE
SbMPK20-1	KYIHTANVYHRDLKPPNILANSNCKLKI	CD FGLAR VAFND TPTTI FWIDYVATR WYRAPE
ScMPK20-1	KYIHTANVYHRDLKPPNILANSNCKLKI	CD FGLAR VAFND TPTTI FWIDYVATR WYRAPE
SbMPK20-2	KYIHTANVYHRDLKPPNILANSNCKLKI	CD FGLAR VAFND TPTTV FWIDYVATR WYRAPE
ScMPK20-2	KYIHTANVYHRDLKPPNILANSNCKLKI	CD FGLAR VAFND TPTTV FWIDYVATR WYRAPE
SbMPK20-4	KYIHTANVYHRDLKPPNILANANCKLKI	CD FGLAR VAFND TPTTV FWIDYVATR WYRAPE
ScMPK20-4	KYIHTANVYHRDLKPPNILANANCKLKI	CD FGLAR VAFND ILQYSG---LLQDG IGPLN
SbMPK20-5	KYIHTANVYHRDLKPPNILANANCKLKI	CD FGLAR VAFND SPTTV FWIDYVATR WYRAPE
ScMPK20-5	KYIHTANVYHRDLKPPNILANANCKLKI	CD FGLAR VAFND SPTTV FWIDYVATR WYRAPE
AlMPK18	KPMHTANVYHRDLKPPNILANANCKLKI	CD FGLAR VAFND TPTTV FWIDYVATR WYRAPE
AlMPK19	KPMHTANVYHRDLKPPNILANANCKLKI	CD FGLAR VAFND TPTTV FWIDYVATR WYRAPE
AlMPK20	KYIHTANVYHRDLKPPNILANANCKLKI	CD FGLAR VAFND TPTTI FWIDYVATR WYRAPE
AlMPK8	VHA---ANVFHRDLKPPNILANADCKLKI	CD FGLAR VAFND APTAI FWIDYVATR WYRAPE
AlMPK15	VHA---ANVFHRDLKPPNILANADCKLKI	CD FGLAR VAFND APTAI FWIDYVATR WYRAPE
AlMPK9	IHT---ANVFHRDLKPPNILANSCKLKI	CD FGLAR VAFND APSAI FWIDYVATR WYRAPE
SbMPK21-1	IHA---ASVFHRDLKPPNILANADCKLKI	CD FGLAR VAFND TPSAI FWIDYVATR WYRAPE
SbMPK21-2	IHA---ANVFHRDLKPPNILANGDCKLKI	CD FGLAR VAFND TPSAI FWIDYVATR WYRAPE
SbMPK17-1	IHA---ANVFHRDLKPPNILANSCKLKI	CD FGLAR VAFND APSAI FWIDYVATR WYRAPE
ScMPK17-1	IHA---ANVFHRDLKPPNILANSCKLKI	CD FGLAR VAFND APSAI FWIDYVATR WYRAPE
SbMPK17-2	MHS---ANVFHRDLKPPNILANSCKLKI	CD FGLAR VAFND SLSAI YWIDYVATR WYRAPE
ScMPK17-2	MHS---ANVFHRDLKPPNILANSCKLKI	CD FGLAR VAFND SLSAI YWIDYVATR WYRAPE
AlMPK17	MHS---ANVFHRDLKPPNILANADCKLKI	CD FGLAR VAFND SPSAV FWIDYVATR WYRAPE
AlMPK16	KYIHTANVYHRDLKPPNILANADCKLKI	CD FGLAR VAFND TPTAI FWIDYVATR WYRAPE
SbMPK16-2	KYIHTANVYHRDLKPPNILANADCKLKI	CD FGLAR VAFND TPTAI FWIDYVATR WYRAPE
ScMPK16-2	KYIHTANVYHRDLKPPNILANADCKLKI	CD FGLAR VAFND TPTAI FWIDYVATR WYRAPE
SbMPK16-1	KYIHTANVYHRDLKPPNILANADCKLKI	CD FGLAR VAFND TPTAI FWIDYVATR WYRAPE
ScMPK16-1	IQLIAANVFHRDLKPPNILANADCKLKI	CD FGLAR VAFND TPTAI FWIDYVATR WYRAPE

Figura 18: Alinhamento das seqüências MPKs de sorgo, cana-de-açúcar e *A. lyrata*. O alinhamento foi feito utilizando o programa ClustalW2 diretamente no servidor (gap open 50 e gap ext 0.5). As regiões VII e VIII pertencem ao subdomínio PK (MPK). O motif ki nase (DxK) e o motif de fosforilação (TxY) foram destacados. As seqüências de cana-de-açúcar foram obtidas em (SUCEST – <http://www.sucest-fun.org>), as de sorgo em (<http://genome.jgi-psf.org/cgi-bin/runAlignment?db=Sorbi1&advanced=1>) e *A. lyrata* em (<http://genome.jgi-psf.org/Araly1/Araly1.home.html>), onde os acrônimos utilizados significam: (At – *Arabidopsis thaliana*, Al – *Arabidopsis lyrata*, Os (*Oryza sativa*, Sb-*Sorghum Bicolor*, Sc- cana-de-açúcar e Pt- *Populus trichocarpa*)

Considerando-se o fato de as análises contemplarem 3 espécies de monocotiledôneas (sorgo, arroz e cana-de-açúcar) e três dicotiledôneas (*A. thaliana*, *A. lyrata* e álamo) e com o objetivo de se destacar os dados relevantes observados para ambos os grupos, a análise filogenética com os componentes de MPKs foi dividida em 2 árvores: uma formada com os componentes MPKs de *A. thaliana* em conjunto com as espécies monocotiledôneas e a outra com as MPKs de *A. thaliana* em conjunto com as espécies dicotiledôneas. As árvores das MPKs podem ser observadas nas figuras 19 e 20.

Os dados provenientes da análise filogenética com as MPKs sugerem que nem todos os componentes MPKs observados para *A. thaliana* identificados nos clados A, B, C e D possuem ortólogos em cana-de-açúcar e sorgo, como é o caso dos componentes AtMPK10 (clado A); AtMPK11, AtMPK5 e AtMPK13 (clado B); AtMPK1 e AtMPK2 (clado C); AtMPK8, AtMPK9, AtMPK15, AtMPK18 e AtMPK19 (clado D). Entretanto, o mesmo grupo de ortólogos foi observado para *A. thaliana* e *lyrata*. A análise filogenética também revela uma relativa distância observada para o membro AIMP10, juntamente com AtMPK10, em relação aos demais membros que compõe o clado A de dicotiledôneas. Uma distância relativa também foi observada para o componente AtMPK13. Não foram identificados ortólogos do gene ScMPK12 em cana-de-açúcar no banco de dados do SUCEST. Nossas análises também revelaram a ocorrência do módulo SbMPK21, não identificado em nenhuma das dicotiledôneas analisadas por este trabalho. Adicionalmente, foi observada uma expansão de genes pertencentes ao clado D, especialmente aqueles que pertencem ao grupo das proteínas MPK20, tanto em sorgo quanto em cana-de-açúcar, não observada em membros dos demais clados.

4.1.2. Eventos de paralogia evidenciam expansão do Clado “D” e análise Filogenética revelam duplicações de genes MPKs que ocorreram após a divergência entre monocotiledôneas e dicotiledôneas

Com o objetivo de se verificar a atividade de expansão dos membros do clado D, especialmente observadas em sorgo e cana-de-açúcar, foi feita uma análise investigativa confrontando os dados previamente obtidos para arroz e

álamo (Hamel et al., 2006) com os observados neste trabalho para sorgo, cana-de-açúcar e *A. lyrata*.

Nossas análises revelam que, em sorgo, *A. lyrata* e cana-de-açúcar, assim como observado em arroz e álamo, foram identificados eventos de paralogia associados ao grupo de sequências MPKs específicas do clado D. A análise do grupo de genes que compõe o clado D, MPKs de sorgo e cana-de-açúcar (MPK20-1, MPK20-2, MPK20-4 e MPK20-5), em conjunto com as MPKs de arroz (MPK20-1, MPK20-2, MPK20-3, MPK20-4 e MPK20-5), reforçam as evidências de que estas duplicações, de fato, ocorreram antes da divergência entre monocotiledôneas e dicotiledôneas. Entretanto, outras aparentemente ocorreram em eventos posteriores, como pode ser observado para as MPKs de sorgo (MPK21-1, MPK21-2 e MPK21-3) e de arroz (MPK21-1 e MPK21-2). Em uma comparação do número total de sequências MPKs identificadas neste clado, também foi observado que o número de MPKs em monocotiledôneas é maior, se comparado ao número de MPKs identificadas nas plantas dicotiledôneas. Tais evidências podem ser observadas nas árvores filogenéticas apresentadas nas Figura 19 e 20.

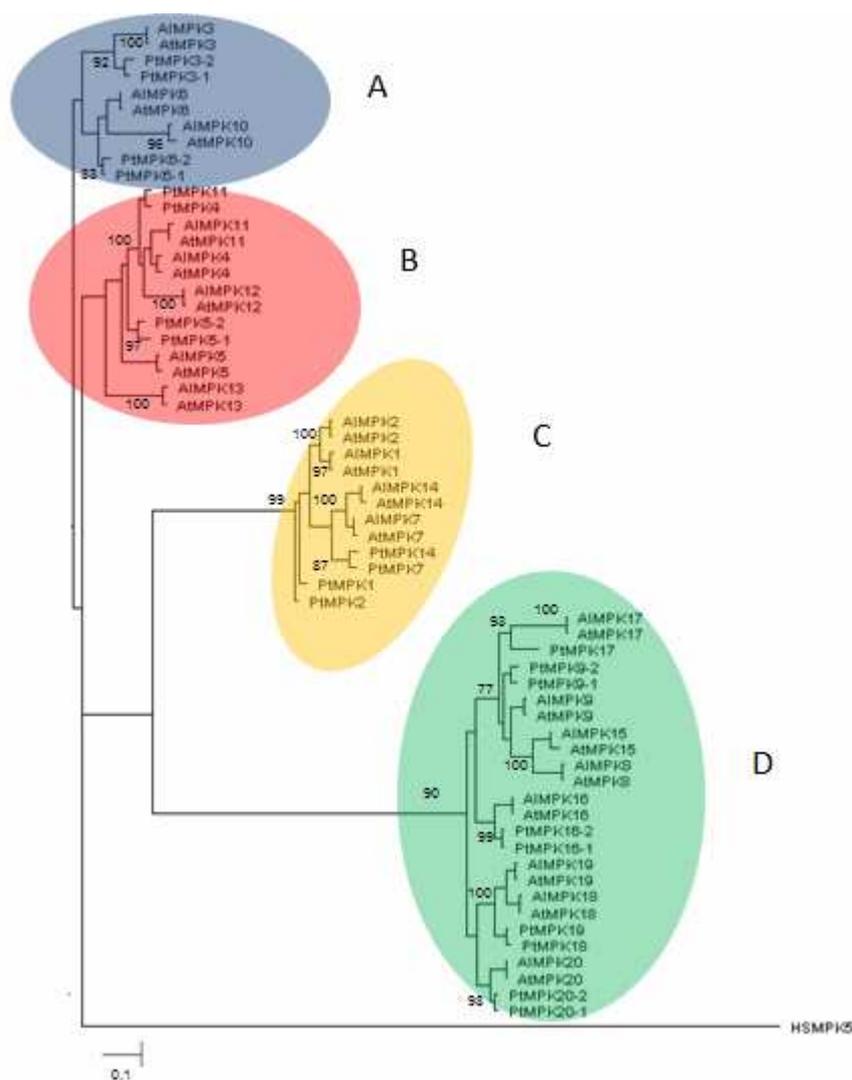


Figura 19. Relacionamentos filogenéticos de genes MPK de *A. Thaliana*, *A. Lyrata* e *populus*. O genoma de *populus* foi obtido em <http://genome.jgi-psf.org/Poptr1/Poptr1.home.html>, usando 20 seqüências de aminoácidos MPK de *Arabidopsis* em consultas diretas. O genoma de *A. Lyrata* foi obtido em <http://genome.jgi-psf.org/Araly1/Araly1.home.html>, da mesma maneira. Só foram aceitas seqüências conônicas de consenso para proteínas kinases serine/threonine, bem como um posicionamento apropriado do loop de ativação motif -TXY-. As seqüências foram alinhadas utilizando o software clustalW com os parâmetros: Pairwise alignment – Gap opening, 35.0, Gap extension, 0.75; Multiple alignment – Gap opening, 50.0, Gap extension, 0.50. O resultado do alinhamento no formato phyml foi submetido ao PHYML online para geração da árvore (<http://atgc.lirmm.fr/phyml/>). (Bootstrap=100). Para identificar as espécies originais de cada MPK, um acrônimo foi incluído antes do nome da proteína: At, *Arabidopsis thaliana*; Al, *Arabidopsis Lyrata*; Pt, *Populus trichocarpa*; Hs, *Homo sapiens*.

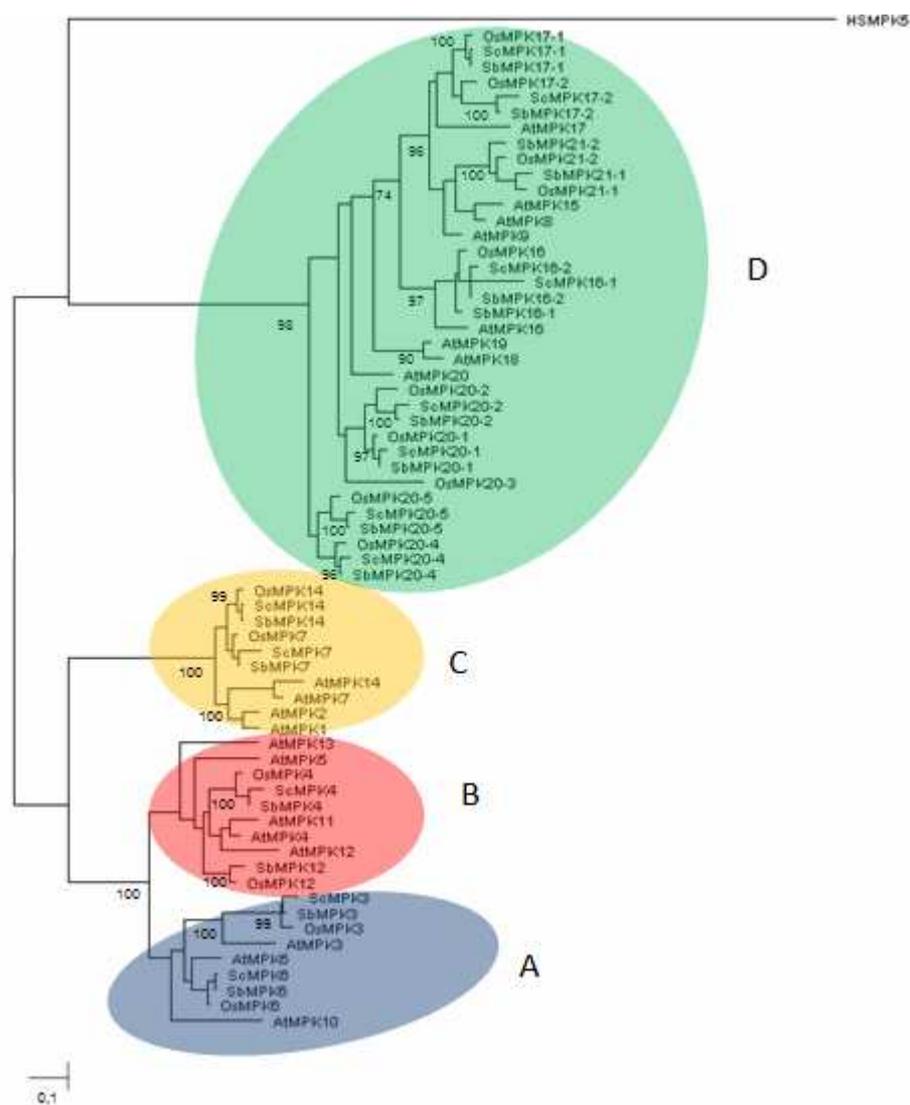


Figura 20. Relacionamentos filogenéticos de genes MPK de *A. Thaliana*, Sorgo, cana-de-açúcar e arroz. O genoma de sorgo foi obtido em <http://genome.jgi-psf.org/cgi-bin/runAlignment?db=Sorbi1&advanced=1>, usando 20 sequências de aminoácidos MPK de *Arabidopsis* em consultas diretas. Os transcritos de arroz foram obtidos em http://www.tigr.org/tdb/e2k1/osa1/data_download.shtml e os de cana-de-açúcar em <http://www.sucest-fun.org/en/database-tools/cane-genome/sucest/cluster-read-search>, por meio de consultas diretas com as MPKs derivadas de *A. Thaliana* e sorgo. Só foram aceitas sequências conônicas de consenso para proteínas kinases serine/threonine, bem como um posicionamento apropriado do loop de ativação motif –TxY–. As sequências foram alinhadas utilizando o software clustaW com os parâmetros: Pairwise alignment – Gap opening, 35.0, Gap extension, 0.75; Multiple alignment – Gap opening, 50.0, Gap extension, 0.50. O resultado do alinhamento no formato phylip foi submetido ao PHYLML online para geração da árvore (<http://atgc.lirmm.fr/phylml/>) (Bootstrap=100). Para identificar as espécies originais de cada MPK, um acrônimo foi incluído antes do nome da proteína: At, *Arabidopsis thaliana*; Al, *Arabidopsis lyrata*; Pt, *Populus trichocarpa*; Hs, *Homo sapiens*.

4.1.3. Ortólogos do AtMPK12 apresentam variação na assinatura kinase (PK)

Com o objetivo de se investigar a variação no domínio kinase (PK) observada na região da assinatura –DxK- que, em módulos MPK12 de sorgo, cana-de-açúcar e *A. lyrata* apresentaram variação –DLR-, foi feita uma análise, separadamente, com ortólogos desses genes em outras espécies, a fim de se verificar se esta variação constitui uma característica típica deste grupo de proteínas pertencentes ao clado B de MPKs. Nossas análises com ortólogo da proteína MPK12 de *A. thaliana*, em conjunto com as MPK12, identificadas por este trabalho, em *A. lyrata* e sorgo, e também em arroz e milho, sendo estas últimas obtidas no banco do NCBI, revelam um novo panorama da região PK. Os dados sugerem que esta assinatura (-DLR-) é típica de genes que compõe o grupo de ortólogos MPK12, tanto em monocotiledôneas quanto em dicotiledôneas e são distintas das demais MPKs identificadas, que apresentam assinatura –DLK-, -DVK- ou –DVK-, descritas como referência na literatura. A Figura 21 mostra o alinhamento feito com as sequências MPK12 de todas as espécies analisadas, em conjunto com os demais componentes pertencentes ao clado B. Tal hipótese é reforçada quando a SbMPK12 é utilizada como sequência de entrada para alinhamento via programa *Blast* junto aos bancos *GenBank* (<http://www.ncbi.nlm.nih.gov/>) e PHYTOZOME v5.0 (<http://www.phytozome.net/>) e, como resultado, foram obtidas as sequências ortólogas de arroz OsMPK12 (Os08g0157000) e milho ZmMPK12 (GRMZM2G123BB6_P01), respectivamente. Tais proteínas, embora encontrem-se depositadas no banco do NCBI e no PHYTOZOME como proteínas kinases, não estão definidas como MPK 12. Nossas análises de filogenia confirmaram que estas proteínas que possuem a assinatura –DLR- pertencem ao clado B e constituem um grupo de ortólogos do gene AtMPK12. O resultado dessa análise pode ser observado na Figura 22.

Em *álamo* e em transcritos de cana-de-açúcar, ortólogos a AtMPK12 não foram identificados. As proteínas AtMPK12 e AIMP12 compartilham cerca de 99% de identidade em relação aos seus domínios, embora o tamanho da sequência das MPK12 de *A. lyrata* seja um pouco maior.

AtMPK4	KYVHSANVLHRDLKPSNLLLNANCDLKGDFGLARTKSETD---FMTEYVVTRWYRAPE
AIMP4	KYVHSANVLHRDLKPSNLLLNANCDLKGDFGLARTKSETD---FMTEYVVTRWYRAPE
PtMPK4	KYVHSANVLHRDLKPSNLFNANCDLKGDFGLARTTSETD---FMTEYVVTRWYRAPE
SbMPK4	KYVHSANILHRDLKPSNLFNANCDLKIADFGGLARTTSETD---IMTEYVVTRWYRAPE
ScMPK4	KYVHSANILHRDLKPSNLFNANCDLKIADFGGLARTTSETD---IMTEYVVTRWYRAPE
OsMPK4	KYVHSANVLHRDLKPSNLFNANCDLKIADFGGLARTTTETD---IMTEYVVTRWYRAPE
AtMPK11	KYVHSANVLHRDLKPSNLLLNANCDLKGDFGLARTKSETD---FMTEYVVTRWYRAPE
AIMP11	KYVHSANVLHRDLKPSNLLLNANCDLKGDFGLARTKSETD---FMTEYVVTRWYRAPE
PtMPK11	KYVHSANVLHRDLKPSNLFNANCDLKGDFGLARTTSETD---FMTEYVVTRWYRAPE
PtMPK5-1	KYVHSANVLHRDLKPSNLLLNANCDLKGDFGLARTTAETD---FMTEYVVTRWYRAPE
PtMPK5-2	KYVHSANVLHRDLKPSNLLLNANCDLKGDFGLARTTSETD---FMTEYVVTRWYRAPE
AtMPK5	KYIHSANVLHRDLKPSNLLLNANCDLKITDFGLARTTSETD---YMTEYVVTRWYRAPE
AIMP5	KYIHSANVLHRDLKPSNLLLNANCDLKITDFGLARTTSETD---FMTEYVVTRWYRAPE
ZmMPK12	KYVHSANVLHRDLRPSNMLLNKCDLKGDFGLARTTTETD---FMMEYVVTRWYRAPE
OsMPK12	KYVHSANVLHRDLRPSNLLLNKCDLKGDFGLARTTNETD---FMMEYVVTRWYRAPE
SbMPK12	KYVHSANVLHRDLRPSNLLLNKCDLKGDFGLARTTTETD---FMMEYVVTRWYRAPE
AtMPK12	KYVHSANILHRDLRPSNVLLNSKNEKIGDFGLARTTSDTD---FMTEYVVTRWYRAPE
AIMP12	KYVHSANILHRDLRPSNVLLNSKNEKIGDFGLARTTSDTD---FMTEYVVTRWYRAPE
AtMPK13	KYIHSANVLHRDLKPSNLVLTNCDLKICDFGLARTSNETE---IMTEYVVTRWYRAPE
AIMP13	KYIHSANVLHRDLKPSNLVLTNCDLKICDFGLARTSNETE---IMTEYVVTRWYRAPE

Figura 21. Alinhamento de membros do clado B, ortólogos da AtMPK12 onde é mostrada a variação no domínio PK. Todas as sequências ortólogas da sequência MPK12 estão em destaque e possuem o domínio -DLR- na região da assinatura kinase. As sequências de cana-de-açúcar foram obtidas em SUCEST – <http://www.sucest-fun.org>, as de sorgo em <http://genome.jgi-psf.org/cgi-bin/runAlignment?db=Sorbi1&advanced=1> e *A. lyrata* em <http://genome.jgi-psf.org/Araly1/Araly1.home.html>. As sequências OsMPK12 e ZmMPK12 foram obtidas do NCBI (<http://www.ncbi.nlm.nih.gov/>). Para identificar as espécies originais de cada MPK, um acrônimo foi incluído antes do nome da proteína: At, *Arabidopsis thaliana*; Al, *Arabidopsis Lyrata*; Pt, *Populus trichocarpa*.

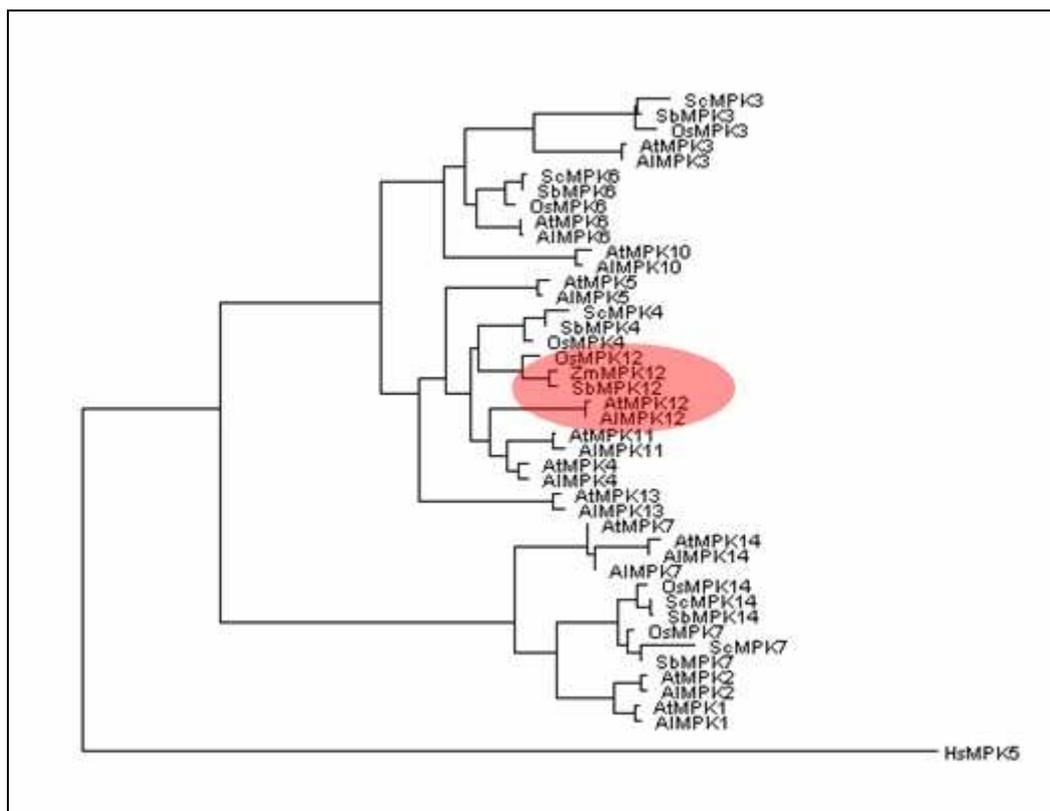


Figura 22. Análise de filogenia conjunta das sequências MPK (-TxY-) revelando que as proteínas do grupo MPK12 são próximas na escala evolutiva. O alinhamento das sequências foi feito com ClustalW2 (gap open=50 gap ext=0.5) O arquivo de saída foi entregue ao programa PhyML para geração da árvore filogenética. As sequências de cana-de-açúcar foram obtidas em SUCEST – <http://www.sucest-fun.org>, as de sorgo em <http://genome.jgi-psf.org/cgi-bin/runAlignment?db=Sorbi1&advanced=1> e *A. lyrata* em <http://genome.jgi-psf.org/Araly1/Araly1.home.html>. Para identificar as espécies originais de cada MPK, um acrônimo foi incluído antes do nome da proteína: At, *Arabidopsis thaliana*; Al, *Arabidopsis Lyrata*; Pt, *Populus trichocarpa*.

4.2. Componentes MKKs

4.2.1. Identificação de componentes MKKs de Sorgo, Cana-de-açúcar e *A. lyrata*

Para o grupo de genes MKKs, as análises nos bancos de dados permitiram identificar 13, 10 e 5 sequências, respectivamente, para sorgo, *A. lyrata* e cana-de-açúcar (Tabela 2 e Tabela 3).

Todas as MKKs identificadas possuem a região canônica, onde ocorre a fosforilação de MKKs por MKKKs em plantas, a típica assinatura *-S/TxxxxxS/T-*. Assim como foi feito com as MPKs, todas as MKKs de *A. lyrata*, sorgo e cana-de-açúcar foram denominadas seguindo a nomenclatura estabelecida inicialmente para *A. thaliana*. Nossas análises identificaram a região do domínio MKK conservada para estas famílias, com exceção dos genes do grupo MKK10 que, curiosamente, para todas as espécies analisadas, revelaram variações de sequência na região do *loop* de ativação da proteína (ver Figura 19). Consultas em bancos de dados públicos (*GenBank*) mostraram que, de um modo geral, os ESTs identificados em cana-de-açúcar para MKKs, assim como para MPKs, possuem identidade acima de 85% com sequências MKKs depositadas no NCBI.

Alguns “clusters” identificados em cana-de-açúcar representam sequências parciais de um mesmo gene de MKK. A identificação segura e consequente montagem da sequência completa do gene foram possíveis graças à análise comparativa com sequências de sorgo, ortólogas para estes genes. Os genes de cana-de-açúcar obtidos por montagem foram MPK7, MKK3 e MKK4. Devido ao elevado grau de ploidia da cana-de-açúcar, a identificação de transcritos correspondentes a diferentes alelos para um mesmo gene é esperada. Os “clusters” que representam possíveis alelos de genes MKK de cana-de-açúcar estão descritos na Tabela 3. O esquema de montagem dos transcritos de cana-de-açúcar, a partir das sequências parciais, pode ser observado no item *Material e Métodos*. Curiosamente, a variação da região PK observada para genes que pertencem ao grupo MPK12, assinatura *-DLR-*, também foi observada para um componente de sorgo, SbMKK10-5. O componente SbMKK5-1 também apresenta uma variação no domínio PK, assinatura *-GIK-*. Entretanto, SbMKK5-1, mas não o SbMKK10-5, parece possuir uma construção adequada no *loop* de ativação da proteína, região *S/TxxxxxS/T* (Ver figura 23).

AlMKK1	FMDGGS	LADLLK	VEKVPEN	----	MLAA	ICKRV	LRGLCY	THHERR	IHR	DLK	PSNLLIN																			
SbMKK1	YMDGGS	LADFLK	TVKTI	PED	----	HLLAA	ICKQV	LKGLI	HLHHER	HIHR	DLK	PSNLLIN																		
ScMKK1-1	YMDGGS	LADFLK	TVKTI	PED	----	HLLAA	ICKQV	LKGLI	HLHHER	HIHR	DLK	PSNLLIN																		
AlMKK2	YMDGGS	LADFLK	SVKTI	PDS	----	YLSA	IFRQV	LQGLI	YLHHR	RIHR	DLK	PSNLLIN																		
AlMKK6	YMDRGS	LADVIR	QVKT	I	LEP	----	YLAV	CKQV	LLGLV	YLHNER	HVIHR	DLK	PSNLLVN																	
SbMKK6	YMDRGS	LADIVK	QVKT	V	LEP	----	YLAV	LCKQV	LEGLL	YLHHER	HVIHR	DLK	PSNLLVN																	
ScMKK6	YMDRGS	LADIVK	QVKT	V	LEP	----	YLAV	LCKQV	LEGLL	YLHHER	HVIHR	DLK	PSNLLVN																	
AlMKK3	YMNGGS	LADILK	VTKKI	I	PEP	----	VLSS	MFHK	LLQGL	SYLHG	VRLVHR	DLK	PANLLIN																	
SbMKK3	YMDGGS	LADVIR	VKKS	I	PEP	----	VLSH	MLQK	GLR	---	YLHE	VRLVHR	DLK	PANVLVN																
ScMKK3	YMDGGS	LADVIR	VKKS	I	PEP	----	VLSH	MLQK	GLR	---	YLHE	VRLVHR	DLK	PANVLVN																
AlMKK4	FMDKG	---	SLEGA	HVWKE	----	QQLA	DL	SRQIL	SGLAY	LHRR	HIVHR	DLK	PSNLLIN																	
SbMKK4	YMDRG	---	SLENH	RIADE	----	HFLA	H	VARQV	LSGI	AYLHRR	HIVHR	DLK	PSNLLID																	
ScMKK4	YMDGG	---	SLENH	RIADE	----	HFLA	H	VARQV	LSGI	AYLHRR	HIVHR	DLK	PSNLLID																	
AlMKK5	FMDQG	---	SLEGA	HVWQE	----	QELA	DL	SRQIL	SGLAY	LHRR	HIVHR	DLK	PSNLLIN																	
SbMKK5-2	YMDGG	---	SLDGR	RIAAE	----	PFLA	D	VARQV	LSGI	AYLHRR	HIVHR	DLK	PSNLLID																	
SbMKK5-1	YMDGG	---	SLNGH	HIIATE	----	PLLA	D	VARQV	LSGI	AYLHRR	HIVHC	G	IK	PSNLLID																
AlMKK10	LMEKGS	LHDALL	LAQV	FTEP	----	MI	STL	ANRIL	QGLR	YV	LQEM	RIVHG	DLK	PSNLLIN																
SbMKK10-7	YMDAGS	LGDLL	RRRRR	----	VGI	PEA	AAV	EA	VAAH	CVV	ALAQ	LHSR	GV	AHL	DVK	PDNLLAN														
SbMKK10-6	YVDAGS	LDRL	LLRR	RRLVGGI	PEA	AA	V	EA	VAAQ	CVV	ALAQ	LHSR	GV	AHL	DVK	PDNLLAN														
SbMKK10-5	LMDAGT	LAGIV	RRRG	----	RG	I	PE	CA	LA	EA	VAAH	VHSR	GV	AHL	D	IK	PDNLLAN													
SbMKK10-2	LMPGGS	LSGL	LRLGR	----	EM	GER	P	IA	A	VARQ	AL	LGLA	L	HAL	R	V	VHR	DLK	PSNLLVG											
SbMKK10-4	LADGGS	LDAV	RTRRG	----	AFA	E	A	A	L	E	VAAQ	A	L	SGLAY	LHARR	I	V	H	L	DLK	PANLLAT									
SbMKK10-3	LVDGGS	LDAVA	ARRG	----	AF	S	E	A	A	L	E	V	T	AQAL	LAGL	A	H	L	QARR	V	VHR	DVK	PANLLVS							
SbMKK10-1	YVPGGS	LDVLR	RCRGL	PED	----	AI	AG	V	A	R	H	V	L	C	G	L	R	L	H	R	L	G	I	V	H	G	DVK	PSNLLVG		
AlMKK7	YMDGGN	LESLR	GAVTEK	----	Q	L	A	G	F	S	R	Q	I	L	K	G	L	S	Y	L	H	S	L	K	L	V	H	R	DLK	PANLLLN
AlMKK8	YMDLGT	LESLS	-GVSEN	----	Q	L	A	L	M	A	R	Q	V	L	E	G	L	N	Y	L	H	E	H	K	I	V	H	R	DLK	PANLLRS
AlMKK9	YMDGGT	LESLR	GGVTEQ	----	K	L	A	G	F	A	K	Q	I	L	K	G	L	S	Y	L	H	A	L	K	I	V	H	R	DLK	PANLLLN

A

AlMKK1	HRGE	----	VKIT	DFGV	SKIL	T	S	S	L	A	N	S	F	V	G	T	Y	P	M	S	P	E	R	I	S	G	S	L	Y	S	N	K	----																
AlMKK2	HRGE	----	VKIT	DFGV	STV	M	T	N	T	A	G	L	A	N	T	F	V	G	T	Y	N	M	S	P	E	R	I	G	N	K	Y	S	N	K	----														
SbMKK1	HRGE	----	VKI	S	D	F	G	V	S	A	I	S	S	S	G	R	-D	T	F	G	T	Y	N	M	A	P	E	R	I	S	G	K	K	H	G	Y	M	----											
ScMKK1	HRGE	----	VKI	S	D	F	G	V	S	A	I	S	S	S	G	R	-D	T	F	G	T	Y	N	M	A	P	E	R	I	S	G	K	K	H	G	Y	M	----											
AlMKK6	HKGE	----	VKI	S	D	F	G	V	S	A	S	I	A	S	S	M	Q	R	D	T	F	V	G	T	Y	N	M	S	P	E	R	I	S	G	S	T	Y	D	S	----									
SbMKK6	RKGE	----	VKIT	DFGV	S	A	V	L	A	S	S	I	G	Q	R	D	T	F	V	G	T	Y	N	M	A	P	E	R	I	S	G	S	S	Y	D	Y	K	----											
ScMKK6	RKGE	----	VKIT	DFGV	S	A	V	L	A	S	S	I	G	Q	R	D	T	F	V	G	T	Y	N	M	A	P	E	R	I	S	G	S	S	Y	D	Y	K	----											
AlMKK3	LKGE	----	PKIT	DFG	I	S	A	G	L	N	S	M	A	M	C	A	T	F	V	G	T	V	T	Y	M	S	P	E	R	I	R	N	D	S	Y	S	P	A	D	I	W	S	L	G	L				
SbMKK3	LKGE	----	AKIT	DFG	V	S	T	G	L	D	N	T	M	A	M	C	A	T	F	V	G	T	V	T	Y	M	S	P	E	R	I	R	N	E	N	Y	S	A	A	D	I	W	S	L	G	L			
ScMKK3-2	LKGE	----	AKIT	DFG	V	S	T	G	L	D	N	T	M	A	M	C	A	T	F	V	G	T	V	T	Y	M	S	P	E	R	I	R	N	E	N	Y	S	A	A	D	I	W	S	L	G	L			
ScMKK3-1	LKGE	----	AKIT	DFG	V	S	T	G	L	D	N	T	M	A	M	C	A	T	F	V	G	T	V	T	Y	M	S	P	E	R	I	R	N	E	N	Y	S	A	A	D	----								
AlMKK4	SAKN	----	VKI	A	D	F	G	V	S	R	I	L	A	Q	T	M	D	P	C	N	S	V	G	T	I	A	Y	M	S	P	E	R	I	N	T	D	L	N	Q	G	L	----	Y	D					
AlMKK5	SAKN	----	VKI	A	D	F	G	V	S	R	I	L	A	Q	T	M	D	P	C	N	S	V	G	T	I	A	Y	M	S	P	E	R	I	N	T	D	L	N	H	G	R	----	Y	D					
SbMKK4	SGRR	----	VKI	A	D	F	G	V	G	R	I	L	N	Q	T	M	D	P	C	N	S	V	G	T	I	A	Y	M	S	P	E	R	I	N	T	D	L	N	D	G	A	----	Y	D					
ScMKK4	SGRR	----	VKI	A	D	F	G	V	G	R	I	L	N	Q	T	M	D	P	C	N	S	V	G	T	I	A	Y	M	S	P	E	R	I	N	T	D	L	N	D	----	Y	D							
SbMKK5-2	SARR	----	VKI	A	D	F	G	V	G	R	I	L	N	Q	T	M	D	P	C	N	S	V	G	T	I	A	Y	M	S	P	E	R	I	N	T	D	L	N	D	G	S	----	Y	D					
SbMKK5-1	SARH	----	VKI	A	E	F	G	V	G	H	I	L	Q	T	M	E	P	S	N	S	V	G	T	I	A	Y	M	S	P	E	Q	I	N	T	N	L	S	D	G	----	Y	D							
SbMKK10-7	TRGE	----	IKI	S	D	F	N	L	S	R	I	L	Y	G	G	S	G	E	-R	L	V	P	I	T	G	G	T	L	M	Y	L	S	P	E	R	F	A	P	N	A	R	A	G	----	P	H	A		
SbMKK10-6	ARGE	----	IKI	S	D	F	N	L	S	R	I	L	Y	G	G	S	G	E	-R	L	V	P	I	T	G	G	T	R	A	Y	F	S	P	E	R	F	A	P	N	A	H	A	G	----	P	H	G		
SbMKK10-5	FRGD	----	IKI	G	D	F	S	V	S	R	I	L	F	G	R	I	D	A	-R	R	K	V	S	V	A	V	G	S	P	M	Y	L	S	P	E	R	F	E	P	D	A	H	A	E	----	P	R	G	
SbMKK10-2	AGGE	----	VKI	A	D	F	G	A	G	K	V	L	R	R	L	D	P	----	C	A	S	Y	V	G	T	A	A	Y	M	S	P	E	R	F	D	P	E	A	Y	S	G	----	D	Y	D				
SbMKK10-4	TAGE	----	IKV	A	D	F	G	I	A	R	V	L	P	H	A	G	----	D	H	C	T	S	Y	V	G	T	A	A	Y	M	S	P	E	R	F	D	P	E	A	H	G	----	H	Y	D				
SbMKK10-3	AAGE	----	DKI	T	D	F	G	I	A	K	V	L	S	R	A	G	----	D	H	C	T	A	Y	E	G	T	A	A	Y	M	S	P	E	R	F	D	T	E	R	H	G	----	H	A	D				
AlMKK10	NKGE	----	VKI	A	D	F	G	A	S	R	I	V	A	G	G	D	Y	G	S	N	----	G	T	C	A	Y	M	S	P	E	R	V	D	P	E	K	W	G	F	G	----	G	E	V					
SbMKK10-1	ROGE	----	VKI	A	D	F	G	A	S	R	H	V	P	S	A	G	G	A	H	R	L	A	G	A	E	G	T	C	A	Y	M	S	P	E	R	L	D	P	E	S	F	G	A	S	P	S	G	A	----
AlMKK7	SRNR	R	I	S	T	G	E	S	P	L	K	F	G	V	S	K	I	M	I	R	S	L	D	Y	C	N	S	V	G	T	C	A	Y	M	S	P	E	R	F	D	S	-V	A	G	E	N	----	S	D
AlMKK9	SRNE	----	VKI	A	D	F	G	V	S	K	I	L	V	R	S	L	D	S	C	N	S	V	G	T	C	A	Y	M	S	P	E	R	F	D	S	E	S	S	G	G	S	----	S	D					
AlMKK8	SKEE	----	VKI	T	D	F	G	V	S	K	I	V	V	K	S	F	N	K	C	N	S	V	G	T	Y	A	Y	M	S	P	E	R	L	D	S	E	A	D	G	V	T	E	E	D	----	K	S	N	

B

Figura 23. Regiões de assinatura conservadas para MKKs: (A) Região da assinatura kinase (PK) em destaque para as sequências MKK de sorgo, *A.lyrata* e Cana-de-açúcar e a (B) região do subdomínio MKK S/TxxxxxS/T em amarelo. O alinhamento foi feito utilizando o programa ClustalW2 (gap open=50 e gap ext=0.5). As regiões destacadas na horizontal ilustram as MKKs que não possuem uma construção adequada no seu loop de ativação. As sequências de cana-de-açúcar foram obtidas em SUCEST – <http://www.sucest-fun.org>, as de sorgo em <http://genome.jgi-psf.org/cgi-bin/runAlignment?db=Sorbi1&advanced=1> e *A. lyrata* em <http://genome.jgi-psf.org/Araly1/Araly1.home.html>. Para identificar as espécies

originais de cada MPK, um acrônimo foi incluído antes do nome da proteína: At, *Arabidopsis thaliana*; Al, *Arabidopsis lyrata*; Pt, *Populus trichocarpa*.

A análise filogenética dos componentes de MKKs foi dividida em 2 grupos: o primeiro grupo com *A. thaliana* em conjunto com as espécies monocotiledôneas e o segundo grupo com *A. thaliana* em conjuntos com as espécies dicotiledôneas. O objetivo foi o destacar os dados relevantes observados para ambos os grupos, conforme feito anteriormente para as MPKs. Nossas análises revelam que módulos MKK10, para todas as espécies observadas, apresentam uma distância relativa dos demais membros do clado a que pertencem (clado D). O mesmo pode ser observado para o componente SbMKK5-1. A análise filogenética com os componentes de MKKs de monocotiledôneas e dicotiledôneas pode ser observada nas figuras 24 e 25. Com o objetivo de se investigar melhor a divergência e expansão apresentada pelos componentes do clado D de MKKs de ortólogos do gene MKK10, uma árvore filogenética parcial com membros de todas as espécies para este clado foi gerada (Figura 26). Nossas análises permitiram verificar que, aparentemente, os componentes do grupo MKK10 guardam uma distância relativa, inclusive entre membros parálogos dentro deste mesmo clado (Osmkk10-1 e SbMKK10-1 em relação aos outros componentes MKK10 da sua espécie). Membros MKK10 apresentam também forte atividade de expansão, principalmente verificado entre as plantas monocotiledôneas. Entretanto, apesar de esse grupo de proteínas, tanto em monocotiledôneas quanto em dicotiledôneas, apresentarem uma construção inadequada do *loop* de ativação, a atividade de expansão observada para módulos MKK10 não foi observada entre ortólogos das dicotiledôneas analisadas.

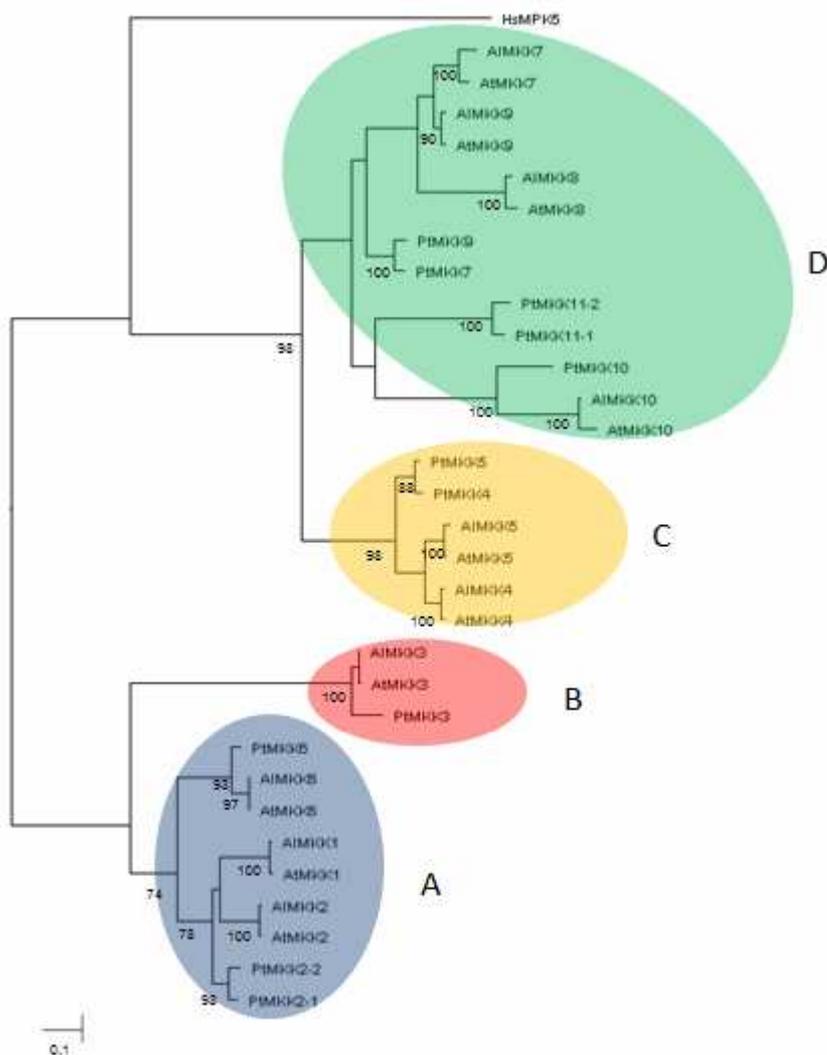


Figura 24. Relacionamentos filogenéticos de genes MKK de *A. Thaliana*, *A. Lyrata* e *populus*. O genoma de *populus* foi obtido em <http://genome.jgi-psf.org/Poptr1/Poptr1.home.html>, usando 10 seqüências de aminoácidos MPK de *Arabidopsis* em consultas diretas. O genoma de *A. Lyrata* foi obtido em <http://genome.jgi-psf.org/Araly1/Araly1.home.html>, da mesma maneira. Só foram aceitas seqüências conônicas de consenso para proteínas kinases serine/threonine, bem como um posicionamento apropriado do loop de ativação motif $-S/TxxxxxS/T-$. As seqüências foram alinhadas utilizando o software clustalW com os parâmetros: Pairwise alignment – Gap opening, 35.0, Gap extension, 0.75; Multiple alignment – Gap opening, 50.0, Gap extension, 0.50. O resultado do alinhamento no formato phylib foi submetido ao PHYLML online para geração da árvore (<http://atgc.lirmm.fr/phyml/>). (Bootstrap=100). Para identificar as espécies originais de cada MKK, um acrônimo foi incluído antes do nome da proteína: At, *Arabidopsis thaliana*; Al, *Arabidopsis Lyrata*; Pt, *Populus trichocarpa*; Hs, *Homo sapiens*.

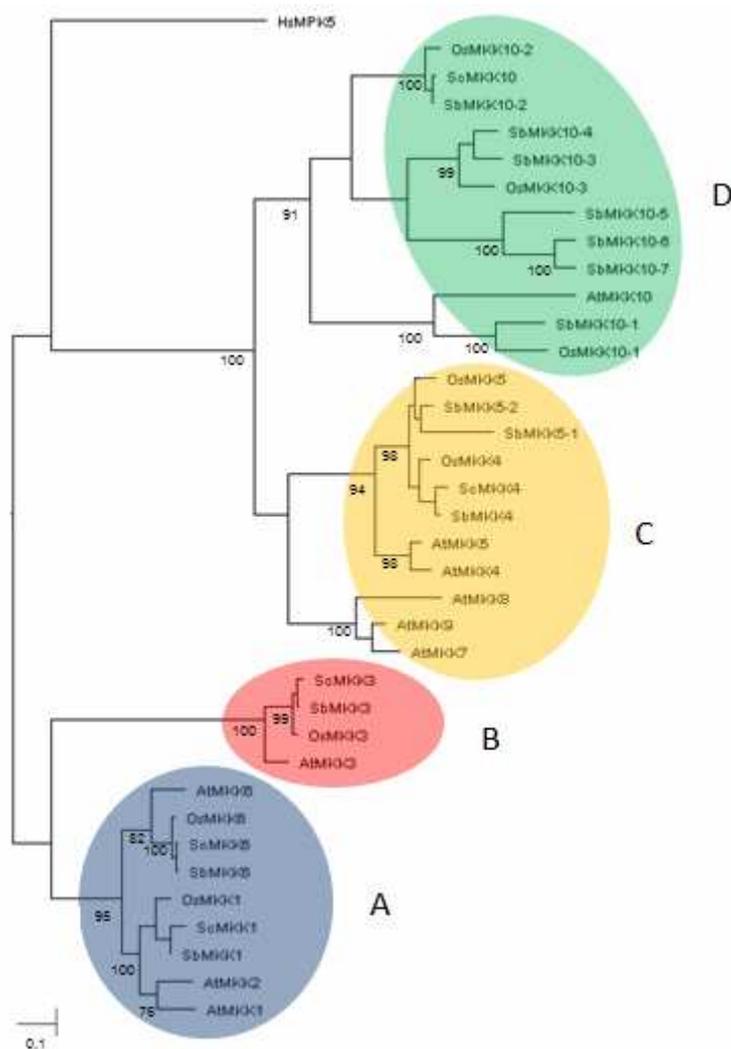


Figura 25. Relacionamentos filogenéticos de genes MKK de A. Thaliana, Sorgo, cana-de-açúcar e arroz. O genoma de sorgo foi obtido em <http://genome.jgi-psf.org/cgi-bin/runAlignment?db=Sorbi1&advanced=1>, usando 10 sequências de aminoácidos MKK de Arabidopsis em consultas diretas. Os transcritos de arroz foram obtidos em http://www.tigr.org/tdb/e2k1/osa1/data_download.shtml e os de cana-de-açúcar em <http://www.sucest-fun.org/en/database-tools/cane-genome/sucest/cluster-read-search>, por meio de consultas diretas com as MKKs derivadas de A. Thaliana e sorgo. Só foram aceitas sequências conônicas de consenso para proteínas kinases serine/threonine, bem como um posicionamento apropriado do loop de ativação motif $-S/TxxxxxS/T-$. As sequências foram alinhadas utilizando o software clustalW com os parâmetros: Pairwise alignment – Gap opening, 35.0, Gap extension, 0.75; Multiple alignment – Gap opening, 50.0, Gap extension, 0.50. O resultado do alinhamento no formato phylip foi submetido ao PHYLML online para geração da árvore (<http://atgc.lirmm.fr/phyml/>). (Bootstrap=100). Para identificar as espécies originais de cada MKK, um acrônimo foi incluído antes do nome da proteína: At, Arabidopsis thaliana; Al, Arabidopsis Lyrata; Pt, Populus trichocarpa; Hs, Homo sapiens.

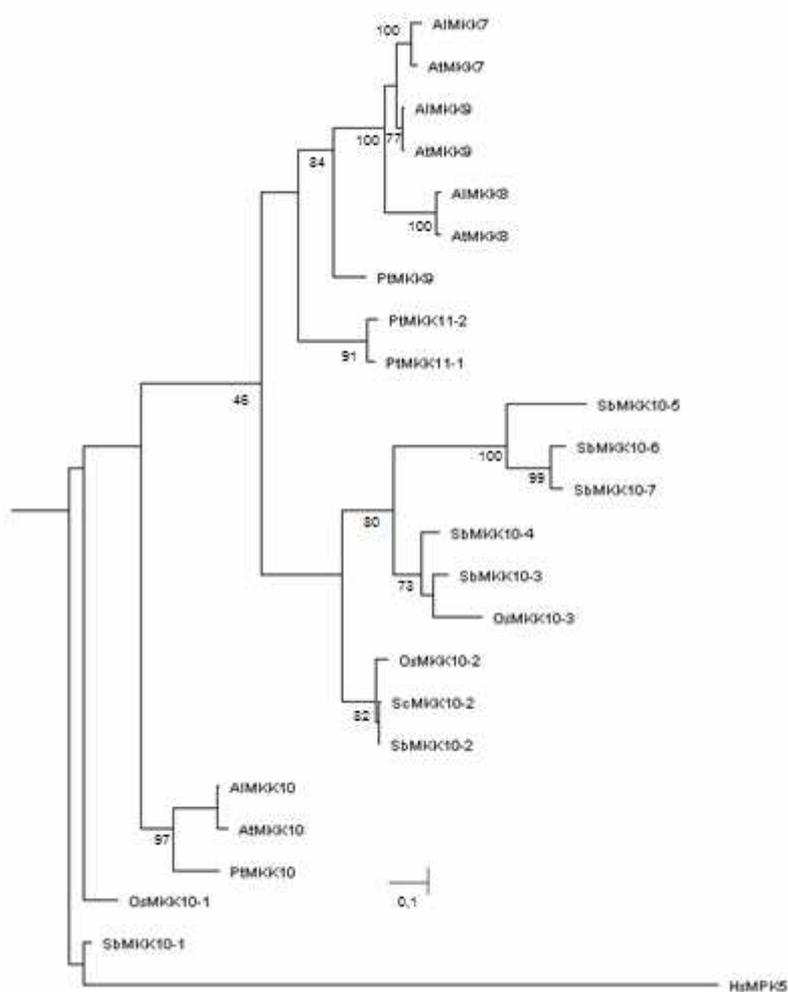


Figure 26. Árvore filogenética com as seqüências de MKKs de sorgo, cana-de-açúcar, *A. lyrata*, *A. thaliana*, álamo arroz. Para obter uma melhor resolução, um alinhamento separado das seqüências de aminoácidos do clado D foi proposto e a árvore filogenética foi construída utilizando a ferramenta PHYML, utilizando o modelo de substituição JTT – bootstrap=100. As seqüências de cana-de-açúcar foram obtidas em SUCEST – <http://www.sucest-fun.org>, as de sorgo em <http://genome.jgi-psf.org/cgi-bin/runAlignment?db=Sorbi1&advanced=1> e *A. lyrata* em (<http://genome.jgi-psf.org/Araly1/Araly1.home.html>). Acrônimos foram utilizados para identificar as espécies (AI – *A. lyrata*, At- *A.thaliana*, Pt – *Populus trichocarpa*, Sc – *Sugarcane* e Os – *Oryza sativa* e Sb – *Sorghum bicolor*)

4.2.2. MKK3 de Sorgo, cana-de-açúcar e *A. lyrata* possuem domínio NTF conservado

Um dos mais interessantes clados MKK é composto pelas sequências do gene MKK3, pertencentes ao Grupo B. Este gene também foi identificado em sorgo, cana-de-açúcar e *A. lyrata*, bem como já havia sido observado no genoma de arroz. Uma característica incomum destes genes é o fato de possuírem uma extensão 3' codificando um domínio NTF ('Fator de Transferência Nuclear') (Ver figura 27).

```

SbMKK3 QKDADARPMCEKLLSHPPFKRYAGTEVLDLAAVVKSVVDPTERLQQAEMLAIHYYLLFNGSDGIWHHMKTFFMEQSTFSFSGRVTVQNDI FDSL NNRKRLGDRPREKI VHVVEKLHC
S-mMKK3 QKDADARPMCEQLLSHPPFKRYAGTEVLDLAAVVKSVVDPTERLQQAEMLAIHYYLLFNGSDGIWHHMKTFFMEQSTFSFSGRVTVQNDI FDSL NNRKRLGDRPREKI VHVVEKLHC
O-mMKK3 QKDADAR-SCQQLLSHPPFKRYENTVLDLAAVVKSVVDPTERLQQAEMLAIVHYLLFNGSDGIWHHMKTFFMEQSTFSFSGRVTVQNDI FDSL NNRKRLGDRPREKI VHVVEKLHC
A-mMKK3 QKDPDARPTADQLLSHPPFKRERVDLATVQSIPTDPTQRLEKLA DMLTIHYYSLFDGFDLWVHAKSLYTETS VTFSPSGKHNT GSTEI PSALS DIRNT LTGDL PSEKL VHVVEKLHC
AlMKK3 QKDPDARPTADQLLSHPPFKRERVDLATVQSIPTDPTQRLEKLA DMLTIHYYSLFDGFDLWVHAKSLYTETS VTFSPSGKHNT GSTEI PSALS DIRNT LTGDL PSEKL VHVVEKLHC

SbMKK3 RANGEDGVAIRVSGSFTVGNQFLVCGDGIQAEGLPSLDELSIDIPSKRVGQFREQFIMLPGNLMSCYYISKQDLYI-----
S-mMKK3 RANDETGVAIRVSGSFTVGNQFLVCGDGIQAEGLPSMDELSIDIPSKRVGQFREQFIKATGEFVWMLLHTRARSVHYPILEHII-----
O-mMKK3 RAHGETGVAIRVSGSFTVGNQFLVCGDGLQAEGLPSLEELSIDIPSKRVGQFREQFIMEPGSVMGCYYILRQDLYI I-----
A-mMKK3 KPCGSGGVIIRAVGFTVGNQFLVCGDGIQAEGLPSFKDLGFDVASERVGRFQEQVVESSGLIGKYFLAQELYI-----
AlMKK3 KPHSGGVIIRAVGFTVGNQFLVCGDGIQAEGLPSFKDLGFDVASERVGRFQEQVVESSGLIGKYFLAQELYI-----

```

Figura 27. Alinhamento do domínio NTF das MKK3 de sorgo, cana-de-açúcar, arroz, *A. thaliana* e *A. lyrata*. A região do transportador nuclear NTF está em amarelo. O alinhamento foi feito com o programa ClustalW2 (gap open 50 gap ext 10.0). As sequências de cana-de-açúcar foram obtidas em SUCEST – <http://www.sucest-fun.org>, as de sorgo em <http://genome.jgi-psf.org/cgi-bin/runAlignment?db=Sorbi1&advanced=1> e *A. lyrata* em <http://genome.jgi-psf.org/Araly1/Araly1.home.html>.

Embora proteínas contendo domínio NTF sejam codificadas em outros genomas eucarióticos, incluindo *A. thaliana*, a combinação verificada em plantas de genes com função MAP Kinase Kinase (MKKs) e que incluem domínio NTF parece ser única.

4.3. Identificação de componentes MPKs/MKKs específicos para monocotiledôneas e dicotiledôneas

Análises filogenéticas revelam componentes MPKs comuns encontrados no genoma de plantas superiores. Entretanto, ao longo do processo evolutivo, alguns módulos aparentemente continuaram se expandindo e deram origem a grupos de genes específicos de plantas monocotiledôneas, outros permaneceram

monofiléticos e só foram preservados para plantas dicotiledôneas. Dentre os componentes MPKs típicos de plantas dicotiledôneas, podemos citar as proteínas do grupo MPK1, MPK2, MPK5, MPK8, MPK9, MPK10, MPK11, MPK13, MPK15, MPK18 e MPK19. Dentre os componentes MPKs encontrados no genoma de ambos os grupos (monocotiledoneas e dicotiledoneas), podemos citar MPK3, MPK4, MPK6, MPK7, MPK12, MPK14, MPK16, MPK17 e MPK20. Também pode ser observado que cerca de 50% dos módulos MKKs também permaneceram monofiléticos. Dentre estes módulos, citamos o MKK2, MKK7, MKK8, MKK9 e MKK11. Curiosamente, vale ressaltar que só foram observados módulos MKK11 em álamo. Em *A. lyrata* e *A. thaliana*, bem como nas demais espécies, este componente não foi identificado. As figuras 28 e 29 mostra de forma conjunta os componente MPK e MKK nas diversas espécies analisadas, onde o símbolo (*) indica aqueles que não foram identificados na espécie. Desta forma, é possível uma visão geral de como estes componentes estão estruturados e organizados nos diferentes clados (A, B, C e D) nas várias espécies. A Figura 30 ajuda a ilustrar os módulos que são comuns e os módulos que são exclusivos de monocotiledôneas e dicotiledôneas.

grupo	<i>A. Thaliana</i>	<i>A. Lyrata</i>	<i>Populus Trichocarpa</i>	<i>Oryza sativa</i>	<i>Sorghum bicolor</i>	<i>Sugarcane</i>
A	AtMPK3	AIMPK3	PtMPK3-1 PtMPK3-2	OsMPK3	SbMPK3	ScMPK3
	AtMPK6	AIMPK6	PtMPK6-1 PtMPK6-2	OsMPK6	SbMPK6	ScMPK6
	AtMPK10	AIMPK10	*	*	*	*
B	AtMPK4	AIMPK4	PtMPK4	OsMPK4	SbMPK4	ScMPK4
	AtMPK5	AIMPK5	PtMPK5-1 PtMPK5-2	*		*
	AtMPK11	AIMPK11	PtMPK11	*	*	*
	AtMPK12	AIMPK12	*	OsMPK12	SbMPK12	*
	AtMPK13	AIMPK13	*	*	*	*
C	AtMPK1	AIMPK1	PtMPK1	*	*	*
	AtMPK2	AIMPK2	PtMPK2	*	*	*
	AtMPK7	AIMPK7	PtMPK7	OsMPK7	SbMPK7	ScMPK7
	AtMPK14	AIMPK14	PtMPK14	OsMPK14	SbMPK14	ScMPK14
D	AtMPK8	AIMPK8	*	*	*	*
	AtMPK9	AIMPK9	PtMPK9-1 PtMPK9-2	*	*	*
	AtMPK15	AIMPK15	*	*	*	*
	AtMPK16	AIMPK16	PtMPK16-1 PtMPK16-2	OsMPK16	SbMPK16-1 SbMPK16-2	ScMPK16-1 ScMPK16-2
	AtMPK17	AIMPK17	PtMPK17	OsMPK17-1 OsMPK17-2	SbMPK17-1 SbMPK17-2	ScMPK17-1 ScMPK17-2
	AtMPK18	AIMPK18	PtMPK18	*	*	*
	AtMPK19	AIMPK19	PtMPK19	*	*	*
	AtMPK20	AtIMP20	PtMPK20-1 PtMPK20-2	OsMPK20-1 OsMPK20-2 OsMPK20-3 OsMPK20-4 OsMPK20-5	SbMPK20-1 SbMPK20-2 * SbMPK20-4 SbMPK20-5	ScMPK20-1 ScMPK20-2 * ScMPK20-4 ScMPK20-5
	*	*	*	OsMPK21-1 OsMPK21-2	SbMPK21-1 SbMPK21-2 SbMPK21-3	*

Figura 28. Visão conjunta dos ortólogos de MPKs *A. thaliana* identificados *A. lyrata*, álamo, arroz, sorgo e cana-de-açúcar.

grupo	<i>A. Thaliana</i>	<i>A. Lyrata</i>	<i>Populus Trichocarpa</i>	<i>Oryza sativa</i>	<i>Sorghum bicolor</i>	<i>Sugarcane</i>
A	AtMKK1	AIMKK1	*	OsMKK1	SbMKK1	ScMKK1
	AtMKK2	AIMKK2	PtMPK2-1 PtMPK2-2	*	*	*
	AtMKK6	AIMKK6	PtMPK6	OsMKK6	SbMKK6	ScMKK6
	AtMKK3	AIMKK3	PtMPK3	OsMKK3	SbMKK3	ScMKK3
	AtMKK4	AIMKK4	PtMPK4	OsMKK4	SbMKK4	ScMKK4
B	AtMKK5	AIMKK5	PtMPK5	OsMKK5	SbMKK5-1 SbMKK5-2	*
C	AtMKK7	AIMKK7	PtMPK7	*	*	*
	AtMKK8	AIMKK8	*	*	*	*
	AtMKK9	AIMKK9	PtMPK9	*	*	*
	AtMKK10	AIMKK10	*	OsMKK10-1	SbMKK10-1	*
D	*	*	*	OsMKK10-2	SbMKK10-2	ScMKK10-2
	*	*	*	OsMKK10-3	SbMKK10-3	*
	*	*	*	*	SbMKK10-4	*
	*	*	*	*	SbMKK10-5	*
	*	*	*	*	SbMKK10-6	*
	*	*	*	*	SbMKK10-7	*
	*	*	PtMPK11-1	*	*	*
	*	*	PtMPK11-2	*	*	*

Figura 29. Visão conjunta dos ortólogos de MKKs *A. thaliana* identificados *A. lyrata*, álamo, arroz, sorgo e cana-de-açúcar

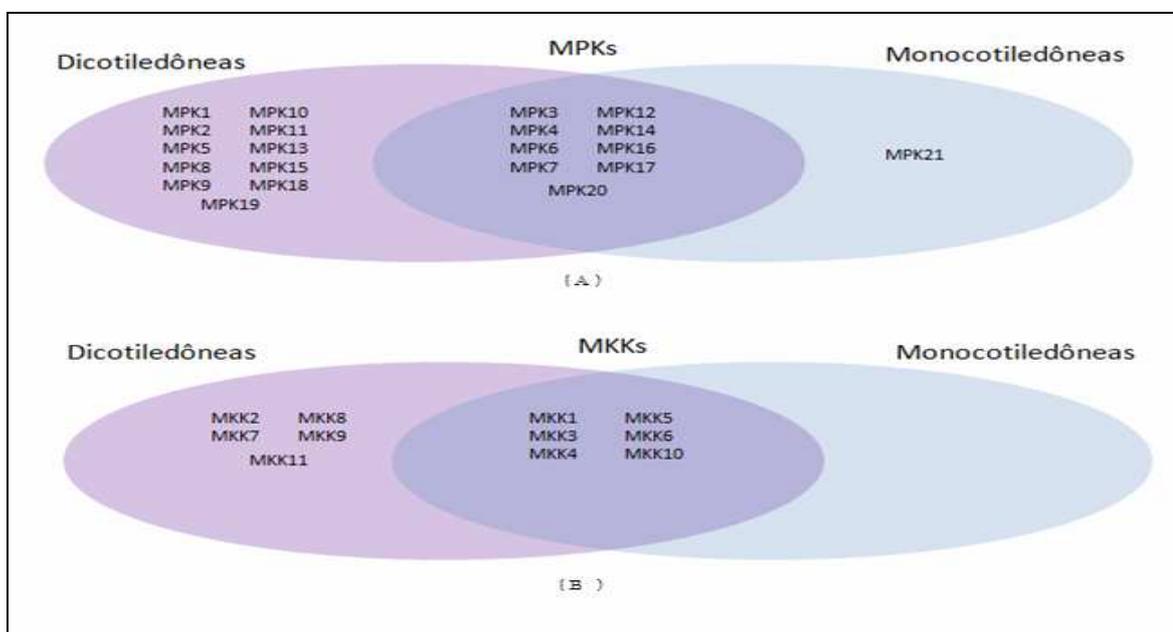


Figura 30. Componentes MPKs/MKKs predominantes em plantas superiores. (a) componentes MPKs encontrados em mono e dicotiledôneas, (b) componentes MKKs encontrados em mono e dicotiledôneas

4.4. *A. thaliana* e *A. lyrata* não apresentam significativas divergências evolutivas para módulos de genes MAP kinases

A análise filogenética realizada com os ortólogos de MPKs/MKKs de *A. thaliana* situou os membros sorgo, cana-de-açúcar e *A. lyrata* em 4 clados, similarmente ao que foi obtido com arroz, *A. thaliana* e álamo (Hamel, 2006). Embora o genoma de *A. lyrata* seja maior que o genoma de *A. thaliana*, a análise filogenética e estrutural de genes MPK/MKK destas espécies permitiram observar que não existem divergências evolutivas significativas entre módulos MPK/MKKs, e que as sequências ortólogas destas duas espécies compartilham cerca de 99% de identidade.

Os genes MPK10 e MPK13, presentes em ambas as espécies, não possuem ortólogos nas demais espécies analisadas e ambas apresentam uma relativa distância das demais MPKs no clado onde estão localizadas.

5. DISCUSSÃO

5.1. Identificação de genes MPKs de sorgo, cana e *A. lyrata*

A família de genes MAP Kinases em plantas reflete um modelo antigo de diversificação, que já havia se estabilizado antes das divergências evolucionárias de monocotiledôneas e dicotiledôneas. Para todas as três espécies, ortólogos/parálogos MPKs/MKKs podem ser identificados em algum grupo de MAP Kinase, tais como os definidos inicialmente para *A. thaliana* (Hamel, 2006).

Pesquisas *in silico* em bancos de dados genômicos de sorgo e *A. lyrata* e no banco de transcrito da cana-de-açúcar (SUCEST, 2010) identificaram, respectivamente, 16, 20 e 13 sequências MPKs. Tais observações reforçam as evidências de que a família de genes MPK seja substancialmente menor em plantas monocotiledôneas se comparadas as plantas dicotiledôneas, conforme proposto por Rodriguez et al. (2010). Uma vez definido o alto grau de conservação estrutural das sequências dos genes MPK, a nomenclatura adotada inicialmente para *A. thaliana* (MAP Group, 2002) e, posteriormente, para arroz e álamo (Hamel, 2006), parece ser adequada para ser utilizada para sorgo, cana-de-açúcar e *A. lyrata*. Cabe destacar, entretanto, que tal modelo proposto não é baseado em evidências quanto à conservação de papéis biológicos.

Com a disponibilidade das MPKs de sorgo e dada à proximidade genético-evolutiva desta espécie com a cana-de-açúcar, espera-se que ambas suportem o mesmo modelo de duplicação gênica e que, portanto, sejam identificados números aproximados de genes ortólogos/parálogos pertencentes

aos mesmos grupos dentro dos clados. As MPKs identificadas no banco de dados genômicos de sorgo ajudaram a nortear as etapas do processo de identificação dos transcritos de cana-de-açúcar correspondentes as sequências (MPK3, MPK7, MPK14, MPK17-1 e MPK20-2), uma vez que foram encontrados mais de um transcrito (ou parte dele) que correspondesse a um mesmo módulo MPK em cana-de-açúcar. A abordagem comparativa de genomas tem sido uma estratégia utilizada em pesquisas que têm se dedicado a um melhor entendimento da fisiologia, metabolismo e genética dessas duas espécies de plantas (Albino et al., 2006; KUITTINEN, H. et al., 2004), uma vez que tem sido demonstrado que o genoma de uma espécie pode ser estudado por meio do entendimento do genoma de espécies correlatas. Dessa forma, a inclusão de dados provenientes do genoma de sorgo constituiu uma estratégia para um melhor entendimento da organização estrutural dos componentes MAP Kinases que foram identificados em cana-de-açúcar a partir do banco de transcrito da cana-de-açúcar (SUCEST, 2005).

5.2. Eventos de paralogia evidenciam expansão MPK no clado “D” e análise filogenética revelam duplicações de genes MPKs que ocorrem após a divergência entre monocotiledôneas e dicotiledôneas

A inclusão de três novas espécies ao conjunto de dados, sendo duas monocotiledôneas, também permitiu uma melhor discussão a respeito da expansão observada no clado D, tanto para módulos de genes MPKs quanto de genes MKKs. A expansão de MPKs do clado D (genes MPK21), observada primeiramente em arroz, mas não em *A. thaliana* e álamo (Hamel, 2006), não permitia responder se seria consequência de uma linhagem-específica ou se representaria eventos de duplicação ocorridos após a divergência de monocotiledôneas e dicotiledôneas. A recente conclusão do sequenciamento do genoma de sorgo tornou possível responder a esta pergunta. Os dados aqui apresentados sugerem que os genes, membros do grupo D de monocotiledôneas, continuam se expandindo, dando origem ao subgrupo de ortólogos do gene MPK21, observado primeiramente em arroz, mas não em álamo e *arabidopsis* (Hamel, 2006) e, identificado por este trabalho, em sorgo. Entretanto, ainda se desconhece as vantagens conferidas às monocotiledôneas em função da notável

expansão observada no clado D. Em consultas realizadas em banco de dados de expressão de arroz (BAR, 2010), pode ser verificado que os genes OsMPK21-1 e OsMPK21-2 são induzidos em vários tecidos e em várias fases do desenvolvimento da planta, notavelmente na fase final da semente. Em cana-de-açúcar, transcritos para o gene MPK21 não foram identificados. Ainda em cana-de-açúcar, foi observada uma variação na assinatura MPK (-TxY) no gene ScMPK20-4 (-RIM-), do grupo das MPKs20, pertencente ao clado D, onde também foi verificada expansão numérica. Pesquisas aos dados de expressão do SUCEST (<http://www.sucest-fun.org/en/database-tools/cane-enome/sucest/cluster-read-search>) sugerem que este gene é expresso em vários tecidos/condições avaliados. Entretanto, não há evidências de que este gene desempenhe atividade MPK e são necessários experimentos *in vitro* para confirmar a atividade MPK para esta proteína.

5.3. Ortólogos do ATMPK12 apresentam variação na assinatura KINASE – DxK-

A análise do grupo de genes MPK12, que pertencem ao clado B das espécies analisadas, mais a inclusão da MPK12 de milho (ZmMPK12) e da MPK12 arroz (OsMPK12), permitiu observar que, curiosamente, a região da assinatura kinase apresenta a variação -DLR- no domínio (PK), ao invés da assinatura -D[L/V/I]K descrita até o momento na literatura (MAPK Group, 2002; Hamel et al., 2006). Ainda não se pode afirmar que a variação observada no domínio PK para este grupo de genes confere à proteína qualquer variação no âmbito da atividade biológico. Embora ainda sejam necessários estudos para confirmação do papel biológico das MPK12 em sorgo, arroz e milho, as análises com ferramentas de bioinformática permitiram identificar esta variante no domínio kinase, aparentemente, típica deste grupo de genes. Pesquisas realizadas com a MPK12 de *A. thaliana* revelam que esta proteína é funcional e interage diretamente com a proteína MPK fosfatase INDOLE-3-BUTYRIC RESPONSE 5 (IBR5). MPK12 em *A. thaliana* (AtMPK12) atua em conjunto com a MPK IBR5, em um cenário muito específico, como reguladores negativos de sinalização de auxinas (Lee et al., 2009). Consultas realizadas nos bancos de expressão de *A. thaliana* e arroz (BAR – 2010) revelam que, tanto AtMPK12 quanto OsMPK12 são

transcritos em inflorescência de ambas as espécies. Em arroz, tais transcritos são mais fortemente expressos na fase S3 e S4 do desenvolvimento de sementes.

5.4. Identificação de genes MKKs de sorgo, cana e *A. lyrata*

Pesquisas *in silico* em bancos de dados genômicos de sorgo, *A. lyrata* e no banco de transcrito da cana-de-açúcar (SUCEST, 2010) identificaram, respectivamente, 13, 10 e 5 sequências MKKs. Em todas as sequências MKKs identificadas foram encontradas as assinaturas pertinentes a família Map kinase kinase, com exceção apenas para as SbMKK5-1 e SbMKK10-5, que apresentaram domínio kinase –GIK- e – DLR-, respectivamente, além das sequências SbMKK1 e ScMKK1, que revelaram uma construção inadequada do domínio MKK, com a ausência de um aminoácido no “loop” de ativação da proteína. Com relação às proteínas do grupo MKK10, assim como observado em outras espécies, a análise filogenética revela intensa atividade de duplicação gênica, embora não haja evidências de que estas proteínas desempenham papel de MKK.

5.5. Identificação de componentes MPKs/MKKs predominantes em plantas dicotiledôneas

Cerca de metade do número de componentes MPKs, observados para o grupo das dicotiledôneas, não foram observados nas espécies de monocotiledôneas aqui analisadas. No âmbito de genes MPKs, foram observados apenas em dicotiledôneas os módulos AtMPK1, AtMPK2, AtMPK5, AtMPK10, AtMPK11, AtMPK8, AtMPK9, AtMPK13, AtMPK15, AtMPK18 e AtMPK19. Apesar de ainda pouco descritos na literatura, nossas análises permitiram observar que tanto MPK10 de *A. thaliana* quanto MPK10 de *A. lyrata* se destacam na árvore filogenética, por apresentarem uma considerável distância em relação aos demais membros MPKs pertencente ao mesmo clado (Clado A). Pesquisas anteriores em *A. thaliana* comprovaram que o gene AtMPK10 é expresso em baixos níveis numa gama de sistemas de ensaios, sugerindo, inclusive, que há uma tendência do gene MPK10 vir a se tornar não funcional em *A. thaliana* (Hamel et al., 2006). Entretanto, não foram obtidas informações nos bancos de expressão a respeito

do papel desempenhado por AtMPK13 em *A. thaliana*, membro do caldo B, sendo também observada uma relativa distância deste componente, na árvore filogenética, em relação aos demais módulos que compõe o mesmo clado.

Com relação aos módulos MKKs que permaneceram monofiléticos (MKK2, MKK7, MKK8, MKK9 e MKK11) e, portanto, apenas observados em dicotiledôneas, sabe-se que em *A. thaliana* o componente AtMKK2 tem sido envolvido em resposta ao frio e a elevada salinidade (Teige et al., 2004). Ainda em *A. thaliana*, o gene AtMKK7 está envolvido no transporte polar de auxinas e resistência basal e sistêmica (Zhang et al., 2007) e AtMKK9 na resposta à etileno (Zhou et al., 2009). Tanto AtMKK7 quanto AtMKK9 pertencem ao clado D. Curiosamente, o gene PtMKK11 foi encontrado apenas em álamo. Não é possível afirmar qual o papel biológico de MKK11 em plantas superiores.

5.6. Análise filogenética revela intensa atividade de expansão de genes pertencentes ao grupo MKK10

Uma análise mais atenta revela que sequências MKK10 de todas as cinco espécies não apresentam uma construção convencional no sítio de ativação da proteína, a assinatura – S/TxxxxxS/T-. Além disso, SbMKK10-5 de sorgo apresenta também a assinatura –DLR-, observada para genes pertencentes ao grupo MPK12. Entretanto, nossas análises verificam que isto constitui uma exceção dentro deste clado, diferente do que foi observado para o grupo das MPK12. Em sorgo, assim como foi verificado em arroz, evidências apontam para registros de intensa atividade de expansão e duplicação gênica (paralogia) para genes MKK10 em monocotiledôneas, mas não em dicotiledôneas. A análise conjunta dos genes MKKs identificados em sorgo, cana-de-açúcar e *A. lyrata*, em comparação com *A. thaliana*, arroz e álamo, revela um panorama onde é possível discutir acerca do papel funcional dos genes pertencentes ao grupo MKK10, localizado no clado D de MKKs. Genes MKK10 em plantas podem representar versões de novas funcionalidades da estrutura canônica MKK, atuando talvez como proteínas acessórias não-catalíticas (Jonak et al., 2002; Whitmarsh et al., 1998). Embora a questão da funcionalidade MKK para este grupo de genes poderia ser amplamente discutível, a análise da expressão (BAR, 2010) revela que, em arroz, MKK10-1 e MKK10-2 são expressos em vários tecidos, sendo

mais induzidos em folha. Em *A. thaliana*, AtMKK10 é expressa na fase final do desenvolvimento da semente e na fase de pólen maduro.

5.7. MKK3 de sorgo, cana e *A. lyrata* possuem domínio NTF conservado

Um dos mais interessantes clados MKK consiste das sequências de ortólogos do componente AtMKK3, que formam o único grupo de genes do clado B de MKK. Tal grupo de proteínas é caracterizado por possuir uma característica incomum aos componentes dos outros clados MKKs, dado o fato de conservarem uma região de extensão 3' codificando um domínio NTF ('Fator de Transferência Nuclear'). Pesquisas têm associado a presença do domínio NTF ao transporte de proteínas núcleo-citoplasma (Steggerda et al., 2002). Em *A. thaliana*, o papel biológico dos genes MKK3 é associado à ativação de genes do clado C na cascata eliciada por patógenos e são dependentes da sinalização de ácido jasmônico (JA) (Rodriguez et al., 2010; Doczi et al., 2007; Takahashi et al., 2007). Pesquisas ainda sugerem que estes papéis são atribuídos provavelmente a habilidade de AtMKK3 de interagir com uma série de MPKs diferentes (AtMPK1, AtMPK2, AtMPK7 e AtMPK14) (Erik et al., 2009). Curiosamente, o genoma *Chlamydomonas* codifica um único gene MKK e este pertence à classe estrutural das MKK3, incluindo o domínio 3'-NTF. Tal observação indica que este arranjo quimérico teve uma longa e bem sucedida história evolutiva na linhagem de eucariotos fotossintéticos (Hamel et al., 2006). Dados derivados de análise em bancos de expressão (BAR) revelam que o transcrito AtMKK3 é significativamente induzido em sementes secas de *A. thaliana* e, em arroz, OsMKK3 é induzido na fase de germinação.

Apesar de o genoma de *A. thaliana* ($n=5$) ser bem menor que o genoma de *A. lyrata* ($n=8$), onde n representa o número de cromossomos da espécie, a análise filogenética e estrutural conjunta das sequências ortólogas dos componentes MPK/MKK, revelou que as duas espécies de *arabidopsis* apresentam o mesmo modelo evolutivo. Nossas análises sugerem que os módulos MPK/MKKS de *A. lyrata* e *A. thaliana* possuem padrão genético-estrutural semelhante para grande parte dos genes MPK/MKK identificados, sugerindo que ortólogos de módulos MKKs/MPKs desempenham papéis biológicos semelhantes.

6. RESUMO E CONCLUSÃO

A análise filogenética feita com as sequências de genes MPKs e MKKs de sorgo, *A. lyrata* e cana-de-açúcar, identificadas por este trabalho, conjuntamente com as sequências de genes MPK e MKKs de arroz, álamo e *A. thaliana*, derivados das pesquisas feitas por Hamel (2006), permitiram desenhar um cenário onde foi possível investigar os padrões e divergências evolutivas de genes MAP Kinases de plantas superiores. A metodologia utilizada foi fundamental dentro do escopo dos resultados obtidos, permitindo uma análise concisa dos dados com as seguintes conclusões:

- Foram identificados 16, 13 e 20 componentes MPKs em sorgo, cana e *A. lyrata*, bem como foram identificados 13, 5 e 10 componentes MKKs em sorgo, cana e *A. lyrata*, respectivamente. Nossas análises permitiram confirmar que em monocotiledôneas o número de genes MPKs e MKKs constituem uma família de genes menos numerosa se comparadas com o número de componentes identificados para dicotiledôneas. Esta característica já havia sido sugerida em trabalhos anteriores (Hamel, 2006);
- A atividade de expansão observada no clado D para componentes MPK permitiu concluir que o grupo MPK21 foi originado após a divergência entre dicotiledôneas e monocotiledôneas.
- Nossas análises concluíram que a variação no segundo aminoácido do motivo (DxK), o domínio –DLR-, observada em componentes do grupo B de MPKs, parecem típicas de membros do grupo das MPK12. O domínio –DLR- foi pela primeira vez identificado por este trabalho.

- Nossas análises levam a crer que alguns componentes do clado D de dicotiledôneas permaneceram monofiléticos e não foram encontrados em monocotiledôneas (MKK7, MKK8, MKK9 e MKK11). Dessa forma, foi possível concluir que membros MKK10 formam os únicos constituintes deste clado em monocotiledôneas.
- A análise conjunta dos genes de MKKs de todas as espécies também permitiu concluir que componentes MKK3 são únicos constituintes do clado B de MKKs tanto de monocotiledôneas quanto de dicotiledôneas. O que permite entender que o domínio NTF parece ser exclusivo destes componentes da família kinase e continuaram sendo preservados durante o processo evolutivo mesmo após a divisão de mono e dicotiledôneas. Isto sugere que estes genes desempenham um papel relevante dentro do contexto evolucionário de plantas superiores.
- A nomenclatura proposta para a anotação de componentes MPK/MKK se mostrou adequada, embora não esteja baseada em similaridades de papéis biológicos.
- *Nossas análises também permitiram verificar que A. lyrata e A. thaliana não possuem divergências relevantes para os módulos MPK/MKKs.*

REFERÊNCIAS BIBLIOGRÁFICAS

- Ahlfors, R., Violetta Macioszek, Jason Rudd, Mikael Brosché, Rita Schlichting, Dierk Scheel, Jaakko Kangasjärvi (2004) Stress hormone-independent activation and nuclear translocation of mitogen-activated protein kinases in *Arabidopsis thaliana* during ozone exposure. *Plant J*, 40:512–522
- Altschul S. F., Madden T. L., Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, 25:3389–3402.
- Andreasson E., Jenkins T, Brodersen P, Thorgrimsen S, Petersen NH, (2005). The MAP kinase substrate MKS1 is a regulator of plant defense responses. *EMBO J*, 24:2579–89.
- BAR – www.bar.utoronto.ca , acessado em junho de 2010.
- Bardwell AJ., Frankson E, Bardwell L. (2009). Selectivity of docking sites in MAPK kinases. *J. Biol. Chem*, 284:13165–73.
- Bogre, L. et al., 1999 A MAP kinase is activated late in plant mitosis and becomes localized to the plane of cell division. *Plant Cell*, 11, 101–113.
- Brader, G. (2007) The MAP kinase kinase MKK2 affects disease resistance in *Arabidopsis*. *Mol. Plant Microbe Interact*, 20:589–596.

- Brodersen P., Petersen M, Bjorn Nielsen H, Zhu S, Newman MA. (2006) Arabidopsis MAP kinase 4 regulates salicylic acid- and jasmonic acid/ethylene-dependent responses via EDS1 and PAD4. *Plant J*, 47:532–46.
- Bush, S.M. and Krysan, P.J., (2007) Mutational evidence that the Arabidopsis MAP kinase MPK6 is involved in anther, inflorescence and embryo development. *J. Exp. Bot*, 58, 2181–2191.
- Cardinale F., Meskiene I, Ouaked F, Hirt H. (2002) Convergence and divergence of stress-induced mitogen-activated protein kinase signaling pathways at the level of two distinct mitogen-activated protein kinase kinases. *Plant Cell*, 14:703–11.
- Cho, S.K. , Clayton T. Larue, David Chevalier, Huachun Wang†, Tsung-Luo Jinn, Shuqun Zhang, and John C. Walker (2008) Regulation of floral organ abscission in Arabidopsis thaliana. *Proc. Natl. Acad. Sci. U. S. A* 105, 15629–15634
- Colest , Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., (1998). Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature*, 393:537-44.
- Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *GenomeRes*. 14:1188–90
- Marenda D.R., A.D. Vrailas, A.B. Rodrigues, S. Cook, M.A. Powers, J.A. Lorenzen, L.A. Perkins, K. Moses (2006) MAP kinase subcellular localization controls both pattern and proliferation in the developing Drosophila wing, *Development* 133: 43-51.
- Doczi R, Brader G, Pettko-Szandtner A, Rajh I, Djamei A, et al. 2007. The Arabidopsis mitogen-activated protein kinase kinase MKK3 is upstream of

group C mitogen-activated protein kinases and participates in pathogen signaling. *Plant Cell*,19:3266–79

Droillard MJ, Boudsocq M, Barbier-Brygoo H, Lauriere C. (2004). Involvement of MPK4 in osmotic stress response pathways in cell suspensions and plantlets of *A. thaliana thaliana*: activation by hypoosmolarity and negative role in hyperosmolarity tolerance. *FEBS Lett*, 574:42–48

Erik Andreasson, Brian Ellis (2009) Convergence and specificity in the Arabidopsis MAPK nexus - *Trends in Plant Science* Vol.15 No.2

Maria D. Ermolaeva, Martin Wu, Jonathan A. Eisen and Steven L. Salzberg (2003) *Plant Molecular Biology*, Volume 51, Number 6, Pages 859-866

Ewing, B.; GREEN, P. (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research*, Woodbury, 8:186-194.

FAPESP - (<http://www.agencia.fapesp.br/materia/9227/especiais/a-vez-da-cana-de-açúcar.htm>). Acessado em jan 2010

Feilner, T., Claus Hultschig, Justin Lee, Svenja Meyer, Richard G. H. Immink, Andrea Koenig, Alexandra Possling, Harald Seitz, Allan Beveridge, Dierk Scheel, Dolores J. Cahill, Hans Lehrach, Jürgen Kreuzberger, and Birgit Kersten (2005) High through-put identification of potential Arabidopsis MAP kinase substrates. *Mol. Cell. Proteomics* 4, 1558–1568

Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, 269:496-512.

Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD. (1995) The minimal gene complement of *Mycoplasma genitalium*. *Science*, 270:397-403.

- Frye, Catherine A. Dingzhong Tang, and Roger W. Innes. (2001) Negative regulation of defense responses in plants by a conserved MAPKK kinase. *Proc. Natl. Acad. Sci. U. S. A.* 98, 373–378
- Gao, L. and Xiang, C-B. , (2008) The genetic locus At1g73660 encodes a putative MAPKKK and negatively regulates salt tolerance in Arabidopsis. *Plant Mol. Biol*, 67:125–134
- Garrington, T. and Johnson, G. (1999) Organization and regulation of mitogen-activated protein kinase signaling pathways. *Curr. Opin. Cell Biol*, 11;211–218
- Genevestigator – (<https://www.genevestigator.com/gv/index>.), acessado em janeiro de 2010
- Globalbioenergy(http://www.globalbioenergy.org/uploads/media/0811Wageningen_Sugarcane_ethanolcontributions_to_climate_change_mitigation_and_the_environment.pdf), acessado em junho de 2009
- Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook, J, Sessions A, Oeller P, Varma H. (2002) A draft sequence of rice genome (*Oryza sativa* L. Ssp. Japonica). *Science*, 296:92-100.
- Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H (1996) Life with 6000 genes. *Science*, 274:546.
- Gordon, D.; ABAJIAN, C., GREEN, P. (1998) Consed: a geographical tool for sequence finishing. *Genome Research*, Woodbury, 8:195-202.
- Gribskov M, Fana F, Harper J, Hope DA, Harmon AC, Smith DW, Tax FE, Zhang G (2001) PlantsP: a functional genomics database for plant phosphorylation. *Nucleic Acids Res*, 29:111–113

- GU, Xun, Zhongqi Zhang, and Wei Huang (2005) Rapid evolution of expression and regulatory divergences after yeast gene duplication. *Proc. Natl. Acad. Sci. U. S. A.* 102, 707–712
- Guinon S, Lethiec F, Duroux P, Gascuel O. (2005) A web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res.* 2005 Jul 1;33(Web Serve):W557-9. PMID: 15980534
- Hagemann C, Rapp UR. (1999) Isotype-specific functions of Raf kinases. *Exp Cell Res*, 253 (1):34-46.
- Hamel, Louis Philippe, Marie-Claude N., Somrudee S., Marie-Josée Morency, Margaret Ellis, Juergen E., Nathalie B., Brad B., Dan K., Justin Lee, Greg Martin, John M., Yuko Ohashi, Dierk Scheel, Jen Sheen, Tim Xing, Shuqun Zhang, Armand Seguin, Brian E. Ellis (2006) Ancient signals: comparative genomics of plant MAPK and MAPKK gene families - *TRENDS in Plant Science*. Vol.11 No.4 April 2006
- Ichimura K, Mizoguchi T, Irie K, Morris P, Giraudat J (1998) Isolation of ATMEKK1 (a MAP kinase kinase kinase)-interacting proteins and analysis of a MAP kinase cascade in Arabidopsis. *Biochem. Biophys. Res. Commun*,43: 253:532
- MAPK Group (Kazuya Ichimura et al.) 2002 Mitogen-activated protein kinase cascades in plants: a new nomenclature. *Trends Plant Sci.* 7, 301–308
- JGI - *Joint Genome Institute* (JGI) - (<http://genome.jgi-psf.org/cgi-bin/runAlignment?db=Sorbi1&advanced=1>), acessado em janeiro de 2010.
- Jonak, C. et al. (2002) Complexity, cross talk and integration of plant MAP kinase signaling. *Curr. Opin. Plant Biol.* 5, 415–424
- Jouannic, S., A. Hamal, A. S. Leprince, J. W. Tregear, M. Kreis and Y. Henry (1999) Plant MAP kinase kinase kinases structure, classification and evolution. *Gene* 233, 1–11

- Jouannic, S. *et al.* (2001) The protein kinases AtMAP3K□1 and BnMAP3K□1 are functional homologues of *S. pombe* Cdc7p and may be involved in cell division. *Plant J.* 26, 637–649
- Kiegerl S, Cardinale F, Siligan C, Gross A, Baudouin E (2000) SIMKK, a mitogen-activated protein kinase (MAPK) kinase, is a specific activator of the salt stress-induced MAPK, SIMK. *Plant Cell*, 12:2247–58
- Kresovich S, Barbazuk B, Bedel A, Borrel A, Buell CR, Burke J. (2005). Toward sequencing the sorghum genome. A U.S. National Science Foundation Sponsored Workshop Report. *Plant Physiology* 138:1898-1902.
- Kuittinen, H, Anita A. de Haan, Claus Vogl, Sami Oikarinen, Johanna Leppälä, Marcus Koch, Tom Mitchell-Olds, Charles H. Langley[§] and Outi Savolainen. Comparing the Linkage Maps of the Close Relatives *Arabidopsis lyrata* and *A. thaliana* (2004). *Genetics*, Vol. 168, 1575-1584, November 2004, Copyright © 2004 doi:10.1534/genetics.103.022343
- Lampard, G. , Cora A. MacAlister, Dominique C. Bergmann (2008) *Arabidopsis* stomatal initiation is controlled by MAPK-mediated regulation of the bHLH SPEECHLESS. *Science* 322, 1113–1116
- Lamparda, Gregory R., Wolfgang Lukowitzb, Brian E. Ellisc and Dominique C. Bergmann (2009) Novel and expanded roles for MAPK signaling in *Arabidopsis* stomatal cell fate revealed by cell-type specific manipulations. *Plant Cell*, (in press).
- Lee JS, Wang S, Sritubtim S, Chen JG, Ellis BE. (2009). *Arabidopsis* mitogen-activated protein kinase MPK12 interacts with the MAPK phosphatase IBR5 and regulates auxin signaling. *Plant J.* 57:975–85
- Lee, J-S. Kyung Won Huh, Apurva Bhargava, and Brian E Ellis (2008) Comprehensive analysis of protein–protein interactions between *Arabidopsis*

MAPKs and MAPK kinases helps define potential MAPK signalling modules. *Plant Signal, Behav.* 3, 1037–1041

Levchenko, A, Jehoshua Bruck, and Paul W. Sternberg (2000) Scaffold proteins may biphasically affect the levels of mitogen-activated protein kinase signaling and reduce its threshold properties. *Proc. Natl. Acad. Sci, U. S. A.* 97:5818–5823

Lewis, T.S., Paul S. Shapiroa and Natalie G. Ahna (1998) Signal transduction through MAP kinase cascades. *Adv. Cancer Res.* 74, 49–139.

Liu, Yidong, Hailing Jin, Kwang-Yeol Yang, Cha Young Kim, Barbara Baker, Shuqun Zhang. (2003) Interaction between two mitogen-activated protein kinases during tobacco defense signaling. *Plant J*,34, 149–160

Madhani, H.D. and Fink, G.R. (1998) The riddle of MAP kinase signaling specificity. *Trends Genet*, 14:151–155

Maere, Steven, Stefanie De Bodt, Jeroen Raes, Tineke Casneuf, Marc Van Montagu, Martin Kuiper, and Yves Van de Peer (2005) Modeling gene and genome duplications in eukaryotes. *Proc. Natl. Acad. Sci. U. S. A.* 102, 5454–5459

Maffei, Massimo E., Axel Mithöfer and Wilhelm Boland (2007) Massimo E., Axel Mithofer and Wilhelm Boland (2007). Before gene expression: early events in plant–insect interaction. *TRENDS in Plant Science* Vol.12 No.7

Mapes, J., Ota, I. (2004) Nbp2 targets the Ptc1-type 2C Ser/Thr phosphatase to the HOG MAPK pathway. *EMBO J*, 23:302–311

MAPK Group. (2002) Mitogen-activated protein kinase cascades in plants: a new nomenclature. *Trends Plant Sci*, 7:301–308

- Melikant, B., C. Giuliani, S. Halbmayer-Watzina, A. Limmongkon, E. Heberle-Bors, C. Wilson (2004) The Arabidopsis thaliana MEK AtMKK6 activates the MAP kinase AtMPK13. *FEBS Lett.* 576, 5–8
- Menges, M., Róbert Dóczy, László Ökrész, Piero Morandini, Luca Mizzi, Mikhail Soloviev, James A. H. Murray, László Bögre (2008) Comprehensive gene expression atlas for the Arabidopsis MAP kinase signaling pathways. *New Phytol.* 179, 643–662
- Meszáros T, Helfer A, Hatzimasoura E, Magyar Z, Serazetdinova L, (2006) The Arabidopsis MAP kinase kinase MKK1 participates in defense responses to the bacterial elicitor flagellin. *Plant*, 48:485–98
- Miao, Ying, Thomas M. Laun, Anja Smykowski and Ulrike Zentgraf (2007) Arabidopsis MEKK1 can take a short cut: it can directly interact with senescence-related WRKY53 transcription factor on the protein level and can bind to its promoter. *Plant Mol. Biol.* 65, 63–76
- Miles GP, Samuel MA, Zhang Y, Ellis BE. (2005). RNA interference-based (RNAi) suppression of AtMPK6, an Arabidopsis mitogen-activated protein kinase, results in hypersensitivity to ozone and mis-regulation of AtMPK3. *Environ. Pollut*, 138:230–37
- Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP, Flanigan, M.J. (2000). whole-genome assembly of Drosophila. *Science*, 287:2196-204.
- Nakagami, Hirofumi, Andrea Pitzschke and Heribert Hirt (2005) Emerging MAP kinase pathways in plant stress signaling. *Trends Plant Sci.* 10, 339–346
- Nishihama, R., Soyano T, Ishikawa M, Araki S, Tanaka H, Asada T, Irie K, Ito M, Terada M, Banno H, Yamazaki Y, Machida Y.. (2002) Expansion of the cell plate in plant cytokinesis requires a kinesin-like protein/MAPKKK complex. *Cell* 109, 87–99

- Parkhill J, Wren BW, Thomson NR, Titball RW, Holden MT, Prentice MB, (2001) Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature*, 413:523-27.
- Petersen M, Brodersen P, Naested H, Andreasson E, Lindhart U. (2000). Arabidopsis map kinase 4 negatively regulates systemic acquired resistance. *Cell* 103:1111–20
- Peyssonnaud, C, Eychene A. (2001) The Raf/MEK/ERK pathway: new concepts of activation. *Biol Cell*, 93 (1-2):53-62.
- Popescu, S. et al. (2009) MAPK target networks in *Arabidopsis thaliana* revealed using functional protein microarrays. *Genes Dev*, 23, 80–92
- Qiu JL, Fiil BK, Petersen K, Nielsen HB, Botanga CJ, (2008). Arabidopsis MAP kinase 4 regulates gene expression through transcription factor release in the nucleus. *EMBO J*. 27:2214–21
- Qiu JL, Zhou L, Yun BW, Nielsen HB, Fiil BK, (2008). Arabidopsis mitogen-activated protein kinase kinases MKK1 and MKK2 have overlapping functions in defense signaling mediated by MEKK1, MPK4, and MKS1. *Plant Physiol*. 148:212–22
- Rodriguez, Maria Cristina Suarez. Morten Petersen, and John Mundy, (2010). Mitogen- Activated Protein Kinase Signaling in Plants Annual Review of Plant Biology Vol. 61: 621-649
- Romeis, T. (2001) Protein kinases in the plant defence response. *Curr. Opin. Plant Biol*. 4:407–414
- Samuel, Marcus A, Balbir K Chaal, Greg Lampard, Beverley R Green, and Brian E. Ellis (2008) Surviving the passage: Non-canonical stromal targeting of an Arabidopsis mitogen-activated protein kinase kinase. *Plant Signal. Behav*. 3, 1–7

- Schaeffer, H.J., Weber, M.J. (1999) Mitogen-activated protein kinases: specific messages from ubiquitous messengers. *Mol. Cell. Biol.* 19, 2435–2444
- Schaffer, R. , Jeff Landgraf, Monica Accerbi, Vernadette Simon, Matt Larson, and Ellen Wisman (2001) Microarray analysis of diurnal and circadian-regulated genes in *Arabidopsis*. *Plant Cell*, 13:113–123
- Seo S, Katou S, Seto H, Gomi K, Ohashi Y. (2007). The mitogen-activated protein kinases WIPK and SIPK regulate the levels of jasmonic and salicylic acids in wounded tobacco plants. *Plant J.* 49:899–909
- Sokal R.R, Michener C.D. (1958) A statistical method for evaluating systematic relationships. *Univ Kansas Sci Bull* 28:1409–1438
- Soyano, Takashi, Ryuichi Nishihama, Kayoko Morikiyo, Masaki Ishikawa, and Yasunori Machida (2003) NQK1/NtMEK1 is a MAPKK that acts in the NPK1 MAPKKK-mediated MAPK cascade and is required for plant cytokinesis. *Genes Dev.* 17, 1055–1067
- Steggerda S.M., Paschal B.M., (2002). Regulation of nuclear import and export by the GTPase *Ran*. *Int. Rev. Cyto.*, 217:41–91
- Stephen, M.K. (2000) Protein phosphatases and the regulation of mitogen-activated-protein kinase signaling. *Curr. Opin. Cell Biol.* 12:186–192
- Sterck, Lieven, Stephane Rombauts, Stefan Jansson, Fredrik Sterky, Pierre Rouzé, Yves Van de Peer (2005) EST data suggest that poplar is an ancient polyploid. *New Phytol.* 167, 165–170
- SUCEST (<http://www.sucest-fun.org/en/database-tools/cane-enome/sucest/cluster-read-search>). Acessado em janeiro de 2005.

- Takahashi F, Yoshida R, Ichimura K, Mizoguchi T, Seo S, (2007) The mitogen-activated protein kinase cascade MKK3-MPK6 is an important part of the jasmonate signal transduction pathway in Arabidopsis. *Plant Cell* 19:805–18
- Tchieu JH, Fana F, Fink JL, Harper J, Nair TM, Niedner RH, Smith DW, Steube K, Tam TM, Veretnik S, Wang D, Gribskov M.(2003) - The PlantsP and PlantsT Functional Genomics Databases. *Nucleic Acids Res.* 2003 Jan 1;31(1):342-4. - PMID: 12520018 [PubMed - indexed for MEDLINE
- Teige M, Scheikl E, Eulgem T, Doczi R, Ichimura K, (2004). The MKK2 pathway mediates cold and salt stress signaling in Arabidopsis. *Mol. Cell* 15:141–52
- Tena, Guillaume, Tsuneaki Asai, Wan-Ling Chiu and Jen Sheen (2001) Plant mitogen-activated protein kinase signaling cascades. *Curr. Opin. Plant Biol.* 4, 392–400
- Vettore, A.L., da Silva, F.R., Kemper, E.L., Arruda, P. (2001). The libraries that made SUCEST. *Genet. Mol. Biol.* 24:1-7.
- Walia A, Lee JS, Wasteneys G, Ellis B. (2009) Arabidopsis mitogen-activated protein kinase MPK18 mediates cortical microtubule functions in plant cells. *Plant J*, 59:565–75
- Wang H, Ngwenyama N, Liu Y, Walker JC, Zhang S. (2007). Stomatal development and patterning are regulated by environmentally responsive mitogen-activated protein kinases in Arabidopsis. *Plant Cell*, 19:63–73
- Wang H, Liu Y, Bruffett K, Lee J, Hause G. (2008). Haplo-insufficiency of MPK3 in MPK6 mutant background uncovers a novel function of these two MAPKs in Arabidopsis ovule development. *Plant Cell*, 20:602–13
- Wang, Degeng , Jeffrey F. Harper, and Michael Gribskov (2003) - Systematic Trans-Genomic Comparison of Protein Kinases between Arabidopsis and

Saccharomyces cerevisiae- *Plant Physiology*, August 2003, Vol. 132, pp. 2152–2165, www.plantphysiol.org © 2003 American Society of Plant Biologists

Whitmarsh, A.J. Davis, R.J. (1998) Structural organization of MAPkinase signaling modules by scaffold proteins in yeast and mammals. *Trends Biochem. Sci.* 23: 481–485

Widmann, Christian, Spencer Gibson, Matthew B. Jarpe, and Gary L. Johnson (1999) Mitogen-activated protein kinase: conservation of a three-kinase module from yeast to human. *Physiol. Rev.* 79, 143–180

Xing, Yu Wensuo Jia, Jianhua Zhang (2008) AtMKK1 mediates ABA-induced CAT1 expression and H₂O₂ production via AtMPK6-coupled signaling in Arabidopsis. *Plant J.* 54, 440–451

Yu J, Hu S, Wang J, Wong GKS, Li S, Liu B, Deng Y, Dai L. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp indica). *Science*, 296: 76-91.

Yu,Jun Yu, Jun Wang, Wei Lin, Songgang Li, Heng Li, Jun Zhou, Peixiang Ni (2005). The genomes of *Oryza sativa*: a history of duplications. *PLoS Biol.* 3, e38

Zhang X, Dai Y, Xiong Y, DeFraia C, Li J (2007). Overexpression of Arabidopsis MAP kinase kinase 7 leads to activation of plant basal and systemic acquired resistance. *Plant J.* 52:1066–79

Zhang, S., Klessig, D.F. (2001) MAPK cascades in plant defense signaling. *Trends Plant Sci.*6:520–527

Zhang, Shuqun, Yidong Liu, Daniel F. Klessig. (2000) Multiple levels of tobacco WIPK activation during the induction of cell death by fungal elicitors. *Plant J.* 23, 339–347

Zhou C, Cai Z, Guo Y, Gan S. (2009). An Arabidopsis mitogen-activated protein kinase cascade, MKK9-MPK6, plays a role in leaf senescence. *Plant Physiol.* 150:167–77

Zou X, Peng T, Pan Z. (2008). Modeling specificity in the yeast MAPK signaling networks. *J. Theor. Biol.* 250:139–55

APÊNDICES

Quadro 1.

Espécie	MP	MKK	Espécie	MP	MKK
<i>Triticum aestivum</i>	AAC28850.1 GI:3396052		<i>Euphorbia esula</i>	AF65766.1 GI:7649153	
	CAE54269.1 GI:40641579		<i>Ipomoea batatas</i>	AAD37790.1 GI:5007038	
	AAX20165.1 GI:60460041		<i>Medicago sativa</i>	CAA57719.1 GI:1204129	CAC69138.1 GI:15528441
	AAO16560.1 GI:27542952			CAB37188.1 GI:4456682	CAC69137.1 GI:15528439
<i>Zea Mays</i>	BAA74733.1 GI:4239887	AAC83393.1 GI:2737973		AAD28617.1 GI:4732123	
	BAA74734.1 GI:4239889	AAK73104.1 GI:14719279	<i>Nicotiana tabacum</i>	BAA09600.1 GI:1136298	CAC24705.1 GI:12331300
<i>Capsicum annum</i>	AAF81419.1 GI:8925321			Q40532.1 GI:2499615	AAG53979.1 GI:12484128
	AAF81420.1 GI:8925323			AAB58396.1 GI:2138340	BAA06731.1 GI:862342
<i>Cicer arietinum</i>	CAB61750.1 GI:6469137			CAA58760.1 GI:634068	AAF67262.1 GI:7682432

The image shows the Clustal Omega web interface. The top section contains various configuration options for sequence alignment. The 'GAP OPEN' parameter is set to 50 and the 'GAP EXTENSION' parameter is set to 0.5. The 'OUTPUT FORMAT' is set to 'phylip'. The bottom section shows a text input field for sequences in FASTA format, with a red circle highlighting the input and a blue arrow pointing to the text 'Formato FASTA'.

YOUR EMAIL

ALIGNMENT TITLE

RESULTS

ALIGNMENT

KTUP (WORD SIZE)

WINDOW LENGTH

SCORE TYPE

TOPDIAG

PAIRGAP

MATRIX

GAP OPEN

NO END GAPS

GAP EXTENSION

GAP DISTANCES

ITERATION

NUMITER

OUTPUT

OUTPUT ORDER

TREE TYPE

CORRECT DIST.

PHYLOGENETIC TREE

IGNORE GAPS

CLUSTERING

Enter or paste a set of sequences in any supported format:

```

>AtMPK1
MATLVDPNPGIRNEGKHYFSMWQTLFEIDTKYMPIKPIGRGAYGVVCSVNSDTNEKVAIKKIHNVYE
NRIDALRTLRELKLLRHLRHENVIALKDVMMPHMKMSFKDVYLVYELMDTDLHQI IKSSQVLSNDHCQ
LNTEL
>AtMPK2
MATPVDPNPGIRNQGKHYFSMWQTLFEIDTKYVPIKPIGRGAYGVVCSVNRRESNERVAIKKIHNVFE
NRIDALRTLRELKLLRHLRHENVVALKDVMANHKRSFKDVYLVYELMDTDLHQI IKSSQVLSNDHCQ
NNNEVSEF

```

Figura 1. Alinhamento com o Clustal. Os parâmetros que devem ser alterados são: *gap open* e *gap extension* quando se está alinhando transcritos. O demais parâmetros se mantêm nos valores *default*.

Please cite:
 "A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood."
 Guindon S., Gascuel O.
 Systematic Biology, 52(5):696-704, 2003.

Server load: 34%

Sequences (PHYLIP format)

Data Type File Example file

Sequence file DNA Amino-Acids

Number of data sets interleaved sequential

Substitution Model

Substitution model LG model is new and tends to outperform JTT and WAG

Tree Searching

Starting tree(s) File BIONJ

Type of tree improvement

Number of random starting tree

Optimise topology yes no

Optimise branch lengths yes no

Branch Support

Compute aLRT Use aLRT to save computing time yes no

Perform bootstrap yes no

Your name

Your email

Figura 2. Ferramenta PhyML. Carga do arquivo com extensão *.aln* (formato PHYLIP) gerado pela ferramenta CLUSTALW. O parâmetro Server Load indica a carga de processamento atual da ferramenta. Quando a ferramenta atinge 100% é preciso aguardar, pois o servidor não tem condições de processar no momento.

Sequências dos clones de cana-de-açúcar submetidos ao NCBI em 11/12/2010.

MPKs

>ScMPK3

SCCCRZ2C04G10.G

PCEPVFLFSVVEGNKPACQGLFPGXKKTMSGGGMDGAPVAEFRPTVTH
 GGRFLQYNIFGNLFEITHKYQPPIMPIGRGAYGIVCSVMNFETREMVAIKKI
 ANAFDNHMDAKRTLREIKLLRHLDHENIIGIRDVIPPPIQAFNDVYIGTELM
 DTDLHHIIRSNQELSEEHCQYFLYQILRGLKYIHSANVIHRDLKPSNLLLNA
 NCDLKICDFGLARPSSSESDMMTEYVVTRWYRAPELLNSTDYAAIDVWSV
 GCIFMELINRQPLFPGRDHMHQMRLITEVIGTPTDDELGFIRNEDARKYMR
 LPQFPRRPFASLFPKVQSVAAALDIERMLTFNPLQRITVEEALHPYLERLH
 DIADEPICTEPFSDFEQQALTEDQMKQLIFNEAIEMNPNFRY

ATATATATATATCCAATCTAATATCGGAAATTGGGGTTCATTTTCGATGGCCTCGTTGAAT
 ATCAGCTGCTTCATTTGGTCTTCTGTTCAGAGCCTGCTGCTCGAAGTTCGAACGAGAATGGC
 TCCGTGCAGATGGGCTCGTCGGCGATGTCTGTGAACCGCTCCAGGTACGGGTGCTCCAGC
 GCCTCTTCAACTGTGATCCTCTGCAGCGGGTTGAAGGTTAGCATCCGCTCGATGAGGTCC
 AGCGCCGCGACGGACTGCACCTTCGGGAACAGGCTCGCGAACGGCCGCGCGGGAACTGC
 GGGAGGTGGCGCATGTACTTCCGCGCGTCTCGTTCCGGATGAACCCGAGCTCATCGTCC
 GTCGGCGTCCCAGTACCTCGGTGATGAGGCGCATCTGGTGCATGTGGTTCGCGTCCGGGG
 AAGAGCGGCTGGCGGTTGATGAGCTCCATGAAGATGCAGCCAACGGACCAGACGTTCGATG
 GCGGCGGAGTAGTTCGGTGGAGTTGAGCAGCAGCTCGGGCGCGCGGTACCACCGGGTCACC
 ACGTACTCCGTTCATCATGTTCGCTCTCGGAGGAGGGGCGCGCCAGCCCCGAAGTCGCAGATC
 TTGAGGTTCGCAGTTGGCGTTTTCAGCAGCAGGTTGCTGGGCTTGAGGTTCGCGGTGGATCACG
 TTGGCGGAGTGGATGTACTTGAGCCCCGAAGGATCTGGTACAGGAAGTACTGGCAGTGC
 TCCGCCGAGAGCTCCTGGTTGGAGCGGATGATGTGGTGGAGGTCCGTGTCCATGAGCTCC
 GTGCCGATGTACACGTTCGTTGAACGCCGCGGGATCGGCGGCGGGATCACGTCCCTGATG
 CCTATGATGTTCTCGTGGTTCGAGGTGCCTCAGCAGCTTGATCTCCCGGAGCGTGCCTTG
 GCGTCCATGTGGTTGTTCGAAGGCGTTGGCGATCTTCTTGATGGCCACCATCTCCCTCGTC
 TCAAAGTTCATCACCGAGCAGACGATCCCGTAGGCGCCGCGCCGATGGGCATGATGGGA
 GGCTGGTACTTGTGCGTGATCTCGAACAGGTTGCCGAAGATGTTGTAAGTGCAGGAACCGG
 CCGCCGTGCGTCACCGTCCGGCCGGAACTCGGCGACCGGAGCTCCGTCCATGCCTCCTCCG
 CTCATCGTCTTCTTCCCCCGGGGAAGAGGCCCTGGCAGGCTGGCTTATTCCCTCGACGA
 CGGAGAAGAGGAAGACCGGCTCACAAGGTACAGGTGCAAGCAGCAAAGCAAATCAGGCA
 ATTTGACTTTATCTTCTCCCTCCCTCG

>ScMPK4

SCVPCL6042B11.G

LSSTPPSREARKEGDEETMDSSSSGGGGAQIKGMATHGGRYVLYNVYGN
 LFEVSSKYAPPPIRPIGRGAYGIVCAAVNSQTGEEVAIKKVGNAFDNHIDAK
 RTLREIKLLRHMDHENILALRTYNAPQLXENFNDVYIVTELMDTDLHQIIRS
 NQPLTDDHCQYFLYQLLRGLKYVHSANILHRDLKPSNLFNANCDLKIADF
 GLARTTSETDLMTEYVVTRWYRAPELLLNCSQYTAIDVWSVGCILGEIVT
 RQPLFPGRDYIQQLKLITELIGSPDDASLGLFRSDNAKRYMKQLPQFPKQD
 FRLRFRNMSPGAVDLLERMLVFDPSRRITVDEALHHPYLASLHDINEEPTC
 PAPFSDFEQPSFTEA

ATCCCCATTCCCTTCCCCAGTCGCCGCCCGAGTCAAAAAGAGGGAAGCTTCTTCCACCACA
 GTCCTTGAGGCCGCCCTCCCATCAACGACTAGACGCTAGTCAAGCAGCAGCACCCACCCT
 AGCAGCCCCACAACCCCTAGGCCCTCTCTCCACACCCCTTCCCAGAGGCGAGGAA
 GGAAGGAGACGAAGAAACCATGGATTCCCTCCCGCGGCGGCGGAGCGCAGATCAAGGG
 TATGGCGACGCACGGGGCCGCTACGTGCTGTACAACGTGTACGGAAACCTCTTCGAGGT
 CTCCTCCAAGTACGCCCCGCCATCCGCCCATCGGCCGCGGCGCCTACGGCATTGTCTG
 CGCGGCTGTCAACTCGCAGACAGGGGAGGAGGTTGCGATCAAGAAGGTTGGCAATGCGTT
 CGACAACCACATCGACGCCAAGCGGACGCTGAGGGAATCAAGCTGCTGCGCCACATGGA
 CCATGAGAACATCCTTGCCCTAAGGACGTATAACGCCCCCACTAGAGAGAACCTTAAT
 GACGTGTACATTGTTACTGAGTTAATGGATACAGATCTCCATCAGATCATACGCTCAAAT
 CAGCCATTGACTGATGATCATTGCCAGTACTTTTTGTATCAGTTGCTAAGAGGACTAAAG
 TATGTGCACTCAGCAAATATATTGCACCGTGATCTAAAGCCGAGCAATTTGTTCCATAAT
 GCAAATTTGTGACCTCAAGATCGCAGACTTTGGGCTTGCAAGGACCACCTCAGAGACAGAT
 CTCATGACGGAGTATGTGGTCACTCGTTGGTACAGGGCACCGGAGCTGCTGTTGAACTGT
 TCACAGTACACTGCCGCCATTGATGTCTGGTCAGTTGGATGCATACTAGGTGAAATCGTT
 ACTCGCCAACCCCTGTTTCCCTGGGCGGGATTACATCCAGCAATTAATAATTGATCACTGAG
 CTCATAGGCTCACCAGATGATGCAAGCCTGGGATTCCCTTCGAAGTGATAATGCAAAAAGA
 TACATGAAACAGCTACCACAGTTTCCAAAGCAGGACTTCCGCCTGCGTTTCCGCAACATG
 TCTCCTGGTGCAGTCGATTTGTTGGAGAGAATGCTTGTGTTTGTATCCAAGCAGACGGATT
 ACAGTTGATCAGGCTTTGCATCACCATACTTTGGCTTCACTTCATGACATCAATGAAGAA
 CCTACCTGCCCTGCACCTTTTCCAGCTTTGATTTTGGAGCAACCATCCTTTACAGAAGCGCAT
 ATAAAAGAACCTTATCTGGAGGGAATCTTTAGCATTAAACCCGGATCCTCCCTACTAATAT
 CAAAAGGATAAGTTATCTGCTAGAGGCATATGAAGATCTCTAGCTCCAGCGGACGCAAGA
 ACTTAGTGTGAGTGTGCATCGATTATTTGTGTGCGGTTTGTGGGCAAACGGGCATGGGT
 TGTGTGCTGCAATAGCAACTCTATTAGGGCTTGTGAGAGCAAACAAACGTCTGTATTCCCTT
 TTCAGATGCCCTGAACCCAGGCTTGTATTATACCGCTTGAATTCATAAACTATTCTTATGT
 ATCTAAGCTTGTAAATTTGTAGACAATTACTTGAACGCCCCCTGTATATCTGAGTTGACA
 TCACCTCCTTATTGTAACATCTCAAACACATTTTATTGTGATTTTAAAAA

>ScMPK6

SCEQRT1030A09.G

IQYNIFGNVFEVTAKYKPPILPIGKAYGIVCSALNSETAEQVAIKKIANAFD
 NKIDAKRTLREIKLLRHMDHENIVAIRDIIPPAQREAFNDVYIAYELMDTDLH
 QIIRSNQALSEEHCQYFLYQILRGLKYIHSANVLHRDLKPSNLLLNNANCDLK
 ICDFGLARTTSETDFMTEYVVTRWYRAPELLNNSSEYTAIDVWSVGCIFM
 ELMDRKPLFPGRDHVHQLRLLMELIGTPNEADLDFVNENARRYIRQLPRH
 ARQSLPEKFPHVQPLAIDLVEKMLTFDPRQRITVEGALAHPYLASLHDISD
 EPVCSMPFSDFEQHALSEEQMKDLIYQEALAFNPDYQCRSSISAGD

CAGCATAAGCGAACGGGGCCATTGGAATTACAGGTCTTTTTTAAACAATTAAGTCGATAA
 ACAAGAATCTTAAGTCGACCAGAAATGTTAACTTTGACGACAGCCACTGAATTCAGGCA
 GCCATCGTTTCATGAACTACCTCTCTTACAGACAACATCGAAAAGTTTTTAAGAAGAAAAG
 TCTGATTAATAACAACATATTGTAAGCACAAGAATCCATAGCTACGCAGAAATTTGACCA
 TCACAAGCTCGCTCTTTTCTACCCTTGAATGACACCGTCTCCAGCCTTTTTAATCAAGGA

GGAATGGAGGATCTTCTGGTAAGGCAATTATTTACAGCTAGCAATCGGAGTTGGTGACTT
TTTTGCAGTGACATCAGGTGGGGAAAAAATCTCCATGTCAATGAAGCCAATCACCGGC
TGAAATTGAACTTCGGCTACTGGTAATCAGGGTTGAATGCAAGAGCCTCTTGGTAGATCA
GATCCTTCATTTGTTCTTCCGACAATGCATGCTGCTCGAAGTCAAAGCTGAAGGGCATTG
AGCAGACTGGCTCATCACTTATGTGCATGAAGTGATGCCAAGTAAGGGTGTGCAAGTGCGC
CTTCAACAGTTATTCTGTCTAGGATCAAAAGTCAGCATCTTTTCCACCAGGTGCGATTG
CTAAAGTTGAACATGTGGAATTTTTTCAGGTAAGGACTGTCTAGCATGCCGGGGAAGTT
GGCGGATATATCTTCTTGCATTTTCATTTACAAAATCAAGATCAGCCTCATTCCGGTGTCC
CAATGAGCTCCATTAGTAGACGTAGCTGATGGACATGGTCTCTTCCAGGAAACAGGGGTT
TTCCGGTCCATCAATTCATAAAATATACAGCCACAGACCACACATCAATAGCAGCAGTAT
ATTCAGAGGAGTTCAACAAAAGCTCTGGTGCTCTATACCATCTTGTGACAACATATTCAG
TCATAAAATCAGTTTCTGAGGTGGTGCGAGCAAGCCCAAAATCACAAATTTGAGGTGAC
AGTTTGCATTTCAAAGAAGATTGCTAGGCTTCAAGTCACGGTGAAGAACATTTGCTGAAT
GTATATACTTCAAGCCACGAAGAATTTGATAAAGAAAATACTGACAGTGCTCCTCTGACA
AAGCTTGATTTGAACGAATAATTTGATGCAGATCAGTATCCATCAATTCATACGCAATAT
ACACATCATTTGAATGCCTCCCTTTGTGCAGGAGGTATGATATCCCTTATTGCAACAATAT
TCTCGTGGTCCATGTGGCGGAGCAGCTTGATCTCGCGGAGCGTGCCTTGGCATCGATCT
TGTTGTGCAAGGCGTTGGCGATCTTCTTGATGGCCACCTGCTCCGCCGTCTCGGAGTTGA
GCGCCGAGCAGACGATGCCGTAGGCGCCCTTGCCGATGGGGAGGATGGGGGGCTTGTACT
TGCGGTGACCTCGAACACGTTACCGAAGATGTTGTACTGGATGA

>ScMPK7 SCCCHR1003D05.G + SCCCLB1C05H07.G

APDLRRRPVDAARAVRRILPDRASPMERRRAADDRTHEHERPLPPEKM
AMMVDPPNGMASQGHYYTMWQTLFEIDTKYVPIKPIGRGAYGIVCSSVN
RETNEKVXDKKINNVFDNRVDALRTLRELKLLRHLRHENVIALKDIMMPAH
RRSFKDVYLVYELMDTDLHQIHKSSQPLSNDHCQYFLFQLLRGLKYLHSAG
ILHRDLKPGNLLVNANCDLKICDFGLARTNNTKGQLMTEYVVTRWYRAPE
LLLCCDNYGTSIDVWSVGCIFAELLGRKPIFPGTECLNQLKLIVNVLGTMSE
ADLEFIDNPKARKYIKSLPYTPGIPLTSMYPQAHPLAIDLLQKMLVFDPSKRI
SVTEALEHPYMSPLYDPSANPPAQVPIDLDIDENXRRRYDPGNDVAGDDP
LPPRGPBKNDHVTGSNEHVAEVCCTRIITVRLFYELTXDYGMSQVITICIG
QSGDLLCKYALNKMCGN

>ScMPK14 SCCCRZ1003E02.G

SSRKMAMMVDPPNGIGSQGHYYTMWQTLFEIDTKYVPIKPIGRGAYGIV
CSSINRETNEKVAIKKIHNVDNRVDALRTLRELKLLRHLRHENVIALKDIM
MPIHRRSFKDVYLVYELMDTDLHQIHKSPQGLSNDHCQYFLFQLLRGLKYL
HSAEILHRDLKPGNLLVNANCDLKICDFGLARTNSSKGQFMTEYVVTRWY
RAPELLLCDNYGTSIDVWSVGCIFAELLGRKPIFPGTECLNQLKLIVNVLG
TMSEADLEFIDNPKARRYIKSLPYTPGVPLASMYPHAHPLAIDLLQKMLIFD
PTKRISVTEALEHPYMSPLYDPSANPPAQVPIDLDIDENISSEMIREMMWQ
EMLHYHPEVVTAISMYEIPLPQDIVR

CATCCCTCCTCGGCTCCTCACGGCACGAGACTCCCTCCCCTCCCCTCTTCTCCCCTCCC
GCGGCTGGAGGTGGTCACACCGAGCGGACCGGAGGCGAGGGATTCCGCCGCGCGCATGGC
GCTGCCGCTTGACCCTAGCTAGCTGCTCCACCCACCCACCCACTCGGCCAGGCGCGCCC
TCCCCTCCCCTTCTCCTCTCGTAAAATGGCGATGATGGTGGATCCTCCGAATGGAATCGG
GAGCCAAGGAAAGCATTACTACACTATGTGGCAGACCTTATTTGAGATAGACACCAAATA
TGTGCCAATCAAGCCATTGGTTCGAGGAGCTTATGGGATAGTTTGTTCATCCATCAATCG
TGAAACAAACGAGAAAGTAGCAATAAAGAAGATACACAATGTTTTTGGACAACCGTGTGGA
TGCACTACGGACCTTGCGGGAGCTGAAACTCCTACGCCATCTCCGCCATGAGAATGTCAT
TGCTTTGAAGGACATAATGATGCCAATACACAGGAGGAGCTTTAAGGATGTATACTTGGT
CTACGAGCTCATGGATACTGATTTGCACCAGATAATCAAATCGCCTCAGGGCCTTTCTAA
TGACCATTGCCAGTATTTCTTTTCCAGTTGCTCCGAGGACTGAAATATCTCCATTGAGC

>ScMPK17-1

SCCCRZ2C04F05.G

AIVHGFRRWFHRRNGSTSGSNQSSVAGEGDDGSSDLEVNEPDLVGLR
 AIRVPKRKMPLPVESHKNSVEMEFFTEYGEASQYQIQEVIGKGSYGVVA
 AAVDTRTGERVAIKKINDVFEHVSDATRILREIKLLRLLRHPDIVEIKHIMLPP
 SRREFQDIYVVFELMESDLHQVIKVNDDLTPHHQFFLYQLLRALKYIHAA
 NVFHRDLKPKNILANSCKLKICDFGLARASFNDAPSAIFWTDYVATRWR
 APELCGSFFSKYTPAIDIWSIGCIFAELLTGRPLFPGKNVVHQLDLITDLLGT
 PSSETLSRIRNEKARRYLSMRKKYPVPFTHKFRNADPLALRLLERLLAFD
 PKDRPTAEEALADPYFASLANVEREPSRHPISKLEFEEFERRKLAKDDVREL
 IYREILEYHPQMLEEYMKGGEQISFLYPSGVDRFKRQFAHLEEHYSKGER
 GSPLQRKHASLPRERVVSKDGNTEQHINDQERSADSVARTTVSPRSE
 DADMNDVKSMSLSSRSYLKSASISASKCVVVTNKHPEDDEIPEEMDDDDVD
 GLSKQVSR

TTTTCTTGACAGGGCGAAATGTGCACTGTAGTGATATCCCCTCAAATGGAAAAAGGGGAT
 GGTACACTGCTGACACTTCGTACGGCAGAAGCAAGCCATTGTAATAAAACTCAAACATAG
 GTCCAGACTCCATTGCAAGAAAACGTCTCATGGTCGACAGGCTGACCATACTTCATTGTT
 CATTCTGTTTTCTCCATGGAAGTGTGCTACAGCTCAGGAATTCTTCATGAATCCATGCA
 GAGCTTCTAGAGACACCATGCCATGCTGGCATGACACGGTGGAAATACAGATGATGAAGA
 AAGGACCATCAGACGGTGGCATATTCCTCTTCTACCCAAGTAAAATCGAACTAGCTGTGG
 GTGTACACAAAATTCAAACCTTGCAATGCAGTGTACCATAGAGTGTTCGCTTTCCAGGC
 TGCCCTCTATATCATCGGCCTGTACCAGGTAGAAAACCTTATGCCAATCGCCTGGACAGTC
 CTGCCTAGGAGTGCATCCTGGAGACCTGTTTAGACAATCCATCGACGTCGTCATCCATTT
 CCTCAGGGATTTTCATCATCCTCTGGGTGTTTATTTGTGACAACGACACACTTGGGAAGCAC
 TTATGCTTGCACCTTCAGGTAGCTCCGAGAGCTTAAGCTCATGGATTTACATCGTTCA
 TGTCACTCTTCTGACCTTGGAGGGCTCACAGTAGTGCGGGCGACAGAATCTGCACCTC
 TCTCCTGGTCAATTAATATGTTGCTCAGTATTACCATCCTTCGATACAAACCACTCTTTCC
 TTGGTAAAGAAGCATGCTTTCTTTGTCAGTGGAGAACCTCTTTCTCCTTTGCTATAATGCT
 CCTCAAGATGTGCAAACCTGCCGTTTGAACCGATCAACTCCACTTGGATAGAGGAAGCTAA
 TCTGTTCTCCACCTTTTCATGTACTCCTCCAGCATCTGTGGATGATACTCCAAAATCTCTC
 GATAGATCAATTCTCTAACATCATCTTTGGCCAGCTTCCTTCTTTCAAATTCAAAACCTCAA
 GTTTAGAAATAGGGTGTCTTGATGGCTCACGTTCCACATTAGCAAGGGATGCAAAGTATG
 GATCAGCTAGAGCCTCTTCTGCAGTAGGCCGATCTTTAGGATCAAATGCAAGTAAACGCT
 CTAGAAGACGAAGAGCCAAAGGATCAGCATTTGCAAATTTATGAGTAAAGGGCACGGGAT
 ATTTTTTCCGCATGCAGCTCAAGTATCTCCTGGCCTTCTCATTTGCAATCCGAGATAACG
 TCTCAGATGATGGAGTTCCAAGAAGATCTGTTATAAGATCTAATTGGTGTACAACATTCT
 TCCCAGGAAATAATGGCCTTCCAGTGAGAAGTTCAGCAAATATGCACCCAATGCTCCAAA
 TATCAATTGCAGGAGTGTATTTGGAGAAAAAGGAACACATAGTTCAGGTGCCCGGTACC
 ACCTTGTGTGCCACATAATCTGTCCAAAATATAGCTGAAGGGGCGTCATTAATGATGCAC
 GGGCAAGTCCAAAGTCACATATTTTCAGTTTGCAGTCCGAGTTAGCCAGTATATTCTTGG
 GCTTTAGGTGCGGATGAAATACATTAGCAGCATGTATGTACTTGAGAGCACGAAGAAGTT
 GGTACAGGAAAAACTGGTGTATGTTCAAGGGTGGAGTTCATCATTACCTTGATCACTTGAT
 GGAGATCGGACTCCATGAGCTCAAAAACAACATAGATATCTTGGAACTCCCTCCGAGAAG
 GAGGGAGCATGATGTGCTTAATCTCAACTATGTCTGGGTGACGAAGCAACCCGAAGGAGCT
 TGATCTCCCGCAGAATGCGCGTGGCATCCGAGACGTGCTCAAACACATCATTGATCTTCT
 TGATAGCAACCCGCTCACCGGTGCGGGTATCTACAGCTGCAGCAACTACTCCGTAGCTCC
 CTTTACCAATGACTTCTTGGATCTGGTACTGGCTTGCCTCTCCATACTCTGTGAAGAACT
 CCATTTCCACCGAGTTCTTCTGTGGCTCTCGACGGGCGAGCGGCATCTTGCCTTGGGCA
 CGCGGATGGCGCGGAGGCCGACGAGATCCGGGTCTCGTTGACCTCGAGATCGGACGAGC
 CGTCGTCCCCCTCGCCGGCGACGGACGACTGGTTGGAGCCGGAGGTGGAGCCGTTCCGGC
 GGTGGAACCAGCGCGGAAGCCGTGGACGATCGCGCCTCCTCCGCCATGCTCCTAGCCC
 TGCGCGCCCACGGCATTGGCCCCGACGCCCGCCCGCCTAGCTATATAACCACCACAGTC
 CACCGCTCCAAGCTCCTCTCCTCCCGTGTCCCGTCCCTCCTCTACTCGCTTGGTGCCGT
 CAG

AGGCTCCACAAAGAGTGCCAACAGCGAGACCAGGAAGGGTTGTTGGTCCAGTAATGCCTT
 ACCAAAATGGAGACACCAAAGACCCCTATGACACACGAAGGTTGGCAATGAACTCAGGAT
 ATCCTCCCAACAACAAATCCCAACAACATTTGGTTATTATCAGACACATGGCAAACCAG
 CTTGCTCCGAGCCATCACAGGCTGAAAGGTATACGCTGCATCAGCAGGCCTATGCCTGCG
 CAAACCGTACAGCTGTGTCTGATGTTGCTCTGGACATGAGAGCACCCCTTTTCATCATC
 TGTCGGCAGGGCAAAAAGGTGACTCTTCTGATAGGCTAACAGCCGAGACAACTTGTACA
 CGAGATCACTCAACGGCATTGCTGCTACTGCAGCAGGAGTGGCAGCAACTACTCACAGAA
 AGGTCGGTGTGTTCCGTTTGGCATGTCAAGCATGTATTAGTGACTGGATCATGGCCGAG
 TTATGGTGAATCCTACGTGCCCTTTGCTGGAGCTATGCTGCCCTTCAGGTGGCTAAACATCA
 CATCGATAATGAAGCGGCCAATCAAACCTTTGGCTGGATGCTGCCAAGTTTAGAGTGCAG
 AGCAGTACAGGGTTCCGTGTGTGGTCAGGAAAACAGGAACCACTGTCGCCGGCATGAAGCG
 TGTCAACATGTTTGTCTATTTAGGGCACCAGCTTGTATCTGCCTGTGCCTGAGAAGTGC
 CCTTGGTGGCACACTGAAGGGGTAGTCCAGCGGCTTGTAACTGGGTAATTTAGGCGCCA
 ATTTCTTTTCTAGAGCGAGGAGGCAGAAGGAAAGGAAGGAACATTAACGTTTTAAATGT
 TGTATAGAGATGTTTCTCAAATATCAATGTGCTACATCAGAATCACTTCAGAAAGATAACC
 AAGATATAGCAAAAATTTGCTTGGTATGAATGTTTGC

>ScMPK20-2

SCSGRT2064G11.G

PSSRALHAVRGEIGRRSDPFAARPPSHVLGRSGACSPDPARARRXRG
 AVPPPQRRRVL SAAEARAAGFSFGTGVGREIRLAGMQDQRNKTSAEAE
 FFTEYGDANRYKIQEVIGKGSYGVVCSAIDLHTRQRVAIKKIHNIFEHVSDA
 ARILREIKLLRLLRHPDIVEIKHIMLPPSRKDFKDIYVVFELMESDLHQVIKAN
 DDLTKEHYQFFLYQLLRALKYIHTANVYHRDLKPKNILANSNCKLKICDFGL
 ARVAFNDTPTTVFWTDYVATRWRAPELCGSFFSKYTPAIDIWSIGCIFAE
 VLTGKPLFPGKNVVHQLDLMTDLLGTPSMDTISRVRNEKARRYLSMRKK
 DPVPFSQKFPSADPLALKLLEKLLAFDPKDRPTAEEALRDPYFKGLSRVER
 EPSCQPIRKVEFD FEHKRMSKEEIRELIFREILEYHPQLLNSYINGTERTTFL
 YPSAVDQFKKQFSLLEESGGNGPSVPTDRKHASLPRTTVVHSNPIPAKEQ
 PLAASSKVDQSLMVRVRIFGRKEVVLEMFPGHLWLHKDKHKQXSVRV
 NGSVMNSGYPPHQIPQAYGYHQMPARLDSTNPSQAMGGYTLQSQQA
 YAYANSKGSPPDVAVNMRAPPFHLPAAGPKNNNPLDRLATDIDIYTRSLNGI
 VAAATASEALALALTETLASAIWHVKDVLADMLATASPLACKK

TCCCCGCTCCGCCGCTGCCCTGCCCTTCCCTTCTCTCCCATCCTCGAGGGCGCTGCACGCGGT
 GCGCGCGGAGATCGGGCGGAGGAGCGACCCCTTCGCCCGCGCGGCCACCGTCCCACGTCTCT
 GGGGCGCTCGGGGGCTTGTCTGCCCCCTGATCCCGCCCGCGCGCGCGCTCTCGAGGGCCG
 GCGGTGCCGCGCCGCGCAGCGTAGGCGGGTGTGAGTGCGGCGGAAGCGAGGGCCGCGGGG
 TTTTCATTTGGCACCGGCGGTGGGGCGGGAGATCCGGCTAGCGGGGATGCAGCAGGACCA
 GCGGAACAAGACTTCTGCAGAGGCCGAATTCTTCACAGAGTATGGGGATGCAAACCGGTA
 CAAGATCCAAGAAGTCATTGGTAAGGGAAGCTATGGGGTGTCTGCTCTGCCATTGATCT
 TCACACTCGGCAGAGAGTGGCGATCAAGAAGATACACAATATCTTTGAGCACGTCTCCGA
 TGCCGCGAGGATCCTCCGTGAGATCAAACCTCTGAGGCTCCTAAGGCACCCTGATATTGT
 TGAGATCAAGCACATTATGTTACCTCCCTCGAGGAAGACTTCAAGGATATTTATGTTGT
 TTTTGAAC TTATGGAGTCTGATCTCCACCAAGTTATAAAGGCCAATGATGACTTGACCAA
 GGAGCATTACCAGTCTTTCTCTATCAGTTACTTCGGGCCCTCAAATACATTACTACTGC
 TAACGTTTATCACCGTGACCTGAAGCCCAAGAATATTTTAGCAAACCTCTAACTGCAAATT
 GAAAATATGTGACTTTGGACTAGCCCGAGTGCATTCAATGATACCCCAACAACAGTCTT
 CTGGACGGATTATGTTGCAACAAGGTGGTACAGAGCTCCAGAGCTCTGTGGATCCTTCTT
 CTCCAAGTATACACCAGCTATTGACATTTGGAGCATTGGATGCATATTTGCTGAGGTGTT
 GACAGGGAAGCCTTTATTTCTGGTAAAAATGTTGTTTCATCAGCTAGATTTGATGACTGA
 TCTTCTAGGCACACCGTCAATGGATAACAATTTCTCGGGTTTCGGAATGAGAAAGCAAGAAG
 GTACTTGAGCAGCATGAGAAAGAAGGACCCTGTTCCATTTTCTCAGAAGTTTCCCAGTGC
 AGATCCTTTGGCATTAAACTATTAGAAAAGCTATTAGCGTTTGGATCCAAAGGACCGGCC
 AACAGCAGAAGAGGCATTGCGTGATCCATACTTCAAAGGTCTTTCCAGGGTTGAAAGAGA
 ACCATCCTGTCAGCCAATCAGAAAAGTGAATTTGACTTTGAGCACAAAAGAATGTCAAA

GGAAGAGATAAGAGAGTTGATATTCCGTGAGATACTGGAATATCACCCACAATTGCTGAA
 TAGCTACATTAATGGCACAGAGAGGACAACCTTTCTCTACCCAAGTGCTGTTGATCAATT
 TAAGAAGCAGTTTTCTCTTCTTGAAGAGAGTGGTGGTAACGGCCCATCAGTTCCAACAGA
 CAGGAAACACGCATCCCTTCCCAGGACTACTGTTGTTCACTCTAATCCAATTCCTGCCAA
 GGAACAACCTCTTGCTGCCTCATCAAAGGTGAGACCAGTCTCTAATGGTTTCGTGTAAGAA
 TCTTTGGGAGAAAGGAAGTGGTCTCGGAAATGTTCCCAGGGCATCTCTGGCTCCACAAGG
 ACTGCAAGCACAAAGCAGCAGGATCAGTAAGAGTTAATGGCTCAGTGATGAATTCAGGGTA
 TCCCTCCTCATCAACAAATCCCACAAGCATAACGGTTACCACCAAATGCCTGCAAGATTGGA
 CAGTACCAACCCATCACAGGCCATGGGAGGTTACACACTGCAGTCGCAGCAGGCTTATGC
 TTATGCCAAACAGCAAAGGCTCGCCCGATGTGGCTGTGAACATGAGAGCACCCCCCTTCCA
 TCTCCCAGCTGGACCAAAGAATAATAACCCATTAGATAGACTAGCAACTGACACCGACAT
 ATACACAAGATCCCTTAAACGGCATTGTTGCTGCTGCCACGGCATCAGTAGGCGCTGGCGC
 TGGCGCTCACCGAAACGTTGGCGTCTGCCATCTGGCATGTCAAGGATGTATTAGCTGCT
 GATATGCTGGCCACAGCTTCTCCAGGACTAGCCTGCAAGAAGTGGCGTATGGTATAAGTA
 GTAATTCATAAGAAGAGGTGATCACATAGAGT

>ScMPK20-4 SCEPLR1030H10.G
 PGISTLSSVREGRRRAGANAGSLTGEERYAAAAGEXMQAPEQLHQQQQ
 RRKGSSEMDFSEYGDANRYKIQEVIGKGSYGVVCSAIDQHTGDKVAIKKI
 HNIFEHLSDAARILREIKLLRLLRHPDIVEIKHIMLPPSRDFKDIYVVFELMD
 TDLHQVIKANDDLTREHHQFFLYQMLRALKYIHTANVYHRDLKPKNILANA
 NCKLKICDFGLARVAFNDILQQYSGRIMLLQDGIGPLNCVDLSFLSIHQLLT
 YGVXCFIAEILTGKPLFPGKNVVHQLDLMTDLLGTPSLDTVSRIRNEKAR
 RYLSSMRKKQPVPFSERFPKADPAALKLLQRLLAFDPKDRPTAEEALADP
 YFKGLAKVEREPSCQPITKMEFEFERRKVTKEDVKELIFREILEYHPQLLKD
 YMNGTEKTNFLYPSAVDNFRQFANLEENGGKGGHAHTASSDRKHVSLPR
 TTTVHSNPIPPNGTSQLPQKIPTARPGRVIGPVIFFENAIAVDPYSQRRVAR
 NPVLPPTTNSLSAYAYHRKSDNSDRELQQELEKDRMQYQPAQRFMDAK
 VVSQMSPDLRSSYYIPKGVPKADGADRAALHSSMIHGIAPFSGIAAVGGG
 GGYNKVNAVQYGVSDVLLVALAISKILSRVEVITWRRVKDVGTNELPIG
 SEPKIQSMEYGLIRIKEDVDIRLKNKYWATVMVLSIGPSFSESNAITPRKR

ACAGAGCCAAAAATCCACAGCTTTGAATTAAGAAGCGGGTTTCACCAGCCGTAATAAT
 ACATAAAATCAAGTCACTTTATATACAGAGCTAAACGTAGGCTCACTATTTTTATTCCCA
 CATGTTTCCCTACCTTGCAAAGTACATAATTTCCCTGTAATAATCGTCTCTGCCCCCTTG
 CTCTTTATAAGCAGCAATCAACTGACGGCAGTCCCTTTTTTTCTCTTTTATTTTACCG
 TTTCCGCGGGGTGATGGCATTGCTTTCCGAAAAGCTAGGTCCAATGGACAGAACCATGAC
 TGTGCCCAATATTTGTTTTTCAATCTGATGTCACACGTCCTCCTTGATCCGAGATCAGA
 GATCCATACTCCATGCTTTGTATCTTTGGCTCAGAGCCAATATGGCAGTTCATTGGTACC
 TACATCTTTTACCCTTCGCCATGTTATTACTTCAACTCTGCTGAGGATTTTGGAGATGG
 CCAATGCAACTAGAACTAGTACATCCTAGAACTCCATATTGAACTGCATTGACCTTGTT
 GTAGCCACCTCCACCTCCTACAGCAGCAATGCCACTAAAAGGGGCGATTCCATGAATCAT
 GCTAGAGTGCAAAGCAGCTCTATCTGCACCATCAGCTTTTGGGACACCTTTTGGTATGTA
 GTAGGAAGACCTCAAGTCAGGTGACATCTGTGATACTACCTTGGCGTCCATGAACCGCTG
 TGCTGGCTGGTACTGCATGCGGTCTTTTCCAGCTCCTGTTGGAGCTCTCTGTCTGAGTT
 GTCTGACTTCCGGTGATAGGCGTATGCTGACAGGTTTCGAGGTGGTTGGAGGAAGCACTGG
 ATTCTCGCCACCCTACGTTGGCTGTATGGATCAACAGCTATTGCGTTCTCAAATGGTAT
 TACTGGACCAATCACTCTTCTGGTCTAGCTGTGGGAATCTTTTGGAGGAAGTTGGGAGGT
 TCCATTTGGAGGGATTGGATTAGAATGAACTGTGGTTGTCTTTGGGAGCGAAACATGCTT
 CCTGTCTGATGAAGCGGTGTGTGCCCCACCCTTTTCTCCATTTTCTCCAAGTTAGCAAA
 TTGCCCTTCGGAAATTTGTCAACAGCACTAGGATATAAAAAGTTTGTTTTCTCAGTTCCATT
 CATGTAATCCTTGAGAAGTTGAGGATGATACTCCAATATCTCCCTGAATATAAGTTCTCT
 GACATCCTCTTTTGTACCTTTCTTCGCTCAAACCTCGAACTCCATTTTGTATTGGCTG
 GCATGATGGTTCTCTCTACTTTTGGCGAGGCTTTAAAATATGGATCAAGGCAATGCCTC
 TTCAGCAGTTGGCCTATCCTTGGGATCAAATGCCAAAAGCCTCTGCAAAAGTTTGGATGC
 AGCAGGACTCTGCTTTAGGGAACCTTTTCAGAAAATGGTACTGGCTGTTTTTCTCATGCT
 ACTCAAGTACCTCCTGGCCTTCTCATTCCGAATTCTGGAACTGTATCAAGTGATGGAGT
 ACCCAAGAGATCAGTCATCAAATCCAACCTGGTGGACCACATTTTTTCCAGGAAACAAAGG

CTTCCCAGTTAGGATCTCCGCAAAAATGCAACCGATACTCCATATGTCAATAGCTGGTGA
 ATACTTAGAAAAGAAAGATCCACACAGTTCAGGGGCCCTATACCATCTTGTAGCAACATA
 ATCCGTCCAGAATACTGTTGTAGGAGTGTCAATTGAATGCCACTCTTGCCAATCCAAAATC
 ACATATTTTAAGTTTACAGTTAGCATTTGGCCAATATATTTCTTAGGCTTCAAATCACGATG
 AATAACATTAGCAGTATGAATGTATTTTCAGTGCACGGAGCATCTGATAGAGAAAGAAGT
 ATGATGCTCTCTGGTTAAATCATCATTTGCCTTGATGACTTGGTGGAGGTCTGTATCCAT
 CAGCTCAAATACAACATAGATATCTTTGAAGTCCCTCCTTGATGGAGGCAGCATTATATG
 CTTGATCTCAACTATATCAGGATGGCGTAATAGCCGAAGTAGTTTGTATCTCACGAAGGAT
 CCGGGCAGCATCAGATAAGTGTCTCGAAGATATTGTGTATTTTCTTGATTGCCACCTTGTC
 ACCGGTATGTTGGTCAATAGCTGAGCAAACAACACCATAACTTCCCTTTGCCAATGACCTC
 TTGGATTTTGTATCTGTTAGCATCACCGTATTCACTGAAGAAGTCCATCTCTGACGAACC
 CTTTCTCCGCTGCTGCTGCTGGTGCAGCTGCTCCGGCGCCTGCATCGCTCCCCGGCCGCC
 GCCGCGTCAGCGCTCCTCCGCCCGTGAGGCTACCCGCATTTGCCCGGCGCGCCGGCCGC
 GTTCCCGCACGGACGAGAGAGTGGAGATCCCTGGCCCTCCCTCCCACCAGAGTCCCGTCC
 CACCCCGCCCTGGAACCTGGATCGAGCCCGCGACGCTGCCAAGCCGGGACGCTCGCTGG
 CCCCAGCTGTGGCTGGCCCCCGCTCGCTCGCTGCGACTGAGCAACTCGCGGCGAGGG
 TGGGGTGGAGAGTTGTGGTGGTGTGGCGTCCGATTCCCTGATGGTGGTGTGGTGAATGG
 TGATGCTGCCG

>ScMPK20-5

SCSGLR1084B06.G

PCLRIGAAGSKQGLSHFSREVFRGCGLTMPPEANAGAMAQAEQRKKNspe
 MEFFSECGDANRYKIQEIIGKGSYGVVCSAIDRHTGEKVAIKKIQNIFEHLS
 DAARILREIKLLRLLRHPDIVEIKHVMLPPSRRDYRDIYVVFELMDTDLHQVI
 KANDDLTKEHHQFFLYQMLRALKYIHTANVYHRDLKPKNILANANCKLKIC
 DFLARVAFNDSPTTVFWTDYVATRWRAPELCGSFFSKYSPAIDMWSI
 GCIFAEILTgkplfpgknvvhqldlmtDLLGTPSADTISQIRNEKARRYLS
 MRRKQPIPFSEKFPNADPSALKLLQRLLAFDPKDRPTAEEALADSyFKGIA
 KVEREPSCQPIskMEFEFEQKFTKEDVKELIFQEILQYHPQLLKDYKNGSE
 KTSFLYPSAVDRFRRQFASLEENGGRNATLDRKHESLPRTTTHSTPIPAK
 EGPTATSQVAQRIPAARPGRVVGPVLPFESASITDQHIARWVARNPVPP
 ATNTSSVYCYHLKSDSSDRQGHQLEREKDRRQYRTGQHLMDAKVAPEM
 ARDIRPSQYYVSRGIPKADLTDRA TLHRSTLHGVAPFNGIERSLVGTVRPV
 FFTMELQVCTKSNALGEGNKMGNQSAARQLQLPFPMPGNSEHAXYLTN
 DAGVRKGFSGGEMACVAFIALVLMsNGDDXRANVVKDLPYAAGSEEEGG
 RQFSRVTACPFRRKGEDYPSEEKAKG

GGCCATTTGGAAAGTTCCACCTATAAACATCATTATATGATTTAAATGTTTATAACTGGA
 ACAGAACATTTCAATTGGTTAAATACAAAATTAAGCGTTGCACTGTACACCTATCCACAA
 ATTTCCCTTAGCCTTCTCTTTCAGATGGATAATCTTCCCCCTTTCTTCGAAAGGGGCAGGC
 AGTGACCCGTGAGAATTGTGACCCGCCCTCCTCTTCACTTCTGCGGCATAAGGAAGATC
 ACTTCAACACGTTGGCCCGAGATCATCCCCATTGGCTCATCAACAAGCAATGAAGGCG
 ACACAGCCATTTCCCCGCCGTAATAATCCTTTCCGGACACCTGCATCGTTGGTCAAGGAT
 GCGCGTGTCCGAATTTCCAGGCATAGGGAAGGGCAGCTGGAGCTGGCGAGCTGCACTCT
 GGTTCCTCATCTTATTACCCTCCCCAAGGGCATTGCTCTTAGTACAAACTTGTAACTCCA
 TAGTGAAGAACACCGGCTGACTGTACCCACCAGCGACCGCTGCGATGCCATTGAAGGGA
 GCAACGCCGTGACGCGTGTCTTATGGAGGGTGGCCCTGTCTGTTAGATCAGCTTTGGGA
 ATGCCCTTGGACATAGTACTGTGACGGCGTATGTCCCGGGCCATCTCAGGTGCTACC
 TTGGCATCCATGAGATGCTGCCCTGTCTATACTGCCTCCGGTCTTTCTCCCGCTCTAGC
 TGATGCCCTGCTGTGCGAGCTGTCCGACTTCAAGTGGTAGCAGTAGACTGACGAGGTG
 TTGGTTGCTGGAGGAACAGCTGGATTCCCTTGCCACCATCGTGCAATGTGTTGATCAGTG
 ATGCTAGCACTCTCGAATGGTAATACTGGGCCTACCCTCTTCCCTGGTCTAGCTGCTGGT
 ATCCTTTGAGCAACCTGGGAGGTTGCTGTTGGACCTTCTTTGCAGGAATTGGAGTGGAG
 TGAATGGTTGTTGTCTTGGGAGAGATTATGCTTCTGTCAAGCGTGGCATTCCCTCCCT
 CCATTTTCTTCAAGCTAGCAAATTCCTACGGAACCTGTCAACAGCACTGGGATATAGA
 AAGCTCGTTTTTTCAGAGCCATCTTGTAACTCTTCAAGTGGAGGATGGTACTGCAGT
 ATCTCCTGGAATATAAGTTCCTTGACATCCTCTTGGTAAACTTCTGTTTCAACTCAAAT
 TCCATTTTCGAAATGGCTGACAGGATGGTTCCTCTCCACTTTTGAATGCCTTTAAAA

TAGGAATCAGCCAACGCCTCTTCAGCAGTTGGTCGATCCTTAGGATCAAATGCTAGAAGC
CTTTGTAAGAGCTTGAGCGCCGAAGGATCTGCATTGGGGAACCTTCTCTGAAAAGGGTATT
GGCTGTTTCCTCCTCATGCTGCTCAGATACCTCCTTGCTTTCTCATTTCGAATCTGTGAA
ATAGTATCTGCTGATGGCGTGCCAGGAGATCAGTCATCAAATCCAACCTGGTGAACCACA
TTTTTACCAGGGAATAAAGGCTTCCCAGTCAAATCTCCGCAAAAATGCAACCAATGCTC
CACATATCAATAGCCGGTGAATACTTACTAAAGAAAGATCCACACAGCTCAGGAGCCCTG
TACCATCTAGTAGCAACATAGTCCGTCCAGAAAACGTGGTAGGAGAGTCATTAATGCA
ACCCTTGCTAGCCAAAATCACATATCTTGAGTTTGCAGTTAGCATTGCTAATATATTT
TTCGGCTTTAAATCACGATGATAAACATTAGCAGTATGAATATATTTCAATGCACGCAGC
ATCTGATATAGAAAGAACGGTGGTGTTCCTTTGTCAAGTCATCGTTAGCCTTAATAACT
TGGTGAAGGTCGGTATCCATGAGTTCAAATACAACGTAAATATCTCTGTAGTCCCTCCTC
GACGGAGGTAACATGACATGTTTGATCTCAACAATATCAGGATGCCTTAGCAGCCGTAGA
AGCTTGATCTCACGGAGGATCCTAGCAGCGTCGGACAGATGCTCGAAGATATTCTGTATC
TTCTTGATTGCTACCTTCTCACCGGTATGTCCGTCAATAGCTGAACAAACAACCCATAA
CTTCCTTTGCCAATGATCTCCTGAATTTTGTATCTGTTGGCATCGCCACATTGCTGAAG
AATCCATCTCTGGTGAATTTCTTCTCCTCTGCTCCGCCTGCGCCATCGCACCGGCATTT
GCCTCCGGCATCGTCAGGCCGCAGCCGCGGAAGACCTCGCGGGAGAAGTGAGAGAGCCCC
TGCTTGCTGCCTGCCGCCCAATCCGCAGGCAAGGACGACAAGGAGTGGAGGGGGAGA

MKKs

>ScMKK1 SCACLR2029C10.G (completa por molecular)
 PCSGEEPGAAAGHTSDGDATEAIKLTLPSEQETTIGKFLTHSGTFKDGDLRV
 NKDGLRIVSQSEGGAPPIEPLDSQLSLDDLDVIKVIKGNVQLVRHK
 FTGQFFALKVIQLNIDECIRKQIAKELKINLSTQCQYVVVFYQCFYVNGVISI
 VLEYMDGGSADFLKTVKTIPEDHLAAICKQVLKGLIHLHHERHIIHRDLKP
 SNILINHRGEVKISDFGVSAIISSSSGRDTFTGTYNMAPERISGKKHGYMS
 DIWSLGLVILECATGIFPFPRESLRTRACDVIALLVSPKSITVRPS

TGATGGTTTCAGATCTCTGTGTATAATGTGCCTCTCATGATGCAAGTGTATCAGTCCTTT
 TAATACCTGCTTACAAATTGCAGCAAGGTGGTCCTCTGGAATGGTTTTAACAGTCTTTAG
 GAAATCGGCAAGAGAGCCACCGTCCATGTACTCCAAAACAATAGAAATGACGCCATTGAC
 ATAAAAGCACTGATAGAACACAACAACATATTGGCACTGTGTTGATAAGTTTATTTTCAA
 CTCCTTGGCAATCTGTTTGGGTATACACTCATCAATATTTAGTTGAATAACCTTCAGAGC
 AAAAACTGGCCAGTAAATTTGTGGCGGACCAATTGCACATTTCCGCTGTTACCTTTCCC
 GATCACTTTTATAACGTCTAGATCATCTAAGCTCAGTTGACTATCCAGTGGCTCTATAGG
 AGGAGCCTCGCCTCCCTCACTCTGCGAGACGATGCGGAGGCCGTCCTTGTGACGCGCAG
 ATCCCCGTCCTTGAACGTCCCGCTATGCGTCAAGAATTTGCCGATGGTGGTCTCTTGGGA
 CGGCAGCGTGAGCTTGATTGGCTTCCGTGGCGTCGCCATCGGAAGTATGTCCGGCGGCGG
 CGCCGGGTTCTCGCCGGAGCACGGAGCAGACGCGCGGGGAAAGGAGGGGAGTGGACGG
 ACGGGCTGGTTGTGCAGCAGAGCTGGTGGTCTCCTTCGGAATCGGAAGT

>ScMKK3 SCCCL4002H05.G + SCJFRT2058E01.G
 SFRDDPASPRTATSIDAVALPGLACPLHPCAGFTPPLSSAAPPSATSSPQP
 STPLIPSVGFASSDGPGGNDRYWAHGSVGEPEGEARAAPGEERGSGGPT
 RGRSAGSDGLTHASELCASLATRLVEGERVSGLRTPKAEHTQEGSNLQR
 RAELLSRHNPRSLRDRIGGIKSTSGYCNPAAG
 ALASAHLDAGLEELKKKLQPLLFDPPDRGGASTRVPLPEETCDSYAISDD
 GTVNLLSRSLGEYNINEHGFHKRSAGPDESDESDEKGYRCASHEMHIFGPI
 GNGASSVVERAIFIPVHRILALKKINIFEKEKRQQILNEMRTLCEACCYPGLV
 EFQGAIFYMPDSGQISIALEYMDGGSADVIRVKKSIPEPVLSHMLQKVLLG
 LRYLHEVRHLVHRDIKPANLLVNLKGEAKITDFGVSTGLDNTMAMCATFVG
 TVTYMSPERIRNENYSYAADIWSLGLTILECATGKFPYDVNEGPNLMLQI
 LDDPSPTPPPDAYSPEFCSFINDCLQKDADARPMCEQLLSHPFIKRYAGT
 EVDLAAYVKSVVDPTERLQKQIAEMLAIHYLLFNGSDGIWHHMKTFFYMEQ
 STFSFSGKVYVGQNDIFDLSNIRKKLKGDRPREKIVHVVEKLHCRANDET
 GVAIRVSGSFIVGNQFLVCGDGIKAEGMPSMDELSIDIPSKRVGQFREQFI
 XATGEFNVMLLHIKARSVHYPIIEHIICKITSSMYREFG

>ScMKK4 SCQSRT2036C06.G+SCEZAM2059E11.G
 GRAAELEGAAAAVAAGHAGAGAPPPGFTLPLPQRDLTSLAVPLPLPLPP
 SSAPSSTSSSGALSAPASLGAPTPTSAGSAPANPPPLCELERVRRIGSG
 AGGTVWMVRHRPTGRPYALKVLYGNHDDAVRRQITREIAILRTAEHPAVV
 RCHGMYEQAGELQILLEYMDGGSLENHRIADEXHFLAHVARQVLSGIAYL
 HRRHIVHRDIKPSNLLIDSGRRVKIADFGVGRILNQTMDPCNSSVGTIAYM
 SPERINTDLNDRGAGDIWSFGLSILEFYLRFPPLGENLGKQGDWAALMC
 AICYSDSPQAPRNASADFKNFISLCLQKNPANRPSATRLLQHPFVAQPQT
 QQAQPQPLAAPPYDGSPPAEMHNSGRFD

ACTCCTCACCCACAGCTGTCTCAGAGTCCAACAAGCCAATGGCTCTCGTCCGCCAGCGC
CGCCAGCTGCCGCACCTCACCTCCCGCTCGACCACTTCGCCCTGCGCCTGCCGCCGAG
CCGCAGCCCACCGCCGCGCCGTCCACCTCCACCTCCGACGCGCGCCTCTCGGACTACGAG
AGGCTCTCCGTTCTCGGCCACGGCAACGGCGGCACCGTGTACAAGGCGCGGCACCGGCGG
TCCGCGCAGCCGGTCGCGCTCAAGCTCTTCGCGGACGGGGACACCTCCGCGGCGCGGAG
GCCGAGATACTCATGCTCGCCGCGGACGCGCCCCACGTTCGATGCGCC