

Universidade Federal do Rio Grande do Sul

**Diversificação funcional em ribonucleases T2 na família
Solanaceae**

Lauís Brisolara Corrêa

Orientadora: Dra. Loreta B. Freitas

Coorientadora: Dra. Claudia E. Thompson

Tese submetida ao Programa de Pós-Graduação
em Genética e Biologia Molecular da UFRGS
como requisito parcial para a obtenção do grau de
Doutor em Ciências.

Porto Alegre, outubro de 2015.

INSTITUIÇÕES E FONTES FINANCIADORAS

Este trabalho foi realizado nas dependências do Laboratório de Evolução Molecular, Departamento de Genética – UFRGS; da Unidade de Biologia Teórica e Computacional, Centro de Biotecnologia – UFRGS; e do Laboratório de Química Teórica e Computacional – UFRGS.

O projeto foi financiado com recursos do CNPq e PPGBM-UFRGS e a bolsa de doutorado foi concedida pelo CNPq.

AGRADECIMENTOS

Agradeço a Deus, que me possibilitou ser o dono dos meus pensamentos e o mestre das minhas ações e o qual me permitiu estar entre as pessoas que amo, guiando-me nas provações de minha vida e dando-me forças para seguir minha jornada.

Aos meus pais, Luiz Carlos Martins Corrêa e Vera Maria Brisolará Corrêa, pelo apoio, carinho e todos os votos de confiança que tem depositado em mim durante toda a minha vida, além do grande exemplo de força, dedicação e responsabilidade.

Ao meu irmão Aluísio, que próximo ou distante de mim, sempre me amparou nos momentos difíceis e vibrou com minhas conquistas. E especialmente agradeço ao meu irmão Ulisses, o qual mesmo dentre os momentos mais escuros desta jornada me amparou encontrando forças e iluminando meu caminho, guiando-me pelas adversidades deste trajeto.

Aos meus primos, que sempre foram presença constante durante toda a minha vida, especialmente as minhas primas Cibele Brisolará de Brisolará, Elisa Brisolará-Fabião, e Lisane Brisolará de Brisolará, pelo companheirismo e pela forte influência e incentivo durante todos os momentos da minha vida.

Aos meus tio e tias, pelo carinho, incentivo, respeito, exemplo e dedicação que demonstraram tanto durante a minha formação acadêmica e escolhas tomadas nesta jornada, quanto em meu desenvolvimento pessoal, como pessoa honesta, digna e afetuosa, que me tornei.

A minha orientadora, Dra. Loreta Brandão de Freitas, pela presença insubstituível e dedicação durante todos estes quatro anos, mostrando-me não apenas a pessoal eloquente, responsável e amável, mas também uma imagem na qual me espelhar como um futuro grande pesquisador. Sua representatividade, responsabilidade, suas palavras, sempre serão fonte inspiradora para mim. Obrigado por estar presente mesmo nos momentos mais difíceis deste caminho, sempre me instigando e me direcionando na direção correta a seguir.

A minha amiga e coorientadora, Dra. Claudia Elizabeth Thompson, por todo o seu apoio e dedicação empregados durante esses anos. Na realidade, é difícil colocar em palavras toda a representatividade da Claudia durante esta jornada, ela realmente foi uma grande amiga, a qual espero esteja presente para o resto de minha vida. Além disso, nossas

discussões saborosas sobre ciência, bioinformática, entre outros assuntos, foram muito importantes na minha formação como indivíduo e cientista.

A minha grande amiga, Helen Nathalia Thompson, pelas suas valiosas aulas e discussões sobre Dinâmica Molecular e Biologia Estrutural.

Aos meus colegas lêmures, do Laboratório de Evolução Molecular, Ana Lúcia Anversa Segatto, Alice Backes da Rosa, Geraldo Mäder, Michel João Ferreira Barros, Gustavo Silva Arias, Lina Maria Caballero Villalobos, Ana Laura de Wallau John, Daniele Munareto Rodrigues, Caroline Turcheto, Priscilla Mena Zamberlan, Jéferson Nunes Fregonezi, Raquel Athayde Kriedt, Ana Luiza Ramos Cazé, Giovanna Câmara Giudicelli, Maikel Reck Kortmann, Verônica Aydos Thode, Marcelo Costa Teixeira, Karina Lima e Luana Castro, o meu mais sincero agradecimento por todo auxílio, amizade e companheirismo nesta jornada.

Aos meus amigos do Núcleo de Bioinformática do Laboratório de Imunogenética, Dinler Amaral Antunes, Marialva Sinigaglia e Gustavo Fioravante.

Agradeço especialmente a alguns grandes amigos que conheci durante meu período na representação discente: Angélica Salatino Oliveira, Ana Paula Cristoph e Maurício Menegatti Rigo.

A minha amiga e colaboradora, Dra. Cláudia Lemelle Fernandes, pela imensa amizade compartilhada, aconselhamento, disponibilidade e colaboração nas análises de biologia estrutural.

A Dra. Ana Tereza Ribeiro de Vasconcelos que durante um breve período me acolheu de braços abertos no Laboratório Nacional de Computação Científica.

Ao Programa de Pós-Graduação em Genética e Biologia Molecular (UFRGS), pelo auxílio em viagens e pelos múltiplos cursos e possibilidades de aperfeiçoamento disponibilizados durante a minha formação.

Aos auxiliares administrativos do PPGBM, Elmo J. Antunes Cardoso e Ellen, os quais sempre foram muito prestativos quando tive dúvidas administrativas e precisei de algo da Pós Graduação.

E finalmente, a todos os professores, colegas e amigos do Departamento de Genética da UFRGS, os quais são muitos, portanto nem todos puderam ser nomeados, contudo sempre terão lugar cativo junto ao meu coração.

SUMÁRIO

RESUMO	6
ABSTRACT	8
INTRODUÇÃO	10
1) Famílias multigênicas	10
2) A família gênica das RNases T2	12
3) Auto-incompatibilidade em plantas	15
4) Os sistemas proteicos de auto-incompatibilidade em plantas	17
5) Justificativa	21
6) Objetivos	21
7) Estrutura da Tese	22
CONSIDERAÇÕES FINAIS	24
REFERÊNCIAS	29

RESUMO

As ribonucleases catalisam a clivagem do RNA e são componentes ubíquos das células, desde procariotos até eucariotos. A família T2 é a classe de RNases mais amplamente distribuída entre os organismos vivos. A ampla distribuição desta família sugere que ela deve ter um papel funcional muito importante na biologia celular destes organismos. Em Solanaceae, há basicamente dois grupos destas proteínas, um representado pelas S-RNases, as quais estão envolvidas com a rejeição do pólen na auto-incompatibilidade gametofítica e um outro grupo mais diverso que é denominado de S-like RNases, com funções muito diversificadas. As S-RNases se apresentam na forma de um gene multialélico, altamente polimórfico, que está contido no locus S. O locus S é composto por uma combinação de proteínas SLF (S-locus F-box), responsáveis pela determinação do fator polínico, e uma S-RNase produzida apenas no pistilo, de forma que estes genes estão fortemente ligados formando o haplótipo S. Esses produtos gênicos interagem possibilitando a rejeição do auto-pólen num fenômeno denominado distinção colaborativa do pólen não próprio. No **Capítulo II**, foram utilizados métodos filogenéticos para determinar os principais agrupamentos de alelos na genealogia de S-RNases do gênero *Solanum*. A topologia da árvore não mostrou sinal filogenético para espécies, contudo, isso era esperado uma vez que a diversificação destes alelos ocorreu anteriormente à diversificação das espécies, gerando um fenômeno denominado de polimorfismos trans-específicos. Além disso, foram realizadas análises de seleção positiva, as quais encontraram um alto número de resíduos com altas probabilidades, indicando que estão sob pressão de seleção positiva darwiniana. Com o intuito de compreender a diversidade estrutural destes alelos, foram construídos modelos teóricos com base em modelagem por homologia dos principais clados da filogenia, já que estas sequências apresentam um elevado grau de polimorfismo. Os resultados mostraram grande variação estrutural na região hipervariável destas sequências, enquanto que regiões conservadas não apresentaram grandes mudanças estruturais e com estruturas secundárias características. No **Capítulo III**, foram realizadas diversas análises com o objetivo de compreender a diversificação estrutural e funcional de RNases T2 na família Solanaceae. As análises filogenéticas mostraram a formação de três principais grupos, sendo um de S-RNases, e os outros dois de S-like RNases. Com relação ao Clado 2, podemos inferir que houveram ao menos dois eventos de duplicação gênica. Além disso, também foram utilizados os métodos de NSSites e branch-site para inferência de seleção positiva como uma forma de

identificar possíveis sinais de diversificação molecular. Muitos resíduos parecem estar sob seleção em ambos os métodos, embora um número maior fosse encontrado no NSsites (41), com estes resíduos localizando-se em regiões mais flexíveis da proteína, enquanto que os aqueles selecionados de acordo com o brach-site (8) estavam situados em posições mais rígidas da estrutura. Em sùmula, os resultados encontrados nos capítulos que compreendem esta tese demonstram que análises teóricas podem contribuir efetivamente de inúmeras maneiras com o intuito de desenvolver uma melhor compreensão dos fenômenos biológicos relacionados com a evolução molecular de famílias multigênicas, além de também contribuir no entendimento dos processos de diversificação de genes multialélicos, utilizando como modelo de estudo a família gênica RNase T2.

ABSTRACT

The ribonucleases catalyze the cleavage of RNA and are ubiquitous components of cells, from prokaryotes to eukaryotes. The T2 family is the most widely category of RNases distributed among living organisms. The wide distribution of this family suggests that it should play an important functional role in cell biology of these organisms. Basically, there are two groups of these proteins in Solanaceae, one represented by S-RNases, which are involved in the rejection of pollen in gametophytic self-incompatibility and a more diverse group, which is termed S-like RNases with very diverse functions. S-RNases are a highly polymorphic and multiallelic gene contained in the S locus. The S locus consists of a combination of SLF proteins (F-box S-locus), responsible for the pollen factor, and one S-RNase expressed only in the pistil, and those genes are tightly linked as an S-haplotype. Products of these genes interact enabling the rejection of self-pollen in a phenomenon called collaborative non-self recognition. In **Chapter II**, we used phylogenetic methods to determine the main clusters of alleles in the genealogy of S-RNase in the *Solanum* genus. The topology of the tree showed no phylogenetic sign to species delimitation, but it was expected since the diversification of these alleles occurred previously the diversification of the species, generating a phenomenon called trans-specific polymorphism. In addition, analyzes of positive selection were carried out, and it resulted in a significant number of residues with high probability, indicating they are under Darwinian positive selection. In order to understand the structural diversity of alleles, theoretical models were constructed based on homology modeling of the major clades found in the phylogeny, since it has a high degree of polymorphism in these sequences. The results showed that major structural variations are located in the hypervariable regions of these sequences, while conserved regions performed without major structural changes and showed stable secondary structures. In **Chapter III**, we used several analyzes to understand the structural and functional diversification of RNase T2 in the Solanaceae family. Phylogenetic analyses showed the clustering of three main groups, one with just S-RNases and the two others composed by S-like RNases. Regarding to Clade 2, we could infer that at least two gene duplication events have occurred. In addition, we used two methods for inference of positive selection, NSsites and branch-site, as a mean to identify possible signals of molecular diversification. Many residues seem to be under selection in both methods, although a higher number was found in NSsites (41), and these residues were located in more flexible regions of the protein, while those selected according to the

branch-site (8) were located at more rigid positions of the structure. In summary, the results found in the chapters of this thesis show that theoretical analyses could effectively contribute in many ways in order to develop a better understanding of biological phenomena related to the molecular evolution of gene families. Also, it contributes to the understanding of the processes related to multiallelic genes diversification, using gene family RNase T2 as a model.

INTRODUÇÃO

1) Famílias multigênicas

Famílias gênicas podem ser definidas tanto como o conjunto de genes evolutivamente relacionados compartilhados por diferentes espécies e às vezes desempenhando funções similares, como o conjunto de genes homólogos dentro de uma espécie, uma família gênica espécie-específica (Martinez, 2011). Estes conjuntos de genes surgem através de duplicação gênica, tanto por duplicação de um segmento cromossômico como por duplicação do genoma como um todo. Dessa maneira, alguns genes duplicados podem permanecer no genoma por um longo período enquanto outros perdem sua função ou são simplesmente excluídos do genoma.

Basicamente, existem dois modelos principais de evolução de genes. O mais tradicional deles propõe que a ocorrência de seleção purificadora pode ser seguida pelo acúmulo de mutações que codificam novas funções benéficas que são então preservadas pela seleção natural, denominado evolução convergente. Enquanto que o outro modelo, baseia-se na evolução divergente, postula que genes recém criados podem ter uma taxa evolutiva acelerada, gerando uma população de diferentes proteínas sobre as quais a seleção darwiniana positiva pode atuar no desenvolvimento de um novo estado funcional. Além destes, vários modelos tem sido sugeridos para explicar os processos envolvidos na formação de novos genes, sendo os principais a duplicação gênica, a transposição de elementos proteicos via elementos móveis, a transferência horizontal de genes, a fusão de genes e a fissão de genes (Betrán & Long, 2002; Jiang *et al.*, 2004; Morgante *et al.*, 2005; Ranz *et al.*, 2007).

Até meados de 1990, pensava-se que as famílias multigênicas evoluíssem principalmente na forma de uma evolução em concerto. Todavia, com o surgimento de diversos estudos evolutivos de genes relacionados as mais diversas funções, isso se mostrou infundado, já que muitas destas famílias gênicas apresentavam distintos padrões evolutivos.

A principal controvérsia com relação ao modelo de evolução em concerto reside na impossibilidade da geração de inovações evolutivas, uma vez que este modelo não é compatível com o surgimento de novos genes funcionalmente diversos daquele ancestral. Muito embora, o modelo de evolução em concerto realmente seja o responsável pelos

padrões encontrados em algumas famílias gênicas, atualmente ele é considerado restrito a algumas poucas famílias mais conservadas (Eirín- López *et al.*, 2012).

Assim, um novo modelo de evolução tem sido proposto, denominado evolução por nascimento e morte de novos genes. De acordo com esse modelo, o nascimento é representado pela duplicação gênica e a morte pela inativação e eliminação de genes por mutações deletérias. Esse fenômeno pode ser representado por um processo Markov tempo-contínuo onde o nascimento aumenta o estado variável por um lado enquanto a morte o reduz pelo outro (Nei *et al.*, 1997; Nei & Rooney, 2005).

Como discutido acima, durante a história evolutiva de um grupo de organismos, genes relacionados as mais diversas funções podem ser duplicados e, nestes casos, após um dado tempo, podem evoluir por diferentes trajetórias de acordo com as pressões seletivas que estarão agindo sobre eles. A redundância criada pela duplicação provavelmente gera uma redução na pressão seletiva sobre estes genes, permitindo que mutações sejam acumuladas de forma mais flexível. Com o tempo, estes genes podem acumular tamanho número de mutações que se tornem afuncionais, estabelecendo o processo que se denomina de pseudogenização. Alternativamente, processos de subfuncionalização, que ocorrem quando ambas as cópias compartilham o mesmo estado funcional, ou de neofuncionalização, quando uma das cópias adquire uma inovação funcional, podem também ocorrer com as múltiplas cópias (Ohno, 1970).

Uma vez duplicados os genes ou genomas, diversas forças evolutivas incluindo seleção, mutação e deriva genética dão forma à evolução do número de genes que compõem uma família gênica. E, apesar de seu significado óbvio, muitas questões permanecem em aberto, por exemplo, que pressão seletiva mantém a função original de uma proteína ou que porções destas moléculas são influenciadas neste processo. A avaliação da fração de substituições nucleotídica sinônimas (silenciosas) em relação àquelas não-sinônimas que levam à substituição de aminoácidos pode trazer alguma luz a estes questionamentos (Wagner, 2001).

De forma geral, existem duas classes de métodos estatísticos para se detectar a ocorrência de seleção positiva em sequências codificadoras de genes. Uma delas é baseada na análise de frequências gênicas utilizando conceitos de genética de populações (Teoria da Neutralidade) enquanto que a outra utiliza testes filogenéticos de seleção positiva com base na razão ω (Yang, 2006). Utilizando esta idéia, Messier e Stewart (1997) utilizaram

um método no qual o objetivo era testar a adaptação molecular nos genes da lisosima de primatas. Essa estratégia baseia-se no fato de que um excesso na razão entre a taxa de mutações não-sinônimas e sinônimas (d_n/d_s) pode ser um forte indicativo de seleção positiva ao nível molecular. Os principais métodos utilizados atualmente são aqueles implementados no pacote PAML (Plylogenetic Analysis of Maximum Likelihood; Yang, 2007), o qual explora mudanças nas taxas evolutivas através da história filogenética de um grupo de espécies ou genes. Os testes filogenéticos detectam seleção positiva com base em um excedente de substituições de aminoácidos e por isso requerem que as sequências sejam muito divergentes, tendo praticamente nenhum poder estatístico quando aplicadas a dados populacionais (Anizimova *et al.*, 2002). Em contraste, os testes de Neutralidade devem preferencialmente ser conduzidos em conjuntos de dados provenientes de populações de uma mesma espécie ou no máximo entre populações de espécies relacionadas. Nesse caso, o intuito do teste é detectar a tendência da fixação de formas possivelmente adaptativas em algumas populações, o que poderia ser relacionado a variações geográficas ou comportamentais envolvidas com dinâmicas evolutivas destas populações. Além disso, existem casos em que genes divergiram há tanto tempo, que os grupos não podem mais ser tratados via uma abordagem em escala populacional. A maioria dos testes também assume um modelo de sítios infinitos, que pode ser quebrado quando sequências de espécies diferentes são comparadas (Yang, 2006).

2) A família gênica das RNases T2

As ribonucleases catalisam a clivagem de RNA, atuando sobre substratos fita-simples, dupla-fita ou híbridos DNA-RNA e são componentes celulares ubíquos. Muitas pesquisas tem focado ribonucleases específicas que agem sobre uma infinidade de processos celulares, incluindo síntese de DNA, processamento de RNA, degradação de RNA nuclear ou citoplasmático, RNAi e defesa antiviral (Kubo *et al.*, 2015).

As células também produzem um conjunto de RNases de atividades gerais que são secretadas normalmente ou concentradas em compartimentos celulares delimitados por membrana, como lisossomos ou vacúolos. Essas enzimas incluem membros das famílias da RNaseA, RNaseT1 e RNaseT2. Sua classificação tem como base tanto a similaridade de sequência como a especificidade por substrato. Curiosamente, alguns trabalhos sugerem que membros das famílias RNaseA (Rosenberg, 2008) e RNaseT2 (Hillwig *et al.*, 2011)

frequentemente desempenham papéis relevantes em uma ampla variedade de funções biológicas. Além disso, as enzimas da família RNaseT2 tem recebido maior atenção desde que sua função ubíqua foi determinada. Membros desta família atuam contra vírus infectantes de células humanas, por isso, diversos estudos enfocam sua estrutura protéica e atividade enzimática (Nakagawa *et al.*, 1999). Com isso, loci desta família gênica têm sido estudados em uma ampla variedade de grupos taxonômicos, como bactérias, leveduras, fungos, animais e plantas (Hayashi *et al.*, 2003; Roiz *et al.*, 2006).

As ribonucleases da família T2 são uma classe específica de endoribonucleases que clivam o RNA fita-simples e exibem diversas funções importantes para a biologia de procaríotos e eucariotos, estando também envolvidas em mecanismos de defesa contra doenças em animais e plantas (Luhtala & Parker, 2010). As ribonucleases da família T2 são RNases do tipo transferase e são classificadas de acordo com sua similaridade com a RNaseT2 de *Aspergillus oryzae* (Irie, 1999; Deshpande & Shankar, 2002). Estas enzimas catalisam a clivagem de RNA fita-simples através de um 2',3'-fosfato cíclico intermediário, produzindo mononucleotídeos ou oligonucleotídeos com um 3'-fosfato terminal.

Três características distinguem as ribonucleases T2 das famílias protéicas das RNasesA e RNasesT1. Primeiro, são mais amplamente distribuídas, sendo encontradas em vírus, protozoários, fungos, leveduras, animais e plantas (Deshpande & Shankar, 2002). Em contraste, as enzimas da família RNaseT1 existem somente em bactérias e fungos enquanto que aquelas da família RNaseA são frequentes em animais. Segundo, o pH ótimo para a atividade das ribonucleases T2 está entre 4 e 5, diferindo com relação as enzimas das famílias RNaseT1 e RNaseA que apresentam um pH ótimo alcalino (entre 7 e 8) ou fracamente ácido (entre 6,5 e 7), respectivamente. Esta atividade ácida das enzimas RNaseT2 é consistente com sua localização preferencial nos vacúolos ou lisossomos e sugere que a clivagem de seus próprios RNAs dentro de um compartimento ácido poderia ser uma função destas ribonucleases. Terceiro, as ribonucleases T2 geralmente clivam quaisquer das quatro bases nitrogenadas, enquanto que membros das famílias RNaseA e RNaseT1 tendem a ser específicos para pirimidinas ou guaninas, respectivamente (Irie, 1999; Deshpande & Shankar, 2002).

Como estas proteínas entram em uma rota secretora, elas geralmente são glicosiladas em células eucarióticas. Entretanto, há casos em que as proteínas RNaseT2 penetram no citoplasma. Por exemplo, em leveduras, a proteína Rny1 é liberada do

vacúolo para o citossol durante o estresse oxidativo (Thompson & Parker, 2009). Além disso, em alguns casos as ribonucleases T2 secretadas podem ser internalizadas por outras células, em um processo frequentemente citotóxico para a célula alvo. Dessa forma, a compartimentalização de ribonucleases T2 e sua subsequente liberação no citossol devem desempenhar um importante papel na modulação de sua atividade catalítica.

A estrutura e mecanismo de clivagem do RNA por ribonucleases T2 já são bem compreendidos. Além disso, há um número significativo de estruturas cristalografadas pertencentes a membros da família T2 de bactérias, plantas e fungos, as quais tem revelado uma estrutura α/β central conservada (Ida *et al.*, 2001; Kawano *et al.*, 2002, 2003; Kimura *et al.*, 2004; de Leeuw *et al.*, 2007; Kobayashi *et al.*, 2010). Duas regiões conservadas de ligação ao substrato, denominadas B1 e B2, tem sido observadas dentro de uma concavidade localizada no centro da estrutura, esta cavidade frequentemente está ocupada por nucleotídeos excisados (Kawano *et al.*, 2002, 2003; Suzuki *et al.*, 2000; Numata *et al.*, 2003; Rodriguez *et al.*, 2008).

A catálise é promovida por um a três resíduos de histidina, usualmente dois, que são encontrados em domínios conservados dessas proteínas (denominados CASI e CASII) os quais estão localizados no sítio ativo da enzima (Roberts *et al.*, 1969; Nishikawa *et al.*, 1987). Semelhante a outras ribonucleases do tipo transferase, a clivagem do RNA pelas enzimas T2 ocorre na forma de dois passos consecutivos: a transfosforilação e a hidrólise (Kurihara *et al.*, 1996).

Em vários casos, membros da família RNaseT2 agem como agentes citotóxicos. Um exemplo de proteínas RNaseT2 induzindo citotoxicidade e modulando a função imune celular vem da análise de certas viroses, como o vírus da febre suína e o vírus da diarreia viral bovina. Nestas viroses, um envelope de glicoproteínas, definido como Erns, tem um domínio de RNaseT2 e atividade de ribonuclease e também é secretado por células infectadas (Rumenapf *et al.*, 1993; Schneider *et al.*, 1993; Hulst *et al.*, 1994).

Membros da família RNaseT2 possuem uma surpreendente diversidade de papéis biológicos. Embora suas funções basicamente recaiam na sua atividade como nucleases, existem alguns genes sem atividade ribonucleotídica que no entanto mostram um padrão de expressão alterado durante experimentos de estresse biótico e abiótico. Apesar disto, não está clara qual seria a função biológica destas proteínas (Hillwig *et al.*, 2010; Macintosh *et al.*, 2010), sugerindo que outras funções possam ser independentes da

atividade de nuclease dessas enzimas e o entendimento destas funções poderia revelar novos papéis específicos para membros da família RNase T2 (Luhtala & Parker, 2010). Outro exemplo interessante de membros da família RNase T2 como agentes citotóxicos vem das bem caracterizadas S-RNases.

Mecanismos de resistência a patógenos em plantas também tem sido associados a aumento na expressão de RNasesT2, como a ação citotóxica produzida pelas células de *Nicotiana tabacum* quando infectadas pelos fungos *Phytophthora parasitica* e *Fusarium oxysporum*, inibindo o alongamento das hifas e reduzindo a infecção parasitária (Hugot *et al.*, 2002).

3) Auto-incompatibilidade em plantas

Darwin já havia reconhecido o fenômeno da auto-incompatibilidade em plantas durante seus estudos, embora por desconhecer os aspectos relacionados aos mecanismos biológicos de herança, denominou o processo como auto-esterilidade em seus trabalhos (Darwin, 1877, 1878). Além disso, ele conduziu diversos experimentos demonstrando os benefícios da fecundação cruzada e os efeitos negativos do endocruzamento em plantas. Com isso, ele percebeu que os benefícios da fecundação cruzada eram suficientes para direcionar a evolução em direção a emergência de um sistema de auto-incompatibilidade (McClure, 2009).

A maioria das angiospermas apresenta flores andróginas ou monóicas. Com isso, diversos mecanismos foram desenvolvidos para prevenir a auto-fecundação. Espécies dióicas são raras, representando apenas 5-10% das espécies de plantas (Armstrong & Filatov, 2008). Entre os mecanismos que dificultam a autopolinização, pode-se citar a hercogamia, que é a separação espacial dos verticilos reprodutivos, e a dicogamia, que é a separação temporal na maturidade dos verticilos reprodutivos (Leite & Machado, 2010). Por causa dessa proximidade espacial e temporal entre os órgãos reprodutivos na flor, muitas angiospermas desenvolveram estratégias moleculares que previnem o endocruzamento e favorecem a heterozigose (Stone *et al.*, 1999).

Estas estratégias são responsáveis pelo reconhecimento e rejeição do próprio pólen, impedindo tanto a autogamia como a geitonogamia (Tantikanjana *et al.*, 2009) e podem ser divididas em duas categorias de acordo com a morfologia floral: homomórficas e

heteromórficas. Espécies com flores homomórficas são mais comuns e apresentam um único padrão morfológico, podendo apresentar sistemas moleculares de auto-incompatibilidade. Já espécies com flores heteromórficas apresentam mais de um padrão floral, dessa maneira garantem que flores com morfologia diferente são capazes de trocar pólen através do polinizador, enquanto que flores com a mesma morfologia são incapazes de trocar pólen entre si. Este mecanismo também é regido por um sistema genético de auto-incompatibilidade. Contudo, diversos fatores mecânicos também estão envolvidos no mecanismo heteromórfico, como o reconhecimento chave-fechadura entre estruturas de ornamentação da exina do pólen com as criptas na superfície estigmática e relação do comprimento dos filetes e do estilete (Stone & Goring, 2001).

Historicamente, a auto-incompatibilidade em plantas tem sido associada a um único locus com múltiplos alelos, o locus S. Com os avanços decorridos da evolução das técnicas de biologia molecular, descobriu-se que na realidade haviam no mínimo dois produtos proteicos responsáveis por este mecanismo, codificados por genes fortemente ligados. De forma que um seja transcrito apenas no pistilo enquanto que o outro é transcrito na antera, no pólen ou ainda no tubo polínico. Esses sistemas de compatibilidade são baseados no reconhecimento de produtos celulares específicos, sendo usado para regular a aceitação ou rejeição do pólen depositado sobre o estigma. Assim, os grãos de pólen incompatíveis são reconhecidos e seletivamente inibidos (Franklin-Tong & Franklin, 2003).

Alguns evolucionistas sugerem que um dos maiores avanços evolutivos envolvidos no grande sucesso adaptativo que causou a diversificação das Angiospermas seja o desenvolvimento de mecanismos de auto-incompatibilidade, já que teoricamente estes sistemas estão associados com uma maior capacidade de conservação da diversidade genética intraespecífica (Barret, 2003). Diversos estudos tem demonstrado que espécies alógamas apresentam uma diversidade genética superior àquelas autógamamente relacionadas. Apesar disso, a autogamia ou a perda do sistema de auto-incompatibilidade podem ter tido um papel evolutivo importante para espécies em processo de expansão populacional, assim como espécies invasoras. Basicamente, existem duas possíveis causas para uma vantagem evolutiva e consequente perda do sistema de auto-incompatibilidade em uma linhagem: a ausência de um polinizador e a falta de parceiros compatíveis. Diversos cenários evolutivos podem estar envolvidos com estes processos, dentre eles a expansão geográfica. Por exemplo, a expansão geográfica pela dispersão de suas sementes

poderia ser limitada pela ausência de parceiros compatíveis ou do polinizador na nova região (Pyšek *et al.*, 2011).

Além disso, a maioria das espécies cultivadas de plantas são autógamas, contudo, frequentemente seus parentes selvagens são auto-incompatíveis. Esse fenômeno se deve a dois fatores: (1) que espécies autógamas são facilmente melhoradas, ou seja, a seleção de caracteres de interesse agrônômico é facilitada pela possibilidade de endocruzamento; e (2) que essas espécies apresentam uma maior homogeneidade morfológica, o que é um pré-requisito para uma boa cultivar (Liu & Burke, 2006).

4) Os sistemas proteicos de auto-incompatibilidade em plantas

As estratégias de auto-incompatibilidade homomórficas são tratadas com maior detalhamento abaixo, já que estas são amplamente distribuídas entre as Angiospermas e tem sido melhor compreendidas do ponto de vista molecular. Elas são basicamente divididas em duas classes, o sistema esporofítico e o gametofítico, em função de como o reconhecimento do fator polínico é realizado no pistilo, se reconhece ambos os alelos presentes na planta doadora de pólen ou se apenas o alelo presente no pólen, respectivamente (Stone & Goring, 2001). Há também uma outra subdivisão baseada no modo de interação dessas proteínas, o qual pode ser por auto-reconhecimento, quando a proteína reconhece e interage com aquela do mesmo haplótipo para a rejeição do próprio pólen. Ou ainda por auto-não-reconhecimento quando a não interação das proteínas do mesmo haplótipo resultam na rejeição do auto-pólen (Sun *et al.*, 2015).

Diversos autores têm sugerido múltiplas origens independentes para os sistema de auto-incompatibilidade, tanto gametofítico como esporofítico (Ride *et al.*, 1999; Fobis-Loisy *et al.*, 2004; Tantikanjana *et al.*, 2009; Aguiar *et al.*, 2015). Entretanto, enquanto algumas famílias de plantas apresentam sistemas de auto-incompatibilidade bem caracterizados aos níveis fisiológico e molecular, a maioria das famílias de plantas ainda não teve o modelo molecular determinado. Isto pode ser também um reflexo da diversidade funcional e molecular dos genes envolvidos na expressão deste fenômeno, de forma que múltiplos modelos proteicos tenham sido recrutados paralelamente para a mesma função (Hiscock *et al.*, 1996).

As S-RNases tem sido bem caracterizadas em sistemas de auto-incompatibilidade gametofítica em dicotiledôneas, sendo encontradas nas famílias Solanaceae, Rosaceae, Rubiaceae, Campanulaceae e Plantaginaceae (Franklin-Tong & Franklin, 2003). Apesar de diversas famílias apresentarem o sistema de auto-incompatibilidade gametofítica governado pela S-RNase, uma das famílias onde os mecanismos fisiológicos e bioquímicos têm sido melhor compreendidos é a família Solanaceae, por isso ela tem sido amplamente usada como modelo para estudos evolutivos da auto-incompatibilidade em plantas.

As S-RNases são proteínas exclusivamente expressas no pistilo, as quais são logo secretadas na matriz extracelular (Anderson *et al.*, 1986; Jahnen *et al.*, 1989). Elas também apresentam um domínio de ligação a pequenas proteínas secretadas durante todo o desenvolvimento do tubo polínico (Luu *et al.*, 2000). Quando um grão de pólen compatível germina, ele produz uma proteína SLF (S-locus F-Box) que possui o domínio complementar àquelas S-RNases secretadas pelo pistilo. Uma vez que as duas proteínas se complexam, ocorre a poliubiquitinação da S-RNase, a ribonuclease não é mais capaz de realizar a degradação de RNAs e é direcionada para ser degradada no proteossomo. Contudo, se um tubo polínico incompatível penetra no pistilo, não ocorre a formação do complexo proteico e, conseqüentemente, as ribonucleases secretadas pelo mesmo penetram no tubo polínico causando o cessamento do crescimento em função da degradação do RNA ribossomal (McClure *et al.*, 2011). Um grande número de alelos é encontrado em espécies de Solanaceae, alcançando até 35 alelos por espécie e podendo apresentar uma diversidade espantosamente alta em sequência de aminoácidos (Nowak *et al.*, 2011).

Estudos pioneiros que investigaram as proteínas do estilete de diferentes genótipos S encontraram proteínas associadas com a auto-incompatibilidade. Análises dessas proteínas estilares de *Nicotiana glauca* resultaram na identificação de uma S-glicoproteína de aproximadamente 32 kDa que exibia uma ligação genética com o locus S (Anderson *et al.*, 1986). Com isso, análises de um grande número de variantes desses genes dentre diferentes membros da família Solanaceae revelaram que suas sequências são altamente polimórficas, variando entre 39% e 98% de similaridade (Anderson *et al.*, 1989; Ai *et al.*, 1990; Clark *et al.*, 1990; Ioerger *et al.*, 1991; Kheyr-Pour *et al.*, 1990; Xu *et al.*, 1990).

Uma observação interessante com relação à evolução dos alelos S nas famílias Solanaceae (Richman & Kohn, 2000) e Rubiaceae (Nowak *et al.*, 2011) é o padrão de evolução trans-específica ou trans-genérica destes. Um alelo encontrado em uma espécie

pode ser mais relacionado geneticamente a um alelo presente em uma outra espécie ou gênero do que a outros alelos presentes em sua própria espécie. Esse fenômeno evidencia que a origem das linhagens alélicas é anterior à diversificação das espécies nas quais atualmente são encontrados. A evolução trans-específica também tem sido reportada para o sistema de auto-incompatibilidade esporofítico das espécies de Brassicaceae (Fobis-Loisy *et al.*, 2004) e em outros genes sabidamente sob seleção balanceadora, incluindo os genes de MHC classe II de vertebrados (Consuegra *et al.*, 2013) e os genes de auto-reconhecimento de alguns fungos (Wu *et al.*, 1998).

Em Papaveraceae, assim como nas espécies de Solanaceae, o sistema de auto-incompatibilidade homomórfico gametofítico é geneticamente controlado por um único locus polimórfico, também denominado locus S (Lawrence *et al.*, 1978; Poulter *et al.*, 2011). Contudo, em contraste ao que ocorre em Solanaceae, durante a resposta de auto-incompatibilidade a inibição do pólen já ocorre na superfície do estigma, antes ou imediatamente após o início da germinação (Franklin-Tong *et al.*, 1992). Muito embora não hajam maiores similaridades entre os sistemas de auto-incompatibilidade gametofítica encontrados nestas duas famílias botânicas. Estudos moleculares com *Papaver rhoeas* tem mostrado que o produto S do pistilo bem como o mecanismo para o reconhecimento do auto-pólen são bem distintos daqueles de Solanaceae (Stone & Goring, 2001).

Em *Papaver*, os determinantes S compreendem as proteínas PrsS codificadas pelo locus S expresso no pistilo e PrpS codificada pelo mesmo locus expresso no pólen (Poulter *et al.*, 2011). A proteína PrsS já foi identificada há algum tempo (Foote *et al.*, 1994; Walker *et al.*, 1996; Kurup *et al.*, 1998), embora só mais recentemente tenha sido denominada como PrsS (*Papaver rhoeas* stigma S determinant) com o objetivo de fornecer uma nomenclatura mais clara (Wheeler *et al.*, 2009). A proteína PrpS (*Papaver rhoeas* pollen S determinant) foi descoberta posteriormente à PrsS por estes mesmos autores e suas sequências indicam que ela é uma proteína pequena (15 kDa) e altamente hidrofóbica. Evidências moleculares sugerem que esta seria uma proteína transmembrana, localizada na membrana plasmática do tubo polínico e que possuiria um domínio externo envolvido com a ligação e reconhecimento da PrsS, sendo responsável pelo desencadeamento de um processo de degeneração da actina e do citoesqueleto do tubo polínico, que acabam por levar à morte destas células (Poulter *et al.*, 2011). Atualmente, mais de 60 alelos já foram descritos para *Papaver*, com um alto grau de polimorfismo entre as sequências de aminoácidos, variando entre 40 a 50% de similaridade (Wheeler *et al.*, 2010).

As espécies da família Poaceae também apresentam um sistema de auto-incompatibilidade homomórfico, contudo nesta família dois loci independentes são responsáveis pelo controle desta característica, os loci S e Z (Bian *et al.*, 2004; Klaas *et al.*, 2011). A inibição do pólen é rápida e ocorre na superfície estigmática (de Nettancourt, 1977). Entretanto, um terceiro locus, independente dos dois loci anteriores, denominado de T, tem sido reportado como um gene modificador que frequentemente pode interferir no sistema de auto-incompatibilidade, causando a quebra deste sistema pelo lado polínico, de forma que a proteína do pólen não é expressa e todos os estigmas aceitam seu pólen (Thorogood & Hayward, 1991; Hayman & Richter, 1992).

A família Brassicaceae também tem sido muito estudada e bem caracterizada quanto ao seu sistema genético de auto-incompatibilidade esporofítica. Este sistema é independente da S-RNase, embora também seja governado por duas proteínas fortemente ligadas. Uma dessas proteínas é produzida no pistilo, a SRK (S locus receptor kinase) e a outra está presente na parede do grão de pólen, a SCR (S locus cysteine rich). Neste sistema o pólen incompatível não penetra o estigma. Quando proteínas SCR e SRK do mesmo grupo de ligação se encontram, desencadeiam um processo fisiológico que culmina com a rejeição do pólen incompatível (Shiu *et al.*, 2004). Este processo é dependente de diversos genes modificadores (Tantikanjana *et al.*, 2009). A origem deste sistema de auto-incompatibilidade tem sido hipotetizada como sendo monofilética (Fobis-Loisy *et al.*, 2004). Há também um alto grau de polimorfismo nestes genes, sendo encontrado até 40% de divergência entre sequências de aminoácidos em proteínas SRK (Nasrallah *et al.*, 1987; Schierup *et al.*, 2001), enquanto que no máximo 30% de divergência é encontrada entre alelos SCR da mesma espécie, provavelmente devido ao pequeno tamanho das proteínas SCR (Watanabe *et al.*, 2000; Kusaba *et al.* 2001). O número de alelos estimados para o locus S é geralmente grande, sendo reportados 22 alelos em *Iberis* (Bateman, 1955), 34 em *Raphanus* (Sampson, 1957), 50 em *Brassica oleracea* (Brace *et al.*, 1994) e 30 em *Brassica campestris* (Nou *et al.*, 1993).

Apesar de Asteraceae também apresentar auto-incompatibilidade esporofítica, estudos tem sugerido uma origem independente para seu sistema e a ausência de homologia entre estes sistemas e os outros anteriormente descritos (Timms *et al.*, 2006). A proposta destes autores para o cenário evolutivo mais provável para as Angiospermas é o recrutamento independente de proteínas múltiplas vezes para o surgimento de sistemas de auto-incompatibilidade esporofítica, com isso, a evolução de diferentes blocos gênicos

teria modulado esta característica em diferentes famílias de plantas, tal qual ocorre na auto-incompatibilidade gametofítica (Uyenoyama, 1995). Contudo, o mecanismo molecular ainda permanece desconhecido nas Asteraceae, muito embora alguns estudos indiquem que seu sistema de auto-incompatibilidade não é compatível com quaisquer dos sistemas já conhecidos, apesar de apresentar algumas similaridades com o sistema encontrado nas Brassicaceae (Allen *et al.*, 2011).

5) Justificativa

Muitos estudos tem sido realizados desde a descoberta dos sistemas de auto-incompatibilidade em plantas com o objetivo de caracterizá-los em nível molecular e entender seu mecanismo fisiológico. Com isso, muitos dados de sequência, estrutura de proteínas e interações proteicas foram acumulados nas últimas duas décadas, o que viabiliza múltiplos estudos evolutivos a partir desses dados. Como não há muitos estudos correlacionando a evolução molecular e a biologia estrutural dessas proteínas em plantas, justifica-se o uso de abordagens evolutivas e de modelagem molecular com enfoque na compreensão dos processos de diversificação funcional dessas proteínas na família Solanaceae. Solanaceae está entre as famílias com maior número de organismos modelo, além de várias espécies com genomas completos sequenciados e o mecanismo de auto-incompatibilidade mais bem conhecido ao nível molecular.

Como discutido acima, há uma grande quantidade de dados experimentais para estas proteínas, contudo ainda há muito a ser explorado com relação aspectos comparativos sobre a diversificação funcional, características estruturais, estratégias evolutivas e relações filogenéticas de genes pertencentes a família gênica das RNasesT2. Apesar da alta diversidade de sequências encontrada nesta família gênica, sua estrutura e função parecem ser conservadas ao longo de sua história evolutiva, com especificidades parecendo estarem mais relacionadas aos órgãos de sua expressão.

6) Objetivos

O objetivo geral deste trabalho foi estudar a diversificação molecular, estrutural e funcional de RNasesT2 na família Solanaceae, bem como analisar os mecanismos

evolutivos pelos quais sua ampla diversificação ocorreu neste grupo e correlacionar a diversidade alélica dentro de seus loci.

Como objetivos específicos tivemos:

- a) Descrever a história evolutiva dos alelos de S-RNases em *Solanum*;
- b) Determinar os resíduos de aminoácidos sob seleção positiva entre alelos de *Solanum*;
- c) Compreender as diferenças estruturais entre os diferentes alelos encontrados em S-RNases de *Solanum*;
- d) Avaliar o impacto estrutural de algumas substituições de aminoácidos tanto em regiões conservadas como em regiões hipervariáveis de S-RNases de *Solanum*;
- e) Descrever a história evolutiva dos genes de RNaseT2 em Solanaceae;
- f) Identificar os resíduos de aminoácidos sob seleção positiva como um sinal de diversificação funcional;
- g) Compreender as diferenças estruturais entre as três classes encontradas nas RNasesT2 de Solanaceae;
- h) Avaliar o impacto estrutural de algumas substituições de aminoácidos na estrutura tridimensional de RNasesT2 em Solanaceae.

7) Estrutura da Tese

Esta tese foi dividida em capítulos que correspondem aos diferentes artigos, resultantes dos experimentos desenvolvidos e contemplando os diferentes objetivos do trabalho.

No primeiro capítulo, o artigo já publicado no periódico *Molecular Genetics and Genomics*, contemplando os objetivos específicos a, b, c e d, descreve padrões estruturais de diversificação alélica em S-RNases presentes no gênero *Solanum*.

No capítulo dois, artigo em preparação para ser submetido à revista *Molecular Biology and Evolution*, foi realizado um estudo sobre os padrões evolutivos e de

divergência estrutural dos membros da família gênica RNase T2 em Solanaceae, referente aos objetivos específicos e, f, g e h.

A última parte da tese é composta por uma discussão dos resultados obtidos nos capítulos anteriores, sintetizando os principais achados em relação ao objetivo geral da Tese.

CONSIDERAÇÕES FINAIS

A presente Tese aborda aspectos teóricos relacionados à evolução molecular e biologia estrutural das S-RNases e S-like RNases na família botânica Solanaceae. Sabe-se que essas proteínas fazem parte da categoria das RNases do tipo transferases, que são classificadas em três principais famílias com base no seu peso molecular e especificidade: RNase A, RNase T1 e RNase T2. As principais características que distinguem essas enzimas são seus pontos isoelétricos, pH ótimo e especificidade ao substrato (de Leeuw *et al.*, 2007). A RNase A é encontrada principalmente em mamíferos e tem preferência por pirimidinas, a RNase T1 está presente em fungos e bactérias, enquanto a RNase T2 encontra-se amplamente distribuída em todos os reinos (Luhtala & Parker, 2010).

A família proteica das RNases T2 possui muitas funções biológicas distintas, dentre as quais podem ser citadas a digestão extracelular de polirribonucleotídeos, a remobilização de fosfato a partir de RNA em condições de deficiência de fosfato, a proteção contra patógenos, a prevenção da auto-fertilização, entre outras, indicando claramente a ocorrência de processos de diversificação funcional. Tradicionalmente, são classificadas de acordo com a similaridade em relação à enzima RNase T2 de *Aspergillus oryzae* (Kawata *et al.*, 1988). Com relação às proteínas T2 RNases de plantas, há três grupos principais classificados de acordo com a similaridade de sequência, função e organização gênica. As Classes I e II englobam as S-like RNases, sendo que as primeiras possuem menos de quatro íntrons e as últimas mais de quatro íntrons. Já a Classe III é constituída pelas S-RNases, que são glicoproteínas extracelulares presentes em, pelo menos, três famílias botânicas distintas: Solanaceae, Rosaceae e Plantaginaceae (Ilgic & Kohn, 2001). Possuem atuação fundamental no sistema de auto-incompatibilidade em plantas e têm sido alvo de diversos estudos nos últimos anos.

Com o objetivo de identificar os genes envolvidos na especificidade do pistilo, foram identificadas proteínas do pistilo que mostravam associação aos haplótipos S e apresentavam diferenças em peso molecular e/ou ponto isoelétrico (Anderson *et al.*, 1986). Estas proteínas foram tratadas como possíveis candidatas para produtos alélicos do gene pistilar responsável pela auto-incompatibilidade e esperava-se que houvesse variações em suas sequências de aminoácidos, em função das suas diferenças moleculares. Assim, Anderson e colaboradores (1986) realizaram a primeira descrição do gene da S-RNase na espécie *Nicotiana glauca*. No entanto, sua natureza bioquímica não havia sido determinada

até então. Isso ocorreu somente após a descrição da primeira sequência pertencente a uma RNase T2 de fungos (Kawata et al. 1988), que compartilhava homologia de seus domínios catalíticos e outros domínios conservados com variantes alélicas de S-RNases (McClure *et al.*, 1989). Adicionalmente, foi demonstrada a atividade ribonucleotídica *in vitro* das S-RNases. Posteriormente, os primeiros experimentos com transformação genética determinaram definitivamente que a S-RNase era a proteína responsável pela auto-incompatibilidade por parte do pistilo, o que foi evidenciado a partir de resultados de ganho e perda de função deste gene (Murfett *et al.*, 1994). Além das S-RNases serem restritas ao pistilo, sua síntese é realizada, principalmente, pelas células do tecido transmissor do estilete. Subsequentemente, são secretadas no espaço extracelular do trato transmissor do estilete, sendo que é mais abundante no seu terço superior, onde a rejeição do tubo polínico incompatível ocorre (Ai *et al.*, 1990).

As S-RNases apresentam um alto grau de diversidade alélica, podendo compartilhar apenas 38% de identidade entre dois alelos que apresentam sequências altamente divergentes (McCubbin & Kao, 2000). Contudo, isso é esperado para uma proteína envolvida neste tipo de reconhecimento. A comparação dessas sequências revelou um padrão uniforme, caracterizado por cinco domínios conservados (C1 a C5) e dois domínios hipervariáveis, HVa e HVb (Ioerger *et al.*, 1991). Alguns estudos de seleção positiva, utilizando métodos baseados na identificação de um aumento estatisticamente significativo na taxa de substituições não-sinônimas em relação às taxas de substituições sinônimas nestas regiões hipervariáveis, relataram uma riqueza de sítios positivamente selecionados nestes domínios. No **Capítulo II** da presente Tese, Brisolara-Corrêa e colaboradores (2015) demonstraram evidências da atuação da seleção positiva nas S-RNases do gênero *Solanum*, corroborando estudos anteriores que utilizaram uma amostragem muito menor e não focada nesse gênero. Além disso, demonstraram diversas mudanças estruturais em algumas regiões das proteínas S-RNases, que foram modeladas computacionalmente para diferentes grupos de S-RNases do gênero *Solanum*, que parecem ter um papel relevante no fenótipo de especificidade pistilar. O fator polínico relacionado a este sistema foi identificado apenas posteriormente (Lai *et al.*, 2002) e seu produto gênico está localizado no citoplasma do tubo polínico (Wang & Xue, 2005). No entanto, ainda não se conhece o modo pelo qual as ribonucleases do sistema S são internalizadas, atravessando a parede celular e a membrana plasmática das células alvo. O domínio F-Box está localizado na região N-terminal da proteína e parece ser necessário para a

funcionalidade da proteína (Meng *et al.*, 2011). Já sua especificidade no reconhecimento de S-RNases tem sido relacionada a um domínio C-terminal de interação entre proteínas (Williams *et al.*, 2015).

Estudos sobre as RNases T2 ganharam um grande impulso após a constatação de que todas as glicoproteínas S-RNases estudadas até então, associadas ao locus S da auto-incompatibilidade gametofítica em plantas, pertenciam a mesma família gênica. Diversos estudos foram realizados posteriormente com intuito de compreender a estrutura gênica e os mecanismos de evolução relacionados às proteínas S-RNases, entre eles a clonagem molecular de genes e a transformação genética para estudos funcionais, bem como experimentos de expressão gênica em diferentes membros deste grupo de genes. No entanto, nenhum estudo relacionou os aspectos evolutivos das S-RNases e *S-like* RNases aos aspectos estruturais. Dessa forma, um dos méritos da presente Tese reside justamente no fato de, pela primeira vez, ter sido avaliado o impacto que a aceleração das taxas de substituições não-sinônimas possui na estrutura tridimensional das proteínas pertencentes às Classes I, II e III das RNases T2 em plantas, em especial, na família botânica Solanaceae.

No **Capítulo III** desta Tese, métodos estatísticos classicamente aplicados à área de evolução molecular bem como métodos teórico-computacionais para simulação de proteínas foram utilizados para o estudo das RNases T2 em Solanaceae. A história evolutiva dessas proteínas indica a formação de três clados principais, estatisticamente bem suportados, um representando as S-RNases e dois referentes às *S-like* RNases. Além disso, diversos aminoácidos foram identificados como estando submetidos à seleção positiva no conjunto de dados que incluía proteínas das Classes I, II e III. A fim de verificar se alguns ramos da árvore filogenética possuíam taxas de substituições não-sinônimas aceleradas, foram utilizados métodos estatísticos representados pelo teste de *branch-site*, implementado pelo pacote PAML. O Clado 3, representado pelas S-RNases, e o Subclado 3 do Clado 2 referente a RNases tipo *S-like* possuem aceleração das taxas de substituição não-sinônimas em relação às taxas de substituição sinônimas. Todos os resíduos positivamente selecionados foram indicados nas estruturas tridimensionais de proteínas do tipo S-RNases e *S-like* RNases. Interessante comentar que a grande maioria dos resíduos de aminoácidos indicados como positivamente selecionados pelo teste *NSsites* do pacote PAML encontram-se nas regiões de maior flexibilidade das proteínas, como evidenciado pelo RMSF (*Root mean square fluctuation*). Enquanto que os resíduos

positivamente selecionados indicados pelo teste *branch-site* encontram-se em regiões mais rígidas das estruturas proteicas. A maior parte dos resíduos sob seleção positiva nas S-RNases estão localizadas nas regiões hipervariáveis HVa e HVb, correspondentes a regiões de *coil* com alta flexibilidade, que têm sido consideradas como as regiões responsáveis pela especificidade alélica. Portanto, não é surpreendente que haja uma pressão seletiva para o aumento da variabilidade de aminoácidos nessas regiões, tendo em vista a sua função biológica. Com relação às *S-like* RNases, nossa hipótese é a de que os resíduos sob seleção positiva nessas proteínas sejam aqueles que levam à diversificação funcional nessas enzimas, funções diferenciadas essas que encontram-se exemplificadas na literatura. Um outro aspecto importante que diferencia estruturalmente as S-RNases das *S-like* RNases refere-se ao maior número de resíduos hidrofóbicos que possuem. A análise de superfície hidrofóbica e hidrofílica acessível ao solvente (SAS) indica uma maior área hidrofóbica em relação à hidrofílica nas S-RNases. Enquanto que nas *S-like* RNases essa relação se inverte, com uma maior superfície hidrofílica acessível ao solvente. Além disso, estão descritas no **Capítulo III** várias outras mudanças estruturais que diferenciam essas classes de enzimas. Todas as análises estruturais presentes nesse Capítulo foram realizadas considerando-se como ponto de partida as estruturas tridimensionais obtidas por difração de raios-X presentes nos bancos de dados. Após simulações de dinâmica molecular por 120 ns, várias modificações estruturais foram evidenciadas nas S-RNases e *S-like* RNases, entre elas, modificações em alfa-hélices e fitas beta.

Um aspecto interessante que não foi avaliado nesta Tese é o papel da N-glicosilação na função das S-RNases, uma vez que possuem um sítio de N-glicosilação conservado representado por Asn28, tendo como referência a estrutura PDB ID: 1I0O. Esse sítio está localizado na superfície da proteína e não está próximo nem do sítio ativo nem das regiões hipervariáveis. Williams e colaboradores (2015) mostraram que as S-RNases mantem sua atividade catalítica no estado não glicosilado, sendo sugerido que são deglicosiladas logo após a sua absorção pelo tubo polínico. Considerando que três diferentes estruturas de N-glicanas compostas de seis a oito resíduos de açúcar podem estar ligadas ao sítio Asn28, seria interessante avaliar o impacto que essas diferentes composições têm na estrutura proteica. Adicionalmente, a dinâmica das S-RNases glicosiladas poderia ser comparada à dinâmica das *S-like* RNases para que possamos ter uma melhor ideia do quanto essa estrutura adicional contribui para a diferenciação entre

essas diferentes classes de enzimas, tendo em vista que as *S-like* RNases não apresentam sítio de glicosilação. Esse seria um ponto importante a ser avaliado em estudos futuros.

Nessa Tese foi realizada uma extensa avaliação do impacto de resíduos de aminoácidos selecionados positivamente na estrutura tridimensional de proteínas RNases T2 das Classes I, II e III da família botânica Solanaceae. Adicionalmente, foram descritos novos aspectos relacionados ao processo de diversificação desses alelos e genes. Isso exemplifica a importância de associarmos estudos evolutivos e estruturais para a compreensão de fenômenos moleculares em famílias multigênicas.

REFERÊNCIAS

- Aguiar B, Vieira J, Cunha AE, Vieira CP (2015) No evidence for Fabaceae Gametophytic self-incompatibility being determined by Rosaceae, Solanaceae, and Plantaginaceae S-RNase lineage genes. *BMC Plant Biology* 15: 129.
- Ai Y, Singh A, Coleman CE, Loeferer TA, Kheyr-Pour A, Kao T-h (1990) Self-incompatibility in *Petunia inflata*: Isolation and characterization of cDNAs encoding three S-allele-associated proteins. *Sex Plant Reprod* 3: 130-138.
- Allen AM, Thorogood CJ, Hegarty MJ, Lexer C, Hiscock SJ (2011) Pollen–pistil interactions and self-incompatibility in the Asteraceae: new insights from studies of *Senecio squalidus* (Oxford ragwort). *Ann Bot* 108: 687-698.
- Anderson MA, Cornish EC, Mau S-L, Williams EG, Hoggart R, Atkinson A, Bonlg I, Grego B, Simpson R, Roche PJ, Haley JD, Penschow JD, Nall HD, Regear GW, Coghlan JP, Crawford RJ, Clarke AE (1986) Cloning of cDNA for a stilar glycoprotein associated with expression of self incompatibility in *Nicotiana glauca*. *Nature* 321: 38-44.
- Anderson MA, McFadden GI, Bernatzky R, Atkinson A, Orpin T, Dedman H, Regear G, Fernley R, Clarke AE (1989) Sequence variability of three alleles of the self-incompatibility gene of *Nicotiana glauca*. *Plant Cell* 1: 483-491.
- Anisimova M, Bielawski JP, Yang Z (2002) Accuracy and power of Bayes prediction of amino acid sites under positive selection. *Mol Biol Evol* 19: 950-958.
- Armstrong SJ, Filatov DA (2008) A cytogenetic view of sex chromosome evolution in plants. *Cytogenet Genome Res* 120: 241-246.
- Barrett SCH (2003) Mating strategies in flowering plants: the outcrossing-selfing paradigm and beyond. *Phil Trans Roy Soc B: Biol Sci* 358: 991-1004.
- Bateman AJ (1955) Self-incompatibility systems in angiosperms. III. Cruciferae. *Heredity* 9: 52-68.
- Betrán E, Long M (2002) Expansion of genome coding regions by acquisition of new genes. *Genetica* 115: 65-80
- Bian XY, Friedrich A, Bai J-R, Baumann U, Hayman DL, Barker SJ, Langridge P (2004) High-resolution mapping of the S and Z loci of *Phalaris coarulescens*. *Genome* 47: 918-930.
- Brace J, King GJ, Ockendon DJ (1994) A molecular approach to the identification of S-alleles in Brassica oleracea. *Sex Plant Reprod* 7: 203-208.

- Clark KR, Okuley JJ, Collins PD, Sims L (1990) Sequence variability and developmental expression of S-alleles in self-incompatible and pseudo-self-compatible *Petunia*. *Pl Cell* 2: 815-826.
- Consuegra S, Ellison A, Allainguillaume J, Pachebat J, Peat KM, Wright P (2013) Balancing selection and the maintenance of MHC supertype variation in a selfing vertebrate. *Proc Royal Society B* 280: 20122854.
- Darwin CR (1877) *The different forms of flowers on plants of the same species*. John Murry, Londres. 352pp.
- Darwin CR (1878) *The effects of cross and self-fertilisation in the vegetable kingdom*, 2a. edição. John Murry, Londres. 471pp.
- de Leeuw M, Roiz L, Smirnoff P, Schwartz B, Shoseyov O, Almog O (2007) Binding assay and preliminary X-ray crystallographic analysis of ACTIBIND, a protein with anticarcinogenic and antiangiogenic activities. *Acta Cryst* 63: 716-719.
- Deshpande RA, Shankar V (2002) Ribonucleases from T2 family. *Crit Rev Microbiol* 28: 79-122.
- Eirín-López JM, Rebordinos L, Rooney AP, Rozas J (2012) The birth-and-death evolution of multigene families revisited. *Genome Dyn.* 7: 170-196.
- Fobis-Loisy I, Miege C, Gaude T (2004) Molecular evolution of the S locus controlling mating in the Brassicaceae. *Plant Biol.* 6: 109-118.
- Foote HCC, Ride JP, Franklin-Tong VE, Walker EA, Lawrence MJ, Franklin FCH (1994) Cloning and expression of a distinctive class of self-incompatibility (S) gene from *Papaver rhoeas* L. *Proc. Natl. Acad. Sci. USA* 91: 2265-2269.
- Franklin-Tong VE, Franklin FCH (2003) The different mechanisms of gametophytic self-incompatibility. *Phil Trans R Soc Lond B* 358: 1025-1032.
- Franklin-Tong VE, Thorlby GJ, Lawrence MJ, Franklin FCH (1992) Recognition signals and pollen response in the incompatibility reaction in *Papaver rhoeas*. In: Mulcahy DL, Bergamini-Mulcahy G, Ottaviano E (eds) *Angiosperm pollen and ovules: Basic and applied aspects*. Springer, New York. pp. 84-93.
- Hayashi T, Kobayashi D, Kariu T, Tahara M, Hada K, Kouzuma Y, Kimura M (2003) Genomic cloning of ribonucleases in *Nicotina glutinosa* leaves, as induced in response to wounding or to TMV-infection, and characterization of their promoters. *Biosci Biotechnol Biochem* 67: 2574-2583.
- Hayman DL, Richter J (1992) Mutations affecting self-incompatibility in *Phalaris coerulescens* Desf. (Poaceae). *Heredity* 68: 495-503.

- Hillwig M, Liu X, Liu G, Thornburg RW, McIntosh GC (2010) *Petunia* nectar proteins have ribonucleases activity. *J Exp Bot* 61: 2951-2965.
- Hillwig MS, Contento AL, Meyer A, Ebany D, Bassham DC, MacIntosh GC (2011) RNS2, a conserved member of the RNase T2 family, is necessary for ribosomal RNA decay in plants. *Proc Natl Acad Sci USA* 108: 1093-1098.
- Hiscock SJ, Kues U, Dickinson HG (1996) Molecular mechanisms of self- incompatibility in flowering plants and fungi - different means to the same end. *Trends Cell Biol.* 6: 421-428.
- Hugot K, Ponchet M, Marais A, Ricci P, Galiana E (2002) A tobacco S-like RNase inhibits hyphal elongation of plant pathogens. *Mol Plant Micr Interact* 15: 243-250.
- Hulst MM, Himes G, Newbiggin E, Moormann RJ (1994) Glycoprotein E2 of classical swine fever virus: expression in insect cells and identification as a ribonuclease. *Virology* 200: 558-565.
- Ida K, Norioka S, Yamamoto M, Kumasaka T, Yamashita E, Newbiggin E, Clarke AE, Sakiyama F, Sato M (2001) The 1.55 Å resolution structure of *Nicotiana alata* SF11-RNase associated with gametophytic self-incompatibility. *J Mol Biol* 314: 103-112.
- Igic B, Kohn JR (2001) Evolutionary relationships among self-incompatibility RNases. *Proc Nat Acad Sci USA* 98: 13167-13171.
- Ioerger TR, Gohlke JR, Xu B, Kao T-H (1991) Primary structural features of the self-incompatibility protein in Solanaceae. *Sex Plant Reprod* 4: 81-87.
- Irie M (1999) Structure-function relationships of acid ribonucleases: lysosomal, vacuolar, and periplasmic enzymes. *Pharmacol Ther* 81: 77-89.
- Jahnen W, Batterham MP, Clarke AE, Moritz RL, Simpson RJ (1989) Identification, isolation and N-terminal sequencing of style glycoproteins associated with self-incompatibility in *Nicotiana alata*. *Plant Cell* 1: 493-499.
- Jiang N, Bao Z, Zhang X, Eddy SR, Wessler SR. 2004. Pack-MULE transposable elements mediate gene evolution in plants. *Nature* 431: 569-573.
- Kawano S, Kakuta Y, Kimura M (2002) Guanine binding site of the *Nicotiana glutinosa* ribonucleases NW revealed by X-ray crystallography. *Biochemistry* 41: 15195-15202.
- Kawano T, Kadono T, Furuichi T, Muto S, Lapeyrie F (2003) Aluminum-induced distortion in calcium signaling involving oxidative bursts and channel regulation in tobacco BY-2 cells. *Biochem Biophys Res Commun* 308: 35-42.
- Kawata Y, Sakiyama F, Tamaoki H (1988) Amino-acid sequence of ribonuclease T2 from *Aspergillus oryzae*. *Eur J Biochem* 176: 683-697.

- Kheyr-Pour A, Bintrim SB, loerger T, Remy R, Hammond S, Kao T (1990) Sequence diversity of pistil S-proteins associated with gametophytic self-incompatibility in *Nicotiana glauca*. *Sex Plant Reprod* 3: 88-97.
- Klaas M, Yang B, Bosch M, Thorogood D, Manzanares C, Armstead IP, Franklin FCH, Barth S (2011) Progress towards elucidating the mechanisms of self-incompatibility in the grasses: further insights from studies in *Lolium*. *Ann Bot* 108: 677-685.
- Kobayashi T, Ogo Y, Aung MS, Nozoye T, Itai RT, Nakanishi H, Yamakawa T, Nishizawa NK (2010) The spatial expression and regulation of transcription factors IDEF1 and IDEF2. *Ann Bot* 105: 1109-1118.
- Kubo K-I, Paape T, Hatakeyama M, Entani T, Takara A, Kajihara K, Tsukahara M, Shimizu-Inatsugi R, Shimizu KK, Takayama S (2015) Gene duplication and genetic Exchange drive the evolution of S-RNase-based self-incompatibility in *Petunia*. *Nat Plants* 1: 1-9.
- Kurihara H, Nonaka T, Mitsui Y, Ohgi K, Irie M, Nakamura KT (1996) The crystal structure of ribonuclease Rh from *Rhizopus niveus* at 2.0 angstrom resolution. *J Mol Biol* 255: 310-320.
- Kurup S, Ride JP, Jordan ND, Fletcher G, Franklin-Tong VE, Franklin FCH (1998) Identification and cloning of related self-incompatibility S genes in *Papaver rhoeas* and *Papaver nudicaule*. *Sex Plant Reprod* 1: 192-198.
- Kusaba M, Dwyer K, Hendershot J, Vrebalov J, Nasrallah JB, Nasrallah ME (2001) Self-incompatibility in the genus *Arabidopsis*: characterization of the S locus in the outcrossing *A. lyrata* and its autogamous relative, *A. thaliana*. *Plant Cell* 13: 627-643.
- Lai Z, Ma W, Han B, Liang L, Zhang Y, Hong G, Xue Y (2002) An F-box gene linked to the self-incompatibility (S) locus of *Antirrhinum* is expressed specifically in pollen and tapetum. *Plant Mol Biol* 50: 29-42.
- Lawrence MJ, Afzal M, Kenrick J (1978) The genetical control of self-incompatibility in *Papaver rhoeas*. *Heredity* 40: 239-253.
- Leite AVL, Machado IC (2010) Reproductive biology of woody species in Caatinga, a dry forest of northeastern Brazil. *J Arid Env* 74: 1374-1380.
- Liu A, Burke JM (2006) Patterns of nucleotide diversity in wild and cultivated sunflower. *Genetics* 173: 321-330.
- Luhtala N, Parker R (2010) T2 Family ribonucleases: ancient enzymes with diverse roles. *Trends Biochem Sci* 35: 253-259.
- Luu D, Qin X, Morse D, Cappadocia M (2000) S-RNase uptake by compatible pollen tubes in gametophytic self-incompatibility. *Nature* 407: 649-651.

- Macintosh GC, Hillwig M, Meyer A, Flagel L (2010) RNase T2 genes from rice and the evolution of secretory ribonucleases in plants. *Mol Genet Genomics* 283: 381-396.
- Martinez M (2011) Plant protein-coding gene families: emerging bioinformatics approaches. *Trends in Plant Sciences* 16: 558-567.
- McClure BA, Haring V, Ebert PR, Anderson MA, Simpson RJ, Sakiyama F, Clarke AE (1989) Style self-incompatibility gene products of *Nicotiana glauca* are ribonucleases. *Nature* 342: 955-957.
- McClure B (2009) Darwin's foundation for investigating self-incompatibility and the progress toward a physiological model for S-RNase-based SI. *J Exp Bot* 60: 1069-1081.
- McClure B, Cruz-García F, Romero C (2011) Compatibility and incompatibility in S-RNase-based systems. *Ann Bot* 108: 647-658.
- McCubbin AG, Kao T-H (2000) Molecular recognition and response in pollen and pistil interactions. *Annu Rev Cell Dev Biol* 16: 333-364.
- Meng X, Sun P, Kao T (2011) S-RNase-based self-incompatibility in *Petunia inflata*. *Annals Bot* 108: 637-646.
- Messier W, and Stewart CB (1997) Episodic adaptive evolution of primate lysozymes. *Nature* 385: 151-154.
- Morgante M, Brunner S, Pea G, Fengler K, Zuccolo A, Rafalski A (2005) Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. *Nat Genet* 37: 997-1002.
- Murfett J, Atherton TL, Mou B, Gasser CS, McClure BA (1994) S-RNase expressed in transgenic *Nicotiana glauca* causes S-allele-specific pollen rejection. *Nature*. 367: 563-566.
- Nakagawa A, Tanaka I, Sakai R, Nakashima T, Funatsu G, Kimura M (1999) Crystal structure of a ribonuclease from the seeds of bitter melon (*Momordica charantia*) at 1.75 Å resolution. *Biochim Biophys Acta* 1433: 253-260.
- Nasrallah JB, Kao TH, Chen C-H, Goldberg ML, Nasrallah ME (1987) Amino-acid sequence of glycoproteins encoded by three alleles of the S locus of *Brassica oleracea*. *Nature* 326: 617-619.
- Nei M, Gu X, Sitnikova T. 1997. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc Natl Acad Sci* 94: 7799-7806.
- Nei M, Rooney AP (2005) Concerted and Birth-and-Death Evolution of Multigene Families. *Annual Review of Genetics* 39: 121-152.
- de Nettancourt D (1977) Incompatibility in angiosperms. Springer-Verlag, Berlin. 230pp.

- Nishikawa S, Morioka H, Kim HJ, Fuchimura K, Tanaka T, Uesugi S, Hakoshima T, Tomita K, Ohtsuka E, Ikehara I (1987) Two histidine residues are essential for ribonuclease T1 activity as is the case for ribonuclease A. *Biochemistry* 26: 8620-8624.
- Nou IS, Watanabae M, Isogai A, Hinata K (1993) Comparison of S-alleles and S-glycoproteins between two wild populations of *Brassica campestris* in Turkey and Japan. *Sex Plant Reprod* 6: 79-86.
- Nowak MD, Davis AP, Anthony F, Yoder AD (2011) Expression and trans-specific polymorphism of self-incompatibility RNases in *Coffea* (Rubiaceae). *PLoS One* 6: e21019.
- Numata K, Kanai A, Saito R, Kondo S, Adachi J, Wilming LG, Hume DA, Hayashizaki Y, Tomita M (2003) Identification of putative noncoding RNAs among the RIKEN mouse full-length cDNA collection. *Genome Res* 13: 1301-1306.
- Ohno S (1970) *Evolution by Gene Duplication*. Springer-Verlag, Berlin. 160pp.
- Poulter NS, Bosch M, Franklin-Tong VE (2011) Proteins implicated in mediating self-incompatibility-induced alterations to the actin cytoskeleton of *Papaver* pollen. *Ann Bot* 108: 659-675.
- Pyšek P, Jarošík V, Chytrý M, Danihelka J, Kühn I, Pergl J, Tichý L, Biesmeijer J, Ellis WN, Kunin WE, Settele J (2011) Successful invaders co-opt pollinators of native flora and accumulate insect pollinators with increasing residence time. *Ecol Monogr* 81: 277-293.
- Ranz, JM, Maurin D, Chan YS, Grotthuss M, Hillier LW, Roote J, Ashburner M, Bergman CM (2007) Principles of genome evolution in the *Drosophila melanogaster* species group. *PLoS Biology* 5: e152.
- Richman AD, Kohn JR (2000) Evolutionary genetics of self-incompatibility in the Solanaceae. *Plant Mol Biol* 42: 169-179.
- Ride JP, Davies EM, Franklin FCH, Marshall DF (1999) Analysis of Arabidopsis genome sequence reveals a large new gene family in plants. *Plant Mol Biol* 39: 927-932.
- Roberts GCK, Dennis EA, Meadows DH, Cohen JS, Jardetsky O (1969) The mechanism of action of ribonuclease. *Proc Natl Acad Sci USA* 62: 1151-1158.
- Rodriguez SM, Panjikar S, Van Belle K, Wyns L, Messens J, Loris R (2008) Nonspecific base recognition mediated by water bridges and hydrophobic stacking in ribonuclease I from *Escherichia coli*. *Protein Sci* 17: 681-690.
- Roiz L, Smirnoff P, Bar-Eli M, Schwartz B, Shoseyov O (2006) ACTIBIND, an actin-binding fungal T2-RNase with antiangiogenic and anticarcinogenic characteristics. *Cancer* 106: 2295-2308.

- Rosenberg HF (2008) RNase A ribonucleases and host defense: an evolving story. *J Leukoc Biol* 83: 1079-1087.
- Rumenapf T, Unger G, Strauss JH, Thiel HJ (1993) Processing of the envelope glycoproteins of pestiviruses. *J Virol* 67: 3288-3294.
- Sampson DR (1957) The genetics of self-incompatibility in the radish. *J Hered* 48: 26-29.
- Schierup MH, Mikkelsen AM, Hein J (2001) Recombination, balancing selection and phylogenies in MHC and self-incompatibility genes. *Genetics* 159: 1833-1844.
- Schneider R, Unger G, Stark R, Schneider-Scherzer E, Thiel HJ (1993) Identification of a structural glycoprotein of an RNA virus as a ribonuclease. *Science* 261: 1169-1171.
- Shiu SH, Karlowski WM, Pan R, Tzeng YH, Mayer KF, Li WH (2004) Comparative analysis of the receptor-like kinase family in *Arabidopsis* and rice. *Plant Cell* 16: 1220-1234
- Stone SL, Arnoldo M, Goring DR (1999) A breakdown of *Brassica* self-incompatibility in ARCI antisense transgenic plants. *Science* 286: 1729-1731.
- Stone SL, Goring DR (2001) The molecular biology of self-incompatibility in flowering plants. *Plant Cell Tissue Organ Cult* 67: 93-114.
- Sun P, Li S, Lu D, Williams JS, Kao T-h (2015) Pollen S-locus F-box proteins of *Petunia* involved in S-RNase-based self-incompatibility are themselves subject to ubiquitin-mediated degradation. *The Plant Journal* 83: 213-223.
- Tantikanjana T, Rizvi N, Nasrallah ME, Nasrallah JB (2009) A dual role for the S-locus receptor kinase in self-incompatibility and pistil development revealed by an *Arabidopsis* *rdr6* mutation. *Plant Cell* 21: 2642-2654.
- Thompson DM, Parker R (2009) The RNase Rny1p cleaves tRNAs and promotes cell death during oxidative stress in *Saccharomyces cerevisiae*. *J. Cell Biol.* 185: 43-50.
- Thorogood, D, Hayward MD (1991) The genetic control of self-compatibility in an inbred line of *Lolium perenne* L. *Heredity* 67: 175-181.
- Timms L, Jimenez R, Chase M, Lavelle D, McHale L, Kozik A, Lai Z, Heesacker A, Knapp S, Rieseberg L, Michelmore R, Kesseli R (2006) Analyses of synteny between *Arabidopsis thaliana* and species in the Asteraceae reveal a complex network of small syntenic segments and major chromosomal rearrangements. *Genetics* 173: 2227-2235.
- Uyenoyama MK (1995) A generalized least-squares estimate for the origin of sporophytic self-incompatibility. *Genetics* 139: 975-992.
- Wagner A (2001) Birth and death of duplicated genes in completely sequenced eukaryotes. *Trends in Genetics* 15: 237-239.

- Walker EA, Ride JP, Kurup S, Franklin-Tong VE, Lawrence MJ, Franklin FCH (1996) Molecular analysis of two functional homologues of the S3 allele of the *Papaver rhoeas* self-incompatibility gene isolated from different populations. *Plant Mol Biol* 30: 983-994.
- Wang H-Y, Xue Y-B (2005) Subcellular localization of the S locus F-box protein AhSLF-S2 in pollen and pollen tubes of self-incompatibility *Antirrhinum*. *J Integrat Plant Biol* 47: 76-83.
- Watanabe MA, Ito A, Takada Y, Ninomiya C, Kakizaki T, Takahata Y, Hatakeyama K, Hinata K, Suzuki G, Takasaki T, Satta Y, Shiba H, Takayama S, Isogai A (2000) Highly divergent sequences of the pollen self-incompatibility (S) gene in class-I S haplotypes of *Brassica campestris* (syn. *rapa*) L. *FEBS Lett* 473: 139-144.
- Wheeler MJ, de Graaf BH, Hadjiosif N, Perry RM, Poulter NS, Osman K, Vatovec S, Harper A, Franklin FC, Franklin-Tong VE (2009) Identification of the pollen self-incompatibility determinant in *Papaver rhoeas*. *Nature* 459: 992-995.
- Wheeler MJ, Vatovec S, Franklin-Tong VE (2010) The pollen S-determinant in *Papaver*: comparisons with known plant receptor and protein ligand partners. *J Exp Bot* 61: 2015-2025.
- Williams JS, Wu L, Li S, Sun P, Kao T-H (2015) Insight into S-RNase-based self-incompatibility in *Petunia*: recent findings and future directions. *Front Plant Sci* 6: 41.
- Wu J, Saupe SJ, Glass N L (1998) Evidence for balancing selection operating at the het-c heterokaryon incompatibility locus in a group of filamentous fungi. *Proc Natl Acad Sci USA* 95: 12398-12403.
- Xu B, Mu J, Nevins DL, Grun P, Kao T-H (1990) Cloning and sequencing of cDNAs encoding two self-incompatibility associated proteins in *Solanum chacoense*. *Mol Gen Genet* 224: 341-346.
- Yang Z (2006) *Computational Molecular Evolution*. Oxford University Press, Oxford, England. 357p.
- Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24: 1586-1591.