**UNIVERSIDADE FEDERAL DE PERNAMBUCO**
**CENTRO DE CIÊNCIAS BIOLÓGICAS**
**PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS**
**DISSERTAÇAO DE MESTRADO**

MARX OLIVEIRA DE LIMA

# PROSPECÇÃO IN SILICO DE ESNAQUINAS NO TRANSCRIPTOMA DE DUAS LEGUMINOSAS DE IMPORTÂNCIA ECONÔMICA: SOJA E FEIJÃO-CAUPI

**RECIFE**
**FEVEREIRO/2014**

**MARX OLIVEIRA DE LIMA**

# PROSPECÇÃO IN SILICO DE ESNAQUINAS NO TRANSCRIPTOMA DE DUAS LEGUMINOSAS DE IMPORTÂNCIA ECONÔMICA: SOJA E FEIJÃO-CAUPI

Dissertaçao apresentada como parte dos requisitos para obtenção do título de Mestre em Ciências Biológicas, na área de concentração Biotecnologia/ Biologia Celular e Molecular.

**Orientadora:** **Drª Valesca Pandolfi**
**Co-orientadora:** **Profª Drª Ana Maria Benko-Iseppon**

**RECIFE**
**FEVEREIRO/2014**

Catalogação na Fonte:
Bibliotecário Bruno Márcio Gouveia, CRB-4/1788

Lima, Marx Oliveira de
Prospecção in silico de esnaquinas no trascriptoma de duas leguminosas de importância econômica: soja e feijão-Caupi / Marx Oliveira de Lima. – Recife: O Autor, 2014.

103 f.: il.

Orientadoras: Valesca Pandolfi, Ana Maria Benko-Iseppon
Dissertação (mestrado) – Universidade Federal de Pernambuco. Centro de Biociências. Programa de Pós-graduação em Ciências Biológicas, 2014.
Inclui referências e anexos

1. Leguminosa 2. Feijão-caupi 3. Soja 4. Peptídeos 5. Bioinformática I. Pandolfi, Valesca (orient.) II. Benko-Iseppon, Ana Maria (coorient.) III. Título.

635.65 CDD (22.ed.) UFPE/CCB-2017-065

# PROSPECÇÃO IN SILICO DE ESNAQUINAS NO TRANSCRIPTOMA DE DUAS LEGUMINOSAS DE IMPORTÂNCIA ECONÔMICA: SOJA E FEIJÃO-CAUPI

Dissertação apresentada ao curso de Mestrado do Programa de Pós Graduação em Ciências Biológicas da Universidade Federal de Pernambuco, como parte dos requisitos obrigatórios para obtenção do título de Mestre em Ciências Biológicas, na área de concentração Biotecnologia/Biologia Celular e Molecular

APROVADO EM: 14/02/2014

**COMISSÃO EXAMINADORA**

_____

Drª Valesca Pandolfi

(Orientadora - Depto. de Genética/UFPE)

_____

Profª Drª Ana Maria Benko Iseppon
(Membro Interno - Depto. de Genética – UFPE)

_____

Prof. Dr Tercilio Calsa Junior
(Membro Externo - Depto. de Genética – UFPE)

_____

Dr. Carlos Henrique Medeiros Castelletti
(Suplente Externo - Pesquisador do Instituto Agronômico de Pernambuco – IPA)

_____

Prof Dr. Ederson Akio Kido
(Suplente Interno- Depto. de Genética – UFPE)

*À minha mãe, por me ensinar*
*que a condição não condiciona,*
*impulsiona.*
*Dedico.*

## AGRADECIMENTOS

Primeiramente agradeço a Deus, Sua mãe Maria santíssima e à Santa igreja, por terem sido meu principal suporte, de quem mesmo imerecidamente recebi todo o necessário.

Às "mulheres da minha vida": mãe (Maria), tia (Neide), avó (Hosana) e irmãs (Thaís e Marina), principal incentivo e motivação de toda a minha vida, pessoas a quem devo tudo que tenho e sou.

Aos demais familiares: Thallys, Thacila, Antônio, Lenildo e em especial Sebastião, pessoas por quem tenho especial estima e apreço.

À minha orientadora Dra Valesca Pandolfi, por toda dedicação e paciência, mas agradeço principalmente por ter acreditado em mim e neste projeto. Um grande exemplo que tenho a seguir.

À Profª Ana Benko, por quem tenho grande admiração, agradeço também por todo suporte seja em forma de conselhos, incentivo ou "puxão de orelha".

Aos grandes amigos: Diego Sotero, Hévila Mendes, João Pacífico e Santelmo. Com quem tive o prazer de conviver neste período e que deram uma enorme contribuição para este trabalho.

À Universidade Federal de Pernambuco (UFPE), por meio do Programa de Pós-Graduação em Ciências Biológicas (PPGCB), pela oportunidade e suporte durante todo o período do Mestrado.

À Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE) pelo apoio financeiro durante todo o processo de Mestrado;

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo suporte através do Programa Nacional de Pós-Doutorado – PNPD/ e Auxílio Financeiro a Pprojeto Educacional e de Pesquisa - AUXPE.

Aos demais companheiros de laboratório: Bruna, Lidiane, Sheyla, Stephani, Pedro, Rômulo, Geyner, Artur e Ana Carolina. Com quem aprendi grandes valores.

Por fim agradeço aos amigos que pude reconhecer neste caminho e com quem dividi grandes alegrias: Rafael Sotero, Rafael Louzada, Jéssica Soares, Camila Pimentel, Kleber Nunes, Michel Chaves, Diego Marques, Bruna Vilela, Eduardo Lopes, Bruno Chaves, Nathalie Cortez, Elisângela Lira, Bruno Chaves, Jorge e Adriana Ferraz e Wemerson Silva.

**Muito Obrigado.**

*“Para vir de tudo ao Todo,*
*hás de deixar-te de todo em*
*tudo”*

São João da Cruz.

## RESUMO

Esnaquinas são peptídeos antimicrobianos (AMPs, do inglês *Antimicrobial peptide*) pertencentes a família Snakin/GASA, compostos de cerca de 90 resíduos de aminoácidos e 12 cisteínas conservadas, cuja expressão pode ser constitutiva ou induzida por patógenos, conforme observado em *Solanum tuberosum*. O presente estudo objetivou catalogar genes codificadores para esnaquinas nos transcriptomas de *Glycine max* e *Vigna unguiculata*, bem como analisar suas estruturas e expressão, com vistas à identificação e triagem de elementos úteis para programas de melhoramento destas culturas e posterior uso biotecnológico dos peptídeos isolados. Utilizando sequências proteicas de membros da família Snakin/GASA, realizou-se uma busca por homólogos nos transcriptomas das culturas, representadas por etiquetas de sequências expressas (EST, Expressed Sequence Tag), RNA-Seq e bibliotecas de SuperSAGE, como também realizadas análises de genômica comparativa. Um total de 48 possíveis genes codificadores para esnaquinas foram identificados, sendo 20 na soja e 28 no feijão, apresentando grande homologia estrutural com genes desta família (Snakin/GASA), distribuídas em 15 dos 20 cromossomos da soja, bem como em 6 dos 11 cromossomos de *Phaseolus vulgaris*, refletindo possíveis rearranjos genômicos que as espécies sofreram. A expressão destes genes parece estar associada ao crescimento e desenvolvimento vegetal, resposta a infecção por patógenos e estresses abióticos. As análises de expressão serial mostraram que tanto para estresses bióticos como para abióticos, a expressão destes genes reflete padrões encontrados para a família. Além disso, através da análise de similaridade foi verificada uma ampla diversidade estrutural destes peptídeos antimicrobianos nas culturas estudadas, evidenciando sua diversidade funcional, potencializando seu uso biotecnológico.



**Palavras-chave:** Peptídeos antimicrobianos. Bioinformática. GENOSOJA. NordEST. Mineração de dados.

## ABSTRACT

Snakins are AMPs - antimicrobial peptides, belonging to the Snakin/GASA family, composed of about 90 aminoacids waste and 12 conserved cysteine, whose expression may be constitutive or induced by pathogens, as observed in *Solanum tuberosum*. The present study aims to catalogue *snakin* genes in the transcriptomes of *Glycine max* and *Vigna unguiculata*, as well as analyze its structure and expression, aiming the identification and screening of useful elements intending a biotechnological use of the isolated peptides. Using protein sequences from Snakin/GASA family members, it was performed a search for homologues in the cultures transcriptomes, represented by expressed sequence tags (EST), RNA-Seq and superSAGE libraries, we also performed a comparative genomic analysis. A total of 48 possible Snakins were identified, 20 on soybean and 28 on bean, presenting great structural homology with genes from this family (Sankin/GASA), disperse into 15 of the 20 soybean chromosomes, as much as 6 of 11 *Phaseolus vulgaris* chromosomes, reflecting possible genomic rearrangements in which the species suffered. The expression of these genes showed an association with growth, plant development, and in response to infection caused by pathogens and abiotic stress. The serial analysis shows that either for biotic or abiotic stress, the gene expression reflects patterns found for the family. Trough the similarity analysis it was verified a wide structural diversity of these antimicrobial peptides on the studied cultures, showing its functional diversity and enhancing its biotechnological use.



**Keywords**: Antimicrobial peptides. Bioinformatic. GENOSOJA. NordEST. Data mining.

# LISTA DE FIGURAS

## REVISÃO BIBLIOGRÁFICA



## CAPÍTULO I

**CAPÍTULO II**

# LISTA DE TABELAS

# LISTA DE ABREVIAÇÕES

| | |
|---|---|
| ABA | Ácido Abscísico (*Abscisic Acid*) |
| AMP | Peptídeos antimicrobianos (*Antimicrobial Peptide*) |
| Avr | Gene de Avirulência |
| BLAST | Ferramenta Básica de Busca e Alinhamento Local (*Basic Local Alignment Search Tool*) |
| BR | Brassinoesteróides (*Brassinosteroids*) |
| CaMV | Vírus do Mosaico da Couve-Flor (*Cauliflower mosaic vírus*) |
| CaSN | *Capsicum annuum* snakin |
| CpSMV | Vírus do mosaico severo do caupi (*Cowpea Severe Mosaic Vírus*) |
| DDBJ | Banco de dados de DNA do Japão (*DNA Data Bank of Japan*) |
| EBI | Instituto Europeu De Bioinformática (*European Bioinformatics Institute*) |
| EMBRAPA | Empresa Brasileira de Agropecuária |
| EST | Etiquetas de Sequencias Expressas (*Expressed Sequence Tags*) |
| FaGASA | *Fragaria ananassa* GASA |
| FBCBP | *French bean chitin-binding protein* |
| FsGASA | *Fagus sylvatica* GASA |
| GA | Ácido Giberélico (*Gibberellic acid*) |
| GASA | *Gibberellin acid-stimulated arabidopsis* |
| GASA | Proteína estimulada por ácido giberélico em Arabidopsis (*Gibberellic Acid Stimulated Arabidopsis*) |
| GAST | *Gibberellic acid-stimulated transcript* |
| GEG | *Gerbera homolog of GAST1* gene |
| Gip | *GA-induced proteins* |
| GsGASA | *Glycine soja* GASA |
| GSL | Proteína estimulada por giberelina (*Gibberellin Stimulated Like*) |
| HR | Resposta Hipersensível (*Hypersensitive Response*) |
| INSDC | Banco de Dados de Sequências Nucleotídicas com Colaboração Internacional (*International Nucleotide Sequence Database Colaboration*) |
| KEGG | Enciclopédia de Genes e Genomas Kyoto (*Kyoto Encyclopedia of Genes and Genomes*) |
| LTP | Proteínas transferidoras de lipídeos (*Lipid Transfer Proteins*) |
| MBP1 | Proteína Básica do Milho (*Maize Baisc Protein*) |
| MyroBP | Proteína Ligante de Mirosinase (*Myrosinase Binding Protein*) |
| NCBI | Centro Nacional de Informações sobre Biotecnologia (*National Center for Biotechnology Binformation*) |
| OsGASR | *Oryza sativa* GA-stimulated transcript-related gene |
| PAP | Peptídeos Antimicrobianos de Plantas (*Plant Antimicrobial peptide*s) |
| PDB | Banco de Dados de proteínas (*Protein Data Bank*) |
| PIR | Informações de Recursos Proteicos (*Protein Information Resource*) |
| PR | Proteínas Relacionadas à Patogênese (*Pathogenesis Related Protein*) |
| PRC | Proteínas Ricas em Cisteínas |
| R | Genes de Resistência |
| ROS | Espécies Reativas de Oxigênio (*Reactive Oxygen Species*) |
| RSI-1 | *Root system inducible*-1 |
| SA | Ácido Salicílico (*Salicilic acid*) |
| SAGE | Análise Serial de Expressão Gênica (*Serial analysis of Gene Expression*) |
| SAR | Resposta Sistêmica Adquirida (*Systemic Acquired Resistance*) |
| StSN | *Solanum tuberosum* snakin |
| ZmGSL | *Zea mays gibberellin-stimulated like* |

# SUMÁRIO

# I. INTRODUÇÃO

No decorrer dos processos evolutivos as plantas, assim como todos os outros organismos, desenvolveram mecanismos físicos (defesa pré-existente) ou induzidos eficientes contra uma grande variedade patógenos frequentemente encontrados no meio ambiente. A cobertura serosa das folhas, a rigidez do caule das arbóreas, conferida pela lignina e até mesmo a parede celular são exemplos dessas "barreiras" contra a ação microbiana (Dangl and Jones 2001; Mysore and Ryu 2004). Contudo, alguns microrganismos ainda conseguem ultrapassar estas barreiras, podendo causar danos irreversíveis ao crescimento, desenvolvimento e produçao da planta. Em face de tal situação, um "arsenal" químico é ativado, resultando na produçao de proteinas, metabólitos secundários e, principalmente, peptídeos antimicrobianos (*antimicrobial peptide* – AMP) os quais desempenham um papel -chave na defesa vegetal, tanto de forma constitutiva quanto induzida por algum tipo de estresse, seja ele biótico ou abiótico (García-Olmedo et al. 1998).

Devido a sua importância na defesa da planta, o repertório de peptídeos antimicrobianos sintetizados é extremamente amplo, com centenas de classes de AMPs distribuídos em diferentes famílias em algumas espécies vegetais, incluindo defensinas, tioninas, proteínas trasnferidoras de lipídeos (LTP, *Lipid-Transfer Proteins*), ciclotídeos, heveinas (hevein-like proteins) e esnaquinas (Silverstein et al. 2007; Maróti et al. 2011). Em plantas, as famílias de AMPs foram descritas de acordo com diversos critérios estruturais, tais como: abundância e arranjo dos diferentes aminoácidos, formação de estruturas terciárias, carga, entre outros (Bulet and Stöcklin 2005; Padovan et al. 2010).

Dentre os AMPs ricos em aminoácidos "cisteínas", as esnaquinas, são um grupo que, apesar do seu recente isolamento e caracterização, vem recebendo crescente destaque (Segura et al. 1999; Berrocal-lobo et al. 2002; Mao et al. 2011), principalmente pela já evidenciada gama de funções, tais como na defesa contra diversos fungos e bactérias patogênicos (Almasia et al. 2008; Faccio et al. 2011), como na defesa contra fatores abióticos, principalmente aqueles que interferem na homeostase dos vegetais (Nahirñak et al. 2012b), assim como também participam nos mecanismos das funções vitais da planta, como crescimento e desenvolvimento (Nahirñak et al. 2012a).

Tais propriedades fazem destes AMPs uma alternativa no desenvolvimento de soluções biotecnológicas que possam servir tanto no desenvolvimento de potenciais agentes terapêuticos, como no desenvolvimento de estratégias ou substituir aquelas já existentes em

programas melhoramento de culturas economicamente importantes, tornando as atividades agrícolas mais sustentáveis e eficientes (Keymanesh et al. 2009).

Diante de tais esforços, a busca e, principalmente, a caracterização de novas sequências de genes desta natureza, a bioinformática surge como um instrumento indispensável, uma vez que suas ferramentas englobam conceitos e aplicações de diversas áreas de conhecimento (incluindo a biologia, química, estatística e ciências da computação), tornando possível um estudo cada vez mais interligado com as informações acerca de estrutura, localização, regulação e expressão dos diversos genes e de seus produtos (Edwards and Batley 2004; Droit et al. 2005).

Neste contexto, não diferente para outros estudos, a bioinformática vem provendo resultados plausíveis na prospecção e caracterização de alguns AMPs (Belarmino et al. 2010; Porto et al. 2012) justificando-se, assim, sua aplicabilidade em análises de esnaquinas em culturas de interesse sócio-econômico, a exemplo de algumas leguminosas, como soja e feijão-caupi.

A soja cultivada [*Glycine max* (L.) Merr.] é hoje o principal produto do agronegócio brasileiro, ocupando grandes áreas para seu cultivo e, consequentemente, demandando uma considerável quantidade de mão-de-obra para o seu manejo, gerando assim, milhares de empregos nas mais diversas regiões do país. A considerável importância econômica, aliada ao seu grande valor nutricional, tem tornado a soja um alvo potencial para exploração biotecnológica (EMBRAPA, 2014).

Outra cultura que merece destaque é feijão-caupi (*Vigna unguiculata*), uma das leguminosas que obteve maior sucesso no cultivo em regiões tropicais e subtropicais, principalmente por apresentar características importantes para a agricultura, tais como maior tolerância diversos fatores ambientais, bem como sua participação no combate a erosão, contribuindo para a melhoria da fertilidade do solo e colaborando para a fixação de nitrogênio, através da associação com bactérias fixadoras (Valenzuela and Smith, 2002).

No Brasil (terceiro maior produtor), o cultivo do feijão-caupi predomina nas regiões Norte e Nordeste, representando 95 a 100% do total das áreas ocupadas nessas regiões. Apesar do aumento da área plantada, principalmente nos últimos anos, diversos fatores abióticos como seca e salinidade, além de fatores bióticos (ataque de pragas e fitopatógenos) têm provocado oscilações na produção total do feijão-caupi em nosso país (RODRIGUES, 2012).

O presente trabalho visou identificar, caracterizar e analisar estruturalmente, através de ferramentas computacionais, sequências de esnaquinas e seus homólogos presentes no genoma expresso da soja e do feijão-caupi. Com as informações geradas será possível

identificar potenciais candidatos no desenvolvimento tanto de novos agentes antimicrobianos (fármacos), como na geração de plantas com maior resistência e/ou tolerância a estresses bióticos e/ou abióticos, durante o processo do melhoramento genético vegetal.

# II. OBJETIVOS

## 2.1. Objetivo Geral:

- Identificar e caracterizar a estrutura dos ortólogos de esnaquinas bem como o padrão de expressão desses genes na soja e no feijão-caupi, bem como estruturar o banco de dados local de AMPs ricos em cisteínas.

## 2.2. Objetivos Específicos:

- Selecionar as principais esnaquinas na literatura e em bancos de nucleotídeos e de proteínas para a identificação de genes homólogos nos transcriptomas das culturas em questão;
- Identificar e caracterizar genes candidatos a esnaquinas no banco de dados públicos e privados, bem como caracterizar e descrever regiões conservadas (domínios e motivos) nas sequências identificadas;
- Estabelecer um perfil da expressão *in silico* dos genes analisados, nos diferentes tecidos, considerando-se a montagem do banco de ESTs das culturas;
- Relacionar o padrão estrutural de esnaquinas com a evolução desses genes em plantas superiores;
- Auxiliar na formação do banco de dados local de AMPs ricos em cisteína, contribuindo com a consolidação do grupo PAP (*Plant Antimicrobial Peptides*).

# III. REVISÃO BIBLIOGRÁFICA

## 3.1. Cultura da SOJA

### 3.1.1. Importância Socio-econômica

Devido às tendências no mercado internacional, o Brasil passou a interessar-se pela cultura da soja como um produto comercial no final da década de 60. Em meados de 1970 os aumentos significativos na produção (refletidos tanto pelo aumento na área cultivada, como na produtividade) e explosão do preço da saca de grãos, a cultura da soja foi consolidada como a principal cultura do agronegócio, despertando ainda mais o interesse dos agricultores brasileiros (Pinto, 2005; EMBRAPA, 2009;). Aliado ao interesse comercial, a soja surge como um dos grãos mais consumidos globalmente, principalmente pela sua reconhecida importância nutricional (humana e animal), servindo como uma valiosa fonte de proteínas, minerais, vitaminas e fibras, disponibilizados na forma de produtos como massas, produtos de carne, cereais, misturas preparadas, bebidas, alimentação para bebês, alimentos dietéticos e ração animal, dentre outros (Kawaga, 1995).

O pontencial e versatilidade dos seus produtos e subprodutos fizeram desta leguminosa uma cultura amplamente utilizada, tanto na agroindústria, como na indústria química e de alimentos. Embora seu uso mais difundido seja na forma de óleo refinado na alimentação, na confecção de antibióticos, desinfetantes, isolação elétrica, inseticidas, cosméticos e pigmentos. A casca da soja também se destaca como uma importante forragem, além, é claro, de seu uso como componente de fibras dietéticas e cereais matinais (Pinto, 2005; Valentini, 2007). Adicionalmente, a soja vem crescendo, também, como fonte alternativa de combustível, na forma de biodiesel, o qual está sendo testado por diversas instituições de pesquisa, como a Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) em diferentes cidades brasileiras (EMBRAPA, 2014).

Na safra 2011-2012 o Brasil chegou à marca produtiva de 74 milhões de toneladas de soja, ocupando a segunda posição no *ranking* mundial dos maiores produtores do grão (ou seja, 10 milhoes de ton menos que os Estados Unidos da América - o primeiro produtor), até o ano de 2017, este cenário deverá ser invertido, consolidando o Brasil como maior produtor e exportador mundial, retendo 34 % de toda a produção de soja, seguido pelos EUA com 30 % e Argentina com 20 % (EMBRAPA, 2014).

### 3.1.2. Principais doenças que acomentem a cultura da soja

Dentre os principais fatores limitantes de rendimento da cultura da soja, as doenças são os mais importantes e de difícil controle. Uma importante causa da incidência de doenças tem sido atribuída ao uso de sementes contaminadas, provenientes de diferentes áreas de produção, uma vez que a maioria dos patógenos da soja é transmitida via sementes (Yorinori, 1990). Estudos apontam que a soja seja afetada por, pelo menos, 100 tipos diferentes de doenças, incluindo várias economicamente importantes. Cerca de 40 doenças causadas por fungos, bactérias, nematoides e vírus afetam a cultura no Brasil, resultando em perdas anuais de, aproximadamente, US$ 2 bilhões (Yorinori, 1990).

Entre as principais doenças causadas por fungos fitopatogênicos destacam-se a mancha "olho-de-rã", causado pelo fungo *Cercospora sojina* Hara; o cancro da haste, causado pelo fungo *Diaporthe phaseolorum* (Cke. & Ell) e a podridão parda da haste [*Phialophora gragata* (Allington & Chamberlain) W. Gams], cujos danos foram mitigados devido ao desenvolvimento de cultivares resistentes. A incidência da podridão vermelha da raiz, causada pelo *Fusarium solani* f.sp. *glycines* (App. & Wollenw.) Snyd. & Hans também mostra-se preocupante na cultura. O oídio (*Microsphaera diffusa* Cke. & Pk.), o complexo de doenças foliares de final de ciclo (*Septoria glycines* Hemmi, *Cercospora kirkuchii* Matsu. & Tomoyasu) e a ferrugem (*Phakopsora pachyrhizi* Sdy. & P. Sdy.) causam desfolha acentuada, causando redução da qualidade e da capacidade de germinação das sementes (Almeida et al., 1997).

Doenças causadas pela infecção do nematoide do cisto da soja (*Heterodera glycines* Ichinohe), bem como do nematoide de galhas [*Meloidogyne javanica* (Treub) Chitwood e *M. incognita* (Kofoid & White) Chitwood] são consideradas uma das mais sérias limitações em determinadas áreas, com poucas cultivares resistentes identificadas até o momento (Almeida et al., 1997).

Exemplos de doenças bacterianas incluem o crestamento bacteriano [*Pseudomonas syringae* pv. *glycinea* (Coeper) Yound, Dyl, e Wilkie.], o fogo selvagem [*P. syringae* pv. *tabaci* (Wolf e Foster) Stevens], a pústula bacteriana [*Xanthomonas axonopodis* pv. *glycines* (Nakano) Dye] e a murcha-de-curtobacterium, causada por *Curtobacterium flaccumfaciens* pv. *flaccumfaciens* (Hedges) Collins & Jones. O principal controle dessas doenças se faz pelo uso de cultivares resistentes, além do uso de sementes sadias e o bom preparo do solo (Sediyama et al., 1996; Almeida et al., 1997).

### 3.1.3. Aspectos genômicos da Soja

A família Fabaceae, também conhecida como Leguminosae, é composta por três subfamílias: Mimosoideae, Caesalpinioideae e Papilionoideae. Esta última (Papilionoideae) a maior das três subfamílias de leguminosas inclui quase todas as espécies economicamente importantes, como a ervilha (*Pisum sativum*), a alfafa (*Medicago truncatula e M. sativa*), feijão comum (*Phaseolus vulgaris*), o feijão-caupi (*Vigna unguiculata*) e, a soja cultivada (Joly, 1975), uma espécie de autofecundação, anual e diploide (2n = 40) (Shoemaker, Schlueter and Jackson, 2008).

Embora existam estimativas de que 80-100 % de todas as angiospermas possuam uma provável origem poliploide (Lockton; Gaut 2005), dados gerados pelo grande avanço no sequenciamento de genomas tem demonstrando que, mesmo espécies atualmente consideradas diploides contêm indícios de poliploidia, ou pelo menos grandes duplicações segmentais, prováveis indícios de eventos ancestrais de poliploidização (Vision, Brown and Tanksley, 2000). Neste contexto, não é de surpreender que também a soja possua um ancestral nestas mesmas condições.

Estudos baseados em mapeamento mostraram que o genoma da soja é um provável paleopoliploide, com supostas regiões homeólogas facilmente identificáveis através de mapeamento com base em hibridização. Duplicações observadas nos mapas físicos indicam que o genoma pode ter passado por dois ou mais eventos de duplicação em grande escala (Shoemaker et al., 1996). De um modo geral, *G. max* possui alto grau de conservação genômica, característica compartilhada por todos os membros da subfamília Papilionoideae, corroborada pela presença de blocos sintênicos compartilhados entre os mesmos, dos quais a soja compartilha em maior grau com *Phaseolus vulgaris* e *Vigna radiata* (Shoemaker et al. 2006).

## 3.2. A cultura do FEIJÃO-CAUPI

### 3.2.1. Importância sócio-econômica

O feijão-caupi (*Vigna unguiculata* (L.) Walp.) também conhecido como feijão-de-corda, é uma eudicotiledônea pertencente à família Fabaceae e à ordem *Fabales*, sendo cultivada desde áreas tropicais e subtropicais de baixa altitude até altitudes acima de 1.300 m, chegando a 1.600 m no Kênia e Camarões (Ehlers and Hall, 1997).

Na região Nordeste, sua maior produtora, o feijão-caupi é a segunda maior cultura agrícola, representando 60 % da área total cultivada no Brasil, podendo ser encontrado do litoral até o sertão. A produção desta leguminosa chegou a 429.375 toneladas, alcançando uma produtividade de 303,5 kg/ha (Freire-Filho et al, 1999).

Trata-se de uma das leguminosas mais importantes para a alimentação humana, principalmente como fornecedora de proteínas, carboidratos, cálcio, ferro e vitaminas, principalmente para a população de baixa renda na África (Nigéria e Niger) e América do Sul (incluindo o Brasil), os maiores produtores e consumidores mundiais (Freire-Filho et al., 1999; FROTA et al., 2009). Sendo esta uma espécie de grande valor estratégico, possuindo um alto potencial produtivo, aliado a uma ampla variabilidade genética e capacidade de adaptação, representando uma fonte de genes para a obtenção de cultivares resistentes/ tolerantes nesta, bem como em outras culturas (Ehlers; Hall, 1997; Freire-Filho et al., 1999).

### 3.2.2. Principais doenças que acometem a cultura do caupi

Apesar de apresentar grande rusticidade, o feijão-caupi é hospedeiro de muitas doenças (causadas por vírus, bactérias, fungos, nematoides, além de sofrer a ação, também de insetos fitófagos), as quais são responsáveis por boa parte das perdas na produtividade. Entre os estresses bióticos, o ataque por vírus apresenta-se como um dos fatores mais amplamente limitantes à sustentabilidade da cultura. Em especial está o vírus do mosaico severo do caupi (CpSMV, *Cowpea Severe Mosaic Vírus*) cuja incidência se estende por toda a América Latina, provocando perdas, em decorrência desta doença, de até 70% na produção brasileira (Rodrigues, 2011).

Doenças causadas por fungos também representam riscos na produção do caupi, provocando danos significativos às plantações. A ferrugem, causada pelo fungo *Uromyces appendiculatus;* a cercosporiose (*Mycosphaerella cruenta*), a sarna (*Elsinoe phaseoli*) e a mela

(*Thanatephorus cucumeris*) são exemplos de doenças fúngicas que igualmente acometem a cultura do caupi. Além desses patógenos, há também aquelas doenças causadas por nematoides, o caso da meloidoginose, causada por membros do gênero *Meloidogyne* (EMBRAPA, 2014). Contudo, há poucas fontes de resistência para as doenças existentes, estando essas frequentemente em diferentes cultivares, ou em variáveis que não atendem às necessidades do mercado (Freire-Filho, 2008).

### 3.3. Mecanismos de defesa vegetal

#### 3.3.1. Resposta hipersensível (HR) e Resistência Sistêmica Adquirida (SAR)

Sendo incapazes de evitar os estresses ambientais fugindo ou alterando fisicamente o ambiente em que vivem, os vegetais respondem às pressões do meio sendo ágeis bioquimicamente, visto que muitas vezes suas barreiras físicas, como a rigidez do caule das lenhosas e a camada serosa das folhas, não são suficientemente eficientes para conter o ataque de certos patógenos ou pragas. Neste sentido, a sobrevivência das plantas muitas vezes depende da agilidade com que respondem aos estímulos externos.

Dessa forma, as plantas estão frequentemente "ajustando" seu metabolismo de acordo com as condições adversas impostas pelo ambiente externo, as quais podem limitar o seu crescimento e desenvolvimento. A resposta dos vegetais frente ao estresse biótico, por exemplo, deve ser eficiente ao ponto de impedir a disseminação do invasor, assim como impedir invasões posteriores. Algumas respostas se limitam ao sítio onde ocorreu a lesão, enquanto que outras ocorrem de forma sistêmica (Freitas and Bered, 2003).

Um dos mecanismos mais eficientes de resposta envolve uma cascata de eventos, e de sinais, os quais proporcionam a defesa da planta frente ao patógeno. Tal mecanismo é conhecido como "*Resposta Hipersensível*" (*Hypersensitive Response,* HR). A resposta, neste caso, é iniciada por um reconhecimento designado "interação gene-a-gene", o qual envolve o reconhecimento específico do patógeno pela planta, levando a um dano celular local (apoptose – morte local de células), assim como na ativação de sinais intermediários que, por sua vez, acionam a via recepção-transdução, levando à regulação gênica responsável pela resistência observada (Freitas and Bered, 2003; Loebenstein, 2010).

Produtos de um conjunto de genes da resposta hipersensível, normalmente precedem a ativação de um processo sinalizador, tornando a planta mais refratária ou resistente a infecções subsequentes a uma gama de patógenos constituindo, assim, a chamada *"Resistência Sistêmica*

*Adquirida"* (*System Acquired Resistance,* SAR). O produto da ativação de muitos desses genes desempenha um papel fundamental no controle do crescimento do patógeno, tanto por reforçar a capacidade de defesa das células hospedeiras, como também por codificar enzimas e metabólitos secundários com atividade antimicrobiana (Freitas and Bered, 2003).

### 3.3.2. Genes R e Genes PR

Algumas vezes esses genes HR codificam enzimas ou proteínas que participam de uma via metabólica específica e, em outras, são os genes que codificam diferentes produtos estruturais ativados. Nestes dois principais grupos gênicos - responsáveis pelos mecanismos de defesa - são encontrados os "*genes de resistência*" – genes *R* (*Resistance, R*) e os genes que codificam as proteínas "*relacionadas á patogenicidade*" – genes *PR* **(***Pathogenesis Related, PR*) (Liu et al., 2009).

Os genes *R* determinam o reconhecimento de patógenos específicos, que expressam um gene de avirulência (*avr*) correspondente. Mais de 20 genes *R* com especificidade de reconhecimento a genes *avr* já foram isolados e caracterizados, tanto em monocotiledôneas quanto em eudicotiledôneas. Uma vez que estão diretamente envolvidas no reconhecimento de proteínas *avr*, as proteínas R se localizam preferencialmente junto à membrana plasmática, interceptando as proteínas de origem patogênica. Muitas dessas proteínas também apresentam um papel secundário de sinalização, ou seja, ativando outros genes de defesa. A despeito da grande diferença entre os modos de ação e invasão dos patógenos, os domínios funcionais das proteínas R são bastante conservados (Freitas and Bered, 2003).

Por outro lado, as proteínas PRs são espécie-específicas, isto é, a indução pode resultar da infecção por patógenos ou, artificialmente, em plantas sadias, por tratamento com produtos químicos; altas concentrações de hormônios vegetais; desordens fisiológicas; ou filtrados tóxicos de fungos fitopatogênicos ou de bactérias. Várias famílias de proteínas PR são induzidas durante a resposta hipersensível (HR) e fazem parte do conjunto de respostas que caracterizam a Resistência Sistêmica Adquirida (Freitas and Bered, 2003).

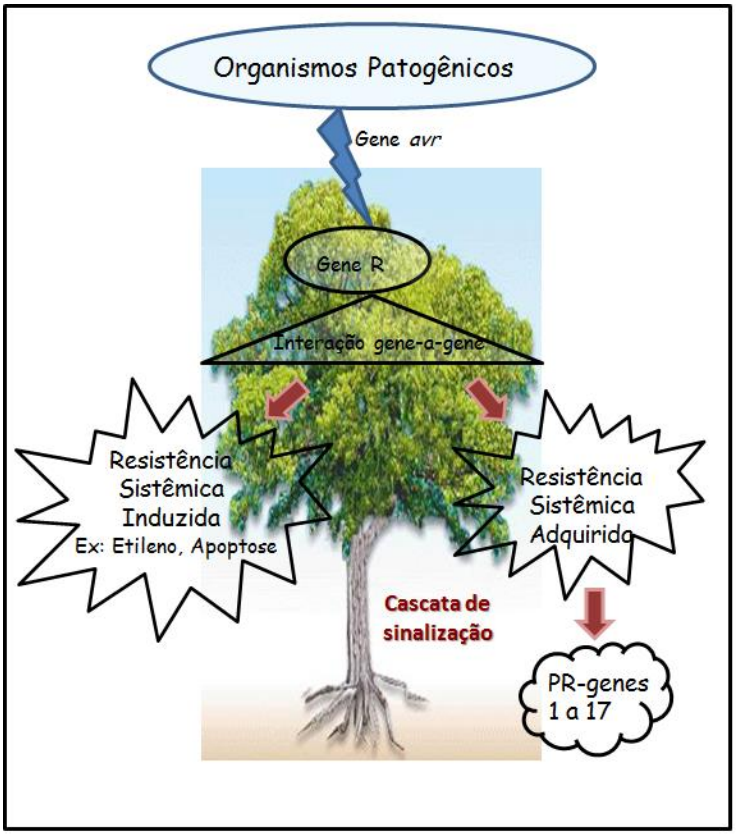



**Figura 1 -** Mecanismos de reconhecimento e resposta da interação planta/patógeno. Organismos patogênicos secretam produtos de genes *avr*, as plantas por sua vez se defendem com genes *R*, estas interações conduzem a uma cascata de eventos (SAR) que ativam os genes *PR* (Adaptado de Benko-Iseppon et al. 2010).

As proteínas PRs apresentam baixo peso molecular (8-100 kDa), localizadas, preferencialmente, nos espaços intercelulares. Estudos de genes *PR* têm demonstrado que as proteínas são sintetizadas com o um grande precursor contendo um peptídeo sinal de aproximadamente 30 resíduos de aminoácidos (Van Loon and Van Strien, 1999). Sào classificadas em famílias com base nas suas sequências de aminoácidos, relações sorológicas e/ou atividade enzimática ou biológica (Van LOON *et al*., 1994).

Até o momento, são conhecidos 17 grupos distintos de proteínas PRs (Van Loon and Van Strien, 1999), sendo a maioria com propriedades bioquímicas intrínsecas ao grupo das hidrolases. Contudo, há outros grupos como as quitinases, laminarinases e aquelas pertencentes a um grupo mais característico - o grupo dos peptídeos antimicrobianos, representados pelas

defensinas (PR-12), as tioninas (PR-13) e as LTPs (PR-14) (Van LOON and Van STRIEN, 1999).

### 3.3.3. Peptídeos Antimicrobianos – AMPs

Uma classe de moléculas que tem recebido uma notável atenção na defesa de plantas são os peptídeos antimicrobianos (*antimicrobial peptide* – AMPs), uma vez que fazem parte do mecanismo constitutivo das defesas vegetal, assim como também induzidos por estresses bióticos e/ou abióticos (García-Olmedo et al., 1998; Benko-Iseppon et al., 2010a).

Os AMPs são caracterizados, principalmente por seu equeno tamanho (em média de 20 a 100 aminoácidos), pela riqueza em aminoácidos do tipo prolina, glicina, histidina, arginina, triptofano ou cisteínas – PRC (Proteínas Ricas em Cisteína) (Bulet *et al*., 2004). Neste último caso (PRC), os resíduos de cisteínas são estabilizados por duas a seis pontes dissulfeto, com cisteínas conservadas também em posição; apresentam carga positiva e um peptídeo sinal (García-Olmedo et al. 1998; Silverstein et al. 2007) (Figura 2).

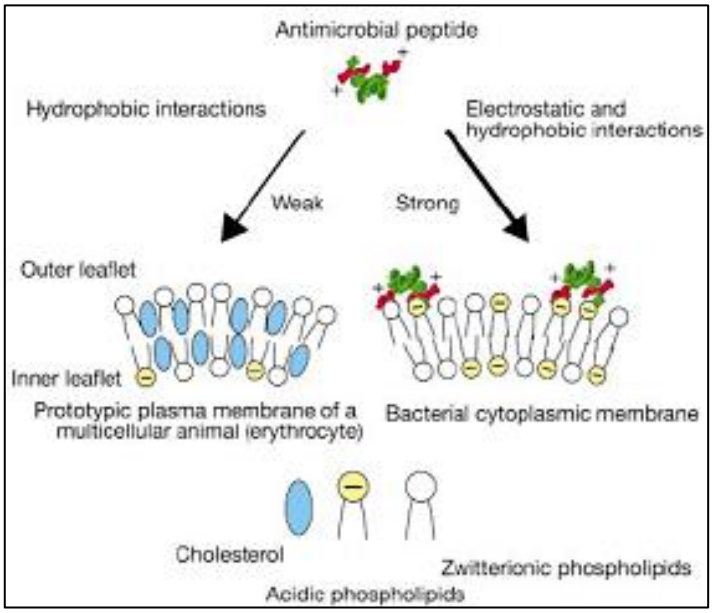



**Figura 2.** Afinidade por membranas dos peptídeos antimicrobianos e seu modo de interação (Zasloff 2002).

Devida a sua importância na defesa vegetal, a expressão de AMPs tem sido amplamente estudada nos mais variados tecidos vegetais, tais como: raízes, folhas, sementes e flores (Lay et al., 2003; Pelegrini et al., 2007). Tais peptídeos podem ser divididos em famílias de proteínas distintas. Estas incluem defensinas (Broekaert et al., 1995), peptídeos *knottin-like*, proteína básica de milho (*Maize Basic Protein* - MBP1) (Duvick et al., 1992), as proteínas transportadoras de lipídios (*Lipid Transfer Protein* - LTP), peptídeos *Hevein-like*, as proteínas ligantes de mirosinase (*Myrosinase Binding Protein* - MyroBP), esnaquinas (*snakins*) (Segura et al., 1999 & Berrocal-Lobo et al., 2002), ciclotídeos, albuminas 2S, entre outros (Manners, 2007, Benko-Iseppon et al., 2010).

Embora sua principal função relatada seja na defesa vegetal contra agentes patogênicos, muitas proteínas ricas em cisteína (PRC) podem agir, também, em processos relacionados com o desenvolvimento da planta, como regulação do crescimento, desenvolvimento de flores e sementes, assim como em resposta a fatores hormonais (Aubert. 1998; Tavares et al. 2008).

#### 3.3.3.1. Esnaquinas (*snakin*)

Entre os peptídeos antimicrobianos mais recentemente descobertos, encontram-se as Esnaquinas (*snakin*). Estes peptídeos receberam esta nomenclatura devido a uma pequena similaridade, de apenas três aminoácidos, com a peçonha de uma serpente (a *desitegrin-like)* (Segura et al. 1999). As esnaquinas também são conhecidas como *gibberellin stimulated-like proteins* (GSL) - peptídeos ricos em cisteínas e identificados pela primeira vez em batata (*Solanum tuberosum*). Sua sequencia de aminoácidos apresenta similaridade com membros da família GASA (*gibberellic acid stimulated in arabidopsis*) sendo, por isso, classificadas como integrantes da família *snakin/GASA* (Nahirñak et al. 2012b).

Estes genes codificam peptídeos com uma sequencia sinal, o qual serve para seu direcionamento subcelular; uma região variável, e um peptídeo maduro, o qual apresenta uma média de 60 aminoácidos, com 12 cisteínas altamente conservadas. Tais aminoácidos auxiliam na manutenção da estrutura tridimensional do peptídeo, além de fornecer estabilidade a molécula quando a planta está sob situação de estresse (Porto and Franco 2013; Meiyalaghan et al. 2014) (Figura 3)

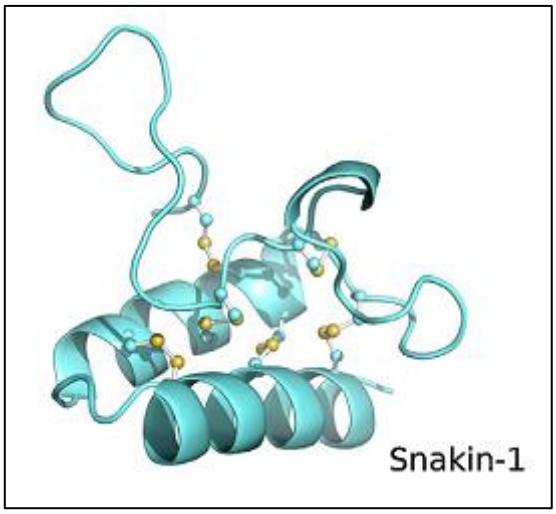



**Figura 3 -** Estrutura teórica e padrão de pontes dissulfeto no peptídeo *snakin-1* de *S. tuberosum*. (PORTO; FRANCO, 2013).

Estudos de expressão evidenciaram a presença de transcritos de esnaquinas nos mais variados tecidos (caule, folhas, flores, sementes e raízes) (Zhang et al. 2009; Zimmermann et al. 2010; Almasia et al. 2010; Guzmán-rodríguez et al. 2013), induzidos por estresses bióticos ou abióticos, principalmente com atividade *in vitro* contra uma variedade de fungos, bactérias e nematoides (Segura et al. 1999; Faccio et al. 2011; Mao et al. 2011). Apesar disso, muitos homólogos da família já tiveram sua expressão relatada em processos biológicos como: divisão, elongação e crescimento celular, floração, embriogênese e sinalização de vias (Aubert et al. 1998; Kotilainen et al. 1999; Ben-Nissan et al. 2004; de la Fuente et al. 2006; Furukawa et al. 2006; Roxrud et al. 2007; Lucau-Danila et al. 2010).

Curiosamente, mais recentemente a atividade apresentada em um homólogo da família (*snakin-Z* de *Ziziphus jujuba*), chamou atenção por conta do seu possível uso no tratamento da doença de Alzheimer em humanos. Juntamente com seu grande potencial como antioxidante, aliada a ótima atividade como inibidor de colinestares, estudos mais esopecíficos são necessários para confirmação de tal propriedade (Zare-zardini et al. 2013)

Assim, novas PRCs estão sendo cada vez mais identificadas e caracterizadas em organismos vegetais, embora estima-se que o número de representantes desta classe esteja muito aquém do que se tem conhecimento, fato este, já demonstrado para as defensinas (Silverstein et al. 2005).

A dificuldade na identificação de novos peptídeos antimicrobianos consiste, em boa parte, devido as características estruturais compartilhadas entre eles, por constituírem moléculas pequenas e pela grande variação de sua estrutura primária. Entretanto, o grande

número de bancos de dados biológicos atualmente disponíveis para o domínio público, aliado a rapidez com que novas ferramentas de bioinformática estão sendo geradas, as análises envolvendo a identificação, caracterização e função de novos AMPs têm sido cada vez mais rápidas e eficientes (Silverstein et al. 2007).

Com base nisso, a aplicação de AMPs tem mostrado resultados promissores tanto na área animal como vegetal onde, através do uso da biotecnologia e engenharia genética, é possível a geração de novas variedades geneticamente modificadas com tais peptídeos e, portanto, servindo como uma alternativa na redução das perdas agrícolas devido ao ataque de patógenos (Sarika et al. 2012).

### 3.4. Banco de dados biológicos

Considerando o volume de dados, cada vez maior e levando-se em consideração a premissa de que tais informações poderiam e deveriam tornar-se acessíveis a todos, teve inicio a construção de bancos de dados públicos e privados, bem como redes de acesso, as quais permitissem o fluxo contínuo de informação entre tais bases. Assim, um banco de dados pode ser definido como uma coleção compartilhada de dados logicamente relacionados, projetado para atender as necessidades de informação de múltiplos usuários (Edwards; Batley, 2004).

Atualmente os principais bancos de dados são o *GenBank*, do Centro Nacional de Informações sobre Biotecnologia – NCBI (*National Center for Biotechnology Binformation*) dos EUA, que comporta sequências de DNA e proteínas; o Instituto Europeu de Bioinformática – EBI (*European Bioinformatics Institute*); o Banco de Dados de DNA do Japão – DDBJ (*DNA Data Bank of Japan*), o Banco de Dados de Proteínas – PDB (*Protein Data Bank*), o qual comporta também informações a cerca da estrutura tridimensional das proteínas; o Recursos de Informações Proteicas - PIR (*Protein Information Resouce*) e a Enciclopédia de Genes e Genomas de Kyoto – KEGG (*Kyoto Encyclopedia of genes and Genomes*), o qual contém informações sobre sequências genômicas de vários organismos, bem como informações relacionadas as suas respectivas vias metabólicas (Prosdoscini, 2002).

Há também o Banco de Dados de Sequências Nucleotídicas com Colaboração Internacional – INSDC (*International Nucleotide Sequence Database Colaboration*), o qual integra o NCBI, EBI e DDBJ. Neste banco (INSDC) é possível encontrar informações a cerca de genes individuais, genomas completos, RNAs, realizar anotações, ESTs, além de sequências sintéticas. Devido à sua designação como sendo um provedor de dados primários, banco

EBI/DDBJ/NCBI tem se tornado a fonte inicial de muitos bancos de dados em biologia molecular (NCBI, 2014).

Em 2004, Wang e Wang, contruiram um banco de dados dedicado exclusivamente à peptídeos antimicrobianos, contudo este banco compilava informações e sequencias de vários tipos de organismos. Foi só em 2009 que Hammami e colaboladores estruturaram o primeiro banco de peptídeos antimicrobianos de plantas, contendo diversas informações sobre 271 peptídeos vegetais.

### 3.4.1. Banco de soja e de feijão-caupi: Uma iniciativa brasileira

#### 3.4.1.1. O Consórcio GENOSOJA

Em 2007, um grupo de instituições brasileiras públicas e privadas iniciou o consórcio GENOSOJA (Consórcio Nacional para Estudos do Genoma da Soja) onde, através de estudos de genômica estrutural e funcional da soja, visou gerar informações que subsidiem um aprimoramento do processo de produção da cultura de soja, enfatizando as análises nos fatores limitantes da produção nacional, como a ocorrência de seca, pragas e patógenos. Tal consórcio consistiu em unir esforços, conduzidos anteriormente de forma individual, de diferentes grupos de pesquisa no Brasil.

Entre as metas do GENOSOJA, está a compreensão da função e dos mecanismos de controle da expressão de genes presentes na soja e envolvidos em diferentes processos de desenvolvimento e/ou defesa contra os mais diversos tipos de estresse. Tais estudos têm facilitado a análise de dados gênicos e de expressão, permitindo inclusive a integração da informação proveniente de várias fontes relacionadas, como a transcriptômica, a proteômica, a metabolômica e a fenômica. Os dados gerados (provenientes de bibliotecas de ESTs, *tags* SuperSAGE, bibliotecas subtrativas, RNA Seq, MicroRNAs, além de dados genômicos) são disponibilizados num banco de dados local, na plataforma GENOSOJA (http://bioinfo03.ibi.unicamp.br/soja/), e em bancos de dados públicos, a exemplo do GenBank, do NCBI. Juntamente com o banco de ESTs do NCBI, o banco conta com quase 1,5 milhão de ESTs, formando 60.747 unigenes, distribuídos entre 30.809 contigs e 29.938 singlets. Já, os dados provenientes das bibliotecas de SuperSAGE permitiram a obtenção de mais de 4 milhoes de tags (Benko-Iseppon et al., 2012; Nascimento et al., 2012).

### 3.4.1.2. O Consórcio NordEST

Em 2005 foi estabelecida a Rede NordEST, com a proposta de efetuar uma análise genômica funcional e estrutural no feijão-caupi (*V. unguiculata*), visando identificar genes candidatos potencialmente úteis para fins de melhoramento desta cultura. No âmbito do referido projeto foram utilizadas ferramentas de genômica expressa (EST, Super/LongSAGE e RNASeq) relacionadas a estresses bióticos (Potyvírus e Mosaico Severo) e abióticos (salinidade e seca) do feijão-caupi. Esta rede contou com a participação de 12 laboratórios parceiros de várias regiões do Brasil sob coordenação da UFPE. O projeto contou, ainda, com a colaboração da Universidade de Frankfurt (Johann Wolfgang Goethe Universität) e da empresa GenXPro ambos na Alemanha, e da Universidade da Califórnia, em Riverside (EUA).

Atualmente o banco local conta com 180.000 ESTs (sendo 27.453 sequências de EST geradas pelo Projeto NordEST, o restante obtidas de bancos públicos como NCBI e Harvest), 21 milhões de sequências SuperSAGE e quase meio milhão de sequencias de RNA-Seq, todas associadas à expressão de genes de resistência a fatores bióticos e abiótico (Benko-Iseppon, comunicação pessoal).

# IV. REFERÊNCIAS


Agizzio, A.P. et al. (2006) The antifungal properties of a 2S albumin homologous protein from passion fruit seeds involve plasma membrane permeabilization and ultrastructural alterations in yeast cells. Plant Sci, v.171, p. 515–522.

Almasia NI, Bazzini AA, Hopp HE, Vazquez-Rovere C (2008) Overexpression of snakin-1 gene enhances resistance to Rhizoctonia solani and Erwinia carotovora in transgenic potato plants. Mol Plant Pathol 9:329–338. doi: 10.1111/J.1364-3703.2008.00469.X

Almasia NI, Narhirñak V, Hopp HE, Vazquez-Rovere C (2010) Isolation and characterization of the tissue and development-specific potato snakin-1 promoter inducible by temperature and wounding. Electron J Biotechnol 13:1–21. doi: 10.2225/vol13-issue5-fulltext-12

Alonso-Ramírez A, Rodríguez D, Reyes D, et al (2009a) Evidence for a Role of Gibberellins in Salicylic Acid-Modulated Early Plant Responses to Abiotic Stress in Arabidopsis Seeds 1. Plant Physiol 150:1335–1344. doi: 10.1104/pp.109.139352

Alonso-Ramírez A, Rodríguez D, Reyes D, et al (2009b) Cross-talk between gibberellins and salicylic acid in early stress responses in Arabidopsis thaliana seeds. Plant Signal Behav 4:750–1. doi: 10.1104/pp.109.139352

Altschul SF, Gish W, Miller W, et al (1990) Basic local alignment search tool. J Mol Biol 215:403–10. doi: 10.1016/S0022-2836(05)80360-2

Aubert D, Chevillard M, Dorne a M, et al (1998) Expression patterns of GASA genes in Arabidopsis thaliana: the GASA4 gene is up-regulated by gibberellins in meristematic regions. Plant Mol Biol 36:871–83.

Avsian-kretchmer O, Gueta-dahan Y, Lev-yadun S, Gollop R (2004) The Salt-Stress Signal Transduction Pathway That Activates the gpx1 Promoter Is Mediated by Intracellular H 2 O 2 , Different from the Pathway Induced by Extracellular H 2 O 2 [ w ]. 135:1685–1696. doi: 10.1104/pp.104.041921.including

Balaji V, Smart CD (2012) Over-expression of snakin-2 and extensin-like protein genes restricts pathogen invasiveness and enhances tolerance to Clavibacter michiganensis subsp. michiganensis in transgenic tomato (Solanum lycopersicum). Transgenic Res 21:23–37. doi: 10.1007/s11248-011-9506-x

Belarmino LC, Capriles PVSZ, Crovella S, et al (2010) EST-Database Search of Plant Defensins – An Example Using Sugarcane , a Large and Complex Genome. Curr Protein Pept Sci 11:1–7.

Benko-Iseppon AM, Galdino SL, Calsa T, et al (2010) Overview on plant antimicrobial peptides. Curr Protein Pept Sci 11:181–8.

Ben-Nissan G, Lee J-Y, Borohov A, Weiss D (2004) GIP, a Petunia hybrida GA-induced cysteine-rich protein: a possible role in shoot elongation and transition to flowering. Plant J 37:229–238. doi: 10.1046/j.1365-313X.2003.01950.x

Ben-Nissan G, Weiss D (1996) The petunia homologue of tomato gast1: transcript accumulation coincides with gibberellin-induced corolla cell elongation. Plant Mol Biol 32:1067–74.

Berrocal-lobo M, Segura A, Moreno M, et al (2002) Snakin-2 , an Antimicrobial Peptide from Potato Whose Gene Is Locally Induced by Wounding and Responds to Pathogen Infection 1. Plant Physiol 128:951–961. doi: 10.1104/pp.010685.1

Bindschedler L V, Whitelegge JP, Millar DJ, Bolwell GP (2006) A two component chitin-binding protein from French bean -- association of a proline-rich protein with a cysteine-rich polypeptide. FEBS Lett 580:1541–6. doi: 10.1016/j.febslet.2006.01.079

Bulet P, Stöcklin R (2005) Insect antimicrobial peptides: structures, properties and gene regulation. Protein Pept Lett 12:3–11.

Bulet P, Stöcklin R, Menin L (2004) Anti-microbial peptides: from invertebrates to vertebrates. Immunol Rev 198:169–84.

Cammue BPA, Bolle MFC De, Terrasi FRG, et al (1992) Isolation and Characterization of a Novel Class of Plant Antimicrobial Peptides from Mirabilis jalapa L. Seeds. J Biol Chem 267:2228–2233.

Carvalho MHC (2008) Drought stress and reactive oxygen species. Plant Signal Behav 3:156–165.

Dangl JL, Jones JDG (2001) Plant pathogens and integrated defence responses to infection. Nature 411:826–833.

De la Fuente JI, Amaya I, Castillejo C, et al (2006) The strawberry gene FaGAST affects plant growth through inhibition of cell elongation. J Exp Bot 57:2401–11. doi: 10.1093/jxb/erj213

Do HM, Lee SC, Jung HW, et al (2004) Differential expression and in situ localization of a pepper defensin ( CADEF1 ) gene in response to pathogen infection , abiotic elicitors and environmental stresses in Capsicum annuum. 166:1297–1305.

Do Nascimento LC, Costa GGL, Binneck E, et al (2012) A web-based bioinformatics interface applied to the GENOSOJA Project: Databases and pipelines. Genet Mol Biol 35:203–11.

Droit A, Poirier GG, Hunter JM (2005) Experimental and bioinformatic approaches for interrogating protein-protein interactions to determine protein function. J Mol Endocrinol 34:263–80.

Edwards D, Batley J (2004) Plant bioinformatics: from genome to phenome. Trends Biotechnol 22:232–7.

Ehlers, JD, Hall AE (1997) Cowpea (Vigna unguiculata (L.) Walp). Field Cro Res. 53:187-204.

Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95:14863–8. doi: 10.1073/pnas.95.25.14863

Faccio P, Hopp E, Favret E, Franzone P (2011) Increased Tolerance to Wheat Powdery Mildew by Heterologous Constitutive Expression of the Solanum chacoense snakin-1 Gene. Czech J Genet Plant Breed 47:135–141.

Ferreira RB, Monteiro S, Regina F, et al (2007) The role of plant defence proteins in fungal pathogenesis. Mol Plant Pathol 8:677–700. doi: 10.1111/J.1364-3703.2007.00419.X

Franco OL (2011) Peptide promiscuity: an evolutionary concept for plant defense. FEBS Lett 585:995–1000. doi: 10.1016/j.febslet.2011.03.008

Freire-Filho FR, Ribeiro VQ, Barreto PD and Santos CAF (1999) Melhoramento genético do caupi (*Vigna unguiculata* (L.) Walp.) na região nordeste. In: Queirós MA de, Goedert CO and Ramos SRR (eds) Recursos genéticos e melhoramento de plantas para o Nordeste Brasileiro. EMBRAPA, CPATSA, Petrolina.

Freire-Filho FR, Rocha M, Ribeiro VQ, Sitollin IM (2008) Avanços e perspectivas para a cultura do feijão-caupi. In: Albuquerque, A. C. S.; Silva, A. G. da (Eds.). Agricultura Tropical: quatro décadas de inovações tecnológicas, institucionais e políticas. Brasília: Embrapa Informação Tecnológica 1:285-250.

Freitas BL, Bered F (2003) In: Genética e evolução vegetal. 1ª ed. Rio Grande do Sul: UFRGS, 463.

Furukawa T, Sakaguchi N, Shimada H (2006) Two OsGASR genes, rice GAST homologue genes that are abundant in proliferating tissues, show different expression patterns in developing panicles. Genes Genet Syst 81:171–80.

García-Olmedo F, Molina a, Alamillo JM, Rodríguez-Palenzuéla P (1998) Plant defense peptides. Biopolymers 47:479–91. doi: 10.1002/(SICI)1097-0282(1998)47:6<479::AID-BIP6>3.0.CO;2-K

Goodstein DM, Shu S, Howson R, et al (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40:D1178–86. doi: 10.1093/nar/gkr944

Guzmán-rodríguez JJ, Ibarra-laclette E, Herrera-estrella L, et al (2013) Analysis of expressed sequence tags ( ESTs ) from avocado seed ( Persea americana var . drymifolia ) reveals abundant expression of the gene encoding the antimicrobial peptide snakin. Plant Physiol Biochem 70:318–324.

Herzog M, Dorne a M, Grellet F (1995) GASA, a gibberellin-regulated gene family from Arabidopsis thaliana related to the tomato GAST1 gene. Plant Mol Biol 27:743–52.

Jianzong P, Liujing LAI, Xiaojing W (2008) PRGL : A cell wall proline-rich protein containning GASA domain in Gerbera hybrida. 51:520–525. doi: 10.1007/s11427-008-0067-z

Kar RK (2011) Plant responses to water stress: role of reactive oxygen species. Plant Signal Behav 6:1741–5. doi: 10.4161/psb.6.11.17729

Keymanesh K, Soltani S, Sardari S (2009) Application of antimicrobial peptides in agriculture and food industry. World J Microbiol Biotechnol 25:933–944. doi: 10.1007/s11274-009-9984-7

Klee EW, Ellis LBM (2005) Evaluating eukaryotic secreted protein prediction. BMC Bioinformatics 6:256. doi: 10.1186/1471-2105-6-256

Ko C-B, Woo Y-M, Lee DJ, et al (2007) Enhanced tolerance to heat stress in transgenic plants expressing the GASA4 gene. Plant Physiol Biochem 45:722–8. doi: 10.1016/j.plaphy.2007.07.010

Kotilainen M, Helariutta Y, Mehto M, et al (1999) GEG participates in the regulation of cell and organ shape during corolla and carpel development in gerbera hybrida. Plant Cell 11:1093–104.

Kovalskaya N, Hammond RW (2009) Expression and functional characterization of the plant antimicrobial snakin-1 and defensin recombinant proteins. Protein Expr Purif 63:12–17. doi: 10.1016/j.pep.2008.08.013

Larkin M a, Blackshields G, Brown NP, et al (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–8. doi: 10.1093/bioinformatics/btm404

Li K-L, Bai X, Li Y, et al (2011) GsGASA1 mediated root growth inhibition in response to chronic cold stress is marked by the accumulation of DELLAs. J Plant Physiol 168:2153–60. doi: 10.1016/j.jplph.2011.07.006

Lin J-Y, Jacobus BH, SanMiguel P, et al (2005) Pericentromeric regions of soybean (Glycine max L. Merr.) chromosomes consist of retroelements and tandemly repeated DNA and are structurally and evolutionarily labile. Genetics 170:1221–30. doi: 10.1534/genetics.105.041616

Liu J et al (2007) Recent progress in elucidating the structure, function and evolution of disease resistance genes in plants. **J of Genet and Genom.**34:765-776.

Loebenstein G (2010) Local Lesions and Induced Resistance. In: Natural and Engineered Resistance to Plant Viruses. 1. ed. EUA: Gad Loebenstein and John P. Carr, 3:73-117.

Lucau-Danila A, Laborde L, Legrand S, et al (2010) Identification of novel genes potentially involved in somatic embryogenesis in chicory (Cichorium intybus L.). BMC Plant Biol 10:122. doi: 10.1186/1471-2229-10-122

Manners JM (2007) Hidden weapons of microbial destruction in plant genomes. Genome Biol 8:225. doi: 10.1186/gb-2007-8-9-225

Mao Z, Zheng J, Wang Y, et al (2011) The new CaSn gene belonging to the snakin family induces resistance against root-knot nematode infection in pepper. Phytoparasitica 39:151–164. doi: 10.1007/s12600-011-0149-5

Marchler-Bauer A, Lu S, Anderson JB, et al (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res 39:D225–9. doi: 10.1093/nar/gkq1189

Maróti G, Kereszt A, Kondorosi E, Mergaert P (2011) Natural roles of antimicrobial peptides in microbes, plants and animals. Res Microbiol 162:363–74. doi: 10.1016/j.resmic.2011.02.005

McHale LK, Haun WJ, Xu WW, et al (2012) Structural variants in the soybean genome localize to clusters of biotic stress-response genes. Plant Physiol 159:1295–308. doi: 10.1104/pp.112.194605

Meiyalaghan S, Thomson SJ, Fiers MW, et al (2014) Structure and expression of GSL1 and GSL2 genes encoding gibberellin stimulated-like proteins in diploid and highly heterozygous tetraploid potato reveals their highly conserved and essential status. BMC Genomics 15:2. doi: 10.1186/1471-2164-15-2

Messing J, Bharti AK, Karlowski WM, et al (2004) Sequence composition and genome organization of maize. Proc Natl Acad Sci 101:14349–14354.

Mohan S, Meiyalaghan S, Latimer JM, et al (2013) GSL2 over-expression confers resistance to Pectobacterium atrosepticum in potato. TAG Theor Appl Genet 127:677–89. doi: 10.1007/s00122-013-2250-2

Moyano-Cañete E, Bellido ML, García-Caparrós N, et al (2013) FaGAST2, a strawberry ripening-related gene, acts together with FaGAST1 to determine cell size of the fruit receptacle. Plant Cell Physiol 54:218–36. doi: 10.1093/pcp/pcs167

Mysore KS, Ryu C-M (2004) Nonhost resistance: how much do we know? Trends Plant Sci 9:97–104. doi: 10.1016/j.tplants.2003.12.005

Nahirñak V, Almasia NI, Fernandez PV, et al (2012a) Potato snakin-1 gene silencing affects cell division, primary metabolism, and cell wall composition. Plant Physiol 158:252–63. doi: 10.1104/pp.111.186544

Nahirñak V, Almasia NI, Hopp HE, Vazquez-rovere C (2012b) Involvement in hormone crosstalk and redox homeostasis Snakin / GASA proteins. Plant Signal Behav 7:1004–1008.

Padovan L, Scocchi M, Tossi A (2010) Structural aspects of plant antimicrobial peptides. Curr Protein Pept Sci 11:210–9.

Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8:785–6. doi: 10.1038/nmeth.1701

Phoenix DA, Dennison SR, Harris F (2013) Antimicrobial Peptides : Their History , Evolution , and Functional Promiscuity. Antimicrobial peptides, 1st edn. Weinheim, pp 1–37

Ponz F, Paz-ares J, Hernandez-lucas C, et al (1986) Cloning and nucleotide sequence of a cDNA encoding the precursor of the barley toxin a-hordothionin. 135:131–135.

Porto WF, Franco OL (2013) Peptides Theoretical structural insights into the snakin / GASA family. Peptides 44:163–167. doi: 10.1016/j.peptides.2013.03.014

Porto WF, Souza V a, Nolasco DO, Franco OL (2012) In silico identification of novel hevein-like peptide precursors. Peptides 38:127–36. doi: 10.1016/j.peptides.2012.07.025

Prosdocini F, Cerqueira GC, Bineck E, et al (2002) Bioinformática: manual do usuário. Biotecnologia, Ciência & Desenvolvimento 29:12-25.

Roxrud I, Lid SE, Fletcher JC, et al (2007) GASA4, one of the 14-member Arabidopsis GASA family of small polypeptides, regulates flowering and seed development. Plant Cell Physiol 48:471–83. doi: 10.1093/pcp/pcm016

Rubinovich L, Weiss D (2010) The Arabidopsis cysteine-rich protein GASA4 promotes GA responses and exhibits redox activity in bacteria and in planta. Plant J 64:1018–27. doi: 10.1111/j.1365-313X.2010.04390.x

Saldanha AJ (2004) Java Treeview--extensible visualization of microarray data. Bioinformatics 20:3246–8. doi: 10.1093/bioinformatics/bth349

Segura A, Moreno M, Madueño F, et al (1999) Snakin-1 , a Peptide from Potato That Is Active Against Plant Pathogens. Mol Plant-Microbe Interact 12:16–23.

Shi L, Gast RT, Gopalraj M, Olszewski NE (1992) Characterization of a shoot-specific, GA3- and ABA-regulated gene from tomato. Plant J 2:153–159.

Shoemaker RC, Schlueter J, Doyle JJ (2006) Paleopolyploidy and gene duplication in soybean and other legumes. Curr Opin Plant Biol 9:104–9. doi: 10.1016/j.pbi.2006.01.007

Shoemaker RC, Schlueter JA, Jackson SA (2008) Soybean Genome Structure and Organization. Genetics and Genomics of Soybean. pp 91–99

Silverstein K a T, Moskal W a, Wu HC, et al (2007) Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants. Plant J 51:262–80. doi: 10.1111/j.1365-313X.2007.03136.x

Silverstein KAT, Graham MA, Paape TD, Vandenbosch KA (2005) Genome Organization of More Than 300 Defensin-Like Genes in Arabidopsis. Plant Physiol 138:600–610. doi: 10.1104/pp.105.060079.600

Simonsen SM, Sando L, Ireland DC, et al (2005) A Continent of Plant Defense Peptide Diversity : Cyclotides in Australian Hybanthus ( Violaceae ). 17:3176–3189. doi: 10.1105/tpc.105.034678.1

Sun S, Wang H, Yu H, et al (2013) GASA14 regulates leaf expansion and abiotic stress resistance by modulating reactive oxygen species accumulation. J Exp Bot 64:1637–1647. doi: 10.1093/jxb/ert021

Tamura K, Peterson D, Peterson N, et al (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–9. doi: 10.1093/molbev/msr121

Tavares LS, Santos MDO, Viccini LF, et al (2008) Biotechnological potential of antimicrobial peptides from flowers. Peptides 29:1842–51. doi: 10.1016/j.peptides.2008.06.003

Taylor BH, Scheuring CF (1994) A molecular marker for lateral root initiation : The RSI-1 gene of tomato ( Lycopersicon esculentum Mill ) is activated in early lateral root primordia. Mol Genet Genomics 243:148–157.

Taylor RD, Koo WW (2011) 2011 Outlook of the U . S . and World Corn and Soybean Industries , 2010-2020. Agribus Appl Econ Rep 2–27.

Torres MA (2010) ROS in biotic interactions. Physiol Plant 138:414–29.

Van Loon LC *et al* (1994) Recommendations for naming plant pathogenesis-related proteins. P Mol Biol. 12:245-264.

Van Loon LC, Van Strien EA (1999) The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. **Physiol and Mol Plant Pathol.** 55:85-97.

Veronese P, Ruiz MT, Coca MA, et al (2003) In Defense against Pathogens . Both Plant Sentinels and Foot Soldiers Need to Know the Enemy. Plant Physiol 131:1580–1590. doi: 10.1104/pp.102.013417.1580

Wanderley-nogueira AC, Belarmino LC, Soares-cavalcanti NM, et al (2012) An overall evaluation of the Resistance ( R ) and Pathogenesis-Related ( PR ) superfamilies in soybean , as compared with Medicago and Arabidopsis. Genet Mol Biol 35:260–271.

Wigoda N, Ben-Nissan G, Granot D, et al (2006) The gibberellin-induced, cysteine-rich protein GIP2 from Petunia hybrida exhibits in planta antioxidant activity. Plant J 48:796–805. doi: 10.1111/j.1365-313X.2006.02917.x

Zare-zardini H, Tolueinia B, Hashemi A, Fesahat F (2013) Antioxidant and Cholinesterase From Ziziphus jujuba Fruits. Am J Alzheimer ' s Dis Other Dementias 4–12. doi: 10.1177/1533317513500839

Zasloff M (2002) Antimicrobial peptides of multicellular organisms. Nature 415:389–395.

Zhang S, Wang X (2008) Expression pattern of GASA, downstream genes of DELLA, in Arabidopsis. Chinese Sci Bull 53:3839–3846. doi: 10.1007/s11434-008-0525-9

Zhang S, Yang C, Peng J, et al (2009) GASA5, a regulator of flowering time and stem growth in Arabidopsis thaliana. Plant Mol Biol 69:745–59. doi: 10.1007/s11103-009-9452-7

Zilli JÉ, Valisheski RR, Rodrigues F, et al (2004) ASSESSMENT OF COWPEA RHIZOBIUM DIVERSITY IN CERRADO AREAS OF NORTHEASTERN BRAZIL. Brazilian J Microbiol 35:281–287.

Zimmermann R, Sakai H, Hochholdinger F (2010) The Gibberellic Acid Stimulated-Like gene family in maize and its role in lateral root development. Plant Physiol 152:356–65. doi: 10.1104/pp.109.149054

**CAPÍTULO I**

**Artigo a ser submetido à revista *Molecular Genetics and Genomics***

---

# Prospection and *in silico* analysis of Snakin in soybean [G*lycine max* l. (Merr.)] trancriptome.

Marx Oliveira-Lima[1]; Ana Maria Benko-Iseppon[1]; Valesca Pandolfi[1*].

[1] Genetics Department, Center of Biological Sciences, Universidade Federal de Pernambuco, Av. Prof. Morais Rego 1235, CEP 50.670-420, Recife, PE, Brazil

Corresponding author*: E-mail: valescapandolfi@gmail.com

**Abstract**

Snakins are antimicrobial peptides (AMPs) composed by ca. 90 amino acid residues and 12 cysteines at conserved positions, its gene may be constitutively expressed or induced by pathogen, as shown for *Solanum tuberosum*. The present study aimed to catalog snakin coding genes in soybean transcriptome as well as analyze their structures and expression sites in order to screen useful elements for biotechnological uses. Using protein sequences of homologues, a search for snakin was performed in soybean transcriptome, represented by expressed sequence tags and Deep SuperSAGE libraries, as well as in the soybean genome. A total of 20 snakin coding genes could be identified in the GENOSOJA database, most of them presenting similarities to unknown proteins of soybean, being distributed in 15 chromosomes with a disposition that reflects possible genomic arrangements undergone by the species. The *in silico* expression evaluation of these genes was associated with tissues under growth and developmental processes, possible target regions of pathogens. The differential expression analysis showed that in general these genes show no differential expression under biotic and abiotic stresses.

**Keywords**: antimicrobial peptides, data-mining, ESTs, GENOSOJA, Deep SuperSAGE.

**Running title:** Analysis *in silico* of snakin in soybean transcriptome

## Introduction

Soybean (*Glycine max* (L.) Merr.) is one of the most important crops in the world, which is a significant source of protein for human diet. It is also a noteworthy oleaginous plant with oil seed content of 18%, which is a raw material by many industries. The soybean market is dominated by USA, Brazil and Argentina, which together corresponded for over than 90% of the world exportations in 2010 (Taylor and Koo 2011). In view of the importance of this crop and the limiting factors to the increase of its production, such as long periods of drought and susceptibility to pathogens, the Soybean Genome Consortium (Genosoja) was created in Brazil, integrating both public and private institutions. This project promotes the search for solutions that may help to solve the domestic problems and production increase. To this end, a database was constructed with millions of sequences including ESTs, SuperSAGE and miRNA (do Nascimento et al. 2012).

The cysteine-rich antimicrobial peptides (AMPs) are efficient components from plant defense system, both constitutive and stress-induced, either biotic or abiotic (Do et al. 2004), emerging as an excellent alternative to mitigate the effects of these problems (García-Olmedo et al. 1998; Silverstein et al. 2007). These peptides can be classified in many different families, with distinct peptides as thionins, defensins, cyclotides and snakins (Ponz et al. 1986; Cammue et al. 1992; Segura et al. 1999; Berrocal-lobo et al. 2002; Simonsen et al. 2005).

Snakins are members of the Snakin/GASA protein family and their homologues have been identified in many plant species like tomato (*Solanum lycopersicum*) (Shi et al., 1990; Herzog et al.1995), petunia (*Petunia hybrida*) (Ben-Nissan and Weiss 1996). *Arabidopsis thaliana* (Aubert et al. 1998; Herzog et al., 1995; Roxrud et al., 2007), potato (*Solanum tuberosum*) (Segura et al., 1999; Berrocal-lobo et al., 2002), common bean (*Phaseolus vulgaris*) (Bindschedler et al., 2006), rice (*Oryza sativa*) (Furukawa et al., 2006; Wang et al., 2009), gerbera (*Gerbera hybrida*) (Kotilainen et al., 1999), strawberry (*Fragaria ananassa*) (de la Fuente et al., 2006), beechnut (*Fagus sylvatica*) (Alonso-Ramírez et al. 2009), maize (*Zea mays*) (Zimmermann et al., 2010), soybean (*Glycine soja*) (Li et al., 2011) and pepper (*Capsicum annum*) (Mao et al., 2011).

These genes encode small proteins, in which can be identified three distinct domains: (1) a signal peptide with 18-29 residues, (2) a variable region that is highly divergent between family members, both in aminoacid composition and sequence length, and (3) a conserved C-terminal region with approximately 60 aminoacids, from which 12 are cysteine residues in conserved positions, the GASA domain (Nahirñak et al., 2012). Studies based on sequence

analysis, phenotypic characterization and expression pattern analysis have suggested that these peptides may be involved in diverse processes, both in development and in response to external factors (Nahirñak et al., 2012; Mao et al. 2011). In the present work, we identified, by *in silico* approaches, homologue sequences of *snakins* in soybean transcriptome and we describe their main features as well.

## Material and Methods

### Screening for snakin homologues

Two complete sequences of Snakin (StSN1 and StSN2), previously obtained from *Solanum tuberosum* (Segura et al., 1999 and Berrocal-lobo et al., 2002, respectively) were used to search for homologues (BLASTp tool; Altschul et al. 1997) in the PhytAMP (http://phytamp.pfba-lab-tun.org/about.php). After that, the UniProt identifiers (IDs) of these and those sequences retrieved from PhytAMP were used as seed to search for orthologous in the SeedServer Software (Guedes et al. http://biodados.icb.ufmg.br/seedserver/), by means of SeedLinkage software (Barbosa-Silva et al., 2008) and KEGG Orthology database (Kanehisa and Goto 2000).

The matches retrieved were subsequently used for alignment through BLASTp in the NCBI database (National Center for Biotechnology Information), in order to obtain a putative function. Subsequently, the Batch CD-Search (Marchler-Bauer et al. 2011) identified the sequences with the complete GASA-domain.

Afterwards, sequences retrieved from SeedServer were used as queries for similarity searches, through tBLASTn algorithm (cut-off value of $\leq e^{-04}$) against the GENOSOJA database (http://lge.ibi.unicamp.br/soybean; Nascimento et al., 2012). In a manual analysis, clusters that matched more than once were eliminated (due to similarity of conserved regions), in order to avoid redundancies.

Those non-redundant clusters were annotated on a local database with their putative function performed through BLASTx against NCBI database, the ORF-Finder tool (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) translated all the selected sequences and the Batch CD-Search identified the conserved domain.

The subcellular localization, the signal peptide localization, the molecular weight (MW) and isoelectric point (pI) (both for the signal peptide and mature peptide) were analyzed

by means of Protcomp (Klee and Ellis 2005), SignalP (Petersen et al., 2011) and JVirgel (Hiller et al., 2006) tools, respectively.

Finally, a multiple alignment with our sequences and the sequences described in previous works as belonging to members from the Snakin/GASA family (Table 1) was performed using the Clustal Omega tool (Sievers et al. 2011).

**Table1 -** Proteins from Snakin/GASA superfamily previously described.

| Description | Organism | UniProt ID | GO annotation | Reference |
|---|---|---|---|---|
| **RSI-1** | *Solanum lycopersicon* | P47926 | Response to GA | (Taylor and Scheuring, 1994) |
| **GAST1** | *Solanum lycopersicon* | P27057 | Response to GA | (Shi et al, 1992) |
| **Gip1** | *Petunia hybrida* | Q43615 | - | (Ben-Nissan and Weiss, 1996) |
| **Gip2** | *Petunia hybrida* | Q9FR10 | - | (Ben-Nissan et al, 2004) |
| **Gip3** | *Petunia hybrida* | Q93WR6 | - | (Ben-Nissan and Weiss, 1996) |
| **Gip4** | *Petunia hybrida* | Q93WR5 | - | (Ben-Nissan et al, 2004) |
| **Gip5** | *Petunia hybrida* | Q93WR4 | - | (Ben-Nissan et al, 2004) |
| **GASA1** | *Arabidopsis thaliana* | P46689 | Response to GA, ABA and BR'S | (Herzog et al, 1995) |
| **GASA2** | *Arabidopsis thaliana* | P46688 | GA mediated signaling pathway | (Herzog et al, 1995) |
| **GASA3** | *Arabidopsis thaliana* | P46687 | Response to GA;<br>GA mediated signaling pathway | (Herzog et al, 1995) |
| **GASA4** | *Arabidopsis thaliana* | P46690 | GA mediated signaling pathway;<br>Cell redox homeostasis. | (Herzog et al, 1995) |
| **GASA5** | *Arabidopsis thaliana* | Q84J95 | GA mediated signaling pathway;<br>Heat acclimation | (Aubert et al, 1998) |
| **GASA6** | *Arabidopsis thaliana* | Q6NMQ7 | GA mediated signaling pathway; | (Aubert et al, 1998) |
| **GASA7** | *Arabidopsis thaliana* | O82328 | GA mediated signaling pathway | (Berrocal-Lobo et al, 2002) |
| **GASA8** | *Arabidopsis thaliana* | O80641 | GA mediated signaling pathway | (Berrocal-Lobo et al, 2002) |
| **GASA9** | *Arabidopsis thaliana* | Q8GWK5 | GA mediated signaling pathway | (Roxrud et al, 2007) |

| | | | | |
|---|---|---|---|---|
| **GASA10** | *Arabidopsis thaliana* | Q8LFM2 | GA mediated signaling pathway | (Roxrud et al, 2007) |
| **GASA11** | *Arabidopsis thaliana* | F4IQJ4 | GA mediated signaling pathway | (Roxrud et al, 2007) |
| **GASA12** | *Arabidopsis thaliana* | Q6GKX7 | GA mediated signaling pathway | (Roxrud et al, 2007) |
| **GASA13** | *Arabidopsis thaliana* | A8MR46 | GA mediated signaling pathway | (Roxrud et al, 2007) |
| **GASA14** | *Arabidopsis thaliana* | Q9LFR3 | GA mediated signaling pathway; Response to GA and ABA; Regulation of ROS metabolic process; Response to salt stress. | (Roxrud et al, 2007) |
| **Snakin-1** | *Solanum tuberosum* | Q948Z4 | Defense response | (Segura et al, 1999) |
| **Snakin-2** | *Solanum tuberosum* | Q93X17 | Defense response | (Berrocal-Lobo et al, 2002) |
| **Snakin-like** | *Phaseolus vulgaris* | Q2YHP6 | - | (Bindschedler et al, 2006) |
| **OsGASR1** | *Oryza sativa* | Q94HA1 | - | (Furukawa et al, 2006) |
| **OsGASR2** | *Oryza sativa* | Q7X885 | - | (Furukawa et al, 2006) |
| **GEG** | *Gerbera hybrida* | Q9XGJ3 | - | (Kotilainen et al, 1999) |
| **FaGAST1** | *Fragraria ananassa* | O49134 | - | (de la Fuente et al, 2006) |
| **FsGASA** | *Fagus sylvativa* | Q0VYL5 | - | (Alonso-Ramírez et al, 2009) |
| **GsGASA** | *Glycine soja* | F1BXA4 | - | (Li et al, 2011) |

Abbreviations: GA – Gibberellic acid; ABA - abscisic acid; BR'S – *Brassinosteroids.*

Soybean Snakin expression based on ESTs and DeepSuperSAGE data.

The expression profile of putative *Snakins* was traced using transcripts (obtained from ESTs and DeepSuperSAGE-based libraries generated from different tissue and accessions of *G. max* under different treatments) downloaded from GENOSOJA Database (http://lge.ibi.unicamp.br/soybean).

In this work, despite the great diversity of ESTs libraries (a total of 65 libraries) available at GENOSOJA database, some EST libraries were grouped according to their tissue/organ (while retaining the main stages of development in individual libraries), resulting in 29 libraries: B01: vegetable buds (field-plants); C0Y: young cotyledons; C06: wounded cotyledons; C07: degenerating cotyledons (10 days old, etiolated seedlings); EN1: endosperm (developing seeds); EP2: seedling epicotyl; F01: floral meristem; F0M: mature flowers; F04: floral meristem mRNA; F05: immature flowers (field-plants); H03: hypocotyl/plumule (germinating seeds); H04: *P. sojae*- infected hypocotyl; H0E: etiolated hypocotyls ; L01: senescing leaf (greenhouse-plants); L03: fully expanded leaves (greenhouse-plants); L04: immature leaves (greenhouse-plants); L0M: mature leaves; L06: leave;drought stressed ; R0R: roots; R03: seedling roots; ; S02: mature seed pods (greenhouse-plants; ; S04: young seeds; S0C: seed coats; S0S: seedlings; S12: seeds (globular-stage embryos); SHS: shoots; SO1: somatic embryos (cultured on MSD 20); S0T: stem; UK1: unknown tissue.

The expression profile of soybean *snakin* was also performed using tags of six DeepSuperSAGE libraries, i.e., four libraries generated from roots of two contrasting soybean genotypes submitted to drought: Embrapa-48 (tolerant to drought, TD) and BR-16 (Susceptible to drought, SD) as compared with non-stressed controls (TC and SC, respectively). The other two libraries from leaves of PI561356 accession (Asian soybean rust-resistant, RR) as compared with the same genotype non-infected (Rust control, RC). Detailed information about all libraries, experimental conditions are available in the Brazilian Soybean Genome Consortium (Benko-Isepponet al 2012; Nascimento et al. 2012).

After counting on different ESTs libraries, the evaluation of reads frequencies that compose each contig was submitted to the Hierarchical Clustering Method (Eisen et al., 1998), which dendrograms were performed with Java Treeview (Saldanha, 2004). Additionally, these contigs were used in a screening of three comparative DeepSuperSAGE libraries.

Anchoring snakin genes in the soybean chromosomes

Using the tBLASTn tool, those peptide sequences from ESTs that presented a complete SNAKIN/GASA domain were aligned against the soybean genome at Phytozome v. 9.1 (Goodstein et al. 2012), to identify their distribution, relative position, and abundance in the soybean virtual chromosomes (provided by GBrowse tool v. 1.1; Stein et al., 2002). Additionally, evaluation of gene structure (number of exons and introns), as well as of the presence of alternative transcripts were also carried out.

**Results**

Search for snakin homologues

Initially, the BLASTp search using two complete sequences of Snakin (SN1 and SN2, IDs: Q948Z4.1 and Q93X17.1, respectively) against the specialized antimicrobial peptide database PhytAMP (http://phytamp.pfba-lab-tun.org/about.php) returned homologous sequences found in four plant families, including: Orchidaceae (*Gymnadenia conopsea*, ID: A3F8U7), Brassicaceae (*Arabidopsis thaliana*, ID: Q1G2Y4), Rosaceae (*Fragraria ananassa*, ID: O49134) and in Asteraceae (*Gerbera hybrida*, ID: Q9XGJ3), all of them exhibiting 12 cysteines residues.

The search performed with these six sequences although *Seed Server* methodology resulted in 52 orthologous sequences, being 47 with complete sequence within their GASA domain. These 47 sequences are distributed in different plant families, where the most representative was Brassicaceae (11 transcrips) followed by Fabaceae (7 transcrips); Solanaceae (6 transcrips); Pinaceae and Poaceae (4 transcrips); Vitaceae, Salicaceae and Euphorbiaceae (3 transcrips); Selaginellaceae and Orchidaceae (1 transcrips) (Figure 1).

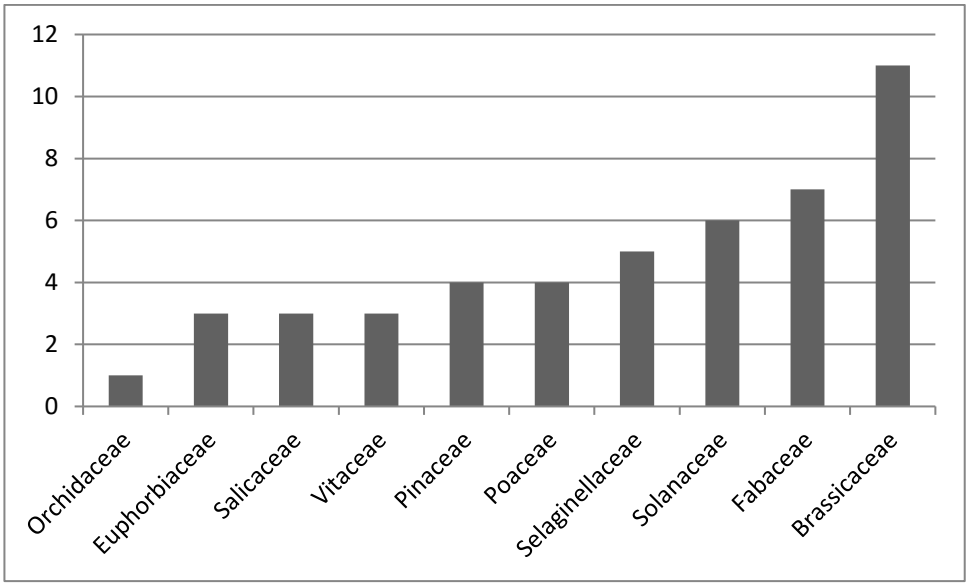



**Figure 1 -** Distribution of the representative genes from Snakin/GASA family found by SeedServer according to their families. The y-axis represents the number of genes and the x-axis represents the taxonomical groups.

Prospection and annotation of snakin in GENOSOJA database

The tBLASTn alignment performed with the 47 orthologues sequences against the GENOSOJA database returned 33 non-redundant sequences within the established parameters (*cut-off* $e^{-4}$) (data not shown) and, of these, 20 sequences exhibited complete GASA-domain.

The results presented by the BLASTx against NCBI demonstrated that the category of the "unknown" proteins was the most representative one, with 14 candidates, all of which being of *Glycine max*. The size of the sequences ranged from 481 to 965 bp, for nucleotide sequences, and from 65 to 138, for amino acid sequences (Table 2).

The virtual 2D-electrophoresis evaluation revealed signal peptide (SP) with isoelectric point (pI) varying between 4.30 (Contig15060) and 9.58 (Contig7785), while the mature peptide (MP) ranged from 5.53 (Contig14720) to 9.32 (Contig9321). Regarding the molecular weight (MW), in the SP there was a low variation between 2.37 (Contig22342) to 2.99 (SJ01-E1-C06-026-A02-UC.F) KDa, while in MP this variation was between 6.88 (Contig15060) and 18.04 (Contig9293) KDa. The Contig14826 showed only MP values due to the absence of a signal peptide (Table 2).

**Table 2 -** Soybean clusters identified thought tBLASTn tool in soybean database. Search conduced with *seed* sequences (*e-value* ≤ $e^{-4}$) reported by the SeedServer (only those with complete GASA-domain) and annotated via BLASTx tool (GenBank). Accession number, organism, description, size of the transcripts in both nucleotide (nt) and aminoacid (aa) as well as its corresponding values for molecular weight (MW) and isoelectric point (pI) from the signal peptide (SP) and mature peptide (MP) are showed.

| | | BLASTx | | Size | | MW | pI |
|---|---|---|---|---|---|---|---|
| **Cluster** | **Acession** | **Organism** | **Description** | **Nt** | **aa** | **SP/MP** | **SP/MP** |
| **Contig1367** | ACU15001.1 | *G. max* | Unknown | 749 | 99 | 2.45/8.07 | 7.00/8.53 |
| **Contig1373** | XP_003549639.1 | *G. max* | snakin-2-like | 965 | 115 | 2.40/10.36 | 8.25/7.93 |
| **Contig14720** | XP_003528161.1 | *G. max* | uncharacterized protein | 804 | 119 | 2.40/10.77 | 9.58/5.53 |
| **Contig14826** | XP_003528161.1 | *G. max* | uncharacterized protein | 802 | 65 | 6.92 | 8.39 |
| **Contig15060** | XP_003525918.1 | *G. max* | snakin-1 | 696 | 88 | 2.77/6.88 | 4.30/8.10 |
| **Contig16419** | ACU14488.1 | *G. max* | Unknown | 608 | 106 | 2.77/9.00 | 8.25/8.50 |
| **Contig18282** | ACU14567.1 | *G. max* | Unknown | 792 | 107 | 2.60/9.33 | 9.27/8.75 |
| **Contig2059** | ACU14224.1 | *G. max* | Unknown | 481 | 106 | 2.46/9.11 | 6.14/7.72 |
| **Contig20912** | ACU15584.1 | *G. max* | Unknown | 809 | 138 | 2.69/12.33 | 6.12/8.78 |
| **Contig22342** | ACU14458.1 | *G. max* | Unknown | 637 | 115 | 2.37/10.31 | 8.25/7.93 |
| **Contig23713** | ACU16624.1 | *G. max* | Unknown | 619 | 90 | 2.76/7.18 | 6.08/8.37 |
| **Contig26289** | XP_003536595.1 | *G. max* | snakin-1-like | 554 | 90 | 2.71/7.24 | 6.08/8.50 |
| **Contig3375** | ACU13692.1 | *G. max* | Unknown | 735 | 99 | 2.43/8.10 | 7.00/8.80 |
| **Contig3731** | ACU14995.1 | *G. max* | Unknown | 791 | 88 | 2.67/6.92 | 6.08/9.28 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Contig627** | ACU14995.1 | *G. max* | Unknown | 783 | 88 | 2.64/6.87 | 6.08/9.09 |

*Table 2 continued*

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Contig7139** | XP_003523021.1 | *G. max* | GRP | 798 | 106 | 2.77/9.01 | 8.25/8.90 |
| **Contig7785** | ACU13226.1 | *G. max* | Unknown | 847 | 117 | 2.41/10.59 | 9.58/6.89 |
| **Contig9293** | ACU16056.1 | *G. max* | Unknown | 935 | 191 | 2.27/18.04 | 8.25/9.29 |
| **Contig9321** | ACU13162.1 | *G. max* | Unknown | 701 | 110 | 2.83/9.38 | 9.26/9.32 |
| **SJ01-E1-C06-026-A02-UC.F** | ACU15001.1 | *G. max* | Unknown | 542 | 94 | 2.99/7.46 | 9.06/8.66 |

In the subcellular localization analysis, it was observed that all sequences possessed the signal to be sent to the extracellular environment, and this data were supported by optimal scores (data not shown). The Clustal Omega alignment evidenced the conservation of some residues in a C-terminal domain what could be observed, especially for the 12 cysteine residues (Figure 2).

Analyses of snakin based on ESTs and DeepSuperSAGE data

A hierarchical clustering analysis of the transcriptional profile of snakin genes, including 29 ESTs libraries of soybean generated from different tissues and development stages (GENOSOJA database: http://bioinfo03.ibi.unicamp.br/soja). Among the organs and tissue, seeds were the most representative libraries (6 libraries, with 1988 reads), presenting 82 % of from total (2,415). The remaining reads was present in tissues as leaves (129 reads), hypocotyls (61 reads), flowers (50 reads), roots (31 reads) and epicotyls (8 reads). Moreover, the pattern of expression, obtained from normalized data, revealed prevalence of transcripts in early stages of development (Figure 3).

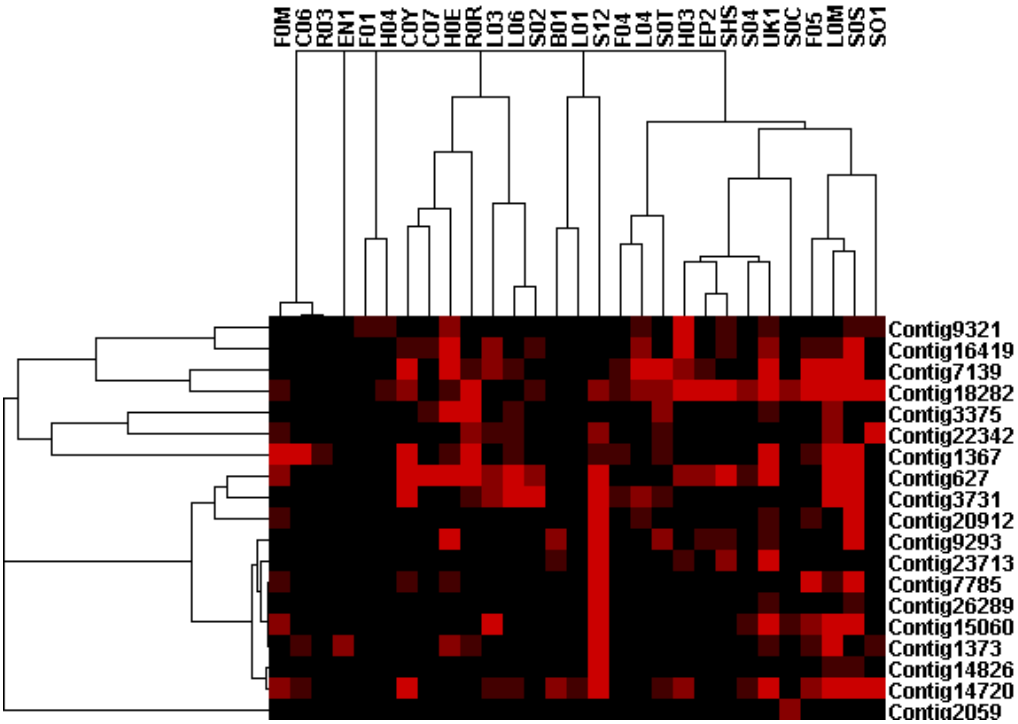



**Figure 3 -** Hierarchical clustering of transcriptional profile in soybean performed using the CLUSTER software. The red colour indicates high expression; the dark red indicates intermediate expression and black indicates no expression.

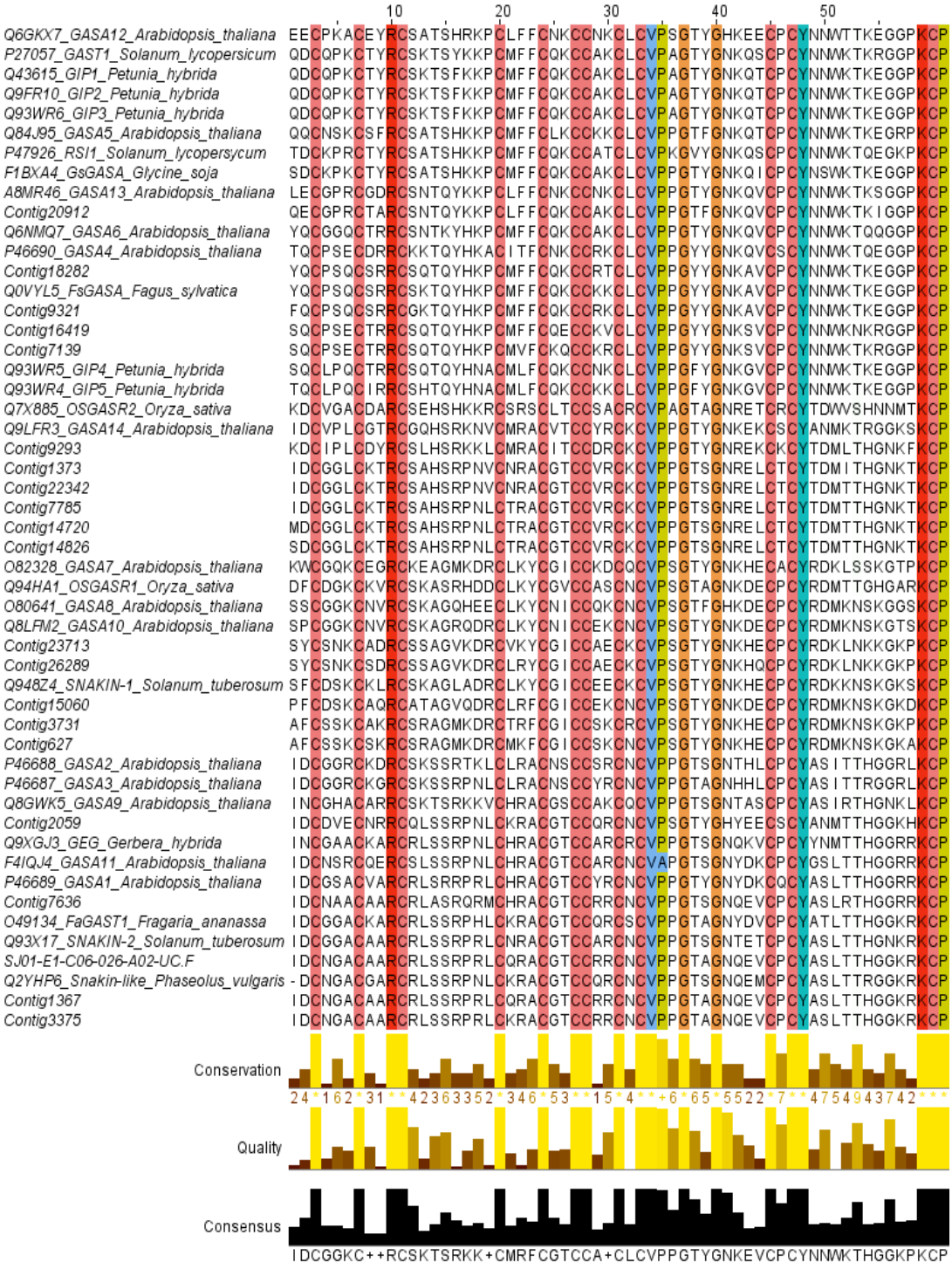



**Figure 2 -** Multiple sequence **a**lignment of complete Snakin/GASA domain, performed with peptide sequences of ESTs and sequences of others plant species previously described. The main conserved residues are indicated by different colours. The rose-shaded cysteine residues are present in all sequences. The other conserved residues also shown in same colour. The numerical index reflects the degree of conservation of the amino acid residues in the alignment. Star (*) indicates amino acids 100% identical in the alignment.

The search in the DeepSuperSAGE libraries resulted in the identification of 81 tags of more than 4,000,000 contained in the database, of which, 45 tags were obtained of contrast RR vs. RC (PI561356 Rust-resistant vs. non-infected Control); 20 tags of contrast TD vs. TC (Embrapa 48, drought-tolerant cultivar vs. non-stressed control) and 16 tags of contrast SD vs. SC (BR16 cultivar, drought-susceptible vs. non-stressed control) (Figure 4).

Regarding exclusive tags in each stress condition, 22 tags were exclusive to drought condition (12 tags exclusive of drought-tolerant cultivar and 9 tags exclusive of drought-susceptible cultivar) and 34 tags were exclusive to Rust-stress condition (Figure 4).

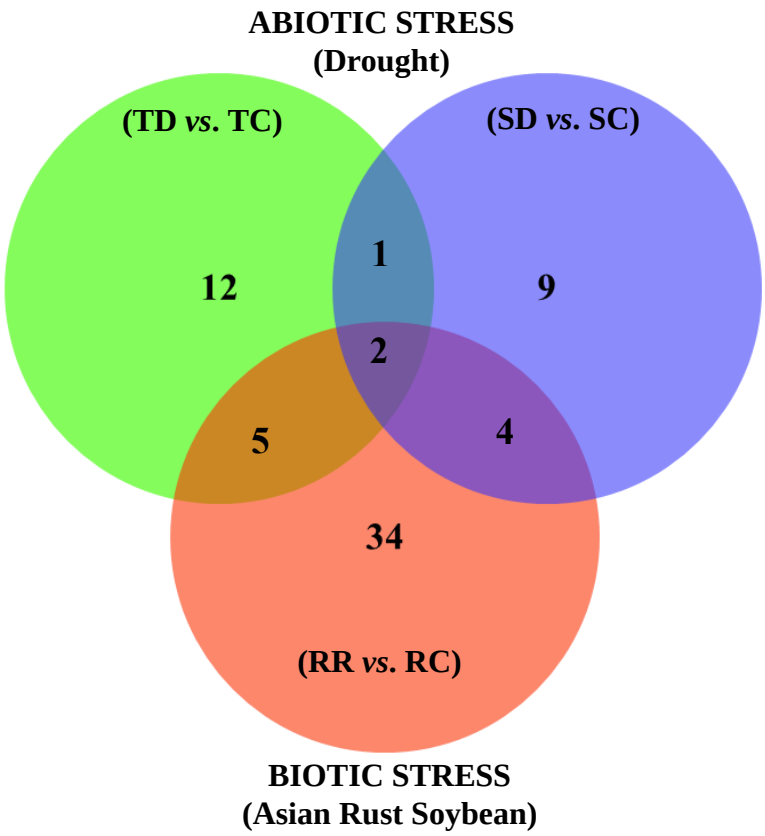



**Figure 4 -** Venn diagram showing the number of DeepSuperSAGE tags matching soybean snakin candidates exclusive in each contrasts: **TD *vs.* TC** (Embrapa 48, drought-tolerant cultivar stressed *vs*. non-stressed control); **SD *vs.* SC** (BR16, drought-susceptible cultivar stressed *vs*. non-stressed control) and **RR *vs.* RC** (PI561356 Rust-resistant stressed *vs*. non-stressed Control).

Among 81 tags, 31 tags (38 %) showed expression differential (up/down- regulated), where: 20 tags (6 up-regulated and 14 down regulated) were present in the contrast RR *vs*. RC (Rust-resistant *vs*. Control); 1 tag (down regulated) in the contrast SD *vs*. SC (drought-susceptible *vs*. control) and 10 tags (1 tag up regulated and 9 down regulated) in the contrast TD *vs*. TC (drought-tolerant *vs*. control) (data not shown).

Physical distribution of snakin sequences on the soybean chromosomes

The alignment performed with the contigs in the soybean genome reported 42 genes (36 primary transcripts and 6 splicing variants) annotated as belonging to the Snakin/GASA family, distributed in 15 of the 20 soybean chromosomes (Figure 5).

The alignment showed some chromosomes as 6, 13, 14 and 17, with a higher prevalence of genes. In the chromosome 17 (Gm17), for example, were observed six representative genes (four primary transcripts and two alternative), whereas the chromosomes 3, 5, 8, 9 and 16 showed only one match (Figure 5).

Regarding the structure of these genes, we found three classes: (1) two exons and one intron, (2) three exons and two introns and (3) four exons and three introns. A total of 14 genes presented the structure with three exons and two introns, other 15 presented structure with four exons and three introns and 7 with two exons and one intron, which were retrieve among the primary transcripts.

In this analysis, we also found the occurrence of one splicing variant localized on chromosomes 9 (Glyma09g37290.1*), 10 (Glyma10g36011.1*), 18 (Glyma18g17490.2*) and 19 (Glyma19g01590.2) and two splicing variants on chromosome 17 (Glyma17g37760.2 and Glyma17g10050.2).

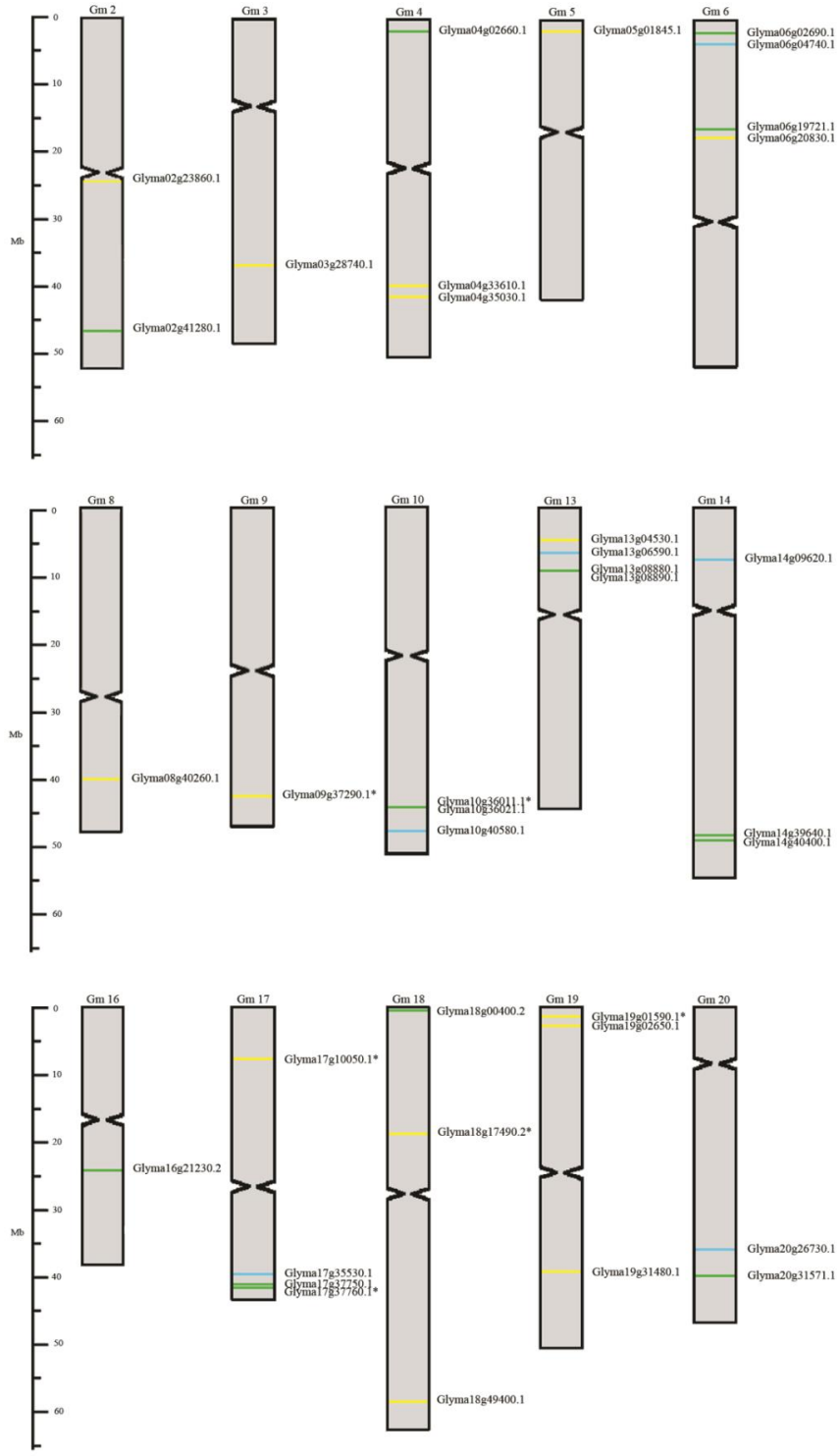



**Figure 5 -** Representative ideogram with the sequences found after the analysis with tBLASTn tool in SoyBase, showing the distribution of the alignments. The blue marks indicate the genes with two exons and one intron; the yellow marks indicate the genes with three exons and two introns and the green marks indicate the genes with four exons and three introns. Sequences marked with asterisk refer to splicing variants.

## Discussion

### Search for snakin homologues

The utilization of the six Uniprot IDs for *snakins* (homologues from *S. tuberosum*, *A. thaliana*, *Ge. hybrida*, *Gy. conopsea* and *F. ananassa*) as seeds in the SeedServer allowed to identify orthologues in different taxa, most of which (90%), with the expected structural features (Snakin/GASA domain). The search for orthologues in the SeedServer also permitted inferences into the possible origin of the Snakin family in the Lycopodiophyta clade (Selaginellaceae), where the most basal *snakin* ortholog was found (Figure 1).

It seems most likely that this family of antimicrobial peptide passed through events of duplication during its evolution, being subsequently settled in upper clades, as it has been observed for other peptides (Franco 2011; Phoenix et al. 2013), although there was no functional annotation to this gene.

As the other groups of antimicrobial peptides, the fixation of this gene may be associated to its wide range of functions, acting both in innate processes (Ben-Nissan et al., 2004; Zhang et al., 2009; Lucau-Danila et al., 2010; Nahirñak et al., 2012) as in induced stresses (Almasia et al., 2008; Balaji & Smart, 2012; Rong et al., 2013), confirming its evolutionary maintenance during the evolutionary process. Such "promiscuity" of functions has been advantageous for plant defense and, therefore, to plant survival and evolution (Franco, 2011), as also, a promising source to be explored in both the pharmaceutical and agricultural sectors (Sarika et al 2012).

### Prospection and annotation in soybean database

The utilization of the *seeds* provided by SeedServer allowed the identification of 20 putative *Snakins* in the soybean expressed genome, considering the adopted cut-off $e^{-4}$. These sequences presented the complete GASA domain, shared by all members of the family, characterized by a cysteine-rich signature (12 cysteines residues of the C-terminus) (Figure 2), in highly conserved positions (XCX3CX3CX8CX3CX2CCX2CXCX11CXCX12CX), which is essential for their biochemical activity and probably responsible for their protein structure (Herzog et al., 1995; Aubert et al., 1998; Jianzong et al., 2008; Sun et al., 2013).

Regarding the subcellular localization, our data showed that the putative *Snakins* of *G. max* were exclusively addressed to the extracellular environment, a role similar to *GASA5*, that acts as a regulator of flowering time and stem growth (Zhang et al. 2009; Moyano-Cañete et al. 2013). Moreover, it is known that Snakin/GASA genes are also involved in other signaling pathways, i.e., hormonal (Nahirñak et al. 2012) and stress response, as already reported for other AMPs (Oard and Enright 2006). According with this evidences a clear divergence of roles may be recognized these peptides, depending on plant physiological state and/or stress condition imposed.

Although a high similarity among these peptides with "unknown" and "uncharacterized" sequences from soybean, they showed conservation in their structure with size, molecular weight and pI very close to the found in other members of the family (Ko et al., 2007; Mao et al., 2011). Besides the most conserved region described for these peptides, the C-Terminal domain (GASA) (Padovan et al., 2010), as observed in the Figure 2, demonstrate a strict conservation, which can be a strong indicative of functional specialization of these genes, also highlighting their importance to the maintenance of the plant homeostasis.

*In silico* expression based on ESTs and SuperSAGE data

The developmental expression pattern of snakin gene was predominantly observed in tissues in early developmental stages (leaves, roots, seeds and cotyledons), which had some tissue-specific relation, where the cellular activity and the action of effectors that regulates the cell division (as gibberellin) is stronger. This pattern was also observed for other members of the Snakin/GASA family (Nahirñak et al., 2012). This fact may indicate that the *snakins* of *Gl. max* follow the same model, as it was observed for homologues involved in elongation of shoots (Shi et al., 1992), corolla and carpel (Kotilainen et al., 1999), as well as in seeds and roots development (Roxrud et al. 2007; Zimmermann et al., 2010).

Beside, as these tissues are also prime targets of many fungal/ bacterial pathogens and nematodes (Berrocal-lobo et al., 2002; Mao et al., 2011), the expression of antimicrobial peptides, like snakins, have been associated as part of the permanent and inducible defense against microbial assault (Veronese et al. 2003; Almasia et al. 2008; Balaji and Smart 2012)

The analyze of DeepSuperSAGE libraries made with ESTs revealed a higher number of matching tags (81), and many contigs were able to map tags in the three comparisons (TD vs. TC; SD vs. SC and RR vs. RC), reinforcing the important role that the *snakins* of soybean exhibit in response both biotic and abiotic stresses.

As expected, most DeepSuperSAGE tags (45 tags) were identified in the biotic stress condition (contrast RR vs. RC). It is known that this fungus often infects leaves, a tissue which has a very common expression of *snakins*, both constitutive and induced (Balaji and Smart 2012; Nahirñak et al., 2012). However, to our knowledge, there are no reports of snakin transcripts associated with Rust (*Phakopsora pachyrhizi*) response. Thus, our data indicated the expression of *snakins* in view of a new specific pathogen.

Regarding the libraries for abiotic stress (drought), the considerable number of tags indicate the presence of crosstalk responses, since the water stress causes osmotic and oxidative disturbs, triggering responses via reactive oxygen species (ROS), reflecting in the induced expression of members from the Snakin/GASA family, a showed for *A. thaliana*, where the *GASA14* modulates the accumulation of ROS. It was also known that members of this family, such as *StSN2*, *GIP4* and *GIP5*, respond to effectors responsive to this stress as the hormone ABA (Nahirñak et al., 2012). Nevertheless, most of these transcripts are down regulated, indicating that the factors involved in growth tend to be suppressed in this situation (Blum, 2011).

Physical distribution of snakin on the soybean chromosomes

The datas about gene structure (presence of introns/ exons) showed that the three structures found in *snakins* orthologues can be observed in members of the Snakin/GASA family. Almasia et al. (2008), isolated a Snakin-1 from *Solanum tuberosum* and identified only one intron and which is associated to the response to pathogens, whereas Berrocal-lobo et al. (2002) and Silverstein et al. (2007) reported genes with two and three introns respectively, associated to both responses, against pathogens and abiotic stress, which makes soybean a promising organism to study these genes, due to the fact that there is a large structural variation and functions.

In comparison with other plants, soybean has a considerable number of snakin candidates in its genome. To this date, 14 representatives were isolated from *A. thaliana* and 10 from *Z. mays* (Herzog et al. 1995; Aubert et al. 1998; Berrocal-lobo et al. 2002; Roxrud et al. 2007; Zimmermann et al. 2010), being the most representative ones (an amount smaller than we have found in soybean), 42 between primary and secondary transcripts, a fact that may have contributed for this abundance can be the large genome of this crop (approximately 975 Mb) in comparison with maize (307 Mb) and Arabidopsis (157 Mb), as highlighted by other authors

(Goodstein et al., 2012; Messing et al., 2004; Wanderley-Nogueira, Belarmino, Soares-cavalcanti, Bezerra-Neto, & Benko-iseppon, 2012).

The genome organization of soybean can explain the clusters formed for these genes, their localization in highly euchromatic regions, and a high gene expression as showed in the heat map, reinforces its importance (Shoemaker et al. 2008).

## Conclusions

We identified sequences of putative new antimicrobial peptides in soybean, which confers valuable resources to biotechnological uses, may especially for improvement of this crop via transgenesis. These data are also important for mapping purposes, considering the distribution and localization of *snakins* in the different soybean chromosomes.

## Acknowledgments

The authors are grateful to National Council for Scientific and Technological Development (CNPq), to National Council for the Improvement of Higher Education (CAPES) and to Foundation for Support of Science and Technology of Pernambuco State - Brazil (FACEPE) for supporting our research.

## References

Almasia NI, Bazzini AA, Hopp HE, Vazquez-Rovere C (2008) Overexpression of snakin-1 gene enhances resistance to Rhizoctonia solani and Erwinia carotovora in transgenic potato plants. Mol Plant Pathol 9:329–338. doi: 10.1111/J.1364-3703.2008.00469.X

Almasia NI, Narhirñak V, Hopp HE, Vazquez-Rovere C (2010) Isolation and characterization of the tissue and development-specific potato snakin-1 promoter inducible by temperature and wounding. Electron J Biotechnol 13:1–21. doi: 10.2225/vol13-issue5-fulltext-12

Alonso-Ramírez A, Rodríguez D, Reyes D, et al (2009a) Evidence for a Role of Gibberellins in Salicylic Acid-Modulated Early Plant Responses to Abiotic Stress in Arabidopsis Seeds 1. Plant Physiol 150:1335–1344. doi: 10.1104/pp.109.139352

Alonso-Ramírez A, Rodríguez D, Reyes D, et al (2009b) Cross-talk between gibberellins and salicylic acid in early stress responses in Arabidopsis thaliana seeds. Plant Signal Behav 4:750–1. doi: 10.1104/pp.109.139352

Altschul SF, Gish W, Miller W, et al (1990) Basic local alignment search tool. J Mol Biol 215:403–10. doi: 10.1016/S0022-2836(05)80360-2

Aubert D, Chevillard M, Dorne a M, et al (1998) Expression patterns of GASA genes in Arabidopsis thaliana: the GASA4 gene is up-regulated by gibberellins in meristematic regions. Plant Mol Biol 36:871–83.

Avsian-kretchmer O, Gueta-dahan Y, Lev-yadun S, Gollop R (2004) The Salt-Stress Signal Transduction Pathway That Activates the gpx1 Promoter Is Mediated by Intracellular H 2 O 2 , Different from the Pathway Induced by Extracellular H 2 O 2 [ w ]. 135:1685–1696. doi: 10.1104/pp.104.041921.including

Balaji V, Smart CD (2012) Over-expression of snakin-2 and extensin-like protein genes restricts pathogen invasiveness and enhances tolerance to Clavibacter michiganensis subsp. michiganensis in transgenic tomato (Solanum lycopersicum). Transgenic Res 21:23–37. doi: 10.1007/s11248-011-9506-x

Belarmino LC, Capriles PVSZ, Crovella S, et al (2010) EST-Database Search of Plant Defensins – An Example Using Sugarcane , a Large and Complex Genome. Curr Protein Pept Sci 11:1–7.

Benko-Iseppon AM, Galdino SL, Calsa T, et al (2010) Overview on plant antimicrobial peptides. Curr Protein Pept Sci 11:181–8.

Ben-Nissan G, Lee J-Y, Borohov A, Weiss D (2004) GIP, a Petunia hybrida GA-induced cysteine-rich protein: a possible role in shoot elongation and transition to flowering. Plant J 37:229–238. doi: 10.1046/j.1365-313X.2003.01950.x

Ben-Nissan G, Weiss D (1996) The petunia homologue of tomato gast1: transcript accumulation coincides with gibberellin-induced corolla cell elongation. Plant Mol Biol 32:1067–74.

Berrocal-lobo M, Segura A, Moreno M, et al (2002) Snakin-2 , an Antimicrobial Peptide from Potato Whose Gene Is Locally Induced by Wounding and Responds to Pathogen Infection 1. Plant Physiol 128:951–961. doi: 10.1104/pp.010685.1

Bindschedler L V, Whitelegge JP, Millar DJ, Bolwell GP (2006) A two component chitin-binding protein from French bean -- association of a proline-rich protein with a cysteine-rich polypeptide. FEBS Lett 580:1541–6. doi: 10.1016/j.febslet.2006.01.079

Bulet P, Stöcklin R (2005) Insect antimicrobial peptides: structures, properties and gene regulation. Protein Pept Lett 12:3–11.

Bulet P, Stöcklin R, Menin L (2004) Anti-microbial peptides: from invertebrates to vertebrates. Immunol Rev 198:169–84.

Cammue BPA, Bolle MFC De, Terrasi FRG, et al (1992) Isolation and Characterization of a Novel Class of Plant Antimicrobial Peptides from Mirabilis jalapa L. Seeds. J Biol Chem 267:2228–2233.

Carvalho MHC (2008) Drought stress and reactive oxygen species. Plant Signal Behav 3:156–165.

Dangl JL, Jones JDG (2001) Plant pathogens and integrated defence responses to infection. Nature 411:826–833.

De la Fuente JI, Amaya I, Castillejo C, et al (2006) The strawberry gene FaGAST affects plant growth through inhibition of cell elongation. J Exp Bot 57:2401–11. doi: 10.1093/jxb/erj213

Do HM, Lee SC, Jung HW, et al (2004) Differential expression and in situ localization of a pepper defensin ( CADEF1 ) gene in response to pathogen infection , abiotic elicitors and environmental stresses in Capsicum annuum ℭ. 166:1297–1305. doi: 10.1016/j.plantsci.2004.01.008

Do Nascimento LC, Costa GGL, Binneck E, et al (2012) A web-based bioinformatics interface applied to the GENOSOJA Project: Databases and pipelines. Genet Mol Biol 35:203–11. doi: 10.1590/S1415-47572012000200002

Droit A, Poirier GG, Hunter JM (2005) Experimental and bioinformatic approaches for interrogating protein-protein interactions to determine protein function. J Mol Endocrinol 34:263–80. doi: 10.1677/jme.1.01693

Edwards D, Batley J (2004) Plant bioinformatics: from genome to phenome. Trends Biotechnol 22:232–7. doi: 10.1016/j.tibtech.2004.03.002

Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95:14863–8. doi: 10.1073/pnas.95.25.14863

Faccio P, Hopp E, Favret E, Franzone P (2011) Increased Tolerance to Wheat Powdery Mildew by Heterologous Constitutive Expression of the Solanum chacoense snakin-1 Gene. Czech J Genet Plant Breed 47:135–141.

Ferreira RB, Monteiro S, Regina F, et al (2007) The role of plant defence proteins in fungal pathogenesis. Mol Plant Pathol 8:677–700. doi: 10.1111/J.1364-3703.2007.00419.X

Franco OL (2011) Peptide promiscuity: an evolutionary concept for plant defense. FEBS Lett 585:995–1000. doi: 10.1016/j.febslet.2011.03.008

Furukawa T, Sakaguchi N, Shimada H (2006) Two OsGASR genes, rice GAST homologue genes that are abundant in proliferating tissues, show different expression patterns in developing panicles. Genes Genet Syst 81:171–80.

García-Olmedo F, Molina a, Alamillo JM, Rodríguez-Palenzuéla P (1998) Plant defense peptides. Biopolymers 47:479–91. doi: 10.1002/(SICI)1097-0282(1998)47:6<479::AID-BIP6>3.0.CO;2-K

Goodstein DM, Shu S, Howson R, et al (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40:D1178–86. doi: 10.1093/nar/gkr944

Guzmán-rodríguez JJ, Ibarra-laclette E, Herrera-estrella L, et al (2013) Analysis of expressed sequence tags ( ESTs ) from avocado seed ( Persea americana var . drymifolia ) reveals

abundant expression of the gene encoding the antimicrobial peptide snakin. Plant Physiol Biochem 70:318–324.

Herzog M, Dorne a M, Grellet F (1995) GASA, a gibberellin-regulated gene family from Arabidopsis thaliana related to the tomato GAST1 gene. Plant Mol Biol 27:743–52.

Jianzong P, Liujing LAI, Xiaojing W (2008) PRGL : A cell wall proline-rich protein containning GASA domain in Gerbera hybrida. 51:520–525. doi: 10.1007/s11427-008-0067-z

Kar RK (2011) Plant responses to water stress: role of reactive oxygen species. Plant Signal Behav 6:1741–5. doi: 10.4161/psb.6.11.17729

Keymanesh K, Soltani S, Sardari S (2009) Application of antimicrobial peptides in agriculture and food industry. World J Microbiol Biotechnol 25:933–944. doi: 10.1007/s11274-009-9984-7

Klee EW, Ellis LBM (2005) Evaluating eukaryotic secreted protein prediction. BMC Bioinformatics 6:256. doi: 10.1186/1471-2105-6-256

Ko C-B, Woo Y-M, Lee DJ, et al (2007) Enhanced tolerance to heat stress in transgenic plants expressing the GASA4 gene. Plant Physiol Biochem 45:722–8. doi: 10.1016/j.plaphy.2007.07.010

Kotilainen M, Helariutta Y, Mehto M, et al (1999) GEG participates in the regulation of cell and organ shape during corolla and carpel development in gerbera hybrida. Plant Cell 11:1093–104.

Kovalskaya N, Hammond RW (2009) Expression and functional characterization of the plant antimicrobial snakin-1 and defensin recombinant proteins. Protein Expr Purif 63:12–17. doi: 10.1016/j.pep.2008.08.013

Larkin M a, Blackshields G, Brown NP, et al (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–8. doi: 10.1093/bioinformatics/btm404

Li K-L, Bai X, Li Y, et al (2011) GsGASA1 mediated root growth inhibition in response to chronic cold stress is marked by the accumulation of DELLAs. J Plant Physiol 168:2153–60. doi: 10.1016/j.jplph.2011.07.006

Lin J-Y, Jacobus BH, SanMiguel P, et al (2005) Pericentromeric regions of soybean (Glycine max L. Merr.) chromosomes consist of retroelements and tandemly repeated DNA and are structurally and evolutionarily labile. Genetics 170:1221–30. doi: 10.1534/genetics.105.041616

Lucau-Danila A, Laborde L, Legrand S, et al (2010) Identification of novel genes potentially involved in somatic embryogenesis in chicory (Cichorium intybus L.). BMC Plant Biol 10:122. doi: 10.1186/1471-2229-10-122

Manners JM (2007) Hidden weapons of microbial destruction in plant genomes. Genome Biol 8:225. doi: 10.1186/gb-2007-8-9-225

Mao Z, Zheng J, Wang Y, et al (2011) The new CaSn gene belonging to the snakin family induces resistance against root-knot nematode infection in pepper. Phytoparasitica 39:151–164. doi: 10.1007/s12600-011-0149-5

Marchler-Bauer A, Lu S, Anderson JB, et al (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res 39:D225–9. doi: 10.1093/nar/gkq1189

Maróti G, Kereszt A, Kondorosi E, Mergaert P (2011) Natural roles of antimicrobial peptides in microbes, plants and animals. Res Microbiol 162:363–74. doi: 10.1016/j.resmic.2011.02.005

McHale LK, Haun WJ, Xu WW, et al (2012) Structural variants in the soybean genome localize to clusters of biotic stress-response genes. Plant Physiol 159:1295–308. doi: 10.1104/pp.112.194605

Meiyalaghan S, Thomson SJ, Fiers MW, et al (2014) Structure and expression of GSL1 and GSL2 genes encoding gibberellin stimulated-like proteins in diploid and highly heterozygous tetraploid potato reveals their highly conserved and essential status. BMC Genomics 15:2. doi: 10.1186/1471-2164-15-2

Messing J, Bharti AK, Karlowski WM, et al (2004) Sequence composition and genome organization of maize. Proc Natl Acad Sci 101:14349–14354.

Mohan S, Meiyalaghan S, Latimer JM, et al (2013) GSL2 over-expression confers resistance to Pectobacterium atrosepticum in potato. TAG Theor Appl Genet 127:677–89. doi: 10.1007/s00122-013-2250-2

Moyano-Cañete E, Bellido ML, García-Caparrós N, et al (2013) FaGAST2, a strawberry ripening-related gene, acts together with FaGAST1 to determine cell size of the fruit receptacle. Plant Cell Physiol 54:218–36. doi: 10.1093/pcp/pcs167

Mysore KS, Ryu C-M (2004) Nonhost resistance: how much do we know? Trends Plant Sci 9:97–104. doi: 10.1016/j.tplants.2003.12.005

Nahirñak V, Almasia NI, Fernandez PV, et al (2012a) Potato snakin-1 gene silencing affects cell division, primary metabolism, and cell wall composition. Plant Physiol 158:252–63. doi: 10.1104/pp.111.186544

Nahirñak V, Almasia NI, Hopp HE, Vazquez-rovere C (2012b) Involvement in hormone crosstalk and redox homeostasis Snakin / GASA proteins. Plant Signal Behav 7:1004–1008.

Padovan L, Scocchi M, Tossi A (2010) Structural aspects of plant antimicrobial peptides. Curr Protein Pept Sci 11:210–9.

Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8:785–6. doi: 10.1038/nmeth.1701

Phoenix DA, Dennison SR, Harris F (2013) Antimicrobial Peptides : Their History , Evolution , and Functional Promiscuity. Antimicrobial peptides, 1st edn. Weinheim, pp 1–37

Ponz F, Paz-ares J, Hernandez-lucas C, et al (1986) Cloning and nucleotide sequence of a cDNA encoding the precursor of the barley toxin a-hordothionin. 135:131–135.

Porto WF, Franco OL (2013) Peptides Theoretical structural insights into the snakin / GASA family. Peptides 44:163–167. doi: 10.1016/j.peptides.2013.03.014

Porto WF, Souza V a, Nolasco DO, Franco OL (2012) In silico identification of novel hevein-like peptide precursors. Peptides 38:127–36. doi: 10.1016/j.peptides.2012.07.025

Roxrud I, Lid SE, Fletcher JC, et al (2007) GASA4, one of the 14-member Arabidopsis GASA family of small polypeptides, regulates flowering and seed development. Plant Cell Physiol 48:471–83. doi: 10.1093/pcp/pcm016

Rubinovich L, Weiss D (2010) The Arabidopsis cysteine-rich protein GASA4 promotes GA responses and exhibits redox activity in bacteria and in planta. Plant J 64:1018–27. doi: 10.1111/j.1365-313X.2010.04390.x

Saldanha AJ (2004) Java Treeview--extensible visualization of microarray data. Bioinformatics 20:3246–8. doi: 10.1093/bioinformatics/bth349

Segura A, Moreno M, Madueño F, et al (1999) Snakin-1 , a Peptide from Potato That Is Active Against Plant Pathogens. Mol Plant-Microbe Interact 12:16–23.

Shi L, Gast RT, Gopalraj M, Olszewski NE (1992) Characterization of a shoot-specific, GA3- and ABA-regulated gene from tomato. Plant J 2:153–159.

Shoemaker RC, Schlueter J, Doyle JJ (2006) Paleopolyploidy and gene duplication in soybean and other legumes. Curr Opin Plant Biol 9:104–9. doi: 10.1016/j.pbi.2006.01.007

Shoemaker RC, Schlueter JA, Jackson SA (2008) Soybean Genome Structure and Organization. Genetics and Genomics of Soybean. pp 91–99

Silverstein K a T, Moskal W a, Wu HC, et al (2007) Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants. Plant J 51:262–80. doi: 10.1111/j.1365-313X.2007.03136.x

Silverstein KAT, Graham MA, Paape TD, Vandenbosch KA (2005) Genome Organization of More Than 300 Defensin-Like Genes in Arabidopsis. Plant Physiol 138:600–610. doi: 10.1104/pp.105.060079.600

Simonsen SM, Sando L, Ireland DC, et al (2005) A Continent of Plant Defense Peptide Diversity : Cyclotides in Australian Hybanthus ( Violaceae ). 17:3176–3189. doi: 10.1105/tpc.105.034678.1

Sun S, Wang H, Yu H, et al (2013) GASA14 regulates leaf expansion and abiotic stress resistance by modulating reactive oxygen species accumulation. J Exp Bot 64:1637–1647. doi: 10.1093/jxb/ert021

Tamura K, Peterson D, Peterson N, et al (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–9. doi: 10.1093/molbev/msr121

Tavares LS, Santos MDO, Viccini LF, et al (2008) Biotechnological potential of antimicrobial peptides from flowers. Peptides 29:1842–51. doi: 10.1016/j.peptides.2008.06.003

Taylor BH, Scheuring CF (1994) A molecular marker for lateral root initiation : The RSI-1 gene of tomato ( Lycopersicon esculentum Mill ) is activated in early lateral root primordia. Mol Genet Genomics 243:148–157.

Taylor RD, Koo WW (2011) 2011 Outlook of the U . S . and World Corn and Soybean Industries , 2010-2020. Agribus Appl Econ Rep 2–27.

Torres MA (2010) ROS in biotic interactions. Physiol Plant 138:414–29. doi: 10.1111/j.1399-3054.2009.01326.x

Veronese P, Ruiz MT, Coca MA, et al (2003) In Defense against Pathogens . Both Plant Sentinels and Foot Soldiers Need to Know the Enemy. Plant Physiol 131:1580–1590. doi: 10.1104/pp.102.013417.1580

Wanderley-nogueira AC, Belarmino LC, Soares-cavalcanti NM, et al (2012) An overall evaluation of the Resistance ( R ) and Pathogenesis-Related ( PR ) superfamilies in soybean , as compared with Medicago and Arabidopsis. Genet Mol Biol 35:260–271.

Wigoda N, Ben-Nissan G, Granot D, et al (2006) The gibberellin-induced, cysteine-rich protein GIP2 from Petunia hybrida exhibits in planta antioxidant activity. Plant J 48:796–805. doi: 10.1111/j.1365-313X.2006.02917.x

Zare-zardini H, Tolueinia B, Hashemi A, Fesahat F (2013) Antioxidant and Cholinesterase From Ziziphus jujuba Fruits. Am J Alzheimer ' s Dis Other Dementias 4–12. doi: 10.1177/1533317513500839

Zasloff M (2002) Antimicrobial peptides of multicellular organisms. Nature 415:389–395.

Zhang S, Wang X (2008) Expression pattern of GASA, downstream genes of DELLA, in Arabidopsis. Chinese Sci Bull 53:3839–3846. doi: 10.1007/s11434-008-0525-9

Zhang S, Yang C, Peng J, et al (2009) GASA5, a regulator of flowering time and stem growth in Arabidopsis thaliana. Plant Mol Biol 69:745–59. doi: 10.1007/s11103-009-9452-7

Zilli JÉ, Valisheski RR, Rodrigues F, et al (2004) ASSESSMENT OF COWPEA RHIZOBIUM DIVERSITY IN CERRADO AREAS OF NORTHEASTERN BRAZIL. Brazilian J Microbiol 35:281–287.

Zimmermann R, Sakai H, Hochholdinger F (2010) The Gibberellic Acid Stimulated-Like gene family in maize and its role in lateral root development. Plant Physiol 152:356–65. doi: 10.1104/pp.109.149054

**CAPÍTULO II**

**Artigo a ser submetido à revista *Functional & Integrative Genomics***

---

# Prospection, expression and comparative genomics: A search for snakin in cowpea transcriptome.

Marx Oliveira-Lima[1]; Ana Maria Benko-Iseppon[1]; Valesca Pandolfi[1*].

[1] Genetics Department, Center of Biological Sciences, Universidade Federal de Pernambuco, Av. Prof. Morais Rego 1235, CEP 50.670-420, Recife, PE, Brazil

Corresponding author*: E-mail: valescapandolfi@gmail.com

**Abstract**

Snakins are AMPs - antimicrobial peptides, belonging to the Snakin/GASA family, composed of about 90 residues and 12 conserved cysteine, which the expression can be constitutive or induced by pathogens, as observed in *Solanum tuberosum*. The present study aims to catalogue and characterize snakin genes based on transcriptome datas from cowpea, as well as analyze its structure and expression, aiming identification and screening of useful elements to programs intending an improvement of those cultures and a late biotechnology use of the isolated peptides. Using protein sequences from Snakin/GASA family members, it was made a search for homologues in the culture's transcriptome, represented by expressed sequence tags (EST), RNA-Seq and superSAGE libraries, also it was made a comparative genomic analysis. A total of 28 possible Snakins were identified on transcriptome, presenting great structural homology with genes fro this family (Sankin/GASA), spread in 6 from 11 chromosomes of *Phaseolus vulgaris*, reflecting possible genomic rearrangements in which the species suffered. The expression of these genes seems to be associated with the response to biotic and abiotic stress. The serial analysis shows that either for biotic or abiotic stress, the gene expression reflects patterns found to the family. Even beyond, trough the similarity analysis it was verified a wide structural diversity of these antimicrobial peptides.

**Keywords**: NordEST; Data mining; Antimicrobial peptides; GASA domain.

**Running title:** prospection, expression and comparative genomics of snakin.

## Introduction

During its evolution, plants have developed a complex pool of biochemical defenses against pathogens, however, for several pathogenic microorganisms, defense mechanisms are still inefficient, leading to disease development (Mysore and Ryu 2004). The class of molecules that have a remarkable function in the plant defense is the antimicrobial peptides (AMPs), since they have been considered as the main weapon of plants both, constitutive and induced by biotic and abiotic stress (García-Olmedo et al. 1998; Benko-Isepponet al. 2010). The AMPs are divided in distinct subfamilies, which include: defensins, thionins, knottin-like, maize basic protein, lipid transfer protein, hevein-like, cyclotides and snakins, for example (Manners, 2007). In general, they have between 20 to 120 aminoacids and can be prolin, glycine, histidine, arginine, tryptophan or cysteine rich (Bulet et al. 2004).

Among the cysteine rich peptides, the Snakins were recently described (Segura et al. 1999; Berrocal-lobo et al. 2002; Mao et al. 2011), as a group of peptides that share large structural similarity with the GASA family members (Ben-Nissan et al. 2004; Furukawa et al. 2006; Wigoda et al. 2006; Zimmermann et al. 2010; Rubinovich and Weiss 2010; Li et al. 2011). This gene family encodes a peptide of low molecular weight, which differs from the others AMPs for possessed 12 conserved cysteines, that stabilize the mature peptide, forming six disulfide bridges (Porto and Franco 2013). Furthermore, these particular AMPs exhibits *in vitro* activity against a wide range of bacteria and fungi (Segura et al. 1999; Berrocal-lobo et al. 2002; Kovalskaya and Hammond 2009; Faccio et al. 2011). Nevertheless its mode of action is still unknown, having few sequences identified in biological databases, in comparison with others AMPs (Manners 2007).

The cowpea bean (*Vigna unguiculata*) is one of the most important crops in the world, mainly due to its nutritional importance for low-income populations (Zilli et al. 2004). Thus, considering the increasing loss of productivity due to biotic and abiotic agents, a Brazilian initiative Consortium have been formed - the NordEST database Project, a functional genomics analyze, structural and comparative of *V.unguiculata,* that contains more than 30000 expressed sequence tags, 20 million of SuperSAGE tags and almost 500 million RNA-Seq sequences (Benko-Iseppon, personal comunication).

Based on these data, this work aims to identify and characterize snakins genes in the transcriptome of this crop, describing their structural and functional patterns.

## Material and Methods

Data Mining and sequence analysis

A manual text mining was performed in order to build a local databank of seed sequences to be used in a search in the NordEST platform, by screening the databases of ESTs (*Expressed Sequence Tags*) and RNA-Seq, of the Santo Inácio and Pingo de Ouro strains, via tBLASTn (Altschul et al. 1990). The results were selected all the matches in the BLAST alignments (e-value ≤ e-4).

**Table 1 -** *Seed* sequences used to screen the NordEST database.

| Description | Organism | UniProt ID | Reference |
|---|---|---|---|
| **RSI-1** | *Solanum lycopersicum* | P47926 | (Taylor and Scheuring 1994) |
| **GAST1** | *Solanum lycopersicum* | P27057 | (Shi et al. 1992) |
| **Gip1** | *Petunia hybrida* | Q43615 | (Ben-Nissan and Weiss 1996) |
| **Gip2** | *Petunia hybrida* | Q9FR10 | (Ben-Nissan et al. 2004) |
| **Gip3** | *Petunia hybrida* | Q93WR6 | (Ben-Nissan and Weiss, 1996) |
| **Gip4** | *Petunia hybrida* | Q93WR5 | (Ben-Nissan et al, 2004) |
| **Gip5** | *Petunia hybrida* | Q93WR4 | (Ben-Nissan et al, 2004) |
| **GASA1** | *Arabidopsis thaliana* | P46689 | (Herzog et al. 1995) |
| **GASA2** | *Arabidopsis thaliana* | P46688 | (Herzog et al. 1995) |
| **GASA3** | *Arabidopsis thaliana* | P46687 | (Herzog et al. 1995) |
| **GASA4** | *Arabidopsis thaliana* | P46690 | (Herzog et al. 1995) |
| **GASA5** | *Arabidopsis thaliana* | Q84J95 | (Aubert et al. 1998) |
| **GASA6** | *Arabidopsis thaliana* | Q6NMQ7 | (Aubert et al, 1998) |
| **GASA7** | *Arabidopsis thaliana* | O82328 | (Berrocal-lobo et al. 2002) |
| **GASA8** | *Arabidopsis thaliana* | O80641 | (Berrocal-Lobo et al, 2002) |
| **GASA9** | *Arabidopsis thaliana* | Q8GWK5 | (Roxrud et al. 2007) |
| **GASA10** | *Arabidopsis thaliana* | Q8LFM2 | (Roxrud et al, 2007) |
| **GASA11** | *Arabidopsis thaliana* | F4IQJ4 | (Roxrud et al, 2007) |
| **GASA12** | *Arabidopsis thaliana* | Q6GKX7 | (Roxrud et al, 2007) |
| **GASA13** | *Arabidopsis thaliana* | A8MR46 | (Roxrud et al, 2007) |
| **GASA14** | *Arabidopsis thaliana* | Q9LFR3 | (Roxrud et al, 2007) |
| **Snakin-1** | *Solanum tuberosum* | Q948Z4 | (Segura et al. 1999) |

*Table 1 continued*

| | | | |
|---|---|---|---|
| **Snakin-2** | *Solanum tuberosum* | Q93X17 | (Berrocal-Lobo et al, 2002) |
| **Snakin-like** | *Phaseolus vulgaris* | Q2YHP6 | (Bindschedler et al. 2006) |
| **OsGASR1** | *Oryza sativa* | Q94HA1 | (Furukawa et al. 2006) |
| **OsGASR2** | *Oryza sativa* | Q7X885 | (Furukawa et al. 2006) |
| **GEG** | *Gerbera hybrida* | Q9XGJ3 | (Kotilainen et al. 1999) |
| **FaGAST1** | *Fragraria ananassa* | O49134 | (de la Fuente et al. 2006) |
| **FsGASA** | *Fagus sylvativa* | Q0VYL5 | (Alonso-Ramírez et al. 2009a) |
| **GsGASA** | *Glycine soja* | F1BXA4 | (Li et al. 2011) |

A manual search for redundancies was carried out in order to select the clusters that matched only once. The annotated nucleotide sequences found in NordEST database were confronted against the non-redundant protein database of NCBI, through reciprocal alignments using the BLASTx tool. On the other hand, the aminoacid sequences were submitted to Batch CD-Search tool (Marchler-Bauer et al. 2011) to identification and evaluation of the conserved domain, thus, only the sequences that possessed complete domain were used in the subsequent analysis.

With the aim to elucidate the structure and function of these peptides, the aminoacid sequences of the orthologues were subjected to the signalP tool (Petersen et al. 2011), in order to infer the signal peptide and the mature peptide, as well as, the molecular weight (MW) and isoelectric point (pI) were visualized using the Protein Calculator v3.3 (Putnam, 2008) whereas the subcellular localization was predicted with PROTCOMP tool (Klee and Ellis 2005).

Expression profile

The *in silico* expression profile was assembled by direct reading of the 9 RNA-Seq libraries from NordEST platform (Table 2). They were grouped by library/tissue of expression and its relative frequencies were traced considering the amount of reads per library. The *snakin* homologues that were found in *Vigna unguiculata* had its expression profile normalized by hierarchical clustering using CLUSTER 3.0 tool (Eisen et al. 1998), the heat map of the profile was generated using the Java TreeView (Saldanha, 2004).

**Table 2 -** Description of the libraries and treatments applied to the related genotypes with RNA-seq database contained in the NordEST.

| Genotype | Library | Description |
|---|---|---|
| ***Santo Inácio*** | RI0 | Root – control |
| ***Santo Inácio*** | RI75 | Root - 75 min after water stress |
| ***Santo Inácio*** | RI150 | Root - 150 min after water stress |
| ***Santo Inácio*** | FI0 | Leaves – Control |
| ***Pingo de Ouro*** | RP0 | Root – Control |
| ***Pingo de Ouro*** | RP75 | Root - 75 min after water stress |
| ***Pingo de Ouro*** | RP150 | Root - 150 min after water stress |
| ***Pingo de Ouro*** | FP75 | Leaves - 75 min after water stress |
| ***Pingo de Ouro*** | FP150 | Leaves - 150 min after water stress |

The NordEST databank is comprised of four SuperSAGE libraries and allowed the generation of fourteen comparisons, including four from leave tissues subjected to mosaic virus (CpSMV - *Cowpea severe mosaic virus*), four inoculated with *Potyvirus Cowpea aphid-borne mosaic virus* (CABMV), two with leaves wounded, two comparisons of SuperSAGE libraries, that comprise experiments from root tissues, submitted to salinity and more two submitted to drought.

To evaluate the snakin-related tags represented in the SuperSAGE libraries, a comparative analysis using all characterized snakin ESTs was carried with a local BLASTn, according to Altschul *et al.* (1990) against a differential expressed TAGs database. For this purpose, the parameters were adjusted and low complexity filter was deactivated. For the results, only tags with identity equal to or larger than 25 bp, without mismatches on the first four bases or gaps in any position were considered.

The differential expressed *Tags* identified in our study were used in further analysis, aiming the identification of activation pattern related with snakin antimicrobial peptide.

*In silico* hybridization

A comparative alignment using tBLASTn tool was carried out against the *Phaseolus vulgaris* genome in phytozome database (Goodstein et al. 2012) with the peptide sequences from ESTs that presented complete domain. Further, matches that reported proteins with Snakin/GASA domain were annotated for quality of alignment and position in virtual chromosomes, show in GBrowser, in order to assemble the digital ideogram. Information about

the structure of the genes (number of exons and introns) as well as verification of the presence of alternative transcripts were also collected.

Phenetic analysis

The analisys of phenetic relations between the seed sequences and the *Vigna unguiculata* protein sequences prospected in the NordEST database, were initially performed using the Clustalw (Larkin et al. 2007) a package from MEGA 5 software (Tamura et al. 2011), where the sequences of C-Terminal domain were aligned. Further, in order to generate the phenetic tree, the Neighbor-Joinig algorithm, considering 2000 replications with bootstrap analysis was employed.

**Results**

Data Mining and sequence analysis

Due to the investigation realized in the NordEST database, we identified a total of 47 sequences, 12 from “Pingo de Ouro”, 20 from “Santo Inácio”, both RNA-seq libraries, and 15 from EST libraries. However, after analyzing the presence and integrity of the conserved domains, we observed 28 sequences with GASA domain and these were submitted to ensuing analysis.

The non-redundant sequences possessed size varying between 481-965 pb for nucleotide sequences and 65 to138 for amino acid sequences. It was shown by the Protein Calculator that the mature peptides possessed a MW ranging between 7.25-16.95 KDa and the pI between 7.12-9.67, whereas the signal peptide showed values for these features between 2.3-3.97 for MW and 4.7-9.58. Only three representatives had the GASA domain incomplete in the N-Terminal region, whereas eight sequences did not indicate any result for signal peptide values (Table 1).

In relation to the subcellular localization analysis, all sequences except for the contig2951 possessed send signal to extracellular environment and all these data were supported by optimal scores (data not show).

**Table 3 -** Clusters identified in NordEST database with tBLASTn tool, using the seed sequences retrieved in the text mining, showing all the non-redundant candidates (e-value 0 - $1e^{-4}$) and with the GASA domain complete. The annotation performed with BLASTx tool in GenBank shows the accession and the size of the transcripts in nucleotide (nt) and aminoacid (aa). Corresponding values for molecular weight (MW) and isoelectric point (pI) from the signal peptide (SP) and mature peptide (MP) are also shown.

| Cluster (NordEST) | Acession (NCBI) | GASA Domain (Situation) | Size (nt/aa) | MW (SP/MP) | pI (SP/MP) |
|---|---|---|---|---|---|
| ***Contig17959*** | ADX36135 | Full | 504/101 | 3.31/7.93 | 5.5/9.09 |
| ***Contig13996*** | ACU15584 | Full | 959/132 | 2.82/11.85 | 9.06/9.09 |
| ***Contig11984*** | XP_003523021 | Full | 992/153 | 16.95 | 9.17 |
| ***Contig14974*** | ACF74297 | Full | 795/129 | 14.36 | 9.05 |
| ***Contig3533*** | XP_003550315 | Full | 937/119 | 13.01 | 8.72 |
| ***Contig14321*** | ACU16624 | Full | 637/99 | 3.73/7.25 | 9.58/8.36 |
| ***Contig1850*** | ACU20508 | N-terminal | 630/124 | 3.62/10.43 | 6.48/9.67 |
| ***Contig4868*** | ACU15001 | Full | 657/102 | 2.69/8.16 | 7.19/8.65 |
| ***Contig2951*** | AFK38851 | Full | 901/217 | 3.29/8.15 | 9.06/8.52 |
| ***Contig9961*** | ACU13692 | Full | 577/106 | 11.44 | 8.65 |
| ***Contig12454*** | ACU13226 | Full | 782/134 | 14.7 | 8.05 |
| ***Contig9985*** | XP_003549639 | Full | 685/131 | 3.97/10.19 | 8.22/7.12 |
| ***PO23707.1*** | XP_003523021 | Full | 796/112 | 2.79/9.62 | 8.22/8.96 |
| ***PO15067.1*** | ACU14995 | Full | 964/89 | 2.87/6.78 | 6.25/8.97 |
| ***PO25827.1*** | ACU16624 | Full | 620/90 | 2.69/7.25 | 6.25/8.36 |
| ***PO13171.1*** | ACU20508 | N-terminal | 591/113 | 2.3/10.43 | 4.7/9.67 |
| ***PO4671.1*** | ACU15001 | Full | 625/99 | 2.39/8.16 | 7.19/8.65 |
| ***PO4671.2*** | ACU15001 | Full | 757/80 | 8.71 | 8.52 |
| ***PO19528.1*** | XP_003549639 | Full | 729/115 | 2.37/10.19 | 8.22/7.12 |
| ***SI39039.1*** | XP_004493436 | Full | 729/121 | 2.8/10.38 | 8.22/8.61 |
| ***SI13272.1*** | XP_003523021 | Full | 941/112 | 2.79/9.62 | 8.22/8.96 |
| ***SI39107.1*** | ACU14995 | Full | 756/124 | 13.56 | 8.82 |
| ***SI31825.1*** | ACU16624 | Full | 728/90 | 2.69/7.25 | 6.25/8.36 |
| ***SI31825.2*** | XP_003536595 | Full | 620/67 | 7.4 | 8.54 |

*Table 3 - continued*

| | | | | | |
|---|---|---|---|---|---|
| ***SI26502.3*** | ACU20508 | N-terminal | 1128/113 | 2.3/10.43 | 4.47/9.67 |
| ***SI36869.1*** | ACU15001 | Full | 777/99 | 2.39/8.16 | 7.19/8.65 |
| ***SI9462.2*** | ACU13692 | Full | 673/99 | 2.58/8.15 | 9.06/8.52 |
| ***SI37871.1*** | XP_003549639 | Full | 1892/115 | 2.37/10.19 | 8.22/7.12 |

Expression profile

In the expression profile, the presence of transcripts were observed in all the 10 RNA-seq libraries screened from the two genotypes, where 9,158 corresponded to "Santo Inácio" genotype and 9,022 to "Pingo de Ouro" genotype, in a total of 1,8180 reads divided among tolerant stressed, sensitive stressed, tolerant control and sensitive control treatments. This allowed the identification of the prevalence and normalizing of their distribution among the different tissues and treatments, represented in the libraries.

The pattern of expression in the "Pingo de Ouro" genotype was almost the same for all libraries, showing differences only between clusters, while "Santo Inácio" genotype showed a marked level of expression in the root treatments, especially in control (Figure 1).

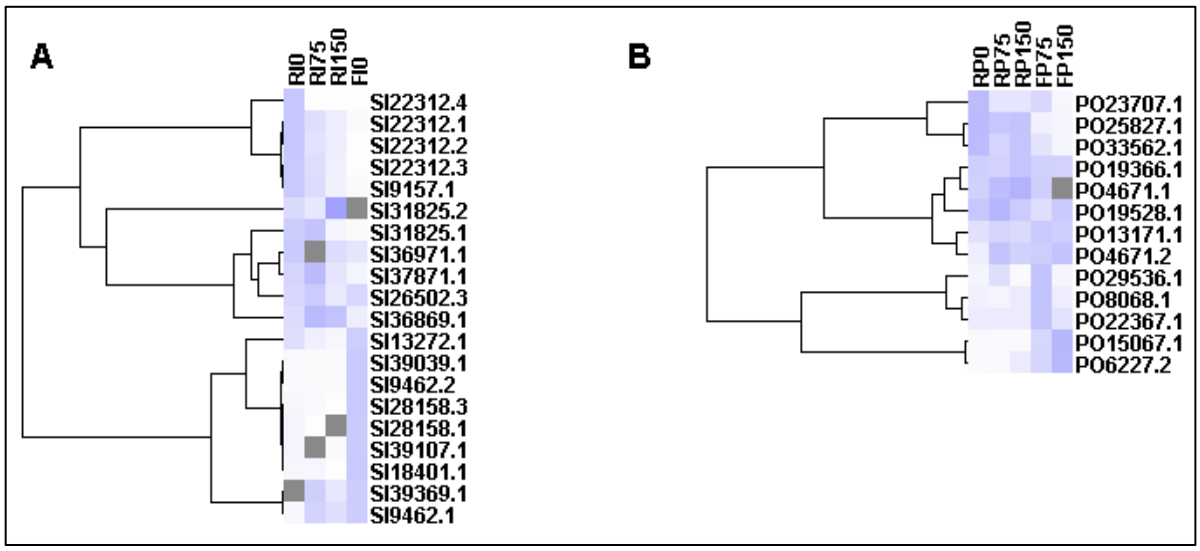



**Figure 1 -** Expression profile obtained after the hierarchical clustering (CLUSTER software). The regions in dark blue indicate high levels of expression, light blue shows lower levels of expression, white indicate absence of expression and grey indicates missing data. (A) shows the pattern of expression from SI (Santo Inácio) genotype; (B) shows the pattern of expression from PO (Pingo de Ouro) genotype.

After the screening of SuperSAGE libraries, we identified a total of 20 tags distributed in the 14 comparisons. It was also observed that the same transcripts appeared in different experiments, for presentation purposes, the treatments were bulked in four main types: infected (eight treatments), injured (two treatments), drought (two treatments) and salt (two treatments), with the absolute number of tags. The infected treatments grouped the highest number of differentially expressed transcripts, follow by drought, injured and salt (Figure 2).

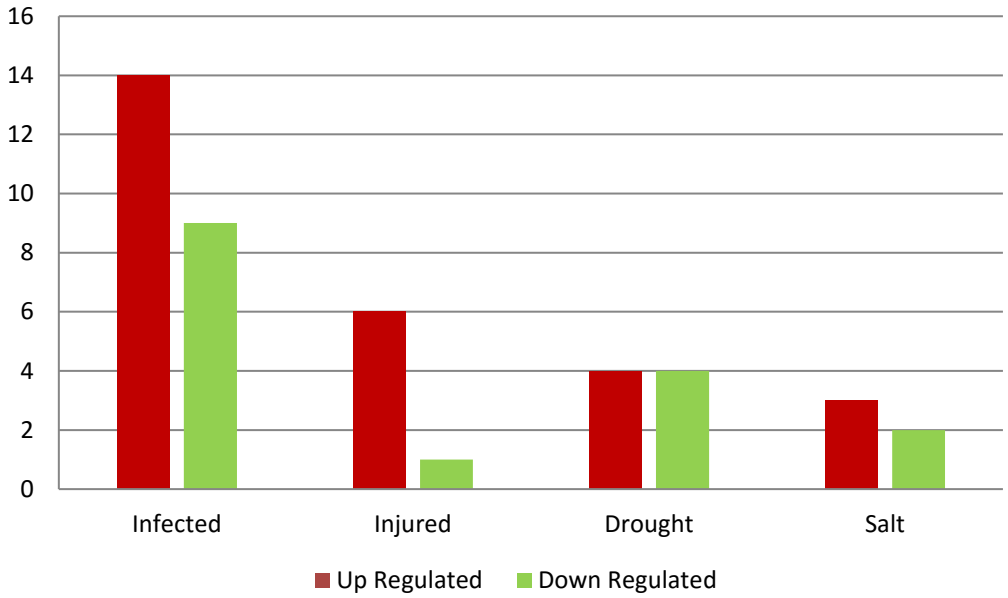



**Figure 2 -** Frequency of differentially expressed tags per set of treatment**.** The dark grey bar represents the up-regulated tags and the light grey bar represents the down-regulated tags.

*In silico* hybridization and comparative gene structure

The comparative analysis of *V. unguiculata* transcripts (12 ESTs) in the *Phaseolus vulgaris* genome allowed the identification of 24 regions, however, Since one of the regions, in the chromosome 1, did not show any homology, we plotted, in the digital ideogram, the 23 loci with identified similarity.

With respect to the genomic distribution of the homologues of snakin/GASA family, the chromosome 1 had more genes than any other, altogether nine genes, whereas, in the chromosomes 7 and 9 were identified four loci and in the chromosomes 3, 4 and 8 were found two genes. Most of these genes formed clusters mainly in the chromosome 9 (Figure 2).

Regarding the structure of the genes, we noticed three different types of arrangement of introns and exons: two exons and one intron, three exons and two introns and four exons and three introns. In total, 10 genes presented the structure with two exons and three introns, being this the most representative structure with no gene exhibiting alternative form (data not show).

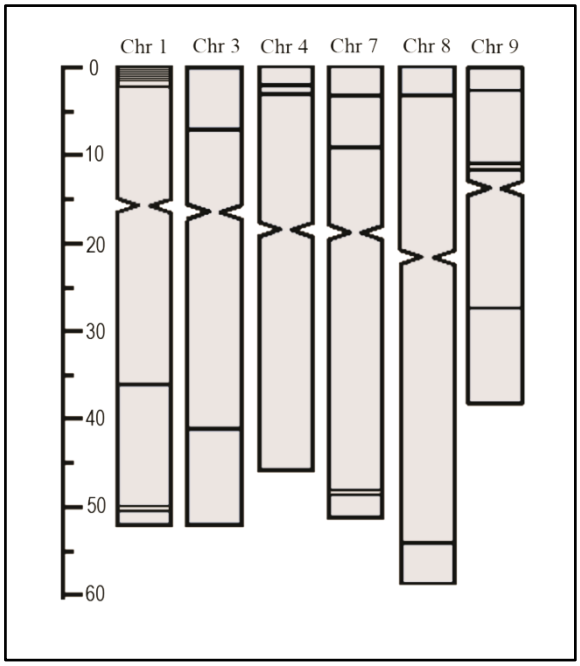



**Figure 3 -** Comparative ideogram of the sequences from cowpea in *P. vulgaris* genome. The 0 to 60 scale refers to genome size in magabases.

Phenetic analysis

The relations of similarity of the sequences found in *V. unguiculata* transcriptome comparatively with the seed sequences resulted in a tree with a topology with four main groups, where the groups 1 and 2 yielded the best statistics support in comparison with the groups 3 and 4 (Figure 3).

Additionally, 20 residues were observed with high conservation, 12 corresponding to the cysteine motif XCX3CX3CX8CX3CX2CCX2CXCX11CXCX12CX and eight to residues of R, V, P/A, G, G, Y, K, P. This pattern was observed in all sequences. Nevertheless, some variable residues were detected in domain region, especially between the residues 10-20 and 40-57 (Figure 4).

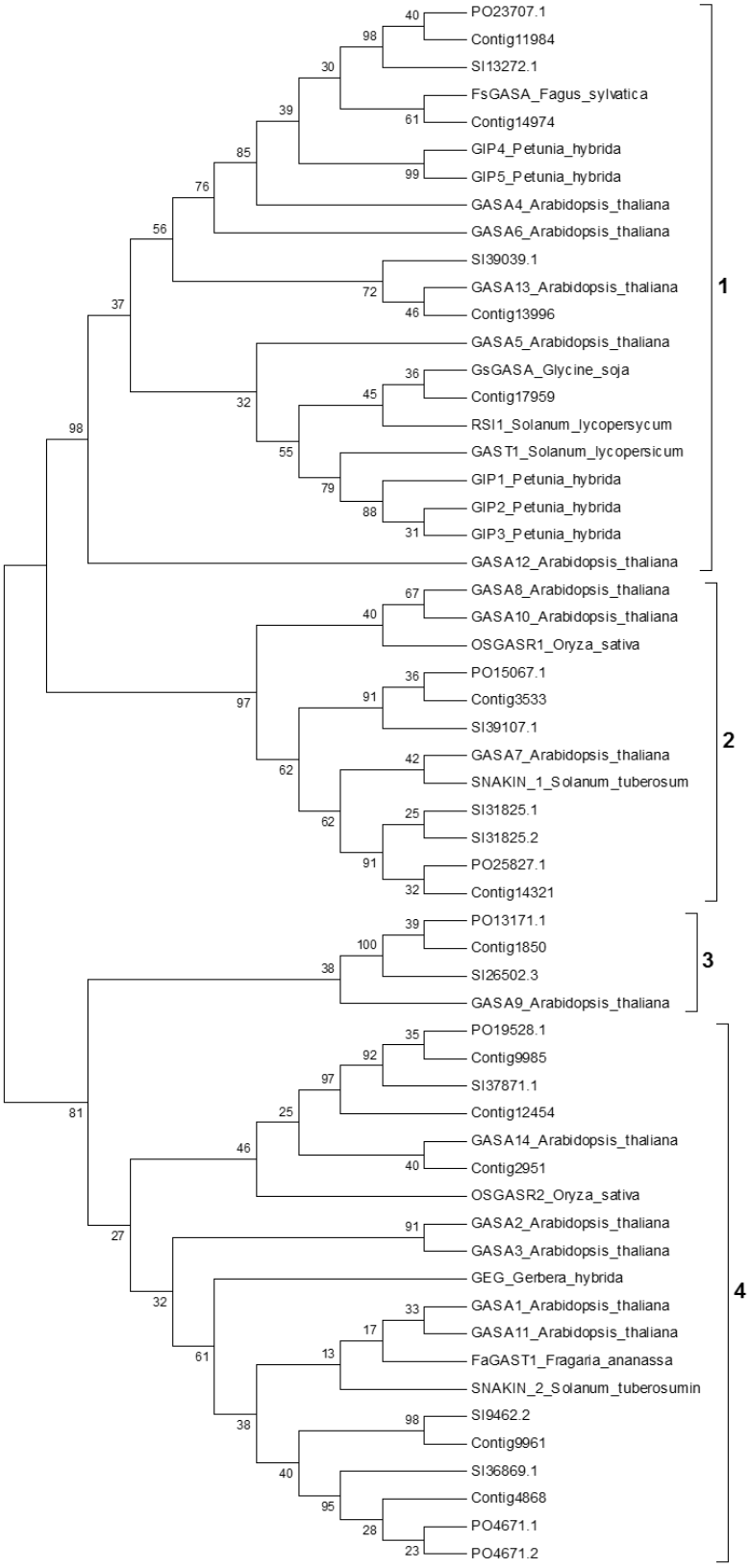



**Figure 4 -** Phenogram made by the comparisons between the seed sequences and the transcripts, where can be observed 4 main groups.

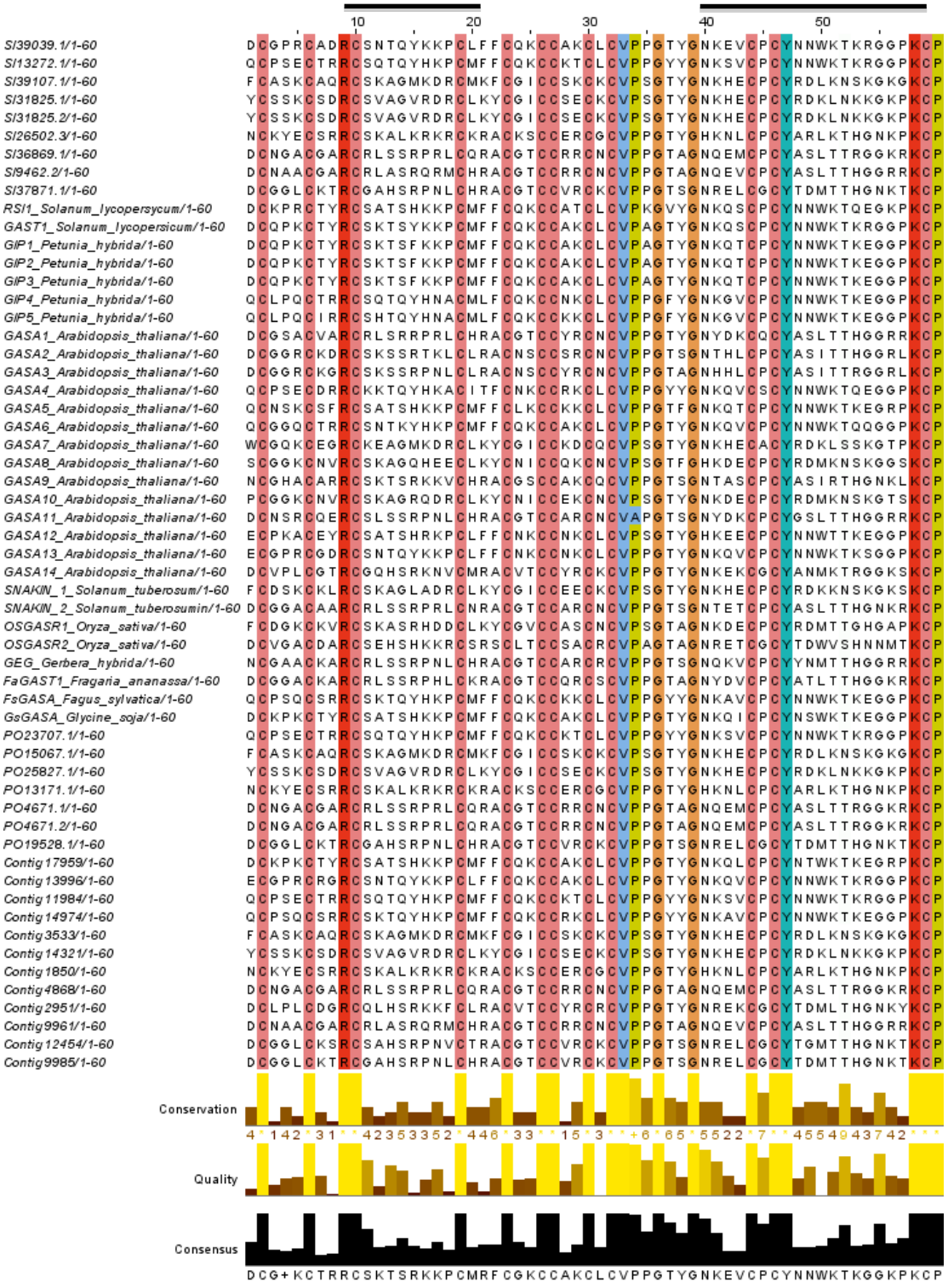



**Figure 5 -** Alignment performed with transcripts peptide sequences and the seeds, showing the main conserved residues. The black bars indicate the most variable sites.

## Discussion

### Data Mining and sequence analysis

The main structural aspects showing by the 28 sequences retrieved in the cowpea expressed genome, considering the cut-off adopted, makes plausible the idea that they were indeed members of the GASA family, once they share the signature cysteine-rich, which contains twelve aminoacids in highly conserved positions: XCX3CX3CX8CX3CX2CCX2CXCX11CXCX12CX, important to maintenance of the structure and function of this peptides, this arrangement highly conserved due to the importance of the disulfide bridges to the 3D conformation of the peptide (Herzog et al. 1995; Aubert et al. 1998; Jianzong et al. 2008; Porto and Franco 2013; Sun et al. 2013),

The subcellular localization showing by our data evidenced that they were almost exclusively addressed to extracellular environment, an information already described in some members of the family (Zhang et al. 2009; Moyano-Cañete et al. 2013) and also for other cysteine-rich antimicrobial peptides (Padovan et al. 2010).

Our findings also show a high rate of conservation in your structure with size, molecular weight and pI very close to the found in other members of the family. The basic value of pI in almost all sequences, was described in snakin-2 of potato and other GASA homologues, as well as in other antimicrobial peptides (Berrocal-lobo et al. 2002; Ferreira et al. 2007; Ko et al. 2007; Mao et al. 2011), which means that these features are very conserved between the AMPs and their homologues, it is known that many AMPs interact with cell membrane, thus the basic characteristic is important in the interaction with this structure.

The rates of molecular weight varied within has been observed for the group (Aubert et al. 1998; Berrocal-lobo et al. 2002; Jianzong et al. 2008). Besides the most conserved region described for these peptides, the C-Terminal domain (GASA) (Nahirñak et al. 2012a) that is shown in the alignment (Figure 5), demonstrate a strict conservation, a strong indication that these genes have passed to process that lead to a functional specialization.

### Expression profile

The expression pattern showed by read counting from RNA-seq libraries for these genes, revealed as was observed in other groups of antimicrobial peptides and other members of this family (Do et al. 2004; Sun et al. 2013), the genes homologues from cowpea play an

important role in response to abiotic stress, this can be observed by the accentuated levels of expression in all the treatments from "Pingo de Ouro" genotype, the tolerant strain. Whereas in the sensitive genotype ("Santo Inácio"), the expression was higher in the control group than in the treated groups, corroborating the idea of the response to abiotic stress. It is known that water stress induce oxidative stress (Carvalho 2008; Kar 2011) and many of the agents involved in the response to biotic stress are also related to the reaction of this stress (Torres 2010), as well as members of the GASA family in *Gerbera hybrida* and *Fagus sylvatica* (Wigoda et al. 2006; Alonso-Ramírez et al. 2009a).

In the treatments of biotic stress was observed higher frequency of up regulated tags, these experiments were carried out with infected leaves, a tissue where was observed for (Faccio et al. 2011 and Almasia et al. 2008) the presence of snakins with antimicrobial activity, however to our knowledge, there is no reports of the response of snakins against virus, being these data the first indicative of this action, it occurs probably because many metabolic pathways that are activated by biotic stress triggers present crosstalk responses with many other pathways, as Reactive Oxygen Species (ROS) (Torres 2010; Nahirñak et al. 2012b; Sun et al. 2013). It is known that ROS are involved also in wounding where snakins play a role in redox regulation, which therefore elevates the expression in wounded tissues (Nahirñak et al. 2012b; Balaji and Smart 2012), the same response is related for salinity and drought stress (Avsian-kretchmer et al. 2004; Carvalho 2008).

*In silico* hybridization and comparative gene structure

*Phaseolus vulgaris* as *Vigna unguiculata* and other members of papilionoid superfamily present evidences of an older shared duplication in their genomes, these duplicated regions are evident in number of similar duplicated genes and in large areas of synteny between chromosomal regions (Shoemaker et al. 2006), the heterogeneous distribution of the orthologues along the genome reinforces the hypothesis that the genomic location of the genes is a consequence of the genomic rearrangements.

Despite of their heterogeneous location, some of the genes were grouped in clusters which also reflects the evolutionary process undergone by the genome of common bean, however this kind of "behavior" seems to be common for defense genes, since it has been observed in genes that response to biotic stress (R and PR genes) were grouped in clusters among the different chromosomes of *Medicago truncatula* and *G. max*. The genomic arrangement in clusters is associated with common ancestry and the diversification of these

genes occurs by pathogen or environmental pressure (Wanderley-nogueira et al. 2012; McHale et al. 2012)

With respect to the location in the chromosomes, most of them were matched in the subtelomeric regions, nevertheless few genes are located in the pericentromeric regions of chromosome 9, which makes plausible the idea that *P. vulgaris* genome underwent similar events to the *G. max* making your pericentromeric regions become structurally and evolutionary labile and with few genes (Lin et al. 2005)

The three gene structures founded in the genome of *P. vulgaris* follow the pattern of the family (Sun et al. 2013; Meiyalaghan et al. 2014). Genes with their structures presenting one intron and two exons are commonly described among the antimicrobial peptides (Silverstein et al. 2007) and were also founded in the snakin/GASA family. In the genes that are components of constitutive defense barriers against pathogens in storage and reproductive plant organs (Segura et al. 1999; Faccio et al. 2011).

About the genes that present two introns and three exons, a similar structure was observed in StSN2 from potato (Berrocal-lobo et al. 2002) and in other AMPs (Silverstein et al. 2007), it is associated with the stages of development and in response to specific pathogens being an important component of both constitutive and induced defense (Mao et al. 2011; Balaji and Smart 2012; Mohan et al. 2013).

Finally, the structure with three introns and four exons is associated with genes that are expressed by hormonal stimulus in regions with activity related to growth (Aubert et al. 1998), but recently Sun et al. 2013 and Nahirñak, Almasia, Hopp, et al. 2012, reported an extra role in reaction to oxidative stress, which makes of these genes a good alternative against abiotic stress.

## Phenetic Analysis

In this comparison the snakin/GASA homologues proved to be highly divergent, in which the sequences are grouped in diversified branches, notwithstanding the phenetic tree maintained the same topology founded in other comparisons of the family (Furukawa et al. 2006; Roxrud et al. 2007; Li et al. 2011), which corroborate the existence of four clusters, however our results show that the basal groups (specially clusters 1 and 2) are better supported than the terminal groups. Perhaps this have occurred due to structural variation of the sequences that possessed only 20 conserved residues, while 40 are highly variable..

This structural diversification in the nature of the sequences, is probably due to the pressure that the plants suffered along the evolutive time in their ecological relations, since this

gene family is related with a range of processes as multiple organ growth (Kotilainen et al. 1999; Furukawa et al. 2006; Roxrud et al. 2007; Zhang et al. 2009; Zimmermann et al. 2010; Nahirñak et al. 2012a), response to hormones (Aubert et al. 1998; de la Fuente et al. 2006; Zhang and Wang 2008; Alonso-Ramírez et al. 2009b; Moyano-Cañete et al. 2013), biotic (Nahirñak et al. 2012b; Sun et al. 2013; Meiyalaghan et al. 2014) and abiotic stresses (Almasia et al. 2008; Mao et al. 2011; Mohan et al. 2013), it was expected that their primary structure exhibit a large variation.

## Conclusions

By using of bioinformatics approaches, could be identified a new antimicrobial peptide of cowpea, these resources will be valuable to develop markers for beans and other plants, specially because of the presence of clusters, founded in the comparative genomic analysis. These findings constitute an important resource for the improvement of the culture and the biotechnological application of these peptides.

## Acknowledgments

The authors are grateful to National Council for Scientific and Technological Development (CNPq), to National Council for the Improvement of Higher Education (CAPES) and to Foundation for Support of Science and Technology of Pernambuco State - Brazil (FACEPE) for supporting our research.

## References

Almasia NI, Bazzini AA, Hopp HE, Vazquez-Rovere C (2008) Overexpression of snakin-1 gene enhances resistance to Rhizoctonia solani and Erwinia carotovora in transgenic potato plants. Mol Plant Pathol 9:329–338. doi: 10.1111/J.1364-3703.2008.00469.X

Almasia NI, Narhirñak V, Hopp HE, Vazquez-Rovere C (2010) Isolation and characterization of the tissue and development-specific potato snakin-1 promoter inducible by temperature and wounding. Electron J Biotechnol 13:1–21. doi: 10.2225/vol13-issue5-fulltext-12

Alonso-Ramírez A, Rodríguez D, Reyes D, et al (2009a) Evidence for a Role of Gibberellins in Salicylic Acid-Modulated Early Plant Responses to Abiotic Stress in Arabidopsis Seeds 1. Plant Physiol 150:1335–1344. doi: 10.1104/pp.109.139352

Alonso-Ramírez A, Rodríguez D, Reyes D, et al (2009b) Cross-talk between gibberellins and salicylic acid in early stress responses in Arabidopsis thaliana seeds. Plant Signal Behav 4:750–1. doi: 10.1104/pp.109.139352

Altschul SF, Gish W, Miller W, et al (1990) Basic local alignment search tool. J Mol Biol 215:403–10. doi: 10.1016/S0022-2836(05)80360-2

Aubert D, Chevillard M, Dorne a M, et al (1998) Expression patterns of GASA genes in Arabidopsis thaliana: the GASA4 gene is up-regulated by gibberellins in meristematic regions. Plant Mol Biol 36:871–83.

Avsian-kretchmer O, Gueta-dahan Y, Lev-yadun S, Gollop R (2004) The Salt-Stress Signal Transduction Pathway That Activates the gpx1 Promoter Is Mediated by Intracellular H 2 O 2 , Different from the Pathway Induced by Extracellular H 2 O 2 [ w ]. 135:1685–1696. doi: 10.1104/pp.104.041921.including

Balaji V, Smart CD (2012) Over-expression of snakin-2 and extensin-like protein genes restricts pathogen invasiveness and enhances tolerance to Clavibacter michiganensis subsp. michiganensis in transgenic tomato (Solanum lycopersicum). Transgenic Res 21:23–37. doi: 10.1007/s11248-011-9506-x

Belarmino LC, Capriles PVSZ, Crovella S, et al (2010) EST-Database Search of Plant Defensins – An Example Using Sugarcane , a Large and Complex Genome. Curr Protein Pept Sci 11:1–7.

Benko-Iseppon AM, Galdino SL, Calsa T, et al (2010) Overview on plant antimicrobial peptides. Curr Protein Pept Sci 11:181–8.

Ben-Nissan G, Lee J-Y, Borohov A, Weiss D (2004) GIP, a Petunia hybrida GA-induced cysteine-rich protein: a possible role in shoot elongation and transition to flowering. Plant J 37:229–238. doi: 10.1046/j.1365-313X.2003.01950.x

Ben-Nissan G, Weiss D (1996) The petunia homologue of tomato gast1: transcript accumulation coincides with gibberellin-induced corolla cell elongation. Plant Mol Biol 32:1067–74.

Berrocal-lobo M, Segura A, Moreno M, et al (2002) Snakin-2 , an Antimicrobial Peptide from Potato Whose Gene Is Locally Induced by Wounding and Responds to Pathogen Infection 1. Plant Physiol 128:951–961. doi: 10.1104/pp.010685.1

Bindschedler L V, Whitelegge JP, Millar DJ, Bolwell GP (2006) A two component chitin-binding protein from French bean -- association of a proline-rich protein with a cysteine-rich polypeptide. FEBS Lett 580:1541–6. doi: 10.1016/j.febslet.2006.01.079

Bulet P, Stöcklin R (2005) Insect antimicrobial peptides: structures, properties and gene regulation. Protein Pept Lett 12:3–11.

Bulet P, Stöcklin R, Menin L (2004) Anti-microbial peptides: from invertebrates to vertebrates. Immunol Rev 198:169–84.

Cammue BPA, Bolle MFC De, Terrasi FRG, et al (1992) Isolation and Characterization of a Novel Class of Plant Antimicrobial Peptides from Mirabilis jalapa L. Seeds. J Biol Chem 267:2228–2233.

Carvalho MHC (2008) Drought stress and reactive oxygen species. Plant Signal Behav 3:156–165.

Dangl JL, Jones JDG (2001) Plant pathogens and integrated defence responses to infection. Nature 411:826–833.

De la Fuente JI, Amaya I, Castillejo C, et al (2006) The strawberry gene FaGAST affects plant growth through inhibition of cell elongation. J Exp Bot 57:2401–11. doi: 10.1093/jxb/erj213

Do HM, Lee SC, Jung HW, et al (2004) Differential expression and in situ localization of a pepper defensin ( CADEF1 ) gene in response to pathogen infection , abiotic elicitors and environmental stresses in Capsicum annuum ℭ. 166:1297–1305. doi: 10.1016/j.plantsci.2004.01.008

Do Nascimento LC, Costa GGL, Binneck E, et al (2012) A web-based bioinformatics interface applied to the GENOSOJA Project: Databases and pipelines. Genet Mol Biol 35:203–11. doi: 10.1590/S1415-47572012000200002

Droit A, Poirier GG, Hunter JM (2005) Experimental and bioinformatic approaches for interrogating protein-protein interactions to determine protein function. J Mol Endocrinol 34:263–80. doi: 10.1677/jme.1.01693

Edwards D, Batley J (2004) Plant bioinformatics: from genome to phenome. Trends Biotechnol 22:232–7. doi: 10.1016/j.tibtech.2004.03.002

Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95:14863–8. doi: 10.1073/pnas.95.25.14863

Faccio P, Hopp E, Favret E, Franzone P (2011) Increased Tolerance to Wheat Powdery Mildew by Heterologous Constitutive Expression of the Solanum chacoense snakin-1 Gene. Czech J Genet Plant Breed 47:135–141.

Ferreira RB, Monteiro S, Regina F, et al (2007) The role of plant defence proteins in fungal pathogenesis. Mol Plant Pathol 8:677–700. doi: 10.1111/J.1364-3703.2007.00419.X

Franco OL (2011) Peptide promiscuity: an evolutionary concept for plant defense. FEBS Lett 585:995–1000. doi: 10.1016/j.febslet.2011.03.008

Furukawa T, Sakaguchi N, Shimada H (2006) Two OsGASR genes, rice GAST homologue genes that are abundant in proliferating tissues, show different expression patterns in developing panicles. Genes Genet Syst 81:171–80.

García-Olmedo F, Molina a, Alamillo JM, Rodríguez-Palenzuéla P (1998) Plant defense peptides. Biopolymers 47:479–91. doi: 10.1002/(SICI)1097-0282(1998)47:6<479::AID-BIP6>3.0.CO;2-K

Goodstein DM, Shu S, Howson R, et al (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40:D1178–86. doi: 10.1093/nar/gkr944

Guzmán-rodríguez JJ, Ibarra-laclette E, Herrera-estrella L, et al (2013) Analysis of expressed sequence tags ( ESTs ) from avocado seed ( Persea americana var . drymifolia ) reveals abundant expression of the gene encoding the antimicrobial peptide snakin. Plant Physiol Biochem 70:318–324.

Herzog M, Dorne a M, Grellet F (1995) GASA, a gibberellin-regulated gene family from Arabidopsis thaliana related to the tomato GAST1 gene. Plant Mol Biol 27:743–52.

Jianzong P, Liujing LAI, Xiaojing W (2008) PRGL : A cell wall proline-rich protein containning GASA domain in Gerbera hybrida. 51:520–525. doi: 10.1007/s11427-008-0067-z

Kar RK (2011) Plant responses to water stress: role of reactive oxygen species. Plant Signal Behav 6:1741–5. doi: 10.4161/psb.6.11.17729

Keymanesh K, Soltani S, Sardari S (2009) Application of antimicrobial peptides in agriculture and food industry. World J Microbiol Biotechnol 25:933–944. doi: 10.1007/s11274-009-9984-7

Klee EW, Ellis LBM (2005) Evaluating eukaryotic secreted protein prediction. BMC Bioinformatics 6:256. doi: 10.1186/1471-2105-6-256

Ko C-B, Woo Y-M, Lee DJ, et al (2007) Enhanced tolerance to heat stress in transgenic plants expressing the GASA4 gene. Plant Physiol Biochem 45:722–8. doi: 10.1016/j.plaphy.2007.07.010

Kotilainen M, Helariutta Y, Mehto M, et al (1999) GEG participates in the regulation of cell and organ shape during corolla and carpel development in gerbera hybrida. Plant Cell 11:1093–104.

Kovalskaya N, Hammond RW (2009) Expression and functional characterization of the plant antimicrobial snakin-1 and defensin recombinant proteins. Protein Expr Purif 63:12–17. doi: 10.1016/j.pep.2008.08.013

Larkin M a, Blackshields G, Brown NP, et al (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–8. doi: 10.1093/bioinformatics/btm404

Li K-L, Bai X, Li Y, et al (2011) GsGASA1 mediated root growth inhibition in response to chronic cold stress is marked by the accumulation of DELLAs. J Plant Physiol 168:2153–60. doi: 10.1016/j.jplph.2011.07.006

Lin J-Y, Jacobus BH, SanMiguel P, et al (2005) Pericentromeric regions of soybean (Glycine max L. Merr.) chromosomes consist of retroelements and tandemly repeated DNA and

are structurally and evolutionarily labile. Genetics 170:1221–30. doi: 10.1534/genetics.105.041616

Lucau-Danila A, Laborde L, Legrand S, et al (2010) Identification of novel genes potentially involved in somatic embryogenesis in chicory (Cichorium intybus L.). BMC Plant Biol 10:122. doi: 10.1186/1471-2229-10-122

Manners JM (2007) Hidden weapons of microbial destruction in plant genomes. Genome Biol 8:225. doi: 10.1186/gb-2007-8-9-225

Mao Z, Zheng J, Wang Y, et al (2011) The new CaSn gene belonging to the snakin family induces resistance against root-knot nematode infection in pepper. Phytoparasitica 39:151–164. doi: 10.1007/s12600-011-0149-5

Marchler-Bauer A, Lu S, Anderson JB, et al (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res 39:D225–9. doi: 10.1093/nar/gkq1189

Maróti G, Kereszt A, Kondorosi E, Mergaert P (2011) Natural roles of antimicrobial peptides in microbes, plants and animals. Res Microbiol 162:363–74. doi: 10.1016/j.resmic.2011.02.005

McHale LK, Haun WJ, Xu WW, et al (2012) Structural variants in the soybean genome localize to clusters of biotic stress-response genes. Plant Physiol 159:1295–308. doi: 10.1104/pp.112.194605

Meiyalaghan S, Thomson SJ, Fiers MW, et al (2014) Structure and expression of GSL1 and GSL2 genes encoding gibberellin stimulated-like proteins in diploid and highly heterozygous tetraploid potato reveals their highly conserved and essential status. BMC Genomics 15:2. doi: 10.1186/1471-2164-15-2

Messing J, Bharti AK, Karlowski WM, et al (2004) Sequence composition and genome organization of maize. Proc Natl Acad Sci 101:14349–14354.

Mohan S, Meiyalaghan S, Latimer JM, et al (2013) GSL2 over-expression confers resistance to Pectobacterium atrosepticum in potato. TAG Theor Appl Genet 127:677–89. doi: 10.1007/s00122-013-2250-2

Moyano-Cañete E, Bellido ML, García-Caparrós N, et al (2013) FaGAST2, a strawberry ripening-related gene, acts together with FaGAST1 to determine cell size of the fruit receptacle. Plant Cell Physiol 54:218–36. doi: 10.1093/pcp/pcs167

Mysore KS, Ryu C-M (2004) Nonhost resistance: how much do we know? Trends Plant Sci 9:97–104. doi: 10.1016/j.tplants.2003.12.005

Nahirñak V, Almasia NI, Fernandez PV, et al (2012a) Potato snakin-1 gene silencing affects cell division, primary metabolism, and cell wall composition. Plant Physiol 158:252–63. doi: 10.1104/pp.111.186544

Nahirñak V, Almasia NI, Hopp HE, Vazquez-rovere C (2012b) Involvement in hormone crosstalk and redox homeostasis Snakin / GASA proteins. Plant Signal Behav 7:1004–1008.

Padovan L, Scocchi M, Tossi A (2010) Structural aspects of plant antimicrobial peptides. Curr Protein Pept Sci 11:210–9.

Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8:785–6. doi: 10.1038/nmeth.1701

Phoenix DA, Dennison SR, Harris F (2013) Antimicrobial Peptides : Their History , Evolution , and Functional Promiscuity. Antimicrobial peptides, 1st edn. Weinheim, pp 1–37

Ponz F, Paz-ares J, Hernandez-lucas C, et al (1986) Cloning and nucleotide sequence of a cDNA encoding the precursor of the barley toxin a-hordothionin. 135:131–135.

Porto WF, Franco OL (2013) Peptides Theoretical structural insights into the snakin / GASA family. Peptides 44:163–167. doi: 10.1016/j.peptides.2013.03.014

Porto WF, Souza V a, Nolasco DO, Franco OL (2012) In silico identification of novel hevein-like peptide precursors. Peptides 38:127–36. doi: 10.1016/j.peptides.2012.07.025

Roxrud I, Lid SE, Fletcher JC, et al (2007) GASA4, one of the 14-member Arabidopsis GASA family of small polypeptides, regulates flowering and seed development. Plant Cell Physiol 48:471–83. doi: 10.1093/pcp/pcm016

Rubinovich L, Weiss D (2010) The Arabidopsis cysteine-rich protein GASA4 promotes GA responses and exhibits redox activity in bacteria and in planta. Plant J 64:1018–27. doi: 10.1111/j.1365-313X.2010.04390.x

Saldanha AJ (2004) Java Treeview--extensible visualization of microarray data. Bioinformatics 20:3246–8. doi: 10.1093/bioinformatics/bth349

Segura A, Moreno M, Madueño F, et al (1999) Snakin-1 , a Peptide from Potato That Is Active Against Plant Pathogens. Mol Plant-Microbe Interact 12:16–23.

Shi L, Gast RT, Gopalraj M, Olszewski NE (1992) Characterization of a shoot-specific, GA3- and ABA-regulated gene from tomato. Plant J 2:153–159.

Shoemaker RC, Schlueter J, Doyle JJ (2006) Paleopolyploidy and gene duplication in soybean and other legumes. Curr Opin Plant Biol 9:104–9. doi: 10.1016/j.pbi.2006.01.007

Shoemaker RC, Schlueter JA, Jackson SA (2008) Soybean Genome Structure and Organization. Genetics and Genomics of Soybean. pp 91–99

Silverstein K a T, Moskal W a, Wu HC, et al (2007) Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants. Plant J 51:262–80. doi: 10.1111/j.1365-313X.2007.03136.x

Silverstein KAT, Graham MA, Paape TD, Vandenbosch KA (2005) Genome Organization of More Than 300 Defensin-Like Genes in Arabidopsis. Plant Physiol 138:600–610. doi: 10.1104/pp.105.060079.600

Simonsen SM, Sando L, Ireland DC, et al (2005) A Continent of Plant Defense Peptide Diversity : Cyclotides in Australian Hybanthus ( Violaceae ). 17:3176–3189. doi: 10.1105/tpc.105.034678.1

Sun S, Wang H, Yu H, et al (2013) GASA14 regulates leaf expansion and abiotic stress resistance by modulating reactive oxygen species accumulation. J Exp Bot 64:1637–1647. doi: 10.1093/jxb/ert021

Tamura K, Peterson D, Peterson N, et al (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–9. doi: 10.1093/molbev/msr121

Tavares LS, Santos MDO, Viccini LF, et al (2008) Biotechnological potential of antimicrobial peptides from flowers. Peptides 29:1842–51. doi: 10.1016/j.peptides.2008.06.003

Taylor BH, Scheuring CF (1994) A molecular marker for lateral root initiation : The RSI-1 gene of tomato ( Lycopersicon esculentum Mill ) is activated in early lateral root primordia. Mol Genet Genomics 243:148–157.

Taylor RD, Koo WW (2011) 2011 Outlook of the U . S . and World Corn and Soybean Industries , 2010-2020. Agribus Appl Econ Rep 2–27.

Torres MA (2010) ROS in biotic interactions. Physiol Plant 138:414–29. doi: 10.1111/j.1399-3054.2009.01326.x

Veronese P, Ruiz MT, Coca MA, et al (2003) In Defense against Pathogens . Both Plant Sentinels and Foot Soldiers Need to Know the Enemy. Plant Physiol 131:1580–1590. doi: 10.1104/pp.102.013417.1580

Wanderley-nogueira AC, Belarmino LC, Soares-cavalcanti NM, et al (2012) An overall evaluation of the Resistance ( R ) and Pathogenesis-Related ( PR ) superfamilies in soybean , as compared with Medicago and Arabidopsis. Genet Mol Biol 35:260–271.

Wigoda N, Ben-Nissan G, Granot D, et al (2006) The gibberellin-induced, cysteine-rich protein GIP2 from Petunia hybrida exhibits in planta antioxidant activity. Plant J 48:796–805. doi: 10.1111/j.1365-313X.2006.02917.x

Zare-zardini H, Tolueinia B, Hashemi A, Fesahat F (2013) Antioxidant and Cholinesterase From Ziziphus jujuba Fruits. Am J Alzheimer ' s Dis Other Dementias 4–12. doi: 10.1177/1533317513500839

Zasloff M (2002) Antimicrobial peptides of multicellular organisms. Nature 415:389–395.

Zhang S, Wang X (2008) Expression pattern of GASA, downstream genes of DELLA, in Arabidopsis. Chinese Sci Bull 53:3839–3846. doi: 10.1007/s11434-008-0525-9

Zhang S, Yang C, Peng J, et al (2009) GASA5, a regulator of flowering time and stem growth in Arabidopsis thaliana. Plant Mol Biol 69:745–59. doi: 10.1007/s11103-009-9452-7

Zilli JÉ, Valisheski RR, Rodrigues F, et al (2004) ASSESSMENT OF COWPEA RHIZOBIUM DIVERSITY IN CERRADO AREAS OF NORTHEASTERN BRAZIL. Brazilian J Microbiol 35:281–287.

Zimmermann R, Sakai H, Hochholdinger F (2010) The Gibberellic Acid Stimulated-Like gene family in maize and its role in lateral root development. Plant Physiol 152:356–65. doi: 10.1104/pp.109.149054

# V. CONCLUSÕES GERAIS

- ✓ Nesta análise foi possível identificar um total de 384 prováveis esnaquinas no transcriptomas de soja e do feijão-caupi, distribuídas com base na similaridade de suas sequências;

- ✓ As esnaquinas identificadas tanto em soja como no feijão-caupi apresentam expressão constitutiva ou induzida tanto por estresse biótico como abiótico;

- ✓ A distribuição genômica de esnaquinas parece resultar de rearranjos cromossômicos sofridos pelas cultivares, refletindo padrões encontrados para genes de defesa em plantas;

- ✓ As esnaquinas hipotéticas de soja são preferencialmente secretadas para o meio extracelular;

- ✓ A estrutura das esnaquinas nestas duas espécies de leguminosas apresentam três conformações estruturais;

- ✓ As sequencias peptídicas constituíram 20 aminoácidos conservados em posição, sendo 12 deles cisteínas, compatível com o padrão para esnaquinas descrito para outras espécies vegetais;

- ✓ Têm-se neste grupo gênico, importantes candidatos para estudos fisiológicos e de defesa vegetal com potencial aplicação no melhoramento destas, e de outras plantas de interesse agronômico e/ou para fins farmacológicos, tendo em vista a ação antimicrobiana destes peptídeos;

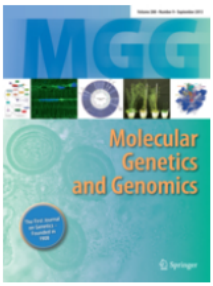
# ANEXOS

**Anexo I.** Instruções para autores: ***Molecular Genetics and Genomics***



Springer www.springer.com

Ciências Biológicas - Cell Biology | Molecular Genetics & Genomics – incl. option to publish open access




Molecular Genetics and Genomics

Chief Editor: Stefan **Hohmann**

ISSN: 1617-4615 (print version)
ISSN: 1617-4623 (electronic version)

Journal no. 438

Instructions for Authors

## Instructions for Authors

TYPES OF PAPERS

Molecular Genetics and Genomics publishes Reviews, Original Papers, and Methodology Papers. Manuscripts for Reviews are usually solicited by the Reviews Editor, but suggestions from authors are encouraged. All submitted manuscripts undergo a rigorous, fair and rapid peer-review.

MANUSCRIPT SUBMISSION

### Manuscript Submission

Submission of a manuscript implies: that the work described has not been published before; that it is not under consideration for publication anywhere else; that its publication has been approved by all co-authors, if any, as well as by the responsible authorities – tacitly or explicitly – at the institute where the work has been carried out. The publisher will not be held legally responsible should there be any claims for compensation.

### Permissions

Authors wishing to include figures, tables, or text passages that have already been published elsewhere are required to obtain permission from the copyright owner(s) for both the print and online format and to include evidence that such permission has been granted when submitting their papers. Any material received without such evidence will be assumed to originate from the authors.

### Online Submission

Authors should submit their manuscripts online. Electronic submission substantially reduces the editorial processing and reviewing times and shortens overall publication times. Please follow the hyperlink "Submit online" on the right and upload all of your manuscript files following the instructions given on the screen.

SPECIFIC REMARKS

## Manuscript submission

Please submit your manuscripts at EditorialManager

EditorialManager

1. Prepare your manuscript exactly according to the author instruction.

2. Go to https://www.editorialmanager.com/mgag/ and log on using your User-ID and password. If you are a first-time user, click on "Register". If you have forgotten your User-ID and/or password, click on "Send Username/Password".

3. Go to the "Submit a Manuscript" and follow the instructions on screen that will guide you through a series of steps to enter all required information.

4. Upload your manuscript as a single Microsoft Word file or PDF file with all tables and figures inserted/imported. If this is not possible, submit text and figures as separate files. The system will guide you through steps to generate a single pdf file for reviewing purposes. Note that PDF files cannot be used for production of the accepted manuscript and source files are required at that stage. See further instructions how to prepare manuscript files below.

No hard copy of the manuscript is required.

5. Peer reviewing will be supervised by the Editorial Office and an appropriate Associate Editor.

6. The Editorial Office will confirm your successful submission by E-mail and contact you as soon as a decision on your manuscript has been taken.

7. When you submit your revised manuscript, please upload ALL source files (even if some of them were not altered). If your manuscript is then accepted, those files will be used directly for production.

## Original Articles and Technical Notes should be written using the following structure

- Introduction

  The introduction should guide the reader and hence contain the following:

  • A short introduction to the field of research

  • The scope and incentives of the work

  • The background, i.e. previous work by your and other research groups strictly on the subject

  • A conclusion summarizing the novelty and the significance of this paper.

  Make sure that the introduction focuses on the background that is relevant for the data presented and the conclusions drawn.
- Materials and Methods

  This section should provide sufficient information to precisely evaluate how the data have been obtained, including the number of repetitions, and should permit a different research group to repeat the experimental work.
- Results

  Data should be presented as precisely and concisely as possible, if appropriate in the form of tables or figures.

  Make sure not to repeat in the text values given in figures or tables.

  The data presentation should follow a logical flow starting with a short description of the incentive of the experiment, followed by a description of the results. Do not

repeat information on the experimental details; those belong to Materials and Methods. Describe and discuss the data, but interpretations and conclusions should be given in the Discussion. Avoid combining Results and Discussion: separate Results and Discussion sections make for easier reading and a better structure of your article.

- Discussion

  The discussion should cover the following aspects:

  • An interpretation of the results and their significance with reference to previous work.

  • A description how the data advance the state-of-the-art

  • The importance of the findings for the research field

  • Any models realistically emerging from the data presented

  • Implications for future research

  Avoid repeating Introduction or Results.

Present your work so that it attracts the interest of referees and readers. This will increase the chances of acceptance and subsequent citation of the published work. Interest is attracted early on in the manuscript by highlighting the research questions addressed, the main results and conclusions derived from the work and their broader scientific impact.

For Review Articles authors should use a suitable structure of their choice. The manuscript should be broken down using informative headings. The Abstract and Introduction should be written so as to attract the reader and point out the broader impact and importance of the field covered.

## TITLE PAGE

### Title Page

The title page should include:

- The name(s) of the author(s)
- A concise and informative title
- The affiliation(s) and address(es) of the author(s)
- The e-mail address, telephone and fax numbers of the corresponding author

### Abstract

Please provide an abstract of 150 to 250 words. The abstract should not contain any undefined abbreviations or unspecified references.

### Keywords

Please provide 4 to 6 keywords which can be used for indexing purposes.

## TEXT

### Text Formatting

Manuscripts should be submitted in Word.

- Use a normal, plain font (e.g., 10-point Times Roman) for text.
- Use italics for emphasis.
- Use the automatic page numbering function to number the pages.
- Do not use field functions.

- Use tab stops or other commands for indents, not the space bar.
- Use the table function, not spreadsheets, to make tables.
- Use the equation editor or MathType for equations.
- Save your file in docx format (Word 2007 or higher) or doc format (older Word versions).

Manuscripts with mathematical content can also be submitted in LaTeX.

LaTeX macro package (zip, 182 kB)

### Headings

Please use no more than three levels of displayed headings.

### Abbreviations

Abbreviations should be defined at first mention and used consistently thereafter.

### Footnotes

Footnotes can be used to give additional information, which may include the citation of a reference included in the reference list. They should not consist solely of a reference citation, and they should never include the bibliographic details of a reference. They should also not contain any figures or tables.

Footnotes to the text are numbered consecutively; those to tables should be indicated by superscript lower-case letters (or asterisks for significance values and other statistical data). Footnotes to the title or the authors of the article are not given reference symbols.

Always use footnotes instead of endnotes.

### Acknowledgments

Acknowledgments of people, grants, funds, etc. should be placed in a separate section before the reference list. The names of funding organizations should be written in full.

## SCIENTIFIC STYLE

- Genus and species names should be in italics.

### Nomenclature abbreviations

Those should follow the rules and recommendations of the IUPAC-IUB Commission on Biochemical Nomenclature (CBN) and of the IUB Commission of Editors of Biochemical Journals (CEBJ). Genus or species names, gene symbols, mathematical formulae and any other words to be given special emphasis should be underlined once for italics. Similar signs have to be clearly distinguished, e.g. 1 (one) and l (letter), or 0 (zero) and O (letter).

### Sequence/nucleotide data

New nucleotide data must be submitted and deposited in the DDBJ/EMBL/GenBank databases and an accession number obtained before the paper can be accepted for publication. Submission to any one of the three collaborating databanks is sufficient to ensure data entry in all. The accession number should be included in the manuscript e.g. as a footnote on the title page:

'Note: Nucleotide sequence data reported are available in the DDBJ/EMBL/GenBank databases under the accession numbers(s)----'.

If requested the database will withhold release of data until publication.

The most convenient method for submitting sequence data is by World Wide Web:

- EMBL via Webin: http://www.ebi.ac.uk/embl/Submission/webin.html
- GenBank via Bankit: http://www.ncbi.nlm.nih.gov/BankIt/
- DDBJ via Sakura: http://sakura.ddbj.nig.ac.jp/

Alternatively, the stand-alone submission tool 'Sequin' is available from the EBI at http://www3.ebi.ac.uk/Services/Sequin and from NCBI at http://www.ncbi.nlm.nih.gov/Sequin/

For special types of submissions (e.g. genomes, bulk submissions etc) additional submission systems are available from the above sites.

Database Contact Information:

- EMBL:
  EMBL Nucleotide Sequence Submissions
  European Bioinformatics Institute,
  Wellcome Trust Genome Campus, Hinxton,
  Cambridge CB10 1SD U.K.;
  Tel.: +44-1223-494400; Fax: +44-1223-494472;
  e-mail: datasubs@ebi.ac.uk
  WWW URL: http://www.ebi.ac.uk
- GenBank:
  National Center for Biotechnology Information,
  National Library of Medicine, Bldg. 38A, Rm 8N-803
  Bethesda, MD 20894, USA
  Tel.: +1-301-4962475; Fax: +1-301-4809241
  e-mail: info@ncbi.nlm.nih.gov
  WWW URL: htpp://www.ncbi.nlm.nih.gov
- DDBJ:
  Center for Information Biology and DNA Data Bank of Japan
  National Institute of Genetics, 111 Yata,
  Mishima, Shizuoka 411-8540, Japan;
  Tel.: +81-559-816853; Fax: +81-559-816849;
  e-mail: ddbj@ddbj.nig.ac.jp
  WWW URL: http://www.ddbj.nig.ac.jp.

### Expression data

MGG requires microarray data to be reported in accordance with the MIAME standards and this to be clearly documented in the Materials and Methods section. Please refer to http://www.mged.org/. In addition, expression data should be deposited in a relevant repository such as:

Gene expression omnibus: http://www.ncbi.nlm.nih.gov/geo/

EBI Microarray Databases: http://www.ebi.ac.uk/Databases/microarray.html

Center for Information Biology gene EXpression database: http://cibex.nig.ac.jp/

Submission should be documented in the manuscript.

## REFERENCES

### Citation

Cite references in the text by name and year in parentheses. Some examples:

- Negotiation research spans many disciplines (Thompson 1990).
- This result was later contradicted by Becker and Seligman (1996).
- This effect has been widely studied (Abbott 1991; Barakat et al. 1995; Kelso and Smith 1998; Medvec et al. 1999).

Reference list

The list of references should only include works that are cited in the text and that have been published or accepted for publication. Personal communications and unpublished works should only be mentioned in the text. Do not use footnotes or endnotes as a substitute for a reference list.

Reference list entries should be alphabetized by the last names of the first author of each work.

- Journal article

  Gamelin FX, Baquet G, Berthoin S, Thevenet D, Nourry C, Nottin S, Bosquet L (2009) Effect of high intensity intermittent training on heart rate variability in prepubescent children. Eur J Appl Physiol 105:731-738. doi: 10.1007/s00421-008-0955-8

  Ideally, the names of all authors should be provided, but the usage of "et al" in long author lists will also be accepted:

  Smith J, Jones M Jr, Houghton L et al (1999) Future of health insurance. N Engl J Med 965:325–329
- Article by DOI

  Slifka MK, Whitton JL (2000) Clinical implications of dysregulated cytokine production. J Mol Med. doi:10.1007/s001090000086
- Book

  South J, Blass B (2001) The future of modern genomics. Blackwell, London
- Book chapter

  Brown B, Aaron M (2001) The politics of nature. In: Smith J (ed) The rise of modern genomics, 3rd edn. Wiley, New York, pp 230-257
- Online document

  Cartwright J (2007) Big stars have weather too. IOP Publishing PhysicsWeb. http://physicsweb.org/articles/news/11/6/16/1. Accessed 26 June 2007
- Dissertation

  Trent JW (1975) Experimental acute renal failure. Dissertation, University of California

Always use the standard abbreviation of a journal's name according to the ISSN List of Title Word Abbreviations, see

www.issn.org/2-22661-LTWA-online.php

For authors using EndNote, Springer provides an output style that supports the formatting of in-text citations and reference list.

EndNote style (zip, 3 kB)

TABLES

- All tables are to be numbered using Arabic numerals.

- Tables should always be cited in text in consecutive numerical order.
- For each table, please supply a table caption (title) explaining the components of the table.
- Identify any previously published material by giving the original source in the form of a reference at the end of the table caption.
- Footnotes to tables should be indicated by superscript lower-case letters (or asterisks for significance values and other statistical data) and included beneath the table body.

ARTWORK AND ILLUSTRATIONS GUIDELINES

For the best quality final product, it is highly recommended that you submit all of your artwork – photographs, line drawings, etc. – in an electronic format. Your art will then be produced to the highest standards with the greatest accuracy to detail. The published work will directly reflect the quality of the artwork provided.

**Anexo II.** Instruções para autores: ***Functional & Integrative Genomics***






Functional & Integrative Genomics

Editor-in-Chief: Rudi **Appels**

ISSN: 1438-793X (print version)
ISSN: 1438-7948 (electronic version)

Journal no. 10142

Instructions for Authors

## Instructions for Authors

### TYPES OF PAPERS

Functional & Integrative Genomics publishes the following types of papers:

- Original Papers
- Reviews
- Short Communication: limited to 1 figure and 1 table of results

### MANUSCRIPT SUBMISSION

#### Manuscript Submission

Submission of a manuscript implies: that the work described has not been published before; that it is not under consideration for publication anywhere else; that its publication has been approved by all co-authors, if any, as well as by the responsible authorities – tacitly or explicitly – at the institute where the work has been carried out. The publisher will not be held legally responsible should there be any claims for compensation.

#### Permissions

Authors wishing to include figures, tables, or text passages that have already been published elsewhere are required to obtain permission from the copyright owner(s) for both the print and online format and to include evidence that such permission has been granted when submitting their papers. Any material received without such evidence will be assumed to originate from the authors.

#### Online Submission

Authors should submit their manuscripts online. Electronic submission substantially reduces the editorial processing and reviewing times and shortens overall publication times. Please

follow the hyperlink "Submit online" on the right and upload all of your manuscript files following the instructions given on the screen.

### Specific Remark

Color figures in Invited Reviews are published free of charge.

## TITLE PAGE

### Title Page

The title page should include:

- The name(s) of the author(s)
- A concise and informative title
- The affiliation(s) and address(es) of the author(s)
- The e-mail address, telephone and fax numbers of the corresponding author

### Abstract

Please provide an abstract of 150 to 250 words. The abstract should not contain any undefined abbreviations or unspecified references.

### Keywords

Please provide 4 to 6 keywords which can be used for indexing purposes.

## TEXT

### Text Formatting

Manuscripts should be submitted in Word.

Use a normal, plain font (e.g., 10-point Times Roman) for text.
Use italics for emphasis.
Use the automatic page numbering function to number the pages.
Do not use field functions.
Use tab stops or other commands for indents, not the space bar.
Use the table function, not spreadsheets, to make tables.
Use the equation editor or MathType for equations.
Save your file in docx format (Word 2007 or higher) or doc format (older Word versions).

Manuscripts with mathematical content can also be submitted in LaTeX.

LaTeX macro package (zip, 182 kB)

### Headings

Please use no more than three levels of displayed headings.

### Abbreviations

Abbreviations should be defined at first mention and used consistently thereafter.

### Footnotes

Footnotes can be used to give additional information, which may include the citation of a reference included in the reference list. They should not consist solely of a reference citation,

and they should never include the bibliographic details of a reference. They should also not contain any figures or tables.

Footnotes to the text are numbered consecutively; those to tables should be indicated by superscript lower-case letters (or asterisks for significance values and other statistical data). Footnotes to the title or the authors of the article are not given reference symbols.

Always use footnotes instead of endnotes.

### Acknowledgments

Acknowledgments of people, grants, funds, etc. should be placed in a separate section before the reference list. The names of funding organizations should be written in full.

## SPECIFIC REMARKS

- Introduction: A detailed historical Introduction should be avoided, but the main problem should be outlined briefly. Continuity with earlier work on the subject should be established by reference to recent papers and reviews.
- Experimental methods must be clearly described. Techniques already published elsewhere need not be extensively described, but should be adequately referenced.
- Results: In short papers, Results may be combined with the Discussion section.
- Discussion should be restricted to interpretation and discussion of the main implications of the experimental data. Excessive speculation is to be avoided.

Nomenclature Abbreviations should follow the rules and recommendations of the IUPAC-IUB Commission on Biochemical Nomenclature (CBN) and of the IUB Commission of Editors of Biochemical Journals (CEBJ).

## SCIENTIFIC STYLE

- Please always use internationally accepted signs and symbols for units (SI units).
- Genus and species names should be in italics.
- Generic names of drugs and pesticides are preferred; if trade names are used, the generic name should be given at first mention.

## REFERENCES

### Citation

Cite references in the text by name and year in parentheses. Some examples:

- Negotiation research spans many disciplines (Thompson 1990).
- This result was later contradicted by Becker and Seligman (1996).
- This effect has been widely studied (Abbott 1991; Barakat et al. 1995; Kelso and Smith 1998; Medvec et al. 1999).

### Reference list

The list of references should only include works that are cited in the text and that have been published or accepted for publication. Personal communications and unpublished works should only be mentioned in the text. Do not use footnotes or endnotes as a substitute for a reference list.

Reference list entries should be alphabetized by the last names of the first author of each work.

Journal article

Gamelin FX, Baquet G, Berthoin S, Thevenet D, Nourry C, Nottin S, Bosquet L (2009) Effect of high intensity intermittent training on heart rate variability in prepubescent children. Eur J Appl Physiol 105:731-738. doi: 10.1007/s00421-008-0955-8

Ideally, the names of all authors should be provided, but the usage of "et al" in long author lists will also be accepted:

Smith J, Jones M Jr, Houghton L et al (1999) Future of health insurance. N Engl J Med 965:325–329

Article by DOI

Slifka MK, Whitton JL (2000) Clinical implications of dysregulated cytokine production. J Mol Med. doi:10.1007/s001090000086

Book

South J, Blass B (2001) The future of modern genomics. Blackwell, London

Book chapter

Brown B, Aaron M (2001) The politics of nature. In: Smith J (ed) The rise of modern genomics, 3rd edn. Wiley, New York, pp 230-257

Online document

Cartwright J (2007) Big stars have weather too. IOP Publishing PhysicsWeb. http://physicsweb.org/articles/news/11/6/16/1. Accessed 26 June 2007

Dissertation

Trent JW (1975) Experimental acute renal failure. Dissertation, University of California

Always use the standard abbreviation of a journal's name according to the ISSN List of Title Word Abbreviations, see

www.issn.org/2-22661-LTWA-online.php

For authors using EndNote, Springer provides an output style that supports the formatting of in-text citations and reference list.

EndNote style (zip, 3 kB)

SPECIFIC REMARKS

Additional kinds of citation

1 Article in electronic journal by DOI (no paginated version):

Slifka MK, Whitton JL (2000) Clinical implications of dysregulated cytokine production. Dig J Mol Med. doi: 10.1007/s801090000086

2 Online database:

Healthwise Knowledgebase (1998) US Pharmacopeia, Rockville. http://www.healthwise.org. Accessed 21 Sept 1998

3 Supplementary material/private homepage:

Doe J (2000) Title of supplementary material. http://www.privatehomepage.com. Accessed 22 Feb 2000

4 University site:

Doe J (1999) Title of preprint. http://www.uni-heidelberg.de/mydata.html. Accessed 25 Dec 1999

5 FTP site Doe J (1999) Trivial HTTP, RFC2169. ftp://ftp.isi.edu/in-notes/rfc2169.txt. Accessed

12 Nov 1999

6 Organization site:

ISSN International Centre (1999) Global ISSN database. http://www.issn.org. Accessed 20 Feb 2000

## TABLES

All tables are to be numbered using Arabic numerals.

Tables should always be cited in text in consecutive numerical order.

For each table, please supply a table caption (title) explaining the components of the table.

Identify any previously published material by giving the original source in the form of a reference at the end of the table caption.

Footnotes to tables should be indicated by superscript lower-case letters (or asterisks for significance values and other statistical data) and included beneath the table body.

## ARTWORK AND ILLUSTRATIONS GUIDELINES

For the best quality final product, it is highly recommended that you submit all of your artwork – photographs, line drawings, etc. – in an electronic format. Your art will then be produced to the highest standards with the greatest accuracy to detail. The published work will directly reflect the quality of the artwork provided.

### Electronic Figure Submission

Supply all figures electronically.

Indicate what graphics program was used to create the artwork.

For vector graphics, the preferred format is EPS; for halftones, please use TIFF format. MS Office files are also acceptable.

Vector graphics containing fonts must have the fonts embedded in the files.

Name your figure files with "Fig" and the figure number, e.g., Fig1.eps.

### Line Art