

AMANDA BORTOLINI SILVEIRA

“Qua-Quine Starch de Arabidopsis thaliana, um gene novo regulado por metilação de DNA e propenso a variação epialélica”

**CAMPINAS
2012**

UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE BIOLOGIA



AMANDA BORTOLINI SILVEIRA

“Qua-Quine Starch de Arabidopsis thaliana, um gene novo regulado por metilação de DNA e propenso a variação epialélica”

Este exemplar corresponde à redação final
da tese defendida pelo(a) candidato (a)
Amanda Bortolini Silveira
e aprovada pela Comissão Julgadora.

Tese apresentada ao Instituto de
Biologia para obtenção do Título de
Doutora em Genética e Biologia
Molecular na área de Genética Vegetal e
Melhoramento.

Orientador: Prof. Dr. Michel Georges Albert Vincentz

CAMPINAS,
2012

FICHA CATALOGRÁFICA ELABORADA POR
ROBERTA CRISTINA DAL' EVEDOVE TARTAROTTI – CRB8/7430
BIBLIOTECA DO INSTITUTO DE BIOLOGIA - UNICAMP

Si39q Silveira, Amanda Bortolini, 1983-
Qua-Quine Starch de *Arabidopsis thaliana*, um gene novo regulado por metilação de DNA e propenso a variação epialélica / Amanda Bortolini Silveira. – Campinas, SP: [s.n.], 2012.

Orientador: Michel Georges Albert Vincentz.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. *Arabidopsis thaliana*. 2. Epigenética. 3. Metilação de DNA. 4. Variação natural. 5. Epialelos.
I. Vincentz, Michel Georges Albert, 1958-. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.

Informações para Biblioteca Digital

Título em Inglês: *Qua-Quine Starch*, a new gene of *Arabidopsis thaliana* regulated by DNA methylation and prone to epiallelic variation

Palavras-chave em Inglês:

Arabidopsis thaliana

Epigenetics

DNA methylation

Natural variation

Epiallele

Área de concentração: Genética Vegetal e Melhoramento

Titulação: Doutor em Genética e Biologia Molecular

Banca examinadora:

Michel Georges Albert Vincentz [Orientador]

Celso Eduardo Benedetti

Paulo Cavalcanti Gomes Ferreira

Renato Vicentini dos Santos

Fabio Tebaldi Silveira Nogueira

Data da defesa: 27-08-2012

Programa de Pós Graduação: Genética e Biologia Molecular

Campinas, 27 de Agosto de 2012

BANCA EXAMINADORA

Prof. Dr. Michel Georges Albert Vincentz (Orientador)


Assinatura

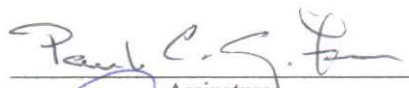
Prof. Dr. Celso Eduardo Benedetti


Assinatura

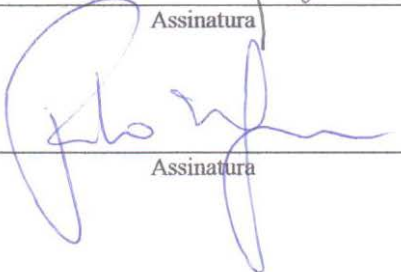
Prof. Dr. Renato Vicentini dos Santos


Assinatura

Prof. Dr. Paulo Cavalcanti Gomes Ferreira


Assinatura

Prof. Dr. Fábio Tebaldi Nogueira


Assinatura

Prof. Dr. Fábio Papes

Assinatura

Prof. Dr. Daniel Scherer de Moura

Assinatura

Profa. Dra. Katlin Massirer

Assinatura

AGRADECIMENTOS

À **FAPESP**, pelo apoio financeiro concedido.

Ao meu orientador **Dr. Michel Vincentz** pelo apoio, excelente orientação e entusiasmo.

Aos **membros da pré-banca e banca** de defesa.

Ao **Dr. Vincent Colot** e a todos os membros de seu laboratório sem os quais este trabalho dificilmente teria sido realizado.

Ao **Dr. Olivier Loudet** e **Charlotte Trontin** pela enorme contribuição a este projeto.

Aos **parceiros do Laboratório de Genética de Plantas** pelas discussões científicas e não científicas, ajuda e amizade.

À **Sandra Lúcia Martins** pelo apoio técnico e ajuda com o cultivo das plantas.

À **Sandra Scarano, Tânia Zambon** e **Gabriela Moraes** pela ajuda administrativa.

Ao **Dr. José Andres Yunes** pelo apoio e oportunidade de trabalho no Centro Infantil Boldrini.

Ao **Joan** pelo companheirismo, suporte e por sempre tentar manter minha motivação e entusiasmo em alta.

Ao meu irmão **André** por sempre estar disposto a ajudar.

À meus pais, **José Geraldo** e **Selma**, pelo apoio incondicional.

À todos o meu mais sincero **MUITO OBRIGADA!**

SUMÁRIO

LISTA DE FIGURAS E TABELAS	vi
RESUMO	viii
ABSTRACT	ix
1. INTRODUÇÃO	
1.1 – Epigenética: conceito e definição	1
1.2 – Metilação de DNA e modelos de estudo	4
1.2.1 – Um panorama da metilação de DNA em <i>A. thaliana</i>	5
1.2.2 – Estabelecimento e manutenção dos padrões de metilação de DNA	8
1.2.3 – Herança transgeracional dos padrões de metilação e epiRILs	12
1.2.4 – Metilação de DNA e o controle da expressão gênica	15
1.2.5 – Importância ecológica e evolutiva da metilação de DNA	18
1.3 – Genes novos – conceito, mecanismos de origem e importância evolutiva	21
1.4 – <i>Qua-Quine Starch</i> de <i>A. thaliana</i> , um gene novo potencialmente relacionado ao metabolismo de amido	22
2. OBJETIVOS	
2.1 – Geral	23
2.2 – Específicos	23
3. RESSULTADOS E DISCUSSÃO	
3.1 – Identificação de variação epialélica para o gene <i>Qua-Quine Starch</i> de <i>A.thaliana</i> .	24
3.2 – Reprodução do manuscrito	
“ Prevalent natural epigenetic variation at a newly formed gene in <i>A. thaliana</i> ” ..	27
3.3 – Análise preliminar dos mecanismos de controle envolvidos no silenciamento transcricional de <i>QQS</i>	55
3.3.1 - <i>QQS</i> está sobre um controle epigenético diferenciado do previamente descrito para os genes <i>FWA</i> e <i>SDC</i> de <i>A. thaliana</i>	55
3.3.2 - RdDM em <i>QQS</i> envolve a participação de <i>DCL2</i> , <i>DCL3</i> e <i>DCL4</i>	56
3.3.3 - RdDM em <i>QQS</i> envolve <i>Pol II</i> ao invés de <i>Pol IV</i> e <i>V</i>	58
4. CONSIDERAÇÕES FINAIS	61
4. PERSPECTIVAS	63
6. REFERÊNCIAS BIBLIOGRÁFICAS	64
7. ANEXOS	71

LISTA DE FIGURAS E TABELAS

INTRODUÇÃO

Figura 1. Organização da cromatina e seu impacto sobre o controle da expressão gênica	2
Figura 2. Exemplos clássicos de alterações de expressão gênica relacionadas a alterações epigenéticas	3
Figura 3. Metilação de DNA em eucariotos	5
Figura 4. Representação esquemática da distribuição cromossômica de genes, elementos repetitivos e metilação de DNA em <i>A. thaliana</i>	7
Figura 5. Representação esquemática da distribuição de metilação de DNA nos contextos CG, CHG e CHH no genoma de <i>A. thaliana</i>	7
Figura 6. Mecanismos de estabelecimento e manutenção dos padrões de metilação nos contextos CG, CHG e CHH em sequências repetidas	10
Figura 7. Distribuição de metilação de DNA nos três contextos de sequência em plantas selvagens, mutantes simples, duplos e triplos para DNA metiltransferases de <i>A.thaliana</i>	12
Figura 8. Construção de linhagens epiRIL Col-0 wt derivadas de <i>ddm1</i>	14
Figura 9. Exemplos de variantes epialélicas em <i>A. thaliana</i>	17
Figura 10. Distribuição de sequências repetidas em relação a genes em <i>Arabidopsis</i> e arroz	18
Tabela 1. Porcentagem de metilação nos três contextos de sequência em plantas selvagens e mutantes simples, duplos e triplos para as DNA metiltransferases de metilação <i>de novo</i> e manutenção	11

RESULTADOS E DISCUSSÃO

PARTE I

Figura 11. Padrão de expressão dos genes <i>QQS</i> e <i>AtbZIP9</i> em linhagens expressando fusões entre a sequência promotora destes genes e o gene repórter <i>GUS</i>	25
Figura 12. Análise de correlação entre <i>AtbZIP9</i> , <i>QQS</i> e a via de sinalização mediada por <i>APL</i>	25
Figura 13. Análise do padrão de expressão de <i>QQS</i> em plantas F1 provenientes de um cruzamento entre Col-0* e Col-0	26

PARTE II

Figure 1. <i>QQS</i> is a target of epigenetic regulation	30
Figure S1. Distribution of DNA methylation at the <i>QQS</i> promoter and 5' UTR sequences ..	30
Figure 2. Spontaneous epigenetic variation at <i>QQS</i>	32
Figure S2. Assessment of <i>QQS</i> DNA methylation rate and transcript accumulation in seedlings of <i>ddm1-2</i> and <i>rdr2-1</i> mutants and Col-0 and Col-0* S1 and S2 single seed descent lines	33
Figure S3. Comparative genome-wide analysis of Col-0 and Col-0*	34
Figure 3. <i>QQS</i> is under autonomous epigenetic control	36
Figure S4. <i>QQS</i> is under autonomous epigenetic control	37
Figure 4. Analysis of <i>QQS</i> DNA methylation and expression profile in <i>A. thaliana</i> accessions representing the worldwide diversity	39
Figure S5. Genetic polymorphisms (SNPs and Indels) at <i>QQS</i> locus and flanking region in accessions carrying methylated and hypomethylated <i>QQS</i> epialleles	40
Figure 5. <i>cis</i> -DNA methylation profiles modulate <i>QQS</i> expression	41
Figure 6. <i>QQS</i> epiallelic variation in wild populations of Central Asia	43
Table S1. Primer List	49

PARTE III

Figura 14. <i>QQS</i> , <i>FWA</i> e <i>SDC</i> são caracterizados por sistemas de silenciamento transcricional diferenciados	56
Figura 15. Padrão de metilação e expressão de <i>QQS</i> em diversos mutantes para proteínas da maquinaria de silenciamento	59
Figura 16. Modelo genético para a manutenção do padrão de metilação de <i>QQS</i> via RdDM envolvendo a atuação de <i>DCL2</i> , <i>DCL3</i> e <i>DCL4</i>	60

RESUMO

Modificações epigenéticas do DNA ou da cromatina atuam principalmente no controle da atividade de elementos de transposição, podendo também silenciar genes, geralmente quando estes estão associados a elementos de transposição ou sequências repetidas. Em plantas, alguns alelos epigenéticos afetando caracteres como morfologia floral, florescimento, estatura ou amadurecimento do fruto foram descritos, revelando o potencial deste tipo de regulação para gerar variabilidade fenotípica herdável não necessariamente vinculada a alterações da sequência de DNA. No entanto, o impacto de mecanismos epigenéticos em processos de evolução adaptativa é ainda bastante desconhecido, em parte, pela falta de informação sobre variação epigenética em populações naturais. Identificamos *Qua-Quine Starch (QQS)* de *Arabidopsis thaliana* como um gene sob um controle epigenético flexível e, portanto, particularmente propenso a variações epialélicas frequentes na natureza. *QQS* é um gene recente, que provavelmente originou-se *de novo* em *Arabidopsis thaliana* em uma região rica em elementos de transposição. Mostramos que *QQS* apresenta-se diferencialmente expresso entre acessos naturais assim como entre indivíduos diretamente coletados na natureza e que estas diferenças de expressão estão negativamente correlacionadas com o nível de metilação de sequências repetidas localizadas em sua região promotora e 5' UTR, não estando relacionadas a variação genética em *cis* ou *trans*. Mostramos ainda que variação epialélica em *QQS* é independente do nível de metilação de transposons vizinhos e que pode ser estavelmente herdada entre gerações. Considerando o impacto potencial de padrões de expressão contrastantes de *QQS* no metabolismo de amido, um importante componente para produção de biomassa e crescimento, sugerimos que variação epialélica em *QQS* possa ter implicações adaptativas. Nossos dados também apontam pela primeira vez uma ligação potencial entre mecanismos epigenéticos e o processo de evolução de genes novos. Propomos que genes novos, especialmente os de origem *de novo*, poderiam ser mais propensos a variar epigeneticamente, o que permite um ajuste fino de seu padrão de expressão até que o estado mais vantajoso seja fixado geneticamente.

ABSTRACT

Epigenetic modifications of DNA or chromatin control of the activity of transposable elements and can also silence genes which are associated to transposons or repetitive sequences. In plants, epigenetic alleles affecting characters such as floral morphology, flowering, stature or fruit ripening have been described, highlighting the potential of this type of regulation in generating heritable phenotypic diversity, not necessarily linked to DNA sequence alterations. However, the impact of epigenetic mechanisms in adaptative evolution is still largely unknown, in part, due to the lack of information about epiallelic variation in natural populations. We have identified *Qua-Quine Starch (QQS)* of *Arabidopsis thaliana* as a gene under flexible epigenetic control and thus particularly prone to epiallelic variation in nature. *QQS* is a recent gene that likely originated *de novo* in *Arabidopsis thaliana* in a transposon-rich region. We show that *QQS* is differentially expressed among natural accessions as well as among individuals directly sampled from the wild and that these expression differences are negatively correlated with the DNA methylation level of repeat sequences located on *QQS* promoter and 5'UTR region and are not correlated with *cis* or *trans* genetic variation. We also show that epiallelic variation at *QQS* is independent of the methylation status of nearby transposable elements and can be stably inherited across generations. Considering the potential impact of contrasting *QQS* expression patterns on starch accumulation, an important component of biomass production and growth, we suggest that epiallelic variation at *QQS* may have adaptative implications. Our data also points for the first time to a potential link between epigenetic mechanisms and the evolution of novel genes. We suggest that novel genes, more specifically those created *de novo*, could be endowed with an increased potential for epigenetic variation and thus for adjusting their expression pattern until the most adaptive state becomes genetically fixed.

1. INTRODUÇÃO

1.1 - Epigenética: conceito e definição

Epigenética literalmente significa “em cima” ou “além da” genética. O termo foi primeiramente usado por C. Waddington em um contexto bastante amplo referindo-se ao estudo da “epigenesis” ou a investigação de como a informação genética de um organismo interage com o ambiente ao longo do desenvolvimento para determinar o fenótipo (Waddington, 1942; Kinoshita & Jacobsen 2012). O uso moderno da palavra como proposto por A. Riggs “estudo de mudanças na função gênica herdáveis mitótica e/ou meoticamente que não podem ser explicadas por mudanças na sequência de DNA” (Russo et al. 1996) ou mais recentemente por A. Bird “adaptação estrutural de regiões da cromatina para registrar, sinalizar ou perpetuar estados de atividade alterados” (Bird 2007) refere-se a como informações herdáveis codificadas além da sequência de DNA, particularmente modificações covalentes do DNA (i.e., metilação) ou mudanças da estrutura da cromatina, alteram a maneira como os genes são expressos (Figura 1). Desta maneira, embora um organismo seja caracterizado por um genoma praticamente invariante entre os diferentes tipos celulares, não existe um único epigenoma, mas sim dezenas a centenas, cada um específico para cada tipo celular ou fase do ciclo celular e que provavelmente muda em resposta a sinais do desenvolvimento, fisiológicos e do ambiente.

O controle epigenético é essencial para o desenvolvimento, diferenciação e comprometimento dos diversos tipos celulares de um organismo, assim como para processos celulares como inativação do cromossomo X em fêmeas de mamíferos (Payer et al. 2011), expressão alelo específica (*imprinting*) e controle de transposons em plantas e mamíferos (Scott & Spielman 2004; Slotkin & Martienssen 2007) e silenciamento dos loci de *mating type* em levedura (Grewal 2000). Alterações epigenéticas podem ter efeitos similares a mutações clássicas e exemplos bastante familiares em diversos organismos incluem alterações na coloração do olho em drosófila (Schotta et al. 2003; Figura 2A) e na cor da pelagem de camundongos (Dolinoy 2008; Figura 2B); o variante pelórico em flores de *Linaria vulgaris* (Cubas et al. 1999; Figura

2C) e o desenvolvimento de doenças como câncer (Sharma et al. 2010). Atualmente, metilação de DNA e modificações de caudas de histonas representam as marcas epigenéticas melhor caracterizadas em eucariotos.

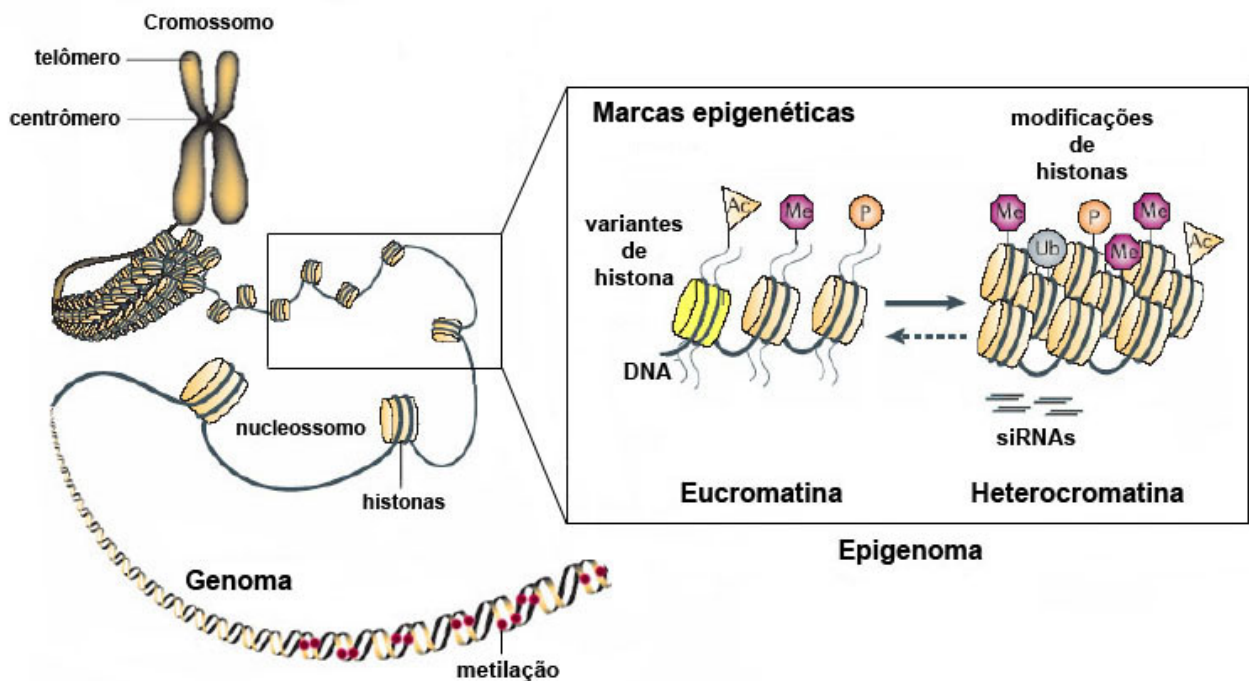


Figura 1. Organização da cromatina e seu impacto sobre o controle da expressão gênica. A unidade fundamental da cromatina é o nucleossomo, composto por DNA enrolado em um núcleo de histonas formado por um tetrâmero H3/H4 e dois dímeros H2A/H2B. Marcas epigenéticas como metilação de DNA, modificações de caudas de histonas, incorporação de histonas não canônicas aos nucleossomos e outros fatores como enzimas remodeladoras de cromatina, proteínas do complexo Polycomb/Trithorax e pequenos RNAs interagem para determinar a conformação da cromatina em resposta a sinais do ambiente ou do desenvolvimento. Estados eucromáticos são caracterizados por cromatina relaxada e ativa transcricionalmente. Estados heterocromáticos correspondem a cromatina condensada e não permissiva a expressão gênica.

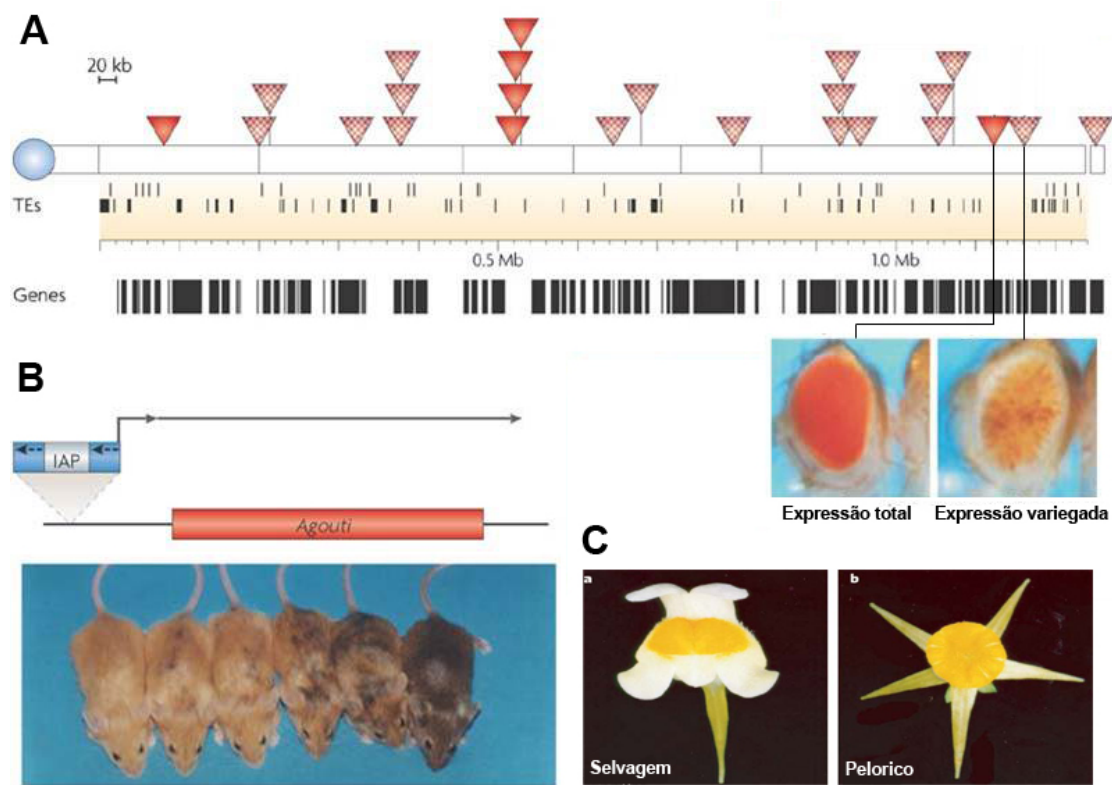


Figura 2. Exemplos clássicos de alterações de expressão gênica relacionadas a alterações epigenéticas. (A) Efeito variegado de posição em drosófila, que descreve o silenciamento de genes eucromáticos quando estes são inseridos próximos a regiões de heterocromatina. Neste caso, inserção do gene *White* em regiões eucromáticas resulta em expressão total do gene e cor de olho vermelha. Em contrapartida, inserção em regiões heterocromáticas contendo elementos de transposição (TEs) resulta em um fenótipo variegado em razão de silenciamento transcricional em número variável de células, o que é indicado pela cor branca. (B) Locus *Agouti* de camundongo no qual a inserção de um retrotransposon (IAP) na sequência promotora do gene influencia sua expressão em função do nível de metilação do retrotransposon, causando variação na coloração da pelagem. (C) Variante pelórico em flores de *Linaria vulgaris* que surge em função de silenciamento do gene *Lcyc* devido a transmissão estável de metilação de DNA neste locus. Ilustrações retiradas de Slotkin & Martienssen 2007 (A e B) e Cubas et al. 1999 (C).

1.2 - Metilação de DNA e modelos de estudo

Em eucariotos, metilação de DNA refere-se à transferência de um grupo metil ao carbono 5 de citosinas através da ação de DNA metiltransferases (Figura 3). Metilação de DNA é uma marca epigenética geralmente repressiva, intimamente associada a modificações de histonas e condensação da cromatina. Apesar de perdida em *S. cerevisiae*, *S. pombe* e *C. elegans* e estar presente em baixíssimos níveis em *D. melanogaster*, metilação de DNA representa um componente central do controle epigenético em alguns fungos, plantas e animais vertebrados, onde é crítica para desenvolvimento normal e manutenção da integridade do genoma através do controle da ativação e movimentação de elementos de transposição (Goll & Bestor 2005; Zemach et al. 2010; Feng et al. 2010; Feng & Jacobsen 2011). Atestando esta importância, mutações em certas DNA metiltransferases em mamíferos resultam em letalidade precoce do embrião (Okano et al. 1999) e em plantas levam a alterações pleiotrópicas da morfologia e do desenvolvimento (Richards 1997; Henderson & Jacobsen 2008). Plantas, no entanto, são mais plásticas e toleram melhor perdas massivas de metilação. Além disto, diferentemente de mamíferos, cujos padrões de metilação são apagados e reestabelecidos durante a gametogênese e início do desenvolvimento, esta informação epigenética em plantas é em grande parte mantida através da meiose, o que as torna um sistema ideal para o estudo desta marca epigenética e sua herança em eucariotos. Além disto, em plantas, as linhagens celulares reprodutivas são estabelecidas tardiamente no desenvolvimento, de maneira que alterações epigenéticas adquiridas ao longo do desenvolvimento podem ser herdadas. Devido a um genoma completo e bem anotado, grande conjunto de ferramentas genéticas e facilidade de manipulação, *Arabidopsis thaliana* (*A. thaliana*) se tornou um dos principais modelos de estudo no campo da epigenética, o que se reflete no fato desta planta possuir um dos metilomas (i.e., mapa de metilação de citosinas do genoma) mais extensivamente caracterizados em eucariotos.

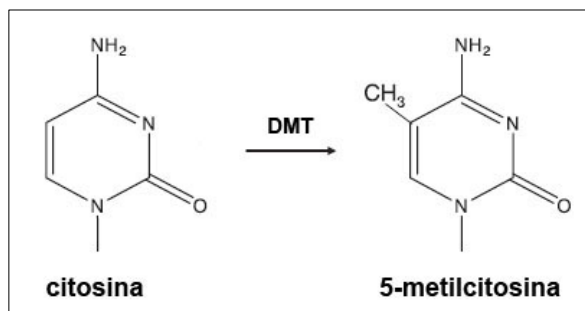


Figura 3. Metilação de DNA em eucariotos. Metilação corresponde a transferência de um grupo metil (CH₃), proveniente de uma molécula de S-adenosilmetionina, ao carbono 5 do anel aromático de citosinas em uma reação catalisada por DNA metiltransferases (DMT).

1.2.1 - Um panorama da metilação de DNA em *A. thaliana*

Apesar de um genoma compacto de aproximadamente 120Mb, *A. thaliana* possui quantidades significativas de heterocromatina (20Mb), rica em elementos de transposição e sequências repetidas e localizada principalmente em regiões centroméricas e pericentroméricas. O resto do genoma (100Mb) abriga regiões de eucromatina distribuídas ao longo dos braços cromossômicos e composta principalmente de genes e suas sequências flanqueadoras (Henderson & Jacobsen 2007; Zhang 2008). Em plantas, metilação de citosinas ocorre em dinucleotídeos CG, como em mamíferos, mas também em sítios CHG e CHH (onde H representa A, T ou C). Em *A. thaliana*, metilação afeta 6-7% das citosinas, sendo que 55% das metilcitosinas ocorrem no contexto CG, 23% no contexto CHG e 22% no contexto CHH. Considerando o número total de citosinas em cada contexto no genoma (13%, 15% e 72% para os contextos CG, CHG e CHH, respectivamente), temos que aproximadamente 24% dos sítios CG estão metilados, contra apenas 7% para CHG e menos de 2% para CHH. Em geral, o nível de metilação global em sítios CG, CHG e CHH varia entre 80-100%, 20-100% e menos do que 20%, respectivamente, o que indica um padrão de metilação bastante homogêneo em sítios CG entre os diversos tipos celulares, variável para sítios CHG e específico de alguns tipos celulares para CHH (Zhang et al. 2006; Cokus et al. 2008; Lister et al. 2008).

Metilcitosinas ocupam cerca de 20% das regiões cromossômicas em *A. thaliana* (Zhang 2008) e estão divididas em dois padrões principais. O primeiro, associado a altos níveis de metilação nos três contextos de sequência em elementos de transposição e outras sequências repetidas, encontrados principalmente em regiões heterocromáticas centroméricas e pericentroméricas, mas também como pequenas manchas entre genes nos braços eucromáticos (Figura 4). O segundo, caracterizado por metilação do tipo CG na porção 3' da região transcrita ou "corpo" de genes codificantes de proteínas, localizados predominantemente em regiões eucromáticas distribuídas ao longo dos braços cromossômicos (Figura 5; Zhang et al. 2006; Cokus et al. 2008; Lister et al. 2008). Funcionalmente estes dois tipos de metilação desempenham papéis bastante distintos. Metilação em elementos de transposição representa um dos principais mecanismos para supressão da atividade destes elementos no genoma, uma vez que perdas massivas de metilação nestas regiões levam a reativação transcricional e aumento da mobilização destes elementos, ameaçando a estabilidade do genoma. Metilação sobre DNA repetitivo é bastante estável entre acessos de *Arabidopsis* (Vaughn et al. 2007) o que evidencia a sua importância na proteção dos genomas contra os efeitos deletérios da atividade de transposons. Em contrapartida, metilação no corpo de genes parece estar associada à expressão ao invés de silenciamento, dada sua preferência por genes com expressão moderada e constitutiva (Zilberman et al. 2007; Zhang 2008). Este tipo de metilação afeta cerca de 20-30% dos genes de *A. thaliana* e é altamente polimórfica entre acessos naturais (Vaughn et al. 2007). Quando perdida não parece levar a alterações de expressão (Zhang et al. 2006; Zilberman et al. 2007), de maneira que sua real função ainda é pouco compreendida. Especula-se que em plantas ela possa representar apenas um subproduto da transcrição (Teixeira & Colot 2009), embora sua aparente preferência por exons ao invés de introns sugira um papel na definição de exons ou controle de *splicing* alternativo (Feng & Jacobsen 2011). Outros trabalhos sugerem um papel na prevenção de transcrição a partir de promotores alternativos crípticos (Zilberman et al. 2007) ou ainda impedindo que estes genes sejam responsivos a fatores do desenvolvimento e ambientais (Aceituno et al. 2008).

Além destes dois padrões principais, metilação de DNA também ocorre em 5% das sequências promotoras de genes em *A. thaliana*, normalmente associada à presença de elementos de transposição ou suas relíquias (Zhang et al. 2006). Em geral, genes com sequências

promotoras metiladas possuem baixa expressão e frequentemente padrão de expressão tecido específico (Henderson & Jacobsen 2007).

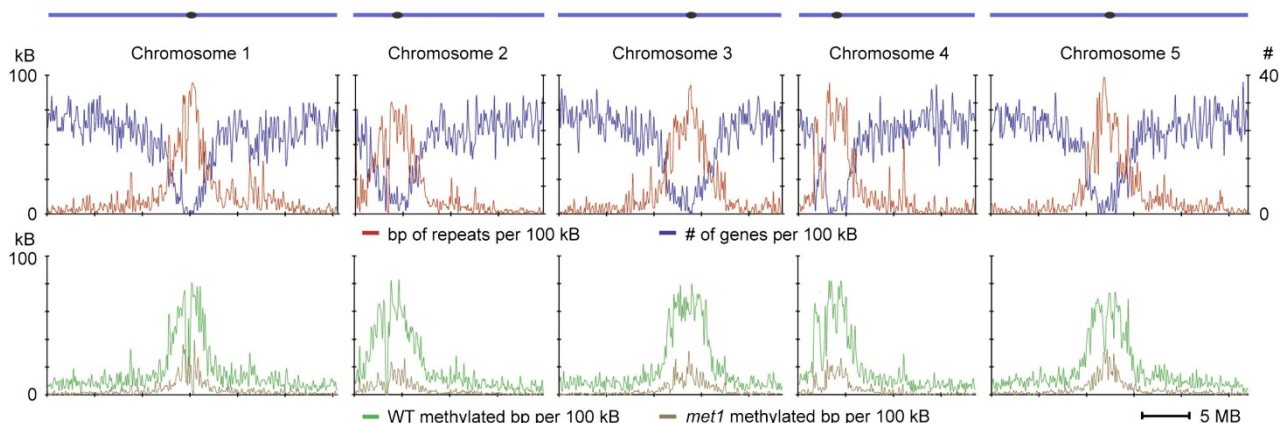


Figura 4. Representação esquemática da distribuição cromossômica de genes, elementos repetitivos e metilação de DNA em *A. thaliana*. Painéis superiores representam o tamanho total de sequências repetidas (eixo y esquerda) e o número de genes (eixo y direita) em uma janela de 100kb. Painéis inferiores, o tamanho total de DNA metilado em plantas selvagem (WT) e mutante *met1* (ver a seguir) no mesmo intervalo. Nota-se que a distribuição dos genes é inversamente correlacionada com a de elementos repetidos e metilação de DNA, que estão intimamente associados às regiões centroméricas e pericentroméricas, identificadas pelos círculos pretos. Figura adaptada de Zhang et al. 2006.

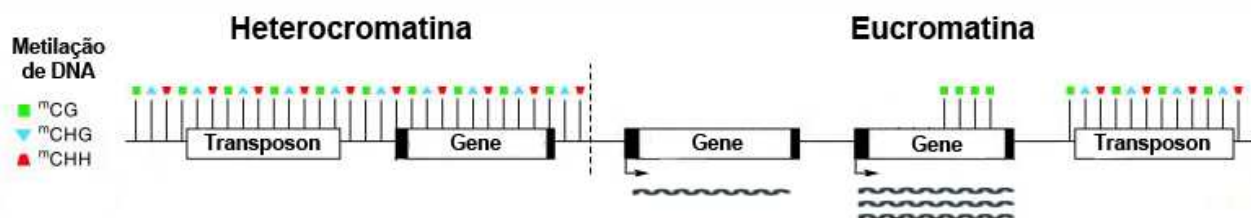


Figura 5. Representação esquemática da distribuição de metilação de DNA nos contextos CG, CHG e CHH no genoma de *A. thaliana*. Regiões heterocromáticas ricas em elementos de transposição e alguns genes silenciados são caracterizadas por alta densidade de metilação de DNA nos três contextos de sequência, assim como elementos de transposição localizados em regiões eucromáticas. No entanto, neste último caso metilação é normalmente restrita ao elemento de transposição e não se espalha para regiões vizinhas como ocorre em zonas heterocromáticas. Genes com expressão moderada tendem a possuir metilação do tipo CG na porção 3' de sua sequência transcrita. Figura adaptada de Feng & Jacobsen 2011.

1.2.2. - Estabelecimento e manutenção dos padrões de metilação de DNA

Em *Arabidopsis*, o padrão de metilação nos três contextos de sequência é inicialmente estabelecido sobre transposons e outras sequências repetidas por uma via complexa conhecida como RdDM (*RNA-directed DNA METHYLATION*), que envolve componentes da maquinaria de siRNAs como uma das quatro enzimas com atividade *RNAse III* de *Arabidopsis* (*DCL3*, *DICER-LIKE III*), uma RNA polimerase dependente de RNA (*RDR2*, *RNA DEPENDENT RNA POLYMERASE 2*), duas RNA polimerases específicas de plantas derivadas de *Pol II* mas especializadas em silenciamento transcricional (*Pol IV* e *Pol V*), uma *de novo* DNA metiltransferase *DRM2* (*DOMAINS REARRANGED METHYLTRANSFERASE 2*), fatores remodeladores de cromatina e outras proteínas recentemente identificadas (Law & Jacobsen 2010; Teixeira & Colot 2010; He et al. 2011; Meyer 2011). De uma maneira simplificada, RNA dupla fita gerado a partir de um transcrito simples fita produzido por *Pol IV* e transformado em dupla fita pela atuação de *RDR2* é reconhecido e clivado por *DCL3* em siRNAs de 24nt. Estes siRNAs de 24nt são então incorporados em um complexo de silenciamento contendo *AGO4* (*ARGONAUTE 4*) que, através de interação com um RNA complementar transcrito por *Pol V*, irá recrutar *DRM2* a sequência de DNA alvo a ser metilada (Figura 6A). Transcritos produzidos tanto por *Pol IV* quanto *Pol V* em associação com o complexo efetor de silenciamento formado por *AGO4* também atuam na amplificação do sinal primário de siRNAs e na produção de siRNAs secundários que irão induzir metilação em regiões adjacentes de maneira a promover a perpetuação da heterocromatina (Figura 6A). Outras proteínas como *SUVH2* e *SUVH9* (*SUPPRESSOR OF VARIATION 3-9 HOMOLOGUE 2* or *9*), que possuem domínios SRA que interagem com DNA metilado, assim como um possível fator de alongação da transcrição *SPT5L* (*SUPPRESSOR OF TY INSERTION 5-LIKE*) e uma proteína remodeladora de cromatina *DRD1* (*DEFECTIVE IN RNA DIRECTED DNA METHYLATION 1*) também parecem atuar nesta via, no entanto, o exato papel destas proteínas ainda não é completamente entendido (Law & Jacobsen 2010; He et al. 2011). Alguns trabalhos sugerem que *AGO6* e *AGO9* sejam parcialmente redundantes com *AGO4* na via de RdDM, embora a atuação destas proteínas seja muito menos predominante que a de *AGO4* (He et al. 2011). Além disto, transcritos dupla fita produzidos por *Pol II* a partir de repetições invertidas ou transcrição bidirecional também podem desencadear a

biogênese de siRNAs e o processo de RdDM (Reinders & Paszkowski 2009; Meyer 2011). Transcritos não codificantes simples fita produzidos por *Pol II* em alguns *loci* do genoma também parecem agir no recrutamento de efetores da via de RdDM para estes *loci*, incluindo *AGO4* e possivelmente *Pol IV* e *Pol V* (Meyer 2011; He et al. 2011). Além de *DRM2*, *A. thaliana* também codifica outra *de novo* DNA metiltransferase identificada como *DRM1*. No entanto, *DRM1* é expressa em níveis muito mais baixos que *DRM2*, sugerindo que *DRM2* seja a *de novo* metiltransferase predominante. Em contraste ao estabelecimento de metilação em elementos de transposição e sequências repetidas, metilação CG presente no corpo de genes codificantes de proteínas é independente de siRNAs e da via de RdDM e provavelmente é resultante da atividade de *Pol II* (Cokus et al. 2008; Lister et al. 2008).

Uma vez estabelecidos, os padrões de metilação podem ser mantidos ao longo dos ciclos celulares através de processos que dependem parcialmente da simetria dos contextos CG e CHG. Manutenção de metilação no contexto CG é dependente da DNA metiltransferase de manutenção *MET1* (*METHYLTRANSFERASE 1*) e uma via que envolve diversas outras proteínas como a enzima remodeladora de cromatina do tipo SWI/SNF2-like *DDM1* (*DECREASED IN DNA METHYLATION 1*) e o fator de domínio SRA *VIM 1* (*VARIATION IN METHYLATION 1*) que parece recrutar *MET1* aos sítios CG hemimetilados (Law & Jacobsen 2010; Figura 6B). Manutenção no contexto CHG é principalmente dependente de *CMT3* (*CHROMOMETHYLASE 3*), uma metiltransferase específica de plantas cuja atividade é intimamente associada a *KRYP/SUVH4* (*KRYPTONITE/ SUPPRESSOR OF VARIATION HOMOLOGUE 4*), a principal metiltransferase de histona que promove dimetilação na posição 9 da lisina da cauda da histona H3 (H3K9me2). *CMT3* provavelmente reconhece H3K9me2 através do seu cromodomínio, enquanto *KRYP/SUVH4* possivelmente se liga a sítios CHG metilados através de seu domínio SRA, estabelecendo um *loop* de *feedback* positivo para manutenção da metilação neste contexto (Figura 6B). *SUVH5* e *SUVH6* também parecem contribuir para manutenção dos padrões de metilação no contexto CHG (Law & Jacobsen 2010). No contexto assimétrico CHH, uma vez que a maquinaria não pode contar com a simetria de sequência para sua manutenção, a metilação é perpetuada por constante metilação *de novo* promovida por *DRM2* e a via de RdDM que garantem reestabelecimento a cada ciclo de replicação (Figura 6B). No entanto, em alguns *loci* metilação CHG e CHH são controladas redundantemente por *CMT3* e *DRM2* via RdDM. Em

outros, *MET1* parece ser requerida para a atuação de *CMT3* e *DRM2* ou possa até mesmo promover a manutenção da metilação nos contextos CHG e CHH (Henderson & Jacobsen 2007; Lister et al. 2008).

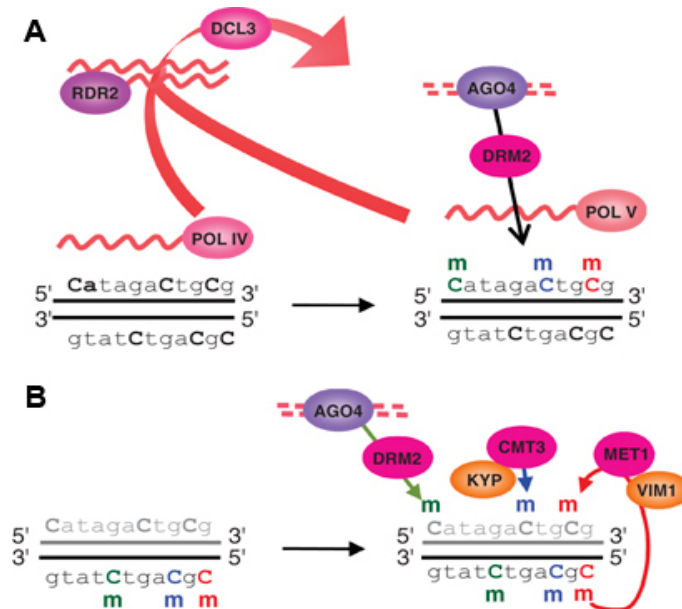


Figura 6. Mecanismos de estabelecimento e manutenção dos padrões de metilação nos contextos CG, CHG e CHH em sequências repetidas. (A) Modelo proposto para a via de RdDM. Transcritos primários produzidos por *Pol IV* a partir de DNA repetitivo são convertidos em RNA dupla fita por *RDR2*, uma RNA polimerase dependente de RNA. Este RNA dupla fita é subsequentemente clivado por *DCL3* em siRNAs de 24nt, que são então carregados em um complexo de silenciamento contendo *AGO4*, que irá recrutar *DRM2* a sequência alvo a ser metilada nos três contextos de sequência através de interação com um transcrito produzido por *Pol V*. Transcritos correspondentes a região alvo produzidos tanto por *Pol IV* quanto *Pol V* podem ser usados para amplificar a produção de siRNAs através de um mecanismo de *feedback* positivo e produzir siRNAs secundários que causam *spreading* da metilação em regiões heterocromáticas. Citosinas em preto e colorido representam sítios não-metilados e metilados, respectivamente (CG, vermelho; CHG, azul; CHH, verde). (B) Modelo proposto para manutenção de metilação de DNA. Após replicação do DNA, manutenção de metilação no contexto CG é feita por *MET1* em uma via que depende do reconhecimento de sítios CG hemimetilados por *VIM1*. No contexto CHG, a manutenção é dependente de *CMT3* que atua em um *loop* de *feedback* positivo com *KRYP/SUVH4*. Metilação no contexto CHH é mantida por atuação constante de RdDM e/ou *CMT3* e *MET1*. Ilustração retirada de Teixeira & Colot 2010.

Em geral, as vias de *RdDM* e manutenção são em grande extensão geneticamente independentes. Mutações em *MET1* eliminam essencialmente metilação CG do genoma, tanto em regiões repetidas quanto no corpo de genes, mantendo quantidades significativas de metilação CHG e CHH (Tabela 1, Figura 7) e siRNAs de 24nt que dirigem metilação *de novo* (Cokus et al. 2008; Lister et al. 2008). Em contrapartida, metilação CG quase não é alterada em mutantes para enzimas da via de RdDM e/ou CMT3 (Tabela 1, Figura 7). No entanto, como já mencionado acima, as diferentes DNA metiltransferases também podem agir redundantemente (Tabela 1 e Figura 7). Em *met1* ou no triplo mutante *met1 drm1 drm2*, por exemplo, hipermetilação de sítios CHG é observada principalmente no corpo de genes em um padrão enviesado para a porção 3' da região transcrita semelhante ao da metilação CG ausente (Cokus et al. 2008; Lister et al. 2008). Este comportamento compensatório, de certa forma, ajuda a explicar a viabilidade de diferentes combinações de mutações como, por exemplo, *met1 cmt3*, *met1 drm1 drm2* ou ainda *drm1 drm2 cmt3*, enquanto o mutante quadruplo *met1 drm1 drm2 cmt3* causa letalidade do embrião (Cokus et al. 2008).

Tabela 1. Porcentagem de metilação nos três contextos de sequência em plantas selvagem (*wild type*) e mutantes simples, duplos e triplos para as DNA metiltransferases de metilação *de novo* (*DRM1* e *DRM2*) e de manutenção (*MET1* e *CMT3*).

% Metilação	CG	CHG	CHH
<i>wild type</i>	24,60%	6,98%	1,70%
<i>met1</i>	0,42%	4,33%	1,11%
<i>drm1 drm2</i>	23,20%	6,60%	1,61%
<i>cmt3</i>	20,63%	1,26%	1,50%
<i>met1 cmt3</i>	0,44%	0,62%	0,74%
<i>met1 drm1 drm2</i>	0,46%	6,82%	1,25%
<i>drm1 drm2 cmt3</i>	19,67%	0,79%	0,95%

Cokus et al. 2008

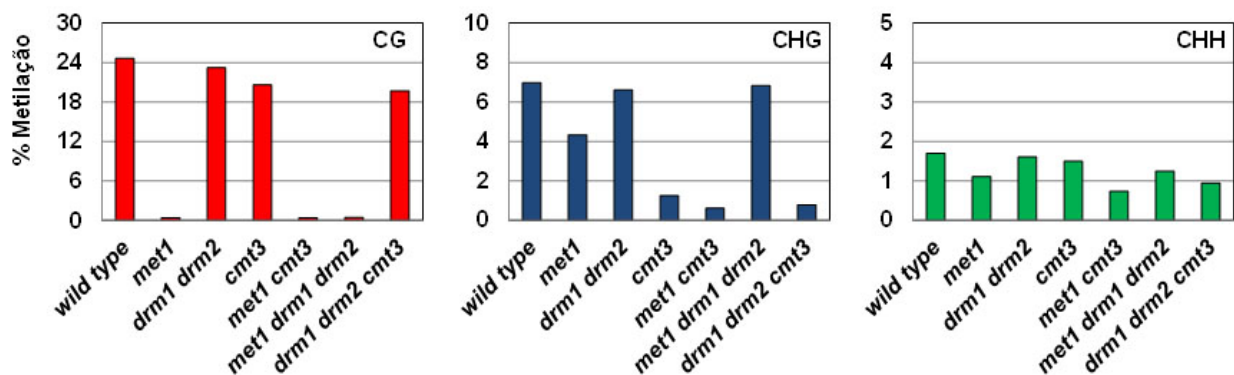


Figura 7. Distribuição de metilação de DNA nos três contextos de sequência em plantas selvagem (*wild type*) e mutantes simples, duplos e triplos para DNA metiltransferases de *A.thaliana* (Cokus et al. 2008).

1.2.3 - Herança transgeracional dos padrões de metilação e *epigenetic Recombinant Inbred Lines* (epiRILs)

DDM1 controla o silenciamento de diversos elementos de transposição, muitos dos quais se tornam ativos em *ddm1*. Plantas, diferentemente de mamíferos que apagam e reestabelecem *de novo* os padrões de metilação de DNA em células germinativas, mantêm seu padrão de metilação através da meiose. Desta maneira, hipometilação induzida em *ddm1* pode ser herdada estavelmente ao longo das gerações mesmo quando o alelo selvagem é recuperado através de retrocruzamento com plantas selvagens. No entanto, esta herança estável é limitada a um conjunto de sequências (alvos “não-remetiláveis”) cujos padrões de metilação nos três contextos de sequência são exclusivamente controlados pela maquinaria de manutenção. Um sistema robusto de remetilação que permite reestabelecer os padrões selvagens de metilação após perdas massivas no contexto *ddm1* pode atuar sobre sequências (alvos “remetiláveis”) que, além da maquinaria de manutenção, também estão sob o controle adicional de siRNAs e a via de RdDM. Para estas sequências “remetiláveis” siRNAs e RdDM pouco contribuem para seu padrão global de metilação, no entanto, são críticos para sua correta remetilação através da manutenção de metilação no contexto CHH, principal fator que distingue sequências “remetiláveis” de “não-remetiláveis”. Desta maneira, a maquinaria de RdDM também atua como um sistema de proteção

contra perdas acidentais de metilação, protegendo o genoma contra os efeitos deletérios da atividade de elementos de transposição. O processo de remetilação não acontece em tecidos vegetativos, sendo dependente da reprodução sexual. Além disto, é um processo gradual que normalmente leva de 2 a 5 gerações para reestabelecer os padrões de metilação selvagem, dependendo do locus do genoma. Acredita-se que um controle baseado exclusivamente nos mecanismos de manutenção possa representar uma transição de um estado inicial controlado por siRNAs e RdDM e que provavelmente afete relíquias ou transposons mais velhos que de alguma maneira perderam a capacidade de mobilização, enquanto RdDM possa ser mais efetivo sobre elementos de transposição ativos recém inseridos no genoma, e, portanto, com alto poder mutagênico (Teixeira et al. 2009; Teixeira & Colot 2010).

Uma vez que plantas *ddm1* acumulam problemas severos de desenvolvimento ao longo das gerações de autofecundação, o que inclui baixa fertilidade (Richards 1997), a análise da herança transgeneracional das alterações de metilação induzidas neste mutante só foi possível através da construção de um conjunto de *epigenetic isogenic Recombinant Inbred Lines* (epiRILs; Johannes et al. 2009), idênticas geneticamente, mas constituídas de epigenomas mosaico onde 75% das regiões cromossômicas são herdadas de um parental selvagem (wt) enquanto os outros 25% são herdados do parental mutante *ddm1* (Figura 8). Desta maneira, locus “remetiláveis” apresentam metilação de DNA em 100% das linhagens enquanto locus “não-remetiláveis” apresentam metilação somente quando a região cromossômica correspondente é herdada do parental selvagem, o que corresponde a 75% do total de linhagens analisadas (Johannes et al. 2009). Remetilação corretiva associada a siRNAs e provavelmente a via de RdDM também foi observada em linhagens epiRIL derivadas do mutante *met1* (Reinders et al. 2009), embora a capacidade de remetilizar e reestabelecer os padrões de metilação selvagem em diferentes *loci* do genoma possa variar entre as duas populações (Reinders & Paszkowski 2009).

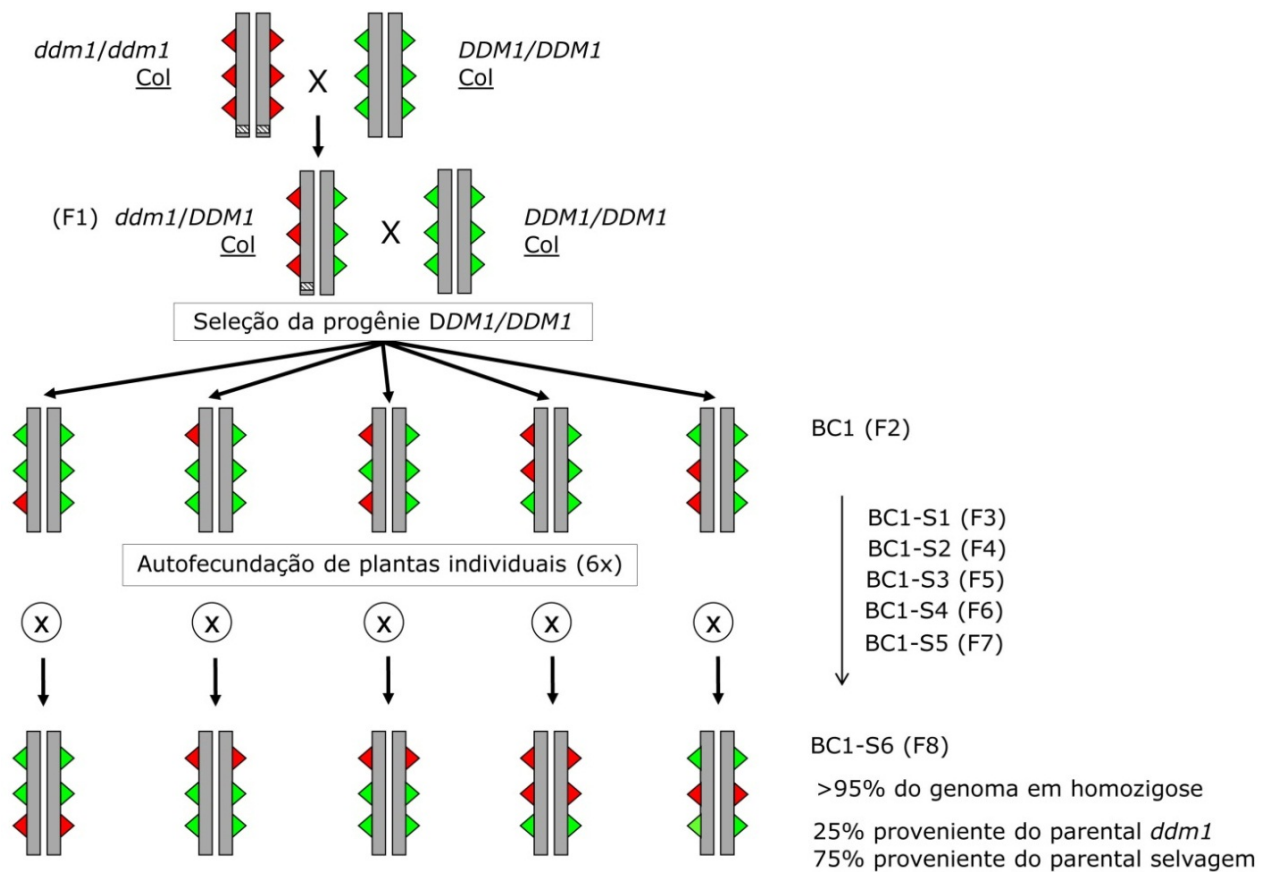


Figura 8. Construção de linhagens epiRIL Col-0 wt derivadas de *ddm1*. Barras cinza representam o genoma de *A. thaliana* e triângulos vermelhos e verdes, o padrão de metilação de DNA em plantas *ddm1* e wt, respectivamente. Para construção desta população, parentais isogênicos Col-0 wt e Col-0 *ddm1* (exceto para o locus *DDM1* localizado no cromossomo 5 e possivelmente por algum evento de transposição que possa ter sido induzido em *ddm1*) foram cruzados para criar um único indivíduo F1 heterozigoto *DDM1/ddm1* (50% do genoma proveniente do parental *ddm1*) que foi então retrocruzado (BC1) com o parental Col-0 wt (*DDM1/DDM1*). Após genotipagem da população F2, diferentes indivíduos com genótipo wt *DDM1/DDM1* (25% do genoma proveniente do parental *ddm1*) foram selecionados e submetidos a seis gerações de autofecundação (S) de maneira que na geração F8 mais de 95% do genoma fosse encontrado em homozigose. Teoricamente diferentes linhagens epiRILs F8 possuem o mesmo genótipo, mas epigenótipos específicos.

1.2.4 - Metilação de DNA e o controle da expressão gênica

Embora a metilação de DNA atue principalmente na supressão da atividade de elementos de transposição e outras sequências repetidas, cerca de 5% dos genes de *A. thaliana* apresentam metilação de DNA em suas sequências promotoras, indicando que esta marca epigenética também pode ser usada para silenciar genes (Zhang et al. 2006). Desta maneira, variação epigenética estável em resposta a fatores ambientais ou estocásticos como erros na maquinaria de controle epigenético podem alterar os padrões espaciais e temporais de expressão gênica independentemente de variação na sequência de DNA, produzindo alelos epigenéticos ou epialelos com consequências fenotípicas imediatas. Em geral, a formação de epialelos é influenciada pela presença de sequências repetidas ou elementos de transposição localizados próximos a genes, capazes de produzir siRNAs e desta maneira recrutar a maquinaria de silenciamento. De acordo com esta observação, a maioria dos epialelos descritos em plantas representam casos de hipermetilação de citosinas associadas a silenciamento transcricional, embora recentemente variação espontânea para perda e ganho de metilação independentemente deste contexto estrutural tenha sido descrita em *A. thaliana* (Becker et al. 2011; Schmitz et al. 2011).

Em *Arabidopsis*, epialelos afetando características morfológicas e do desenvolvimento foram identificados em mutantes para proteínas da maquinaria de manutenção e metilação *de novo* ou como subprodutos de *screens* mutagênicos. O gene *FWA* (*FLOWERING WAGENING*) codifica um fator de transcrição homeobox cuja expressão ectópica resulta em um fenótipo dominante de atraso de florescimento (Soppe et al. 2000). Apesar de expresso no gametófito feminino e endosperma da semente onde sofre *imprinting*, durante o desenvolvimento vegetativo a expressão de *FWA* é silenciada pela presença de metilação de DNA em uma região específica de sua sequência promotora onde são encontradas duas repetições em *tandem* derivadas de um retroelemento do tipo *SINE* (Figura 9A; Kinoshita et al. 2006). Silenciamento de *FWA* é consistente entre acessos de *A. thaliana* (Vaughn et al. 2007), no entanto, este gene torna-se transcionalmente ativo devido a perda de metilação CG induzida nos mutantes *met1* e *ddm1* (Figura 9A). Em contrapartida, perda de metilação não-CG neste locus não está associada a alterações de expressão, embora seja necessária a remetilação corretiva (Lister et al. 2008;

Teixeira et al. 2009). O gene *SDC* (*SUPPRESSOR OF drm1 drm2 cmt3*) codifica uma proteína do tipo F-box cuja expressão é responsável pelas alterações de desenvolvimento observadas no triplo mutante *ddc* (*drm1 drm2 cmt3*) como folhas curvadas e baixa estatura. Expressão de *SDC* também é silenciada em função de metilação de DNA em sua sequência promotora dirigida a um conjunto de sete repetições diretas em *tandem* (Henderson & Jacobsen 2008). No entanto, diferentemente de *FWA*, silenciamento de *SDC* é dependente de *DRM2* e *CMT3* (Lister et al., 2008) e é reestabelecido já na primeira geração de retrocruzamento com plantas selvagens (Henderson & Jacobsen 2008). Um terceiro exemplo de locus epimutável em *A. thaliana* inclui o fator de transcrição *SUP* (*SUPERMAN*), que normalmente é demetilado e expresso ao longo do desenvolvimento. No entanto, epialelos hipermetilados e silenciados de *SUP* identificados como *clk* (*clark kent*) e que induzem alterações de morfologia floral, foram identificados em *screens* mutagênicos e contra-intuitivamente em *met1* (Jacobsen & Meyerowitz 1997; Kakutani 2002). Variantes epialélicas resultantes de eventos de duplicação gênica também foram descritos para os genes *PAI* (*PHOSPHORIBOSYLANTHRANILATE ISOMERASE*; Bender & Fink 1995) e *AtFOLT1* (*A. thaliana FOLATE TRANSPORTER 1*; Durand et al. 2012) em alguns acessos de *A. thaliana*. No acesso Col-0, três genes *PAI* (*PAI1-3*) são encontrados em diferentes locos do genoma. Em contrapartida, em *Ws* uma duplicação invertida do gene *PAI1* (*PAI1-PAI4*) resulta na produção de um transcrito dupla fita que gera siRNAs capazes de induzir metilação em *cis* e em *trans* na região de homologia entre os genes desta família (Figura 9B; Bender & Fink 1995; Melquist & Bender 2003). De maneira similar, silenciamento epigenético do transportador de folato *AtFOLT1* em alguns acessos de *A. thaliana* em razão de duplicação deste gene induzida por um evento de transposição, determina incompatibilidade genética com acessos em que o gene está presente em cópia única devido a falta de folato e fertilidade reduzida (Durand et al. 2012).

Fora de *A. thaliana*, epialelos foram descritos em linária, tomate, melão e arroz. Em linária, surgimento do variante pelórico, caracterizado por flores que apresentam simetria radial ao invés de bilateral, é atribuído a hipermetilação da sequência promotora do gene *CYC* (*CYCLOIDEA*; Figura 2C; Cubas et al. 1999). Em tomate, hipermetilação e silenciamento do gene *Cnr* (*Colourless non-ripening*), possivelmente influenciada por um transposon localizado próximo ao locus (Paszkowski & Grossniklaus 2011), leva a alterações de coloração e amadurecimento do fruto (Manning et al. 2006). Da mesma maneira, inserção de um transposon

na sequência promotora do fator de transcrição *CmWIP1* (*Cucumis melo WIP1*) de melão resulta em silenciamento transcricional deste gene e consequente alteração do desenvolvimento dos órgãos sexuais florais, com a formação de plantas que ao invés de produzirem flores femininas e masculinas produzem somente flores femininas. (Martin et al. 2009). Em arroz, metilação de DNA dirigida a repetições em *tandem* localizadas na sequência promotora do gene *DWARF1* leva a silenciamento deste gene e produção de plantas com estatura reduzida (Miura et al. 2009).

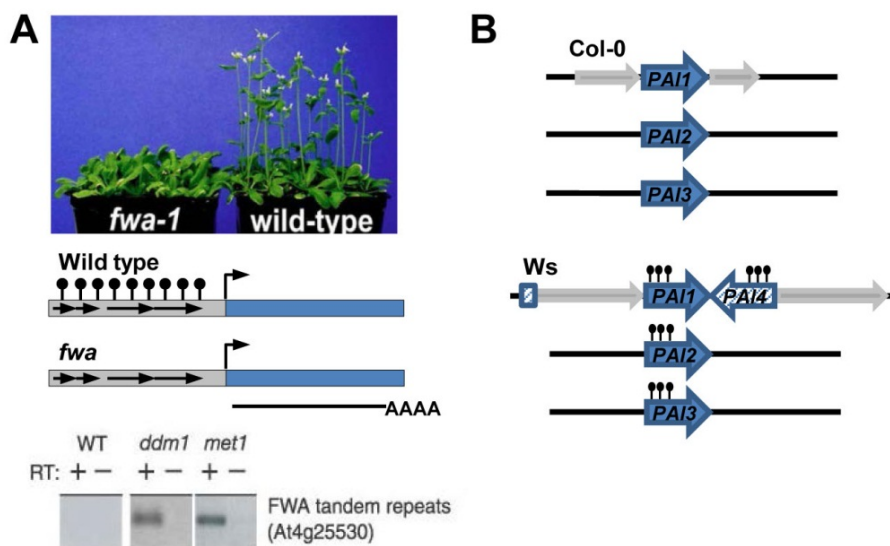


Figura 9. Exemplos de variantes epialélicas em *A. thaliana*. (A) O fator de transcrição *FWA* é silenciado ao longo do desenvolvimento vegetativo em função de metilação de DNA em sua sequência promotora dirigida a repetições diretas em *tandem* derivadas de um retrotransposon do tipo SINE. Em *met1* e *ddm1* perda de metilação CG nesta região leva a reativação transcricional de *FWA* com consequente atraso de florescimento. (B) Família de genes *PAI*. Em *Ws* uma alteração estrutural resultante de um evento de duplicação invertida de *PAII* produzindo um quarto alelo *PAI4* leva a formação de um transcrito dupla fita e siRNAs que irão induzir metilação em *cis* e em *trans* na região de homologia entre estes genes.

Apesar do número de epialelos descritos até o momento ser ainda pequeno, acredita-se que a metilação de DNA possa representar um mecanismo bastante utilizado para controle da expressão gênica em plantas. Embora genes estejam raramente associados a sequências repetidas em *Arabidopsis*, em genomas maiores e mais ricos em repetições esta característica parece ser a

regra ao invés da exceção. Em arroz, por exemplo, uma porcentagem bem maior de genes está intimamente associada a sequências repetidas (Figura 10; Zhang et al. 2008). De acordo com esta informação, embora mutação em *RDR2* tenha poucos efeitos sobre a expressão de genes (Kurihara et al. 2008) e desenvolvimento de *A. thaliana*, mutação na proteína homóloga de milho *MOPI* (*MEDIATOR OF PARAMUTATION 1*) leva a alterações substanciais de expressão gênica (Jia et al. 2009) e a diversas anormalidades desenvolvimentais (Dorweiler et al. 2000).

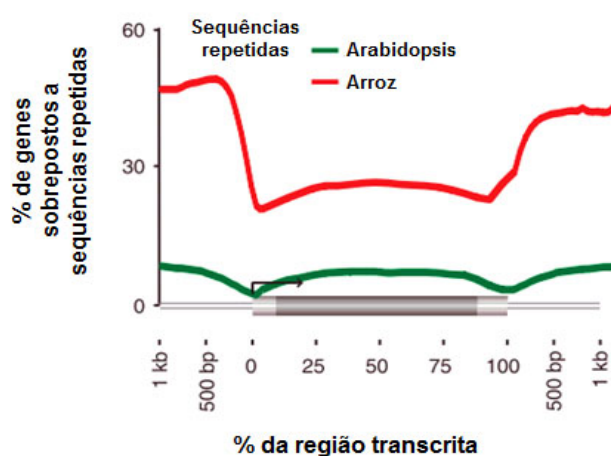


Figura 10. Distribuição de sequências repetidas em relação a genes em *Arabidopsis* e arroz. Figura retirada de Zhang et al. 2008.

1.2.5 - Importância ecológica e evolutiva da metilação de DNA

A teoria evolutiva é baseada na suposição de que a única fonte de variação fenotípica herdável em populações naturais sobre a qual a seleção natural pode atuar é genética. No entanto, como demonstrado acima, variantes epigenéticas também podem se comportar como mutações clássicas e gerar variação fenotípica herdável em caracteres importantes ecologicamente, mesmo na ausência de alterações na sequência de DNA. Ao contrário de polimorfismos genéticos, modificações epigenéticas são reversíveis e além de ocorrerem estocasticamente (Becker et al. 2011; Schmitz et al. 2011) podem também ser moduladas por estímulos ambientais (Dolinoy

2008; Ito et al. 2011; Paszkowski & Grossniklaus 2011). Estas observações aqueceram debates sobre o real papel de variação epigenética no processo de evolução adaptativa e a necessidade de se considerar epialélos em genética de populações e teoria evolutiva. No entanto, o número de exemplos de variantes epialélicas ainda é pequeno e os epialélos naturais identificados até o momento, pelo menos em plantas, estão limitados aqueles associados a alterações fenotípicas marcantes. Consequentemente, o impacto de mecanismos epigenéticos em populações naturais é ainda pouco compreendido (Kalisz & Puruggana 2004; Bossdorf et al. 2008; Richards 2008; Schmitz & Ecker 2012). Permanecem para serem respondidas questões como qual a frequência e estabilidade de variação epigenética em populações naturais e como esta variação está distribuída entre e intra populações; até que ponto variação epigenética pode ser influenciada por variação genética em *cis* ou *trans*; qual a importância relativa de variação epigenética pura em relação a variação genética na determinação de caracteres importantes ecologicamente e finalmente, com que frequência fatores ambientais podem induzir alterações estáveis em marcas epigenéticas.

Recentemente, dois estudos complementares quantificaram a taxa de variação espontânea no padrão de metilação global do genoma de *A. thaliana* em *mutation accumulation lines* (MA) propagadas durante 30 gerações a partir de um ancestral comum em condições de crescimento controladas e teoricamente sem seleção (Becker et al. 2011; Schmitz et al. 2011). Ambos os estudos mostraram que, em geral, o padrão de metilação de citosinas é herdado de forma estável, principalmente em regiões ricas em elementos de transposição e sequências repetidas. No entanto, variações epigenéticas acontecem e acumulam gradualmente ao longo das gerações de maneira similar a mutações genéticas. Os estudos mostraram que cerca de 1,6% (Schmitz et al. 2011) a 6,2% (Becker et al. 2011) das citosinas no contexto CG do genoma são susceptíveis a variações dinâmicas no padrão de metilação, o que determina taxas de epimutação ordens de magnitude mais altas do que mutação na sequência de DNA ($4,4 \times 10^{-4}$ polimorfismos epigenéticos por sítios CG por geração contra 7×10^{-9} substituições de base por sítio por geração; Schmitz et al. 2011). Variação na taxa de metilação de citosinas em alguns *loci* foi constante em diferentes linhagens analisadas, indicando que alguns sítios podem se comportar como *hotspots* de variação epigenética, embora as taxas de reversão sejam altas. Além disto, ambos os estudos encontraram evidências para alterações de metilação levando a mudanças no padrão de expressão gênica. Estas observações indicam que instabilidades epigenéticas podem fazer parte de um

sistema plástico de resposta a condições ambientais, caso a seleção possa estabilizar estes estados epigenéticos de maneira que eles se tornem uma fonte de epialelos herdáveis e potencialmente adaptativos. De acordo com estas observações, simulações mostraram que variação fenotípica e plasticidade geradas por instabilidade epigenética podem ser benéficas em ambientes variáveis e ser alvo de seleção (Feinberg & Irizarry 2010). Nesta mesma linha de evidências, também foi demonstrado que populações de plantas vivendo em ambientes contrastantes desenvolveram alta variabilidade epigenética, sugerindo que algumas espécies possam usar variação epigenética para se adaptar a diversos ambientes (Lira-Medeiros et al. 2010). Da mesma maneira, fenótipos observados em epiRILs derivadas de *met1* e *ddm1* se assemelham a variação encontrada entre acessos naturais de *Arabidopsis* (Reinders et al. 2009; Roux et al. 2011).

Metilação de citosinas pode também gerar variação genética, uma vez que citosinas metiladas tem uma maior taxa de deaminação espontânea para timina. Desta maneira, mutantes de sequência com novos efeitos fenotípicos poderiam surgir e eventualmente serem revelados se demetilados, funcionando de uma maneira semelhante a *HSP90* em *Drosophila* que mascara variação genética críptica (Queitsch et al. 2002). Variação genética associada a mudanças epigenéticas também pode ocorrer devido à mobilização de elementos de transposição, que podem se inserir dentro ou próximo a genes, causar quebras cromossômicas, recombinação ilegítima e rearranjos genômicos (Slotkin & Martienssen 2007).

Em conclusão, plantas e outros organismos provavelmente procuram um balanço entre manter os padrões epigenéticos suficientemente estáveis, de forma a evitar os efeitos deletérios na expressão gênica e estrutura do genoma decorrentes da atividade de elementos de transposição, mas, por outro lado, mantê-los suficientemente flexíveis para induzir variação epigenética e/ou genética de maneira a permitir adaptação rápida a novas condições ambientais.

1.3 - Genes novos – conceito, mecanismos de origem e importância evolutiva

Genes novos são definidos como genes de origem recente em uma escala de tempo evolutiva e que existem em apenas uma ou algumas espécies relacionadas. Os mecanismos de geração de genes novos geralmente dependem de duplicação de um gene ancestral ou rearranjos gênicos seguidos de mutação e seleção natural. Desta maneira, os genes originados a partir deste mecanismo mantêm certa similaridade de sequência com genes na mesma espécie ou em diferentes espécies (Zhang 2003). Entretanto, análises recentes mostram que todo genoma possui também certa fração de genes novos órfãos, que não possuem similaridade de sequência com qualquer proteína ou peptídeo presente em bancos de dados e que não podem ser associados a nenhum gene conhecido (Gollery et al. 2006; Cai et al. 2008; Heinen et al. 2009; Toll-Riera et al. 2009; Xiao et al. 2009). Acredita-se que genes órfãos possam surgir por rearranjos e/ou duplicações seguidos de divergência rápida, domesticação de transposons ou por evolução *de novo* de regiões não codificantes, ou seja, quando sequências não gênicas são transcritas, adquirem ORFs e começam a ser traduzidas (Donoghue et al. 2011; Tautz & Domazet-Lošo 2011). Em geral, genes novos são caracterizados por peptídeos pequenos e de função desconhecida, evolução rápida, nível de expressão baixo e normalmente tecido específico, além de altamente responsivos a condições de estresse (Donoghue et al. 2011). Apesar do papel que estes genes desempenham nos organismos ser ainda pouco entendido, estudos recentes em diferentes organismos sugerem que genes novos possam desempenhar papéis importantes em diversos processos biológicos e que, em certos casos, estejam relacionados com as novidades evolutivas de cada espécie ou grupo sendo, portanto, críticos para diversificação e adaptação dos organismos durante a evolução (Cai et al. 2008; Heinen et al. 2009; Khalturin et al. 2009; Xiao et al. 2009; Chen et al. 2010; Carvunis et al. 2012).

1.4 – *Qua-Quine Starch* de *A. thaliana*, um gene novo regulado por metilação de DNA e potencialmente relacionado ao metabolismo de amido

Qua-Quine Starch (QQS) é um gene novo que provavelmente originou-se *de novo* em uma região heterocromática do genoma de *A. thaliana*. De acordo com sua origem recente, *QQS* codifica uma proteína de apenas 59 aminoácidos e sua expressão é responsiva a fatores do desenvolvimento, perturbações genéticas e do ambiente (www.genevestigator.com; Donoghue et al. 2011). Algumas linhas de evidência ainda sugerem que este gene possa ter um efeito sobre o metabolismo do amido. Primeiramente, a expressão de *QQS* é bastante induzida nos mutantes *AtssIII* (*A. thaliana SUCROSE SYNTHASE III*) e *mex1* (*MALTOSE EXCESS 1*) que apresentam acúmulo de amido (Li et al. 2009). Além disto, linhagens transgênicas silenciando *QQS* através de RNAi apresentam cerca de 20-30% mais amido em folhas do que plantas selvagens no final do período de luz, possivelmente devido ao aumento da síntese deste polímero (Li et al. 2009). Por outro lado, plantas superexpressando *QQS* apresentam crescimento mais lento, folhas cloróticas e uma pequena redução do conteúdo total de amido (Seo et al. 2011). *QQS* também parece ser regulado positivamente pelo fator de transcrição *IDD14* (*INDETERMINATE DOMAIN 14*). Ligação de homodímeros de *IDD14* a sequência promotora de *QQS* é significativamente reduzida por frio em função da transcrição de uma versão truncada de *IDD14* que não possui o motivo de ligação ao DNA, mas ainda é capaz de heterodimizar com a forma funcional da proteína. Consequentemente, sob condições de baixa temperatura a expressão de *QQS* é reduzida e amido acumula. É proposto que a regulação de *QQS* por um variante truncado de *IDD14* seja importante na coordenação do fornecimento de carbono e crescimento na estratégia de adaptação ao frio, diminuindo a taxa de quebra de amido durante a noite para compensar a falta de carbono decorrente da redução da taxa de fotossíntese durante o dia (Seo et al. 2011).

Adicionalmente, a expressão de *QQS* também parece ser regulada por metilação de DNA (Kurihara et al. 2008; Lister et al. 2008). Dado o pequeno número de exemplos de genes controlados por mecanismos epigenéticos em *A. thaliana*, decidimos iniciar um estudo visando a caracterização do controle epigenético que atua sobre *QQS*.

2. OBJETIVOS

2.1 Geral

Caracterização do controle epigenético que atua sobre o gene *QQS* de *A. thaliana*.

2.2 Específicos

- Caracterização dos mecanismos de controle envolvidos no silenciamento transcricional de *QQS*.
- Análise de correlação entre padrão de metilação e expressão de *QQS*.
- Avaliação do nível de influência de alterações genéticas em *cis* ou *trans* em epialelos de *QQS*.
- Análise da herança transgeneracional de epialelos de *QQS*.
- Avaliação do estado epigenético de *QQS* em acessos naturais de *A. thaliana*.
- Avaliação fenotípica de epialelos de *QQS*.

3. RESULTADOS E DISCUSSÃO

PARTE I

3.1 - Identificação de variação epialélica para o gene *QQS* de *A. thaliana*

Visando a caracterização dos genes alvos do fator de transcrição do tipo bZIP *AtbZIP9* (*At5g24800*) de *A. thaliana*, comparamos os transcriptomas de plantas selvagens e plantas *knockout* para *AtbZIP9* (*atbZIP9-1*) através de microarranjos de DNA. *QQS* (*At3g30720*) destacou-se como um dos genes mais reprimidos nesta comparação, o que foi confirmado por análises quantitativas de expressão através de PCR em tempo real. Esta observação, associada a outros fatores como um padrão de expressão preponderante no sistema vascular similar ao de *AtbZIP9* (Figura 11) e a presença de motivos de ligação em sua sequência promotora característicos de fatores de transcrição do tipo bZIP, nos motivaram a investigar mais detalhadamente uma possível interação entre *AtbZIP9* e *QQS*.

AtbZIP9 é um gene expresso especificamente nos feixes de floema e uma versão ativadora constitutiva deste gene é capaz de promover alterações de desenvolvimento e/ou integridade da parede celular deste tecido (Silveira et al. 2007). Uma vez que *QQS* também está fortemente expresso nestas células (Zhao et al. 2005), decidimos avaliar a expressão destes dois genes em plantas mutantes para o gene *APL* (*ALTERED PHLOEM DEVELOPMENT*), necessário à diferenciação das células do floema. Os resultados mostraram que tanto a expressão de *AtbZIP9* quanto a de *QQS* não são alteradas em função da mutação, indicando que estes genes são independentes de *APL*. Entretanto, surpreendentemente, observamos que a expressão de *QQS* estava cerca de 27 vezes mais alta em plantas Columbia-0 (Col-0) tipo selvagem de nosso laboratório (identificado a partir daqui como Col-0*) com relação a expressão deste gene em plantas Col-0 fenótipo selvagem (*APL/APL* ou *APL/apl*) ou fenótipo mutante *apl* (*apl/apl*) provenientes de autofecundação de um indivíduo heterozigoto *APL/apl* (Figura 12A). Em contrapartida, a expressão de *AtbZIP9* era semelhante em todos os genótipos analisados (Figura 12A). Estes resultados nos permitiram inferir que *QQS* possivelmente estivesse induzido no

genótipo Col-0* utilizado na comparação contra *atbzip9-1*, o que foi posteriormente confirmado utilizando-se sementes Col-0 selvagem provenientes de outro laboratório (Figura 12B). Desta maneira, a repressão de *QQS* observada em *atbzip9-1*, na verdade, representava uma indução deste gene no controle selvagem Col-0*.

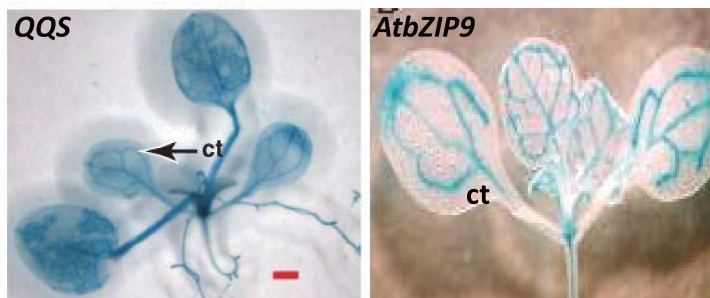


Figura 11. Padrão de expressão dos genes *QQS* (retirado de Li et al. 2009) e *AtbZIP9* em linhagens expressando fusões entre a sequência promotora destes genes e o gene repórter *GUS*. Ct, cotilédone.

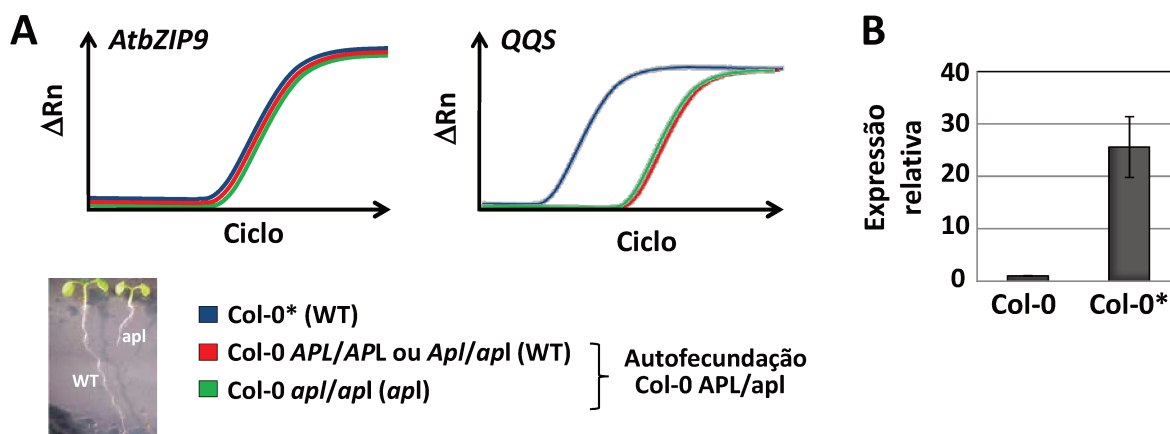


Figura 12. Análise de correlação entre *AtbZIP9*, *QQS* e a via de sinalização media da por *APL*. (A) Padrão de expressão de *AtbZIP9* e *QQS* em plantas Col-0* e plantas com fenótipo selvagem (WT) ou *apl* provenientes de autofecundação de uma indivíduo heterozigoto para a mutação *apl*. (B) Padrão de expressão de *QQS* em plantas Col-0 tipo selvagem e Col-0*.

De maneira a avaliar se alterações genéticas poderiam estar influenciando o padrão de expressão de *QQS* em Col-0*, sequenciamos a região correspondente ao locus *QQS* (promotor e região transcrita) de Col-0*. A sequência obtida foi idêntica à sequência referência de Col-0, indicando que polimorfismos genéticos no locus não eram a causa da indução de expressão. Além disto, uma mutação em *trans* também era improvável uma vez que plantas F1 provenientes de um cruzamento entre Col-0* e Col-0 apresentaram um nível de expressão intermediário entre o de Col-0* e Col-0 (Figura 13). No caso de uma mutação recessiva, esperaríamos níveis de expressão semelhantes ao de Col-0 e no caso de uma mutação dominante, níveis de expressão semelhantes ao de Col-0*. Estas observações, associadas ao fato de *QQS* ser um gene possivelmente regulado por metilação de DNA (Lister et al. 2008; Kurihara et al. 2008), indicaram que a alteração de expressão de *QQS* em Col-0* pudesse ser causada por alterações epigenéticas. De fato, quantificação do nível de metilação de *QQS* em Col-0* mostrou uma diminuição significativa de seu nível de metilação em relação a plantas Col-0, o que será discutido em detalhes na próxima seção.

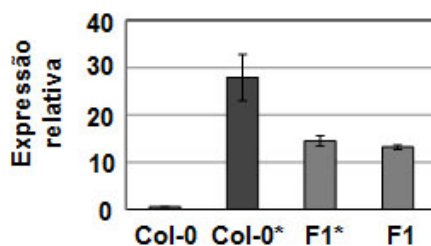


Figura 13. Análise do padrão de expressão de *QQS* em plantas F1 provenientes de um cruzamento entre Col-0* e Col-0. F1*, Col-0* parental feminino e Col-0 parental feminino; F1, Col-0 parental feminino e Col-0* parental masculino.

A partir desta descoberta iniciamos um estudo detalhado visando a caracterização do controle epigenético que atua sobre o gene *QQS*. Mais especificamente, buscamos entender alguns aspectos mecanísticos deste controle e avaliar se epialelos de *QQS* também poderiam existir na natureza e se sim, qual seria sua frequência, estabilidade e impacto adaptativo. Estes tópicos serão discutidos a seguir na forma de um manuscrito e alguns resultados complementares.

PARTE II

3.2 – Reprodução de manuscrito

Prevalent Natural Epigenetic Variation at a Newly Formed Gene in *Arabidopsis thaliana*

Amanda Bortolini Silveira, Charlotte Trontin, Sandra Cortijo, Joan Barau, Luiz Eduardo Vieira
Del Bem, Olivier Loudet, Vincent Colot and Michel Vincentz

ABSTRACT

Epigenetic variation can contribute to plant phenotypic diversity, but because few natural epimutations have been documented, the relevance of this source of heritable phenotypes to adaptation and evolution is still unclear. Here we report a case of prevalent natural epigenetic variation in *Arabidopsis thaliana*, which concerns a gene involved in starch metabolism that likely originated *de novo* in a transposon-rich region since the divergence with *Arabidopsis lyrata*. We show that expression of this gene, named *Qua-Quine Starch (QQS)*, varies considerably among *Arabidopsis thaliana* accessions as well as within populations directly sampled from the wild and that this variation correlates negatively with the DNA methylation status of repeat sequences located within the 5' end of the gene. Furthermore, we demonstrate that *QQS* epiallelic states can be inherited across multiple generations and independently of any DNA sequence changes. Taken together, our findings indicate that stable epigenetic variation can be prevalent and that it could provide natural populations with an increased ability to explore the adaptive landscape.

Keywords: DNA methylation, natural variation, epiallele, *Arabidopsis*

INTRODUCTION

DNA mutations are the main known source of heritable phenotypic variation. Yet, epigenetic alterations such as gain or loss of DNA methylation can also affect gene expression and be stably transmitted across multiple generations without changes in the DNA sequence. Epimutations are therefore also a source of inherited phenotypes and indeed, numerous stable DNA methylation variants affecting a wide range of characters have been described in plants [1-6]. However, the prevalence as well as the adaptive and evolutionary significance of epigenetic variation is still unknown.

In most instances, the generation of epimutations appears to be associated with the presence of structural features near or within genes, such as direct or inverted repeats or transposable element (TE) insertions, which act as units of DNA methylation through the production of small interfering RNAs (siRNAs). Examples of “epimutable” loci in *Arabidopsis thaliana* (*A. thaliana*) include the *PAI* and *AtFOLT* genes, which have suffered siRNA-producing duplication events in some accessions [6-9], the *FWA* locus, which contains SINE-derived siRNA-producing tandem repeats at its 5’ end [10,11], as well as the *SDC* gene, the promoter of which is also composed of tandem repeats [4]. So far, repeat-associated “epimutable” loci have mainly been found in the methylated form in nature [4-7,11-13], which reflects, at least in part, that DNA methylation is particularly well-maintained over repeat sequences among accessions [14-15]. Thus, despite a few examples of spontaneous gain or loss of DNA methylation affecting gene expression [16,17], it is unclear if “epimutable” loci can ever exist in alternatively methylated states that are stable enough to spread into natural populations.

Here we report a case of prevalent natural epigenetic variation that concerns a *de novo* originated gene in *A. thaliana* [18]. We show that this gene, named *Qua-Quine Starch (QQS)*, is differentially expressed among accessions, as well as within wild populations and that these differences do not correlate with DNA sequence changes, but rather are correlated with the DNA methylation status of repeat sequences located at the promoter and 5’ UTR region of the gene. Furthermore, we demonstrate that DNA methylation and expression levels of *QQS* are stably inherited across generations, which therefore designates *QQS* epialleles as *bona fide* epigenetic variants. Given the potential impact of contrasting *QQS* expression levels on starch metabolism

[19,20], we propose that epigenetic variation at *QQS* could provide natural populations with increased opportunities for fine-tuning adaptation to changing environmental conditions.

RESULTS

***QQS* Is a Novel Gene Embedded within a TE-Rich Region of the *A. thaliana* Genome and is Negatively Regulated by DNA Methylation**

QQS is a gene (*At3g30720*) of *A. thaliana* with no similarity to any other gene or nucleotide sequence present in GenBank. Thus, *QQS* most likely originated *de novo* in *A. thaliana* after its divergence from *A. lyrata* (its closest relative, for which genome sequence is available), estimated to have occurred ~10 million years ago [18,21,22]. *QQS* is located on a region that is enriched in TE sequences (Figure 1A). It encodes a protein of 59 amino acids that has recently been characterized as a new component of the starch metabolism network [19,20]. Furthermore, analysis of synonymous versus (vs.) non-synonymous substitutions within the *QQS* coding sequence suggests that it is under purifying selection, as would be expected for a functional gene (dS/dN=1.704; p-value<0.045; *Supplementary Information (SI) Materials and Methods*). In the Columbia (Col-0) accession, the promoter and 5' UTR sequences of *QQS* are mainly methylated over two distinct repeat elements matching 20 to 24 nucleotides siRNAs (Figure 1B and Figure S1). Moreover, publically available data and RT-qPCR indicate that steady state levels of *QQS* mRNAs are increased in *met1* (*METHYLTRANSFERASE 1*), *ddc* (*DOMAINS REARRANGEMENT METHYLTRANSFERASE 1* and *2* and *CHROMOMETHYLASE 3*), *ddm1-1* (*DECREASE IN DNA METHYLATION 1*) and *rdr2-1* (*RNA-DEPENDENT RNA POLYMERASE 2*) mutants, in which DNA methylation of the repeat elements is erased or reduced (Figure S2A; refs. [23,24]). We therefore conclude that *QQS* expression is negatively regulated by DNA methylation in the Col-0 accession.

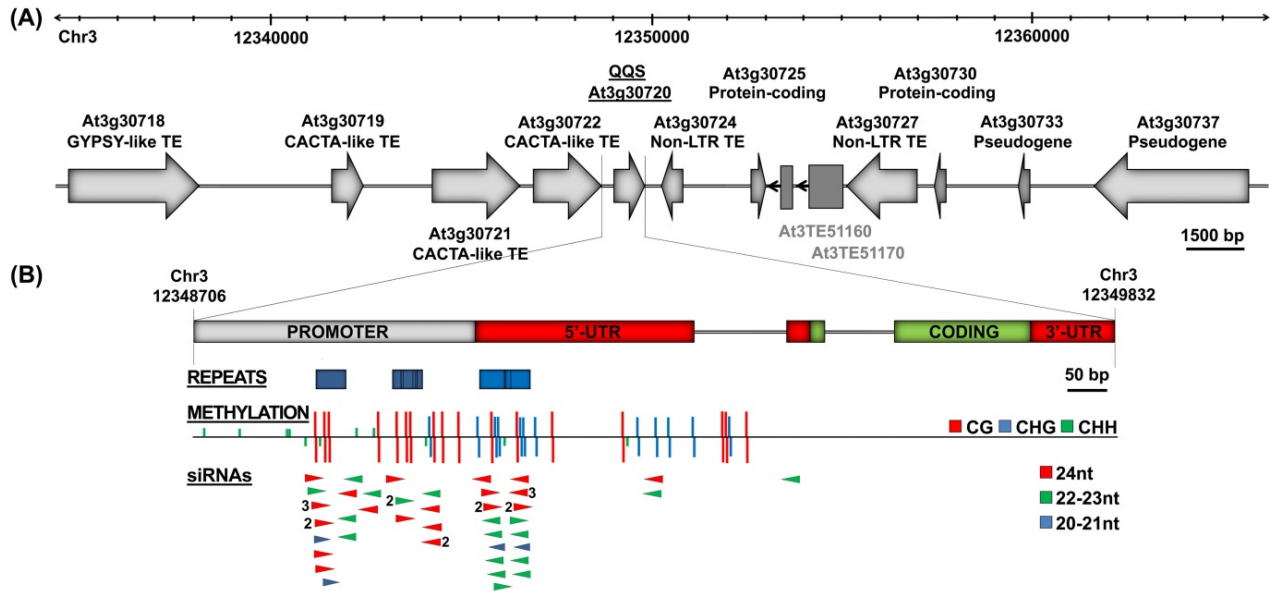


Figure 1. *QQS* is a target of epigenetic regulation. (A) Genomic structure of the *QQS* locus in the Col-0 accession (30kb window). Gray arrows indicate annotated TEs/ pseudogenes/ genes and gray boxes represent two predicted TEs. (B) Magnified view of the *QQS* gene showing methylation of cytosine residues (vertical lines) and locus-specific sense and antisense siRNAs (triangles, with numbers referring to copy number) matching repeat regions (blue boxes) found in the *QQS* promoter (2 units of 37 nucleotides sharing 81% identity and spaced by 57 nucleotides) and 5' UTR (2 direct repeat segments of 31 nucleotides sharing 97% identity). DNA methylation data was compiled from refs. [23,47] and siRNA data from refs. [23,48,49].

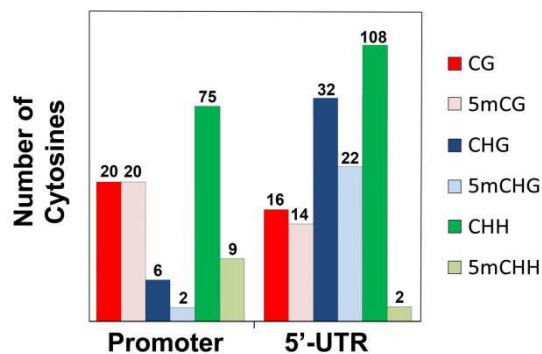


Figure S1. Distribution of DNA methylation at the *QQS* promoter and 5' UTR sequences. Data is presented as the total number of unmethylated (C) and methylated cytosines (5mC) at the three sequence contexts (CG, CHG and CHH) for both DNA strands. DNA methylation data was compiled from refs. [5,6].

Spontaneous and Induced Epigenetic Variation at *QQS*

Epiallelic variation at *QQS* was first observed when we unexpectedly identified a Col-0 laboratory stock (hereafter referred to as Col-0*) in which increased *QQS* expression was correlated with decreased DNA methylation levels at *QQS* promoter and 5' UTR sequences (Figure 2A). The sequence of a 1.2kb region covering the *QQS* gene (Figure 1B) was determined from Col-0* and found to be identical to the *QQS* sequence of the reference Col-0. This therefore excluded local *cis*-regulatory DNA polymorphisms at the *QQS* locus as being responsible for DNA methylation loss in Col-0*. Additionally, comparative genomic hybridization (CGH) analysis and genome-wide DNA methylation profiling using methylated DNA immunoprecipitation (MeDIP) assays (*SI Materials and Methods*) revealed no major differences between Col-0 and Col-0*, indicating that they have essentially the same genome and methylome (Figure S3). More detailed investigation of *QQS* epigenetic status in seedlings derived from the selfing of individual Col-0* plants (S1) and their single seed descent progeny (S2) revealed a range of *QQS* epiallelic variants and a strong anti-correlation between the degree of DNA methylation of the promoter region of *QQS* and expression levels of the gene (Figure 2B). DNA methylation of the 5' UTR was also anti-correlated with gene expression, although not to the same extent (Figure 2B). This suggests that the DNA methylation state of the promoter is the primary determinant of *QQS* expression level, with DNA methylation of the 5' UTR being only secondary. Remarkably, *QQS* expression and DNA methylation states appeared to be stably transmitted from S1 generation to S2 (Figure S2B). These findings indicate that *QQS* can adopt multiple epiallelic states in the Col-0* stock, each of which can be inherited for at least one generation.

To explore further the inheritance of *QQS* epialleles, a random sample of 19 *ddm1*-derived epigenetic Recombinant Inbred Lines (epiRILs) were also analyzed together with the Col-0 wild-type (wt) and Col-0 *ddm1* parental lines that were used to produce the F1 and backcross F2 [25]. High DNA methylation/low expression and low DNA methylation/high expression of *QQS* were observed in 14 and 5 epiRILs, respectively (Figure 2C), which is consistent with Mendelian segregation of the highly methylated/lowly expressed Col-0 wt and lowly methylated/highly expressed Col-0 *ddm1* parental *QQS* epialleles (75%/25% expected because of backcrossing rather than selfing of the F1; $\chi^2=0.017$, $p<0.05$). This demonstrates

that *QQS* epialleles can be stably inherited for at least eight generations, which is the number of generations used to propagate the epiRILs [25].

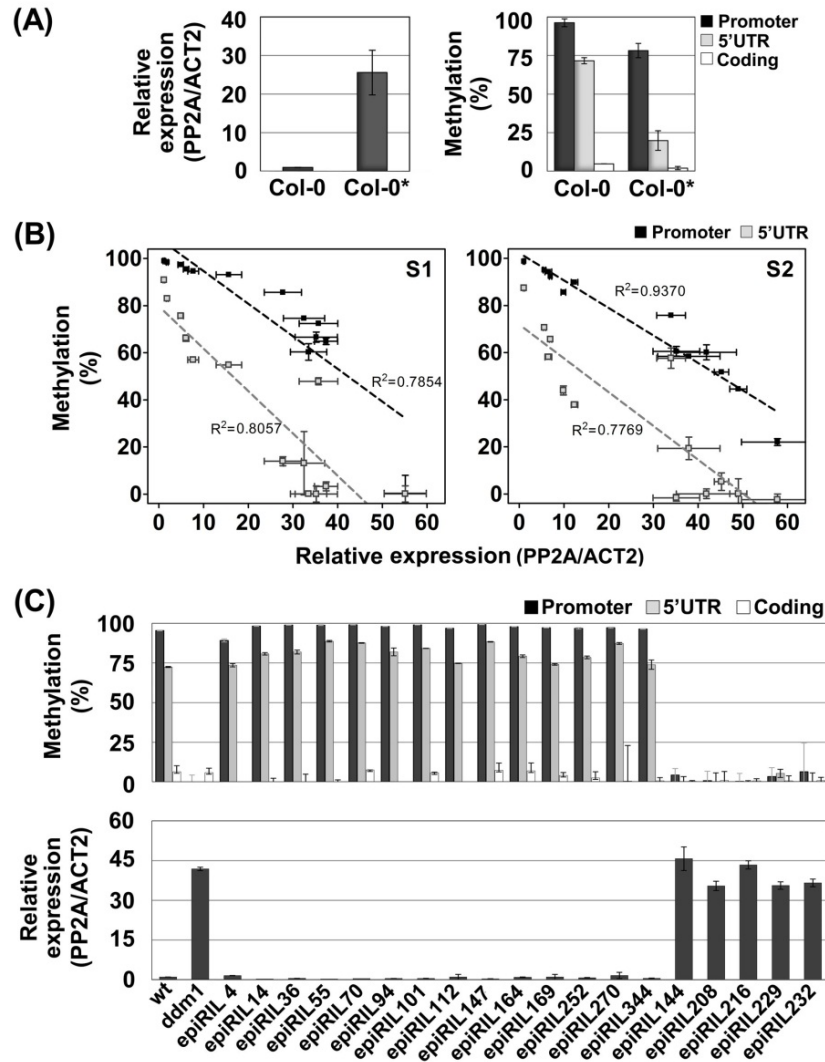


Figure 2. Spontaneous epigenetic variation at *QQS*. (A) Analysis of *QQS* DNA methylation and expression profiles in seedlings of Col-0 and Col-0* S0 population. (B) Analysis of the anti-correlation between *QQS* DNA methylation and expression levels in seedlings of Col-0 and Col-0* S1 and S2 generation single seed descent lines. (C) Analysis of *QQS* DNA methylation and expression profiles in seedlings of *ddm1*-derived epiRIL lines. Error bars represent standard deviation observed in three biological replicates (A, B and C – expression; A - methylation) or two technical replicates (B and C - methylation).

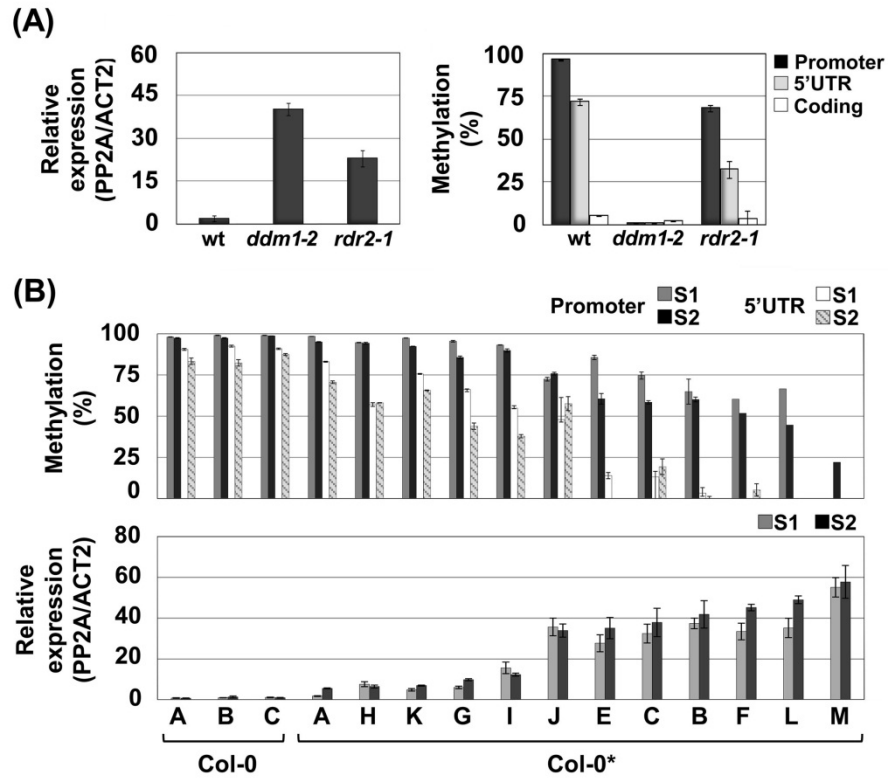


Figure S2. Assessment of *QQS* DNA methylation rate and transcript accumulation in seedlings of **(A)** *ddm1-2* and *rdr2-1* mutants and **(B)** Col-0 and Col-0* S1 and S2 single seed descent lines. Error bars represent standard deviation between two (A- methylation) or three (A and B - expression) biological replicates or two technical replicates (B – methylation).

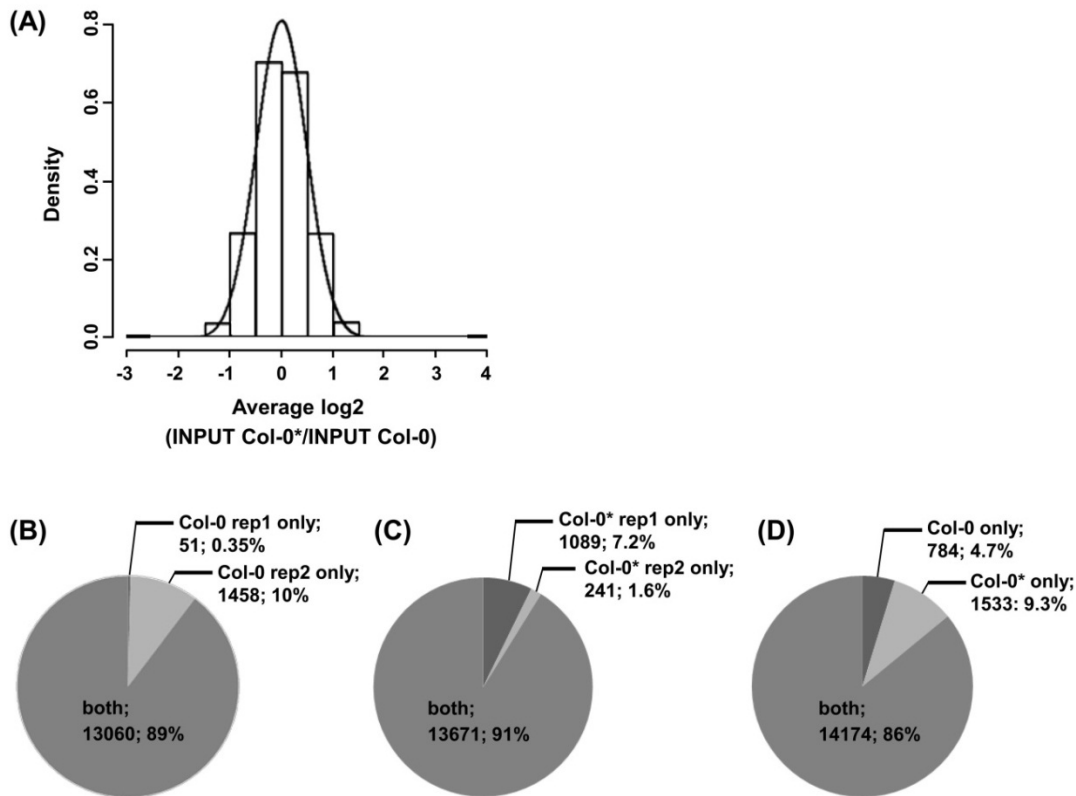


Figure S3. Comparative genome-wide analysis of Col-0 and Col-0*. (A) CGH of Col-0* against Col-0 represented as the average of the log₂ ratio of the signal for the INPUT Col-0* over INPUT Col-0. Only one normal distribution is observed using the normalmixEM function of the mixtools package on R with an expected number of gaussians of two. (B)-(D) MeDIP assays. Representation of the proportion of domains that are methylated (B) only in one replicate of Col-0 or in both, (C) only in one replicate of Col-0* or in both and (D) only in Col-0 or in Col-0* or in both.

Supplemental text Figure S3. The CGH analyses of Col-0* and Col-0 show no decrease or increase in copy number in Col-0*, suggesting that they correspond to the same accession. This is contrasted with the situation in Col-0 vs. Cvi-0 and Col-0 vs. C24 where 6.0 and 5.5% of tiles showed significant CGH polymorphisms respectively [7].

The comparison of the domains that are methylated in Col-0* or in Col-0 or in both shows that 86% of the domains are methylated in both and thus that their methylomes are very similar. This is very comparable to the result obtained for two biological replicates of Col-0 or of Col-0* where 89% and 91% of the domains are methylated in the two replicates, respectively. This result

indicate that there is no more differences between the methylomes of Col-0* and Col-0 than between two biological replicates.

***QQS* Is under Autonomous Epigenetic Control**

In order to define the boundaries of the methylated region involved in the differential expression of *QQS*, we investigated the degree to which DNA methylation of *QQS* and of flanking TEs are independent from each other. The methylation pattern of the three highly methylated TEs *At3g30722*, *At3g30721* and *At3g30718* located upstream of *QQS*, as well as of the poorly methylated TE *At3g30724* located downstream of *QQS* (Figure 3A), was determined in *ddm1*-derived epiRIL lines carrying contrasting *QQS* epialleles (Figure 2C). No correlation between *QQS* and nearby TE methylation pattern could be established (Figure 3B). Indeed, while *QQS* remained demethylated when inherited from the *ddm1* parent, the upstream TEs *At3g30722*, *At3g30721* and *At3g30718* regained wt DNA methylation levels (Figure 3B), presumably because of their efficient targeting by RNA-directed DNA Methylation (RdDM; ref. [26]). Furthermore, T-DNA insertions located ~2.8kb upstream of the *QQS* promoter (GABI-Kat 755C03/*At3g30721* locus) and 653bp downstream of the *QQS* 3'UTR (SALK 003185C/*At3g30724* locus) or a transposon insertion located within the second coding exon of *QQS* (WiscDsLoxHs077_09G) did not appreciably alter DNA methylation of the gene (Figure 3A and C). In contrast, a T-DNA insertion located 153bp upstream of the transcription start site of *QQS* (GABI-Kat 522C07) was associated with a drastic demethylation of both the promoter region and 5' UTR, as well as with an increased expression of the gene (Figure 3A and C), but had no impact on the DNA methylation of the flanking TEs (Figure S4A and B). Taken together, these results indicate that epigenetic variation at *QQS* is independent of the DNA methylation status of surrounding sequences, which thus reinforces the notion that this variation is primarily determined by the promoter region of the gene.

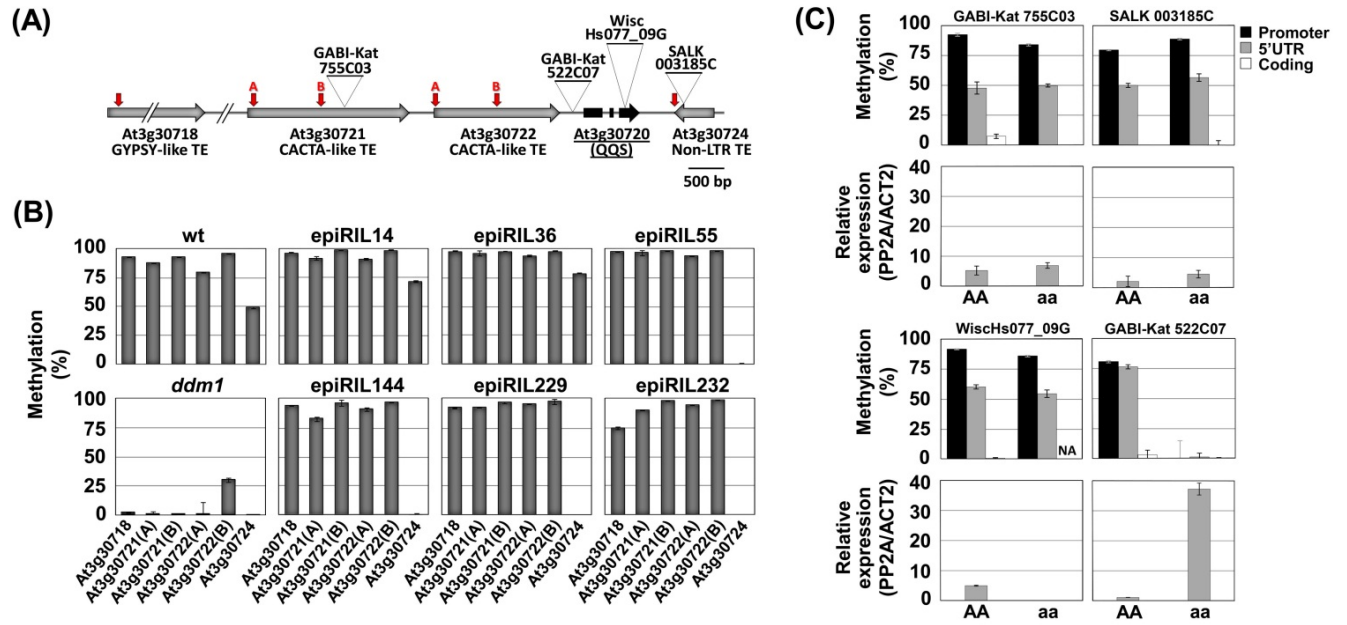


Figure 3. *QQS* is under autonomous epigenetic control. (A) Schematic representation of *QQS* genomic region (16kb) showing TE distribution, TDNA/Transposon-insertion sites (triangles; GABI-Kat 755C03, GABI-Kat 522C07, WiscDsLoxHs077_09G (WiscHs077_09G) and SALK 003185C) and McrBC-qPCR primer pairs (vertical arrows; A and B represent different primer pairs designed for the same element). (B) Quantification of *QQS*-nearby TEs methylation profile in epiRIL lines that had inherited a wt (epiRILs 14, 36 and 55) or *ddm1*-derived *QQS* epiallele (epiRILs 144, 229, 232). (C) Evaluation of *QQS* methylation and expression profile in mutant lines carrying the T-DNA/transposon insertions represented in (A). AA and aa represent wt and T-DNA-homozygous individuals, respectively, coming from the selfing of one heterozygous (Aa) plant. NA, not analyzed. Error bars represent standard deviation observed in two technical replicates (B) or three biological replicates (C).

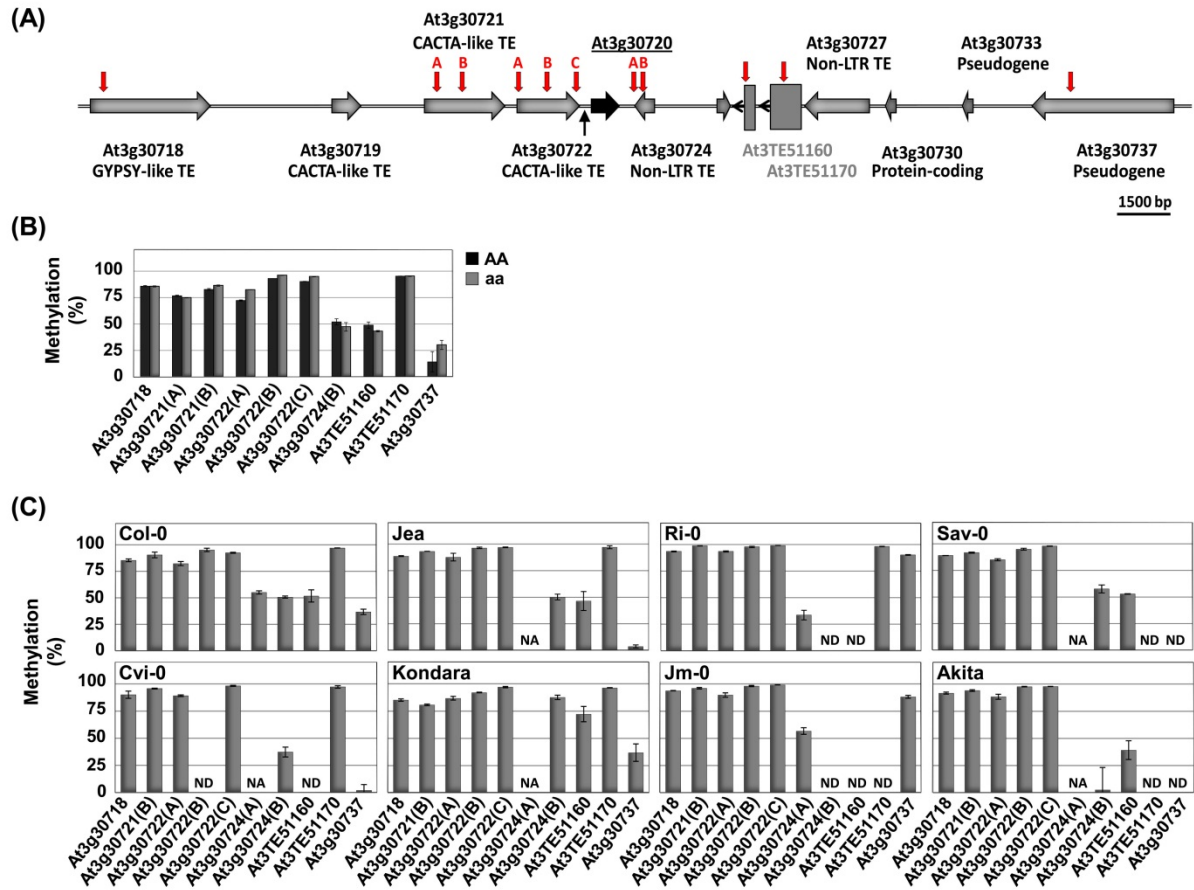


Figure S4. *QQS* is under autonomous epigenetic control. (A) Schematic representation of *QQS* genomic region (30kb). Red arrows indicate McrBC-qPCR primer pairs used to quantify methylation profile of *QQS*-nearby TEs; A, B and C represent different primer pairs designed for the same element. GABI-Kat 522C07 T-DNA insertion (153bp upstream of *QQS* transcription start site) is indicated by a black vertical arrow. (B) Analysis of the methylation profile of *QQS* flanking region in GABI-Kat 522C07 T-DNA insertion mutant. AA and aa represent wt and T-DNA-homozygous individuals, respectively, coming from the selfing of one heterozygous (Aa) plant. (C) Assessment of the DNA methylation profile of *QQS* flanking region in Col-0 (methylated *QQS* epiallele), Jea, Ri-0, Sav-0, Cvi-0, Kondara and Akita (hypomethylated *QQS* epiallele) accessions. 'NA': not analyzed; 'ND': not determined (samples whose methylation status could not be measured precisely due to possible polymorphisms at the primer annealing site). Error bars represent standard deviation observed in two technical replicates.

***QQS* Exhibits Epigenetic Variation among Natural Accessions**

We next explored the possibility that *QQS* is subject to epigenetic variation in nature. To this end, we first analyzed *QQS* expression and DNA methylation in 36 natural accessions representing the worldwide diversity [27]. *QQS* was methylated and lowly expressed in 29 accessions, but unmethylated and highly expressed in seven accessions (Figure 4). Moreover, no correlation between DNA methylation/expression patterns and relatedness of accessions was found (Figure 4), therefore indicating that epigenetic variation at *QQS* is prevalent in nature. To rule out that DNA sequence polymorphisms among *QQS* alleles could be responsible for the differences in DNA methylation/expression observed between accessions, a 2.8 kb interval encompassing the *QQS* gene and its flanking regions was sequenced from seven accessions carrying a hypomethylated/highly expressed epiallele and from three accessions carrying a methylated/lowly expressed epiallele. Although several SNPs and indels were identified, no correlation between any specific sequence alterations and *QQS* DNA methylation states could be established (Figure S5). Moreover, the finding that identical *QQS* sequences in Kondara and Shahdara accessions were respectively associated with low DNA methylation/high expression and high DNA methylation/low expression (Figure 4 and Figure S5), indicates that epiallelic variation at *QQS* is independent from local *cis*-DNA sequence polymorphisms and also corroborates the anti-correlation between DNA methylation and *QQS* expression levels.

Next, seedlings of the Col-0 and Shahdara accessions were grown in the presence of the DNA methylation inhibitor 5-aza-2'-deoxycytidine (5-aza-dC). In both accessions, treatment with 5-aza-dC resulted in reduced DNA methylation and increased expression of *QQS* (Figure 5A). As a control, seedlings of Jea, Kondara and Cvi-0 accessions in which *QQS* is demethylated were also grown in the presence of 5-aza-dC and no further reduction of DNA methylation or increase in *QQS* expression were observed (Figure 5A). Moreover, expression of *QQS* in F1 hybrids derived from crosses between Col-0 (methylated *QQS*) and Jea or Kondara accessions (hypomethylated *QQS*), was always higher for the epiallele inherited from the hypomethylated parent (Figure 5B). Also, treatment of Col-0 vs. Kondara F1 individuals with 5-aza-dC reduced dramatically this expression imbalance, most likely as a consequence of demethylation of the Col-0-derived *QQS* allele (Figure 5C). Taken together, these results clearly demonstrate that DNA methylation is causal in repressing *QQS* expression.

As expected, epigenetic variation at *QQS* in the different accessions was found to be independent of the DNA methylation status of the surrounding TEs (Figure S4C), as in the epiRILs (Figure 3B). Furthermore, analysis of a Cvi-0 vs. Col-0 Recombinant Inbred Line (RIL) population indicated that *QQS* expression level is controlled by a large-effect local-expression quantitative trait locus (local-eQTL) likely acting in *cis* (<http://qtlstore.versailles.inra.fr/>; ref. [28]). Thus, like the Col-0 wt and Col-0 *ddm1* *QQS* epialleles, which are stably inherited for at least eight generations in the epiRILs (Figure 2C), the Col-0 and Cvi-0 *QQS* epialleles are stably inherited across at least five generations in the RILs.

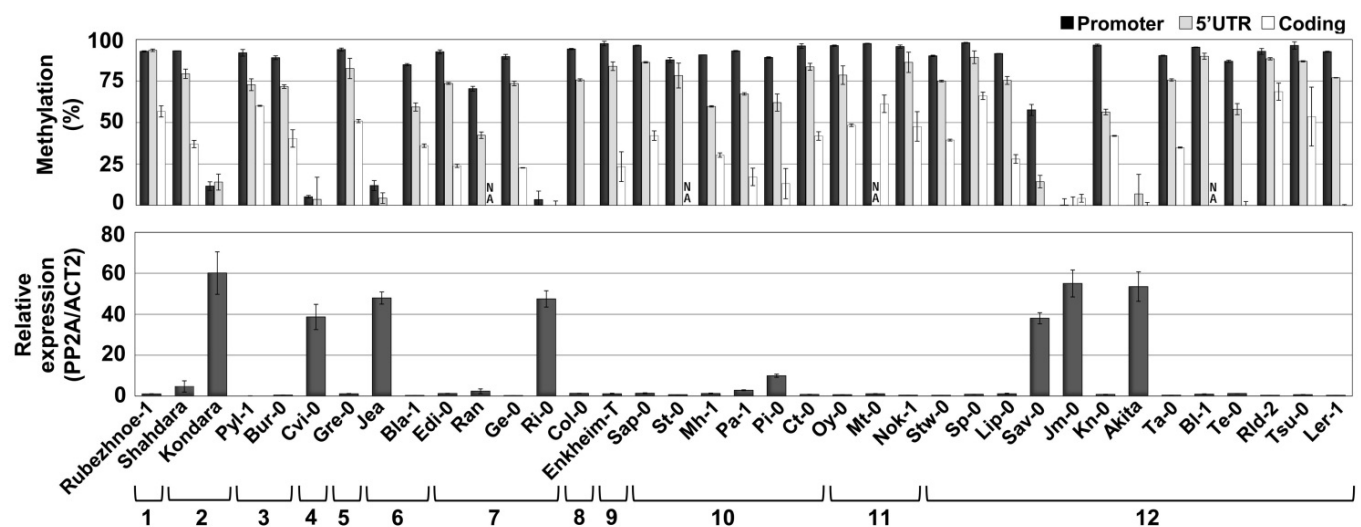


Figure 4. Analysis of *QQS* DNA methylation and expression profile in *A. thaliana* accessions representing the worldwide diversity. Accessions are organized into clades 1 to 12 according to genetic relatedness [38]. Error bars represent standard deviation observed in two (methylation) or three (expression) biological replicates.

	-1709	-1687	-1672	-1586	-1566	-1557	-1522	-1519	-1413	-1411	-1363	-1323	-1324	-1351	-1323	-1311	-1309	-1257	-1251	-1250	-1247	-1231	-1158	-1155	-1112	-1074	-1068	-1061	-1059	-1014	-988	-987	-938	-902	-841	-832		
Col-0	C	G	G	G	G	T	G	G	A	T	G	G	C	C	G	C	A	G	G	G	G	C	C	C	A	G	A	C	C	C	C	C	C	C	C	C	C	G
Pyl-1	C	G	G	G	G	T	G	G	A	T	G	G	C	C	G	C	A	G	G	G	G	C	C	C	A	G	A	C	C	C	C	C	C	C	C	C	C	G
Mh-1	C	G	G	G	G	T	G	G	A	T	G	G	C	C	G	C	A	G	G	G	G	C	C	C	A	G	A	C	C	C	C	C	C	C	C	C	C	G
Shahdara	T	G	G	G	G	T	G	G	A	T	G	G	C	C	G	C	A	G	G	G	G	C	C	C	A	G	A	C	C	C	C	C	C	C	C	C	C	T
Kondara	T	G	G	G	G	T	G	G	A	T	G	G	C	C	G	C	A	G	G	G	G	C	C	C	A	G	A	C	C	C	C	C	C	C	C	C	C	T
Cvi-0	T	G	G	G	G	T	G	G	A	T	G	G	C	C	G	C	A	G	G	G	G	C	C	C	A	G	A	C	C	C	C	C	C	C	C	C	C	T
Jea	T	G	G	G	G	T	G	G	A	T	G	G	C	C	G	C	A	G	G	G	G	C	C	C	A	G	A	C	C	C	C	C	C	C	C	C	C	T
Ri-0	T	G	G	G	G	T	G	G	A	T	G	G	C	C	G	C	A	G	G	G	G	C	C	C	A	G	A	C	C	C	C	C	C	C	C	C	C	T
Sav-0	T	G	G	G	G	T	G	G	A	T	G	G	C	C	G	C	A	G	G	G	G	C	C	C	A	G	A	C	C	C	C	C	C	C	C	C	C	T
Jm-0	T	G	G	G	G	T	G	G	A	T	G	G	C	C	G	C	A	G	G	G	G	C	C	C	A	G	A	C	C	C	C	C	C	C	C	C	C	T
Akita	T	G	G	G	G	T	G	G	A	T	G	G	C	C	G	C	A	G	G	G	G	C	C	C	A	G	A	C	C	C	C	C	C	C	C	C	C	T
	-817	-725	-705	-704	-703	-702	-701	-700	-699	-698	-697	-696	-695	-694	-693	-692	-689	-684	-683	-682	-681	-512	-487	-468	-466	-404	-373	-356	-344	-342	-341	-340	-339	-338	-337	-336		
Col-0	C	T	G	T	T	C	A	T	T	T	A	T	T	T	T	T	T	T	T	T	T	C	G	T	T	G	A	C	C	T	T	A	A	A	G	A	A	
Pyl-1	C	T	G	T	T	C	A	T	T	T	A	T	T	T	T	T	T	T	T	T	T	C	G	T	T	G	A	C	C	T	T	A	A	A	G	A	A	
Mh-1	C	T	G	T	T	C	A	T	T	T	A	T	T	T	T	T	T	T	T	T	T	C	G	T	T	G	A	C	C	T	T	A	A	A	G	A	A	
Shahdara	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T		
Kondara	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T		
Cvi-0	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T		
Jea	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T		
Ri-0	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T		
Sav-0	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T		
Jm-0	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T		
Akita	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T		
	-335	-334	-333	-332	-331	-330	-329	-328	-327	-326	-325	-324	-323	-322	-321	-320	-319	-318	-317	-316	-315	-314	-313	-312	-311	-310	-309	-308	-306	-305	-304	-301	-297	-265	-259			
Col-0	A	A	C	A	A	G	A	C	T	C	A	C	A	C	C	G	T	C	C	A	G	C	A	T	T	G	A	A	A	A	A	G	T	C	A	C	C	
Pyl-1	A	A	C	A	A	G	A	C	T	C	A	C	A	C	C	G	T	C	C	A	G	C	A	T	T	G	A	A	A	A	A	G	T	C	A	C	C	
Mh-1	A	A	C	A	A	G	A	C	T	C	A	C	A	C	C	G	T	C	C	A	G	C	A	T	T	G	A	A	A	A	A	G	T	C	A	C	C	
Shahdara	A	A	C	A	A	G	A	C	T	C	A	C	A	C	C	G	T	C	C	A	G	C	A	T	T	G	A	A	A	A	A	G	T	C	A	C	C	
Kondara	A	A	C	A	A	G	A	C	T	C	A	C	A	C	C	G	T	C	C	A	G	C	A	T	T	G	A	A	A	A	A	G	T	C	A	C	C	
Cvi-0	A	A	C	A	A	G	A	C	T	C	A	C	A	C	C	G	T	C	C	A	G	C	A	T	T	G	A	A	A	A	A	G	T	C	A	C	C	
Jea	A	A	C	A	A	G	A	C	T	C	A	C	A	C	C	G	T	C	C	A	G	C	A	T	T	G	A	A	A	A	A	G	T	C	A	C	C	
Ri-0	A	A	C	A	A	G	A	C	T	C	A	C	A	C	C	G	T	C	C	A	G	C	A	T	T	G	A	A	A	A	A	G	T	C	A	C	C	
Sav-0	A	A	C	A	A	G	A	C	T	C	A	C	A	C	C	G	T	C	C	A	G	C	A	T	T	G	A	A	A	A	A	G	T	C	A	C	C	
Jm-0	A	A	C	A	A	G	A	C	T	C	A	C	A	C	C	G	T	C	C	A	G	C	A	T	T	G	A	A	A	A	A	G	T	C	A	C	C	
Akita	A	A	C	A	A	G	A	C	T	C	A	C	A	C	C	G	T	C	C	A	G	C	A	T	T	G	A	A	A	A	A	G	T	C	A	C	C	
	+201	+192	+176	+158	+141	+130	+119	+94	+82	+78	+74	+73	+72	+71	+70	+69	+68	+67	+66	+65	+64	+63	+62	+61	+60	+59	+58	+57	+56	+55	+54	+53	+52	+51	+50			
Col-0	T	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T			
Pyl-1	T	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T			
Mh-1	T	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T			
Shahdara	T	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T			
Kondara	T	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T			
Cvi-0	T	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T			
Jea	T	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T			
Ri-0	T	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T			
Sav-0	T	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T			
Jm-0	T	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T			
Akita	T	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T			
	+495	+514	+524	+548	+573	+595	+607	+610	+611	+612	+613	+614	+615	+616	+617	+618	+619	+620	+621	+622	+623	+624	+625	+626	+627	+628	+629	+630	+631	+632	+633	+634	+635	+636	+637	+638		
Col-0	T	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G			
Pyl-1	T	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G			
Mh-1	T	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G			
Shahdara	T	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G			
Kondara	T	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G			
Cvi-0	T	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G			
Jea	T	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G			
Ri-0	T	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G			
Sav-0	T	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G			
Jm-0	T	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G			
Akita	T	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G			

Figure S5. Genetic polymorphisms (SNPs and Indels) at *QQS* locus and flanking region (1.5kb upstream and 0.6kb downstream of the *QQS* transcription initiation and termination site, respectively) in accessions carrying methylated (Col-0, Pyl-1, Mh-1, Shahdara) and hypomethylated (Kondara, Cvi-0, Jea, Ri-0, Sav-0, Jm-0 and Akita) *QQS* epialleles. Nucleotide positions are numbered relative to the adenosine of the *QQS* translation initiation site (Position +1).

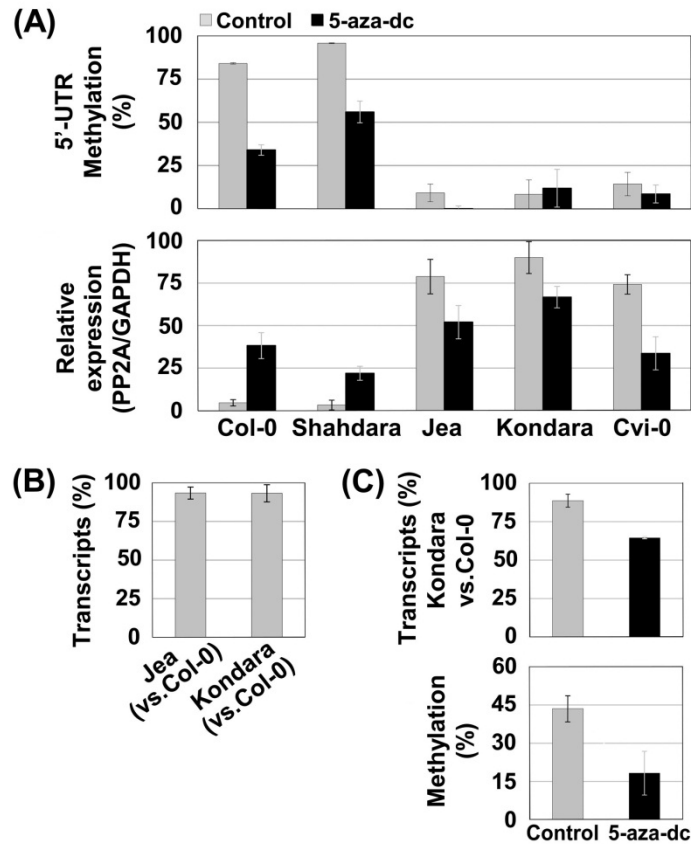


Figure 5. *cis*-DNA methylation profiles modulate *QQS* expression. (A) Effect of the methylation inhibitor 5-aza-dC on *QQS* DNA methylation profile and expression in Col-0, Shahdara, Jea, Kondara and Cvi-0 accessions. (B) Pyrosequencing quantification of *QQS* allele-specific expression in F1 individuals issued from the crosses Col-0 vs. Jea or Col-0 vs. Kondara. Data is expressed as the % of total transcripts originating from the non-Col-0 allele (Jea or Kondara). (C) Effect of 5-aza-dC on *QQS* allele-specific expression (pyrosequencing quantification, expressed as in Fig. 3B) and DNA methylation profile in Col-0 x Kondara F1 individuals. Error bars represent standard deviation observed in two technical replicates (A-methylation) or at least 3 biological replicates (A-expression, B, C).

Wild Populations from Central Asia Exhibit Epigenetic Variation at *QQS*

We next asked whether epigenetic variation at *QQS* could also be observed in natural environments under selective situations or if such variation only emerged in the laboratory, where accessions are grown under controlled growth conditions with relaxed selective pressures. To this end, *QQS* expression and DNA methylation were analyzed in plants grown from seeds directly collected from the wild in Tajikistan, Kyrgyzstan and Iran (Neo-Shahdara, Zalisky and Anzali populations, respectively). Results indicated widespread *QQS* epiallelic variation, both among and within the diverse wild populations that were sampled (Figure 6A). Further analysis of *QQS* epiallelism was conducted in the offspring (after two single seed descent generations) of 25 Neo-Shahdara individuals. These individuals were randomly sampled among a single patch of several thousands of plants that presumably represent the direct descendants of the Shahdara accession. Furthermore, based on ten microsatellite and one SNP markers, we observed that the 25 Neo-Shahdara individuals could be structured into two genetically distinct subpopulations. Subpopulation 'one' contained 16 individuals, all of which harbored a highly methylated *QQS* allele. In contrast, subpopulation 'two' was composed of seven individuals with poorly methylated and two individuals with highly methylated *QQS* sequences (Figure 6B). Whether the ancestral state of *QQS* in subgroup 'two' is demethylated and Neo-2 and Neo-20 represent remethylation events, or whether this group has witnessed an event of DNA methylation loss which has not fully invaded the subpopulation remains to be determined.

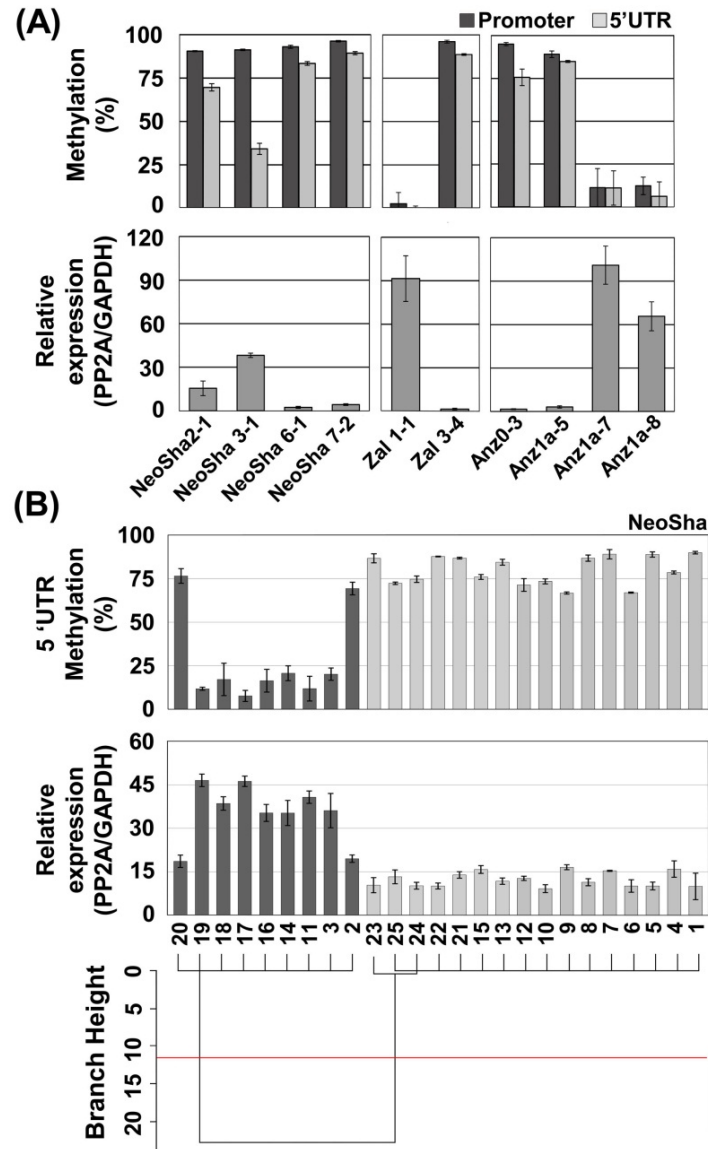


Figure 6. *QQS* epiallelic variation in wild populations of Central Asia. (A) Evaluation of *QQS* expression and DNA methylation profile in plants grown from seeds directly collected in the field from different individuals within NeoShahdara (NeoSha), Zalisky (Zal) and Anzali (Anz) wild populations. For each line, one to three sibling plants were tested and gave similar results so that only one is represented per individual. (B) Analysis of *QQS* epiallelic frequency among 25 individuals from the NeoSha wild population. Plants analyzed here are obtained from seeds produced after two single seed descent generations. On the dendrogram obtained with AWclust from a matrix of 11 genotypic markers and 25 Neo-Shahdara individuals, the line cuts across the two main clusters identified. Error bars represent standard deviation observed between two technical replicates (A, B).

DISCUSSION

QQS is a protein-coding gene that likely originated *de novo* within a TE-rich region of the *A. thaliana* chromosome three after *A. thaliana* diverged from its relative *A. lyrata* (Figure 1A). Whereas expression of *QQS* is clearly under developmental and environmental control (Arabidopsis eFP browser, <http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi> and Genevestigator, <https://www.genevestigator.com/gv/>; refs. [18,29,30]), our findings demonstrate that it is also subject to prevalent epigenetic regulation in nature. We have shown that *QQS* can assume a range of heritable DNA methylation/expression states (Figure 2B, Figure S2B, Figure 4, Figure 5 and Figure 6) that are apparently independent of the epigenetic status of neighboring TEs and of DNA sequence variation in *cis* or *trans* (Figure 2C, Figure 3, Figure S4 and Figure S5). In fact, the recurrent and independent appearance of demethylated/expressed *QQS* epigenetic variants in the Col-0* laboratory stock (Figure S2B), as well as in diverse accessions and wild populations (Figure 4, Figure 6), favors the notion that this gene is a hotspot of epigenetic variation. Supporting this view, spontaneous loss of DNA methylation at *QQS* has also been recently observed in mutation accumulation lines [15]. Furthermore, DNA methylation data and results of insertional mutagenesis (Figure 2B and Figure 3 A and C) suggest that *QQS* promoter and 5' UTR sequences, which contain short repeat elements matching siRNAs (Figure 1B), are responsible for epigenetic variation at *QQS*.

Cytosine methylation at *QQS* concerns CG, CHG and CHH sites, which is the pattern expected for sequences with matching siRNAs (Figure 1B and Figure S1). All three types of methylation sites likely contribute to silencing of *QQS*, as judged by the reactivation of *QQS* in *met1*, *ddm1*, *ddc* and *rdr2* mutant backgrounds (Figure S2A; ref [23,24]). Yet, among the different DNA methyltransferases targeting DNA methylation at *QQS*, *MET1* may play a more prominent role, given that DNA methylation is only fully erased at this locus in *met1* mutant plants [23]. *QQS* demethylated epiallelic variants would thus arise through either spontaneous [16] or stress-induced [31] incorrect maintenance of DNA methylation and would be stably inherited for multiple generations due to the concomitant loss of matching siRNAs, which would prevent efficient remethylation and silencing of *QQS* [26-32]. Indeed, although we could not

detect *QQS* siRNAs by Northern blot analysis, presumably because of their low abundance, deep sequencing data indicate that they accumulate less in *met1* mutant plants than in wt Col-0 [23].

Few genes have been shown so far to be subject to heritable epigenetic variation in *A. thaliana* [2,4,6,7,16,17,33] and *QQS* is unique among these in exhibiting this type of variation in nature. This therefore raises the question as to what distinguishes *QQS* from other genes that can vary epigenetically, such as *FWA*, which is also characterized by a lower abundance of matching siRNAs when it is hypomethylated and expressed [10,23]. Three distinguishing features of *QQS* are the fact that it arose not by gene duplication, but *de novo* through a still unknown process (but see ref.[34], that it was formed recently and that it is localized within a TE-rich environment. Whether it is the combination of some or all of these features that renders *QQS* particularly prone to heritable epigenetic variation remains to be determined. In any case, this suggests a scenario by which genes arising *de novo* would have increased possibilities for epigenetic variation and thus for adjusting their expression pattern until the most adaptive epigenetic state becomes fully stabilized (i.e. genetically assimilated) through DNA sequence changes [35]. Evolutionary forces favoring the capacity to maintain or provide increased epigenetic variability which could match specific environmental alterations may also be at play [36]. Although speculative, these scenarios are worth considering in view of the potential impact of contrasting *QQS* expression levels on starch metabolism [19,20], which is an important component of biomass production and growth [37]. Differentially methylated epialleles, by providing a range of expression options, may fine-tune starch metabolism, potentially contributing to adaptive responses. Although an effect of Col-0* *QQS* epialleles on starch accumulation was not observed under stable growth chamber conditions, it remains to be determined whether this kind of process could operate under selective natural conditions, which is a critical issue in light of the range of *QQS* epigenetic variation found in wild populations such as in Neo-Shahdara (Figure 6).

MATERIALS AND METHODS

Plant material and growth conditions

A. thaliana accessions were obtained from the INRA Versailles collection (dbsgap.versailles.inra.fr/vnat/, www.inra.fr/vast/collections.htm; refs. [27,38,39]). Insertion lines were obtained from the GABI-Kat at University of Bielefeld, Germany (GABI-Kat 755C03 and 522C07, ref. [40]), the ABRC at Ohio State University (SALK 003195C) and University of Wisconsin, Madison, US (WiscDsLoxHs077_09G, ref. [41]). Seeds of *ddm1-2* [42], *rdr2-1* [43] and *ddm1*-derived epiRIL lines [25] were provided by V.Colot. All mutants were genotyped by PCR using the primers described in Table S1. For expression and DNA methylation assays, seedlings were grown in vitro (MS/2 media supplemented with 0,7% sucrose) for eight days in a culture room (22°C, 16 hours light/8 hours dark cycle, 150µmol s⁻¹ m⁻²). Treatment with 5-aza-2'-deoxycytidine was performed as described in ref. [6].

RT-qPCR analysis of *QQS* expression levels

Total RNA was isolated as described in ref. [44] and cDNA was synthesized using oligo(dT) primers and IMPROM II reverse transcriptase (Promega). Real time PCR reactions were run on an Applied Biosystems 7500 Real-Time PCR System using Platinum SYBR green (Invitrogen). *QQS* expression levels relative to *Actin2/PP2A* or *PP2A/GAPDH* internal references were calculated using the formula $(2^{-(Ct_{QQS} - \text{mean } Ct_{\text{internal references}})}) * 100$. Primers are listed in Table S1.

Analysis of DNA methylation by McrBC-qPCR

Total DNA was isolated using Qiagen Plant DNeasy kit following the manufacturer's recommendations. Digestion was carried out overnight at 37°C with 200ng of genomic DNA and 2 to 8 units of McrBC enzyme (New England Biolabs). Quantitative PCR was performed as described above on equal amounts (2ng) of digested and undigested DNA samples using the primers described in Table S1. Results were expressed as percentage of molecules lost through McrBC digestion $(1 - (2^{-(Ct_{\text{digested sample}} - Ct_{\text{undigested sample}})})) * 100$. As a control, the percentage of DNA methylation for *At5g13440*, which is unmethylated in wt, was estimated in all analysis.

Allele-specific expression assays

To assess the relative contribution of each allele to the population of mRNA in F1 individuals from reciprocal crosses between Col vs. Jea and Col vs. Kondara, a single pyrosequencing reaction using the primers described in Table S1 was set up on a SNP polymorphic between the *QQS* parental coding sequences (Figure S5; position +285). Pyrosequencing was performed on F1 cDNA, as well as on 1/1 pools of parents cDNA to establish the allelic contribution to *QQS* expression. F1 genomic DNA is used as pyrosequencing control to normalize against possible pyrosequencing biases. Anything significantly driving allele-specific expression in hybrids is -by definition- acting in *cis*, since F1 nuclei contain a mix of all *trans*-acting factors [45,46].

SI MATERIALS AND METHODS

Overall codon-based Z-test of purifying selection

Available *QQS* coding-sequences (464 different accessions) were downloaded from the “Salk Arabidopsis 1001 Genomes” database (<http://signal.salk.edu/atg1001/index.php>). *A. suecica* *QQS* sequence was also included in the analysis. The aligned sequences were used to calculate the probability of rejecting the null hypothesis (H_0) of strict-neutrality ($dN=dS$; where dN = number of nonsynonymous and dS = number of synonymous substitutions per site) in favor of the alternative hypothesis of purifying selection (H_A ; $dS > dN$). The analysis was done using the MEGA5 software under the Nei-Gojobori method [1] with the variance of the difference calculated by the bootstrap method with 100 replicates. Our overall analysis of 465 sequences rejected H_0 in favor of H_A ($dS/dN=1.704$; $p\text{-value}<0.045$).

Comparative genome hybridization (CGH)

CGH experiments were performed for Col-0* vs. Col-0 using Arabidopsis whole-genome tiling NimbleGen arrays. The normalmixEM function of the mixtools package on R was used to find the normal distribution for the distribution of the Col-0*/Col-0 ratio with an expected number of gaussians of two. A Hidden Markov model [2] was used to find regions that change in copy number.

Analysis of genome wide DNA methylation (MeDIP-Chip)

DNA was extracted using DNeasy Qiagen kit and MeDIP-chip was performed on 1.8µg of DNA as previously described in ref. [3]. The methylated tiles were identified using the ChIPmix method [4]. Probes methylated in one line only (Col-0 or Col-0*) were used to create domains. Domains contain at least three consecutive or nearly consecutive (400 nt min, with one gap of 200nt max) tiles with identical methylation patterns.

ACKNOWLEDGMENTS

We thank F.K.Teixeira and members of the Colot group for valuable assistance, insights and discussions. We thank M. Canut for help with the ASE assays. A.B.S. was supported by a PhD studentship from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) and S.C. by a PhD studentship from the French Ministry of Research. This work was supported by grants from FAPESP (BIOEN Program) to M.V, CNRS/FAPESP to V.C. and M.V. and the European Union Network of Excellence Epigenesys to V.C.

AUTHOR CONTRIBUTIONS

A.B.S., C.T., O.L, V.C. and M.V. designed research; A.B.S., C.T., S.C., J.B. and L.E.V.D.B. performed research; A.B.S., C.T., S.C., O.L, V.C. and M.V. analyzed the data; A.B.S., V.C. and M.V. wrote the paper.

Table S1. Primer List

RT-qPCR				
AGI	Name of the primer pair	Forward primer	Reverse primer	Genotypes
At3g30720 (QQS)	QQS_1	AAGACCAATAGAGACAGGAA	CCTGATGTAGAAAGTGTGAGG	All, with the exception of WiscDslxHs077_09G on Fig.3
	QQS_2	CACCTTCACATCAGGTGTCG	AAGGCCCAATATCAGTAGTTG	
	PP2A_1	CATGTTCCAAACTCTTACCTG	GTTCTCCAAACCGCTTGGT	
	PP2A_2	TTTGTGAAGCTGTAGACCG	CGAGTTCAGGGTTTAAATGCG	
	ACT2	GTACAACCGGATTGTGCTGG	CAAGGTCAAGACGGAGGATG	
At1g13440(GAPDH)	GAPDH	TTGGTGACAAACAGGTCAAGC	AAACTTGTGCGTCAATGCAATC	Fig. 2, Fig. 3, Fig. 4, Fig. S2 Fig.5 and Fig. 6
McrBC-qPCR				
AGI	Name of the primer pair	Forward primer	Reverse primer	Genotypes
At3g30720(QQS)	Promoter(F1A/R1A)	TCACTCGGATTGATGTCGTG	AGGAGACGAAACAGACAAAATC	Fig. 2, Fig. 3 (with the exception of GABI-Kat_522C07), Fig. 4 (Col-0, Gre-0, Cti-1, Enkheim-0, Jea, Ler-1, Nok-1, Pa-1, Pi-0, Mh-1, Sav-0, Mt-0, Akita, Kondara, Sha, Tsu-0), Fig. 5, Fig. 6, Fig. S2
	Promoter (F1A/R1B)	TCACTCGGATTGATGTCGTG	AGGAGATGAAACAGACAAAATC	
	Promoter (F1B/R1B)	GCCAAATTAGAATGTTTCACTCG	AGGAGATGAAACAGACAAAATC	
	Promoter (F1C/R1A)	TCCAAGCTTTGCCAAACGATC	AGGAGACGAAACAGACAAAATC	
	5'-UTR (F2A/R2A)	TCTGTACGCCATTGAAGAAAC	GATAAGGTTTGGGTACAGATC	
	5'-UTR (F2A/R2B)	TCTGTACGCCATTGAAGAAAC	GATAAGGTTTGGGTACAGATC	Fig. 4 (Cvi-0, Bl-1, Bla-1, Ge-0, Pyl-1, Edi-0, Oy-0, Te-0, Jm-0, Lip-0, Rubezhnoe, Sap-0, Rld-2, Stw-0)
	5'-UTR (F2B/R2B)	TCTGTACGCCATTGAAGAAAC	GATAAGGTTTGGGTACAGATC	
	5'-UTR (F2B/R2C)	TCTGTACGCCATTGAAGAAAC	GATAAGGTTTGGGCACAGATC	
	Coding (F3A/R3A)	AAGGTTCAATTTTGCCTCACAC	AAGGCCCAATATCAGTAGTTG	
	Coding (F3B/R3A)	AAGGTTCAATTTGCGCTCACAC	AAGGCCCAATATCAGTAGTTG	
At5g13440	At5g13440	ACAAGCCCAATTTTGTCTGAGC	ACAACAGTCCGAGTGTCTAGGT	All
	At3g30722(A)	GCCGTAGTAACCGTCAGGAA	AGACATTTTATCTGTTAAGTGG	
	At3g30722(B)	CTGTAGAAATGGGGTTTCATC	CCTCCATAGTGGCGAATCAC	
	At3g30722(C)	GTAGACTACAAGTACCAACTC	AAGAGTTGACGAGTCCGTCG	
	At3g30721(A)	CTCTGGAGCATCAATTAAGTTTG	ACTTCAAACTCCATACCTCTGAT	
At3g30721	At3g30721(A)	GAGAAACCTTCGCTTTGGTC	GGGATCAACATAGTCAACATG	Fig.3 (epiRILs), Fig. S4
	At3g30718	GTCTAGATATCCAGGGGATG	CTCTGAACATCAACATGTGC	
	At3g30724(A)	TCCTGTGCTATTGATACTCAC	GACAAAACAAAGTCTGATCGATG	
	At3g30724(B)	TCCTGTGCTATTGATACTCAC	GATGTTTTTGCAGTCAATGAAAC	
	AT3TE51160	CTTTACTTACAAAGTAGATGAGC	CTCGAGATTGACTCTTTTGAG	
At3TE51170	At3TE51170	CCGTGTGATACCCGTCCTCGT	GGTCGATAGTAATACGAGAGA	Fig. S4
	At3g30737	AGTGTCTCGAACGTTGTGTCG	GTTACAGAAGATTCCTATTTGTG	
Pyrosequencing				
QQS_pyro_F1 (PCR)		TCAAAATGAGGGTCATATCATGG		
QQS_pyro_R1-biotin (PCR)		ATTGGATACAATGGCCCTATAACT		
QQS_pyro_S1 (Pyrosequencing)		GATATTTGGGCTTATCAC		

REFERENCES

1. Cubas P, Vincent C, Coen E (1999) An epigenetic mutation responsible for natural variation in floral symmetry. *Nature* 401: 157–161.
2. Soppe WJJ, Jacobsen SE, Alonso-blanco C, Jackson JP, Kakutani T, et al. (2000) The late flowering phenotype of *fwa* mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Mol Cell* 6: 791–802.
3. Manning K, Tör M, Poole M, Hong Y, Thompson AJ, et al. (2006) A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nat Genet* 38: 948–952.
4. Henderson IR, Jacobsen SE (2008) Tandem repeats upstream of the Arabidopsis endogene *SDC* recruit non-CG DNA methylation and initiate siRNA spreading. *Genes Dev* 22: 1597–1606.
5. Miura K, Agetsuma M, Kitano H, Yoshimura A, Matsuoka M, et al. (2009) A metastable *DWARF1* epigenetic mutant affecting plant stature in rice. *Proc Natl Acad Sci USA* 106: 11218–11223.
6. Durand S, Bouché N, Strand EP, Loudet O, Camilleri C (2012) Rapid establishment of genetic incompatibility through natural epigenetic variation. *Curr Biol* 22: 326–331.
7. Bender J, Fink GR (1995) Epigenetic control of an endogenous gene family is revealed by a novel blue fluorescent mutant of Arabidopsis. *Cell* 83: 725–734.
8. Melquist S, Bender J (2003) Transcription from an upstream promoter controls methylation signaling from an inverted repeat of endogenous genes in Arabidopsis. *Genes Dev* 17: 2036–2047.
9. Molnar A, Melnyk CW, Bassett A, Hardcastle TJ, Dunn R, et al. (2010) Small silencing RNAs in plants are mobile and direct epigenetic modification in recipient cells. *Science* 328: 872–875.
10. Lippman Z, Gendrel A-V, Black M, Vaughn MW, Dedhia N, et al. (2004) Role of transposable elements in heterochromatin and epigenetic control. *Nature* 430: 471–476.
11. Kinoshita Y, Saze H, Kinoshita T, Miura A, Soppe WJJ, et al. (2006) Control of *FWA* gene silencing in *Arabidopsis thaliana* by SINE-related direct repeats. *Plant J* 49: 38–45.
12. Rangwala SH, Elumalai R, Vanier C, Ozkan H, Galbraith DW, et al. (2006) Meiotically stable natural epialleles of *Sadhu*, a novel Arabidopsis retroposon. *Plos Genet* 2: e36.
13. Martin A, Troadec C, Boualem A, Rajab M, Fernandez R, et al. (2009) A transposon-induced epigenetic change leads to sex determination in melon. *Nature* 461: 1135–1138.

14. Vaughn MW, Tanurdzić M, Lippman Z, Jiang H, Carrasquillo R, et al. (2007) Epigenetic natural variation in *Arabidopsis thaliana*. Plos Biol 5: e174.
15. Zhang X, Shiu S, Cal A, Borevitz JO (2008) Global analysis of genetic, epigenetic and transcriptional polymorphisms in *Arabidopsis thaliana* using whole genome tiling arrays. Plos Genet 4: e1000032.
16. Becker C, Hagmann J, Müller J, Koenig D, Stegle O, et al. (2011) Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. Nature 480: 245–249.
17. Schmitz RJ, Schultz MD, Lewsey MG, O'Malley RC, Urich MA, et al. (2011) Transgenerational epigenetic instability is a source of novel methylation variants. Science 334: 369–373.
18. Donoghue MTA, Keshavaiah C, Swamidatta SH, Spillane C (2011) Evolutionary origins of *Brassicaceae* specific genes in *Arabidopsis thaliana*. BMC Evol Biol 11: 47.
19. Li L, Foster C, Gan Q, Nettleton D, James MG, et al. (2009) Identification of the novel protein *QQS* as a component of the starch metabolic network in *Arabidopsis* leaves. Plant J 58: 485–498.
20. Seo PJ, Kim MJ, Ryu J-Y, Jeong E-Y, Park C-M (2011) Two splice variants of the *IDD14* transcription factor competitively form nonfunctional heterodimers which may regulate starch metabolism. Nat Commun 2: 303.
21. Beilstein MA, Nagalingum NS, Clements MD, Manchester SR, Mathews S (2010) Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. Proc Natl Acad Sci USA 107: 18724–18728.
22. Hu TT, Pattyn P, Bakker EG, Cao J, Cheng J-F, et al. (2011) The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. Nature Genet 43: 476–481.
23. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, et al. (2008) Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. Cell 133: 523–536.
24. Kurihara Y, Matsui A, Kawashima M, Kaminuma E, Ishida J, et al. (2008) Identification of the candidate genes regulated by RNA-directed DNA methylation in *Arabidopsis*. Biochem Biophys Res Commun 376: 553–557.
25. Johannes F, Porcher E, Teixeira FK, Saliba-Colombani V, Simon M, et al. (2009) Assessing the impact of transgenerational epigenetic variation on complex traits. Plos Genet 5: e1000530.
26. Teixeira FK, Heredia F, Sarazin A, Roudier F, Boccara M, et al. (2009) A role for RNAi in the selective correction of DNA methylation defects. Science 323: 1600–1604.

27. McKhann HI, Camilleri C, Bérard A, Bataillon T, David JL, et al. (2004) Nested core collections maximizing genetic diversity in *Arabidopsis thaliana*. *Plant J* 38: 193–202.
28. Cubillos FA, Yansouni J, Khalili H, Balzergue S, Elftieh S, et al. (2012) Expression variation in connected recombinant populations of *Arabidopsis thaliana* highlights distinct transcriptome architectures. *BMC Genomics* 13: 117.
29. Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV, et al. (2007) An “electronic fluorescent pictograph” browser for exploring and analyzing large-scale biological data sets. *Plos One* 2: e718.
30. Hruz T, Laule O, Szabo G, Wessendorp F, Bleuler S, et al. (2008) Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Adv Bioinformatics* 2008: 420747.
31. Paszkowski J, Grossniklaus U (2011) Selected aspects of transgenerational epigenetic inheritance and resetting in plants. *Curr Opin Plant Biol* 14: 195–203.
32. Teixeira FK, Colot V (2010) Repeat elements and the *Arabidopsis* DNA methylation landscape. *Heredity* 105: 14–23. doi:10.1038/hdy.2010.52.
33. Jacobsen SE, Meyerowitz EM (1997) Hypermethylated *SUPERMAN* epigenetic alleles in *Arabidopsis*. *Science* 277: 1100–1103.
34. Carvunis A-R, Rolland T, Wapinski I, Calderwood MA, Yildirim MA, et al. (2012) Proto-genes and *de novo* gene birth. *Nature*: doi: 10.1038/nature11184.
35. Fujimoto R, Sasaki T, Kudoh H, Taylor JM, Kakutani T, et al. (2011) Epigenetic variation in the *FWA* gene within the genus *Arabidopsis*. *Plant J* 66: 831–843.
36. Feinberg AP, Irizarry RA (2010) Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proc Natl Acad Sci USA* 107: 1757–1764.
37. Sulpice R, Pyl E, Ishihara H, Trenkamp S, Steinfath M, et al. (2009) Starch as a major integrator in the regulation of plant growth. *Proc Natl Acad Sci USA* 106: 10348–10353.
38. Simon M, Simon A, Martins F, Botran L, Tisné S, et al. (2012) DNA fingerprinting and new tools for fine-scale discrimination of *Arabidopsis thaliana* accessions. *Plant J* 69: 1094–1101.
39. Kronholm I, Loudet O, Meaux JD (2010) Influence of mutation rate on estimators of genetic differentiation - lessons from *Arabidopsis thaliana*. *BMC Genet* 11: 33.
40. Kleinboelting N, Huep G, Kloetgen A, Viehoveer P, Weisshaar B (2012) GABI-Kat Simple Search: new features of the *Arabidopsis thaliana* T-DNA mutant database. *Nucleic Acids Res* 40: D1211–D1215.

41. Woody ST, Austin-Phillips S, Amasino RM, Krysan PJ (2007) The WiscDsLox T-DNA collection : an Arabidopsis community resource generated by using an improved high-throughput T-DNA sequencing pipeline. *J Plant Res* 120: 157–165.
42. Vongs A, Kakutani T, Martienssen RA, Richards EJ (1993) *Arabidopsis thaliana* DNA methylation mutants. *Science* 260: 1926-1928.
43. Xie Z, Johansen LK, Gustafson AM, Kasschau KD, Lellis AD, et al. (2004) Genetic and functional diversification of small RNA pathways in plants. *PloS Biol* 2: E104.
44. Oñate-Sánchez L, Vicente-Carbajosa J (2008) DNA-free RNA isolation protocols for *Arabidopsis thaliana*, including seeds and siliques. *BMC Res Notes* 1: 93.
45. Wittkopp PJ, Haerum BK, Clark AG (2004) Evolutionary changes in *cis* and *trans* gene regulation. *Nature* 430: 85–88.
46. Zhang X, Richards EJ, Borevitz JO (2007) Genetic and epigenetic dissection of *cis* regulatory variation. *Curr Opin Plant Biol* 10: 142–148.
47. Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, et al. (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* 452: 215–219.
48. Fahlgren N, Sullivan CM, Kasschau KD, Chapman EJ, Cumbie JS, et al. (2009) Computational and analytical framework for small RNA profiling by high-throughput sequencing. *RNA* 15: 992–1002.
49. Gregory BD, O'Malley RC, Lister R, Urich M a, Tonti-Filippini J, et al. (2008) A link between RNA metabolism and silencing affecting Arabidopsis development. *Dev Cell* 14: 854–866.

SI REFERENCES

1. Nei M, Gojoborit T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3: 418-426.
2. Seifert M, et al. (2009) Array-based genome comparison of Arabidopsis ecotypes using hidden Markov models. *Proceedings of the 2nd International Conference on Bio-inspired systems and signal processing* 3-11.
3. Cortijo S, Wardenaar R, Colome-Tatche M, Johannes F and Colot V (2012) in *Plant Epigenome: Understanding and Analysis*, eds McKeown P, Spillane C (Springer, New York), *In press*

4. Martin-Magniette ML, Mary-Huard T, Berard C, Robin S (2008) ChIPmix: mixture model of regressions for two-color ChIPchip analysis. *Bioinformatics* 24: I181-I186.
5. Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, and Jacobsen SE (2008) Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* 452: 215-219.
6. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, and Ecker JR (2008) Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* 133: 523-536.
7. Moghaddam A.B, Roudier F, Seifert M, Berard C, Magniette MLM, Ashtiyani RK, Houben A, Colot V, Mette MF (2011) Additive inheritance of histone modifications in *Arabidopsis thaliana* intra-specific hybrids. *Plant J* 67: 691-700.

PARTE III

3.3 - Análise preliminar dos mecanismos de controle envolvidos no silenciamento transcricional de *QQS*

3.3.1 - *QQS* está sob um controle epigenético diferenciado do previamente descrito para os genes *FWA* e *SDC* de *A. thaliana*

Em diversos *loci* do genoma *MET1*, *CMT3* e *DRM2* trabalham cooperativamente ou em competição, frequentemente possuindo efeitos sobrepostos sobre metilação CG, CHG e CHH. Em *A. thaliana* os mecanismos de controle epigenético que atuam sobre os genes *FWA* e *SDC* foram bem caracterizados. *QQS* possui uma estrutura semelhante à destes genes, caracterizada pela presença de sequências repetidas na porção 5' do gene associada a metilação de DNA nos três contextos de sequência. Entretanto, uma comparação dos efeitos de mutação em *MET1* e no triplo mutante para *DRM1*, *DRM2* e *CMT3* (*ddc*) indicou que *QQS* está sob um controle diferenciado do previamente descrito para *FWA* e *SDC* (Lister et al. 2008). Silenciamento de *FWA* é exclusivamente dependente de metilação do tipo CG e da maquinaria de manutenção, uma vez que este gene possui poucos sítios CHG e CHH metilados e perda de metilação neste contexto não tem efeito sobre a expressão deste gene. Em contrapartida, em *met1*, perda de metilação CG induz forte expressão de *FWA* (Figura 14). Diferentemente de *FWA*, silenciamento transcricional de *QQS* e *SDC* é dependente tanto da maquinaria de metilação *de novo* quanto da maquinaria de manutenção. Em *ddc*, tanto *QQS* quanto *SDC* apresentam perda de metilação do tipo CHG e CHH, o que resulta em ativação transcricional mesmo na presença de metilações do tipo CG (Figura 14). No entanto, apesar de ambos os genes estarem transcricionalmente ativos em *met1*, *MET1* atua de maneira diferenciada nestes genes. Em *met1*, *SDC* perde metilações do tipo CG, mas mantém metilações do tipo CHG e CHH. Em contrapartida, mutações em *met1* induzem a perda completa de metilação nos três contextos de sequência em *QQS*, sugerindo que para este gene a manutenção de um esqueleto de metilação CG seja importante ao recrutamento da maquinaria de *RdDM* e *CMT3* (Figura 14). De maneira similar a mutações em *MET1*, mutações simples em *DDM1* ou em conjunto com *RDR2* também induzem a perda total de

metilação em *QQS* e reativação transcricional (Figura 15). É possível que a necessidade de manutenção de um perfil de metilações CG para recrutamento de proteínas efetoras da via de RdDM ao locus *QQS* se deva ao envolvimento de *RDM1* no controle transcricional deste gene. *RDM1* é uma proteína recentemente identificada como efetora da via de RdDM. *RDM1* parece se ligar a DNA hemimetilado e pode se associar a *DRM2* e *AGO4* no núcleo celular. Desta maneira, acredita-se que *RDM1* possa ser o elo que ligue a produção de siRNAs com metilação pré-existente ou metilação *de novo* (Gao et al. 2010).

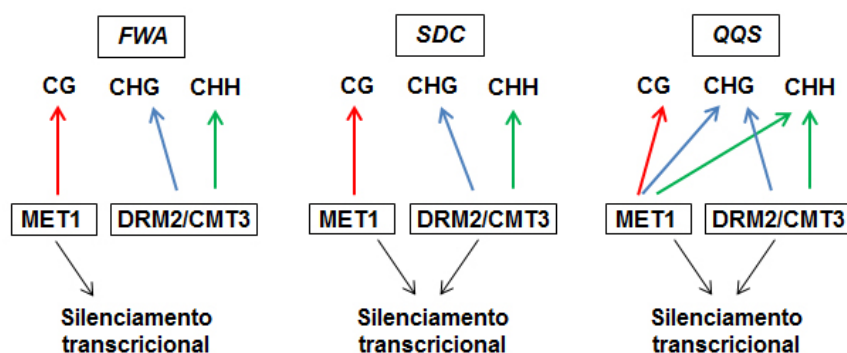


Figura 14. *QQS*, *FWA* e *SDC* são caracterizados por sistemas de silenciamento transcricional diferenciados (Lister et al. 2008).

3.3.2 – RdDM em *QQS* envolve a participação de *DCL2*, *DCL3* e *DCL4*

RdDM em *FWA* e *SDC*, assim como para grande parte dos transposons e sequências repetidas do genoma, depende da atuação de *Pol IV*, *RDR2* e *DCL3* para biogênese de siRNAs heterocromáticos de 24nt e de *AGO4* e *Pol V* para o recrutamento de *DRM2* a sequência de DNA a ser metilada, assim como para amplificar o sinal de siRNAs (Figura 6).

De acordo com a atuação de RdDM no silenciamento transcricional de *QQS*, siRNAs de 24nt representam a classe de pequenos RNAs mais representativa neste locus (Figura 1B - Parte II). Além disto, mutações em *RDR2* induzem ativação transcricional de *QQS* devido à perda de metilação em sua sequência promotora e 5'UTR (Figura 15; Kurihara et al. 2008). No entanto, surpreendentemente, mutações em *DCL3* ou *AGO4* parecem não afetar o padrão de metilação e

expressão de *QQS* (Figura 15). Em contrapartida, mutações em *DCL2* e *DCL4* alteram significativamente o padrão de metilação e expressão deste gene, indicando um papel direto destas proteínas no silenciamento transcricional de *QQS*. De acordo com esta observação, siRNAs de 21 e 22-23nt cognatos de processamento por *DCL4* e *DCL2*, respectivamente, são observados na sequência promotora e 5'UTR deste gene (Figure 1B - Parte II). Apesar do efeito marcante de *DCL2* e *DCL4* sobre a metilação e expressão de *QQS*, curiosamente, mutações duplas para *dcl2 dcl3*, *dcl2 dcl4* e *dcl3 dcl4* parecem não afetar a metilação e o silenciamento transcricional deste gene, indicando que *DCL3* também possui um papel no controle epigenético de *QQS* e que possivelmente *DCL2*, *DCL3* e *DCL4* possam interagir no processamento do transcrito dupla fita produzido a partir deste locus.

DCL2 e *DCL4* normalmente atuam no silenciamento pós-transcricional de RNAs virais. *DCL2* também é conhecida pelo seu papel na biogênese de *nat*-siRNAs (*natural antisense* siRNAs) e *DCL4*, pelo seu papel na biogênese de *ta*-siRNAs (*trans acting* siRNAs; Jamalkandi & Masoudi-Nejad 2009). No entanto, algumas linhas de evidência já indicavam um possível papel redundante e/ou compensatório de *DCL2* e *DCL4* na manutenção de marcas epigenéticas com *DCL3*. *DCL3* representa a principal proteína envolvida com processamento de siRNAs dependentes de *RDR2*. No entanto, Gasciolli et al. 2005 mostrou que na ausência de *DCL3* transcritos dupla fita gerados a partir de *RDR2* podem ser processados por *DCL2* e *DCL4*. De acordo com esta observação, silenciamento transcricional de *QQS* não é dependente de *RDR6* (*RNA DEPENDENT RNA POLYMERASE 6*), que atua nas vias clássicas de processamento envolvendo *DCL2* e *DCL4* (Figura 16; Jamalkandi & Masoudi-Nejad 2009). Neste mesmo trabalho também foi observado que, apesar do mutante *dcl3* apresentar um desenvolvimento normal, duplos mutantes *dcl2 dcl3* e *dcl3 dcl4* exibem uma série de fenótipos desenvolvimentais estocásticos que se assemelham aos fenótipos observados ao longo de gerações de autofecundação de *ddm1*, sugerindo que estas alterações possam ser resultantes de perda de marcas epigenéticas. Além disto, Henderson et al. 2006 mostrou que mutação em *dcl3* possui um efeito fraco sobre o processo de metilação *de novo* de *FWA* e que em mutantes duplos *dcl2 dcl3* ou *dcl3 dcl4* o efeito observado é semelhante ao de *dcl3*. No entanto, no triplo mutante *dcl2 dcl3 dcl4*, falha no silenciamento de *FWA* resulta no fenótipo característico de atraso de florescimento

correspondente a expressão constitutiva deste gene, o que demonstra o papel redundante de *DCL2*, *DCL3* e *DCL4* na metilação *de novo* de *FWA* (Henderson et al. 2006).

Baseando-se nestas informações e no modelo de processamento hierárquico em repetições invertidas endógenas sugerido por Dunoyer et al. 2010, propusemos um modelo genético para explicar o mecanismo da via de RdDM que atua sobre o gene *QQS*. Por este modelo, o padrão de metilação de *QQS* seria hierarquicamente controlado por *DCL2*, *DCL3* e *DCL4*. Metilação seria principalmente controlada por *DCL2* e *DCL4* que seriam capazes de reprimir a atividade de *DCL3*. *DCL3* por sua vez, apesar de não ativa na presença de *DCL2* e *DCL4*, também teria um papel inibitório sobre estas enzimas, possivelmente pela competição por cofatores como proteínas ligantes de RNA dupla fita (Dunoyer et al. 2010; Figura 16). Sendo assim, em *dcl2* perda de metilação é observada uma vez que a atividade inibitória de *DCL3* não permite que *DCL4* compense a deficiência de *DCL2*, o mesmo acontecendo para o mutante *dcl4*. No entanto, no mutante duplo *dcl2 dcl3* ou *dcl3 dcl4*, devido a ausência de *DCL3*, *DCL4* e *DCL2*, respectivamente, seriam capazes de manter o padrão selvagem de metilação e expressão de *QQS*. Em contrapartida, na ausência de *DCL2* e *DCL4* no mutante duplo *dcl2 dcl4*, *DCL3* seria capaz de manter o silenciamento transcricional de *QQS* (Figura 16).

3.3.3 – RdDM em *QQS* envolve *Pol II* ao invés de *Pol IV* e *V*

Além do processamento diferencial por *DCL2*, *DCL3* e *DCL4*, observamos que a produção de siRNAs em *QQS* é independente de *Pol IV* e *Pol V*, as duas RNA polimerases de plantas especializadas em silenciamento transcricional que evoluíram a partir de *Pol II*, o que está de acordo com os dados de Kurihara et al. 2008 que também identifica *QQS* como um locus para o qual o padrão de metilação é independente de *Pol IV*. Indução da expressão de *QQS* em um mutante hipomórfico para a segunda subunidade maior de *Pol II* (Zheng et al. 2009) sugere que no caso de *QQS* a via de RdDM possa ser dependente de *Pol II*. Em alguns loci do genoma transcritos não codificantes produzidos por *Pol II* são requeridos para silenciamento transcricional, atuando como moldes para recrutar os efetores da via de RdDM a sequência de

DNA alvo a ser metilada, incluindo *AGO4*, *DRM2*, *Pol IV* e *Pol V* (Zheng et al. 2009). No entanto, no caso de *QQS Pol II* parece atuar sozinha na via de RdDM, substituindo completamente o papel de *Pol IV* e *Pol V*. De acordo com a atuação de *Pol II* no silenciamento transcricional de *QQS*, *RDM1* que parece fazer o link entre a metilação do tipo CG em *QQS* e o recrutamento da maquinaria de RdDM, é capaz de interagir com *Pol II* (Gao et al. 2010).

Em conclusão, estes resultados, apesar de preliminares, ressaltam a versatilidade das vias de controle epigenético envolvendo siRNAs em plantas e, além de confirmarem o papel redundante de *DCL2*, *DCL3* e *DCL4* no processamento de siRNAs heterocromáticos relacionados a silenciamento transcricional, apontam para uma via de RdDM dependente exclusivamente de *Pol II* ainda não descrita. Caracterização do padrão de metilação e expressão de *QQS* em mutantes para outras proteínas envolvidas na via de RdDM em *Arabidopsis*, análise do padrão de metilação contexto de sequência específico através de sequenciamento bisulfito e identificação do transcrito que desencadeia a produção de RNA dupla fita para entrada na via serão importantes para elucidar de maneira clara a via de silenciamento que atua sobre este gene.

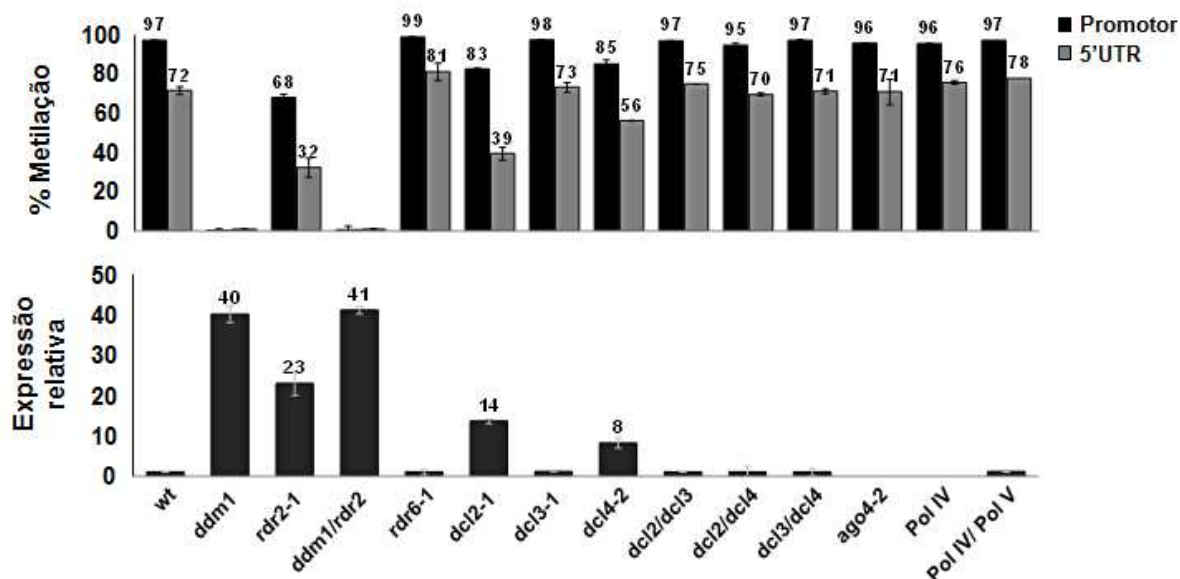


Figura 15. Padrão de metilação (McrBC-qPCR) e expressão (qRT-PCR) de *QQS* em diversos mutantes para proteínas da maquinaria de silenciamento.

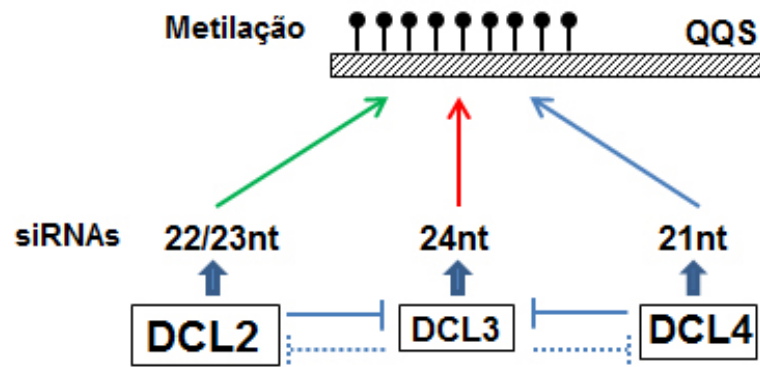


Figura 16. Modelo genético para a manutenção do padrão de metilação de *QQS* via RdDM envolvendo a atuação de *DCL2*, *DCL3* e *DCL4*. O padrão de metilação de *QQS* é mantido hierarquicamente por *DCL2* e *DCL4*, seguido de *DCL3*. *DCL2* e *DCL4* são capazes de reprimir a atividade de *DCL3* e *DCL3* é capaz de reprimir a atividade de *DCL2* e *DCL4*.

4. CONSIDERAÇÕES FINAIS

QQS é um gene novo que evoluiu em uma região heterocromática do genoma de *A. thaliana*. É possível que a atividade de elementos de transposição na vizinhança de *QQS* possa ter contribuído para o controle epigenético de sua transcrição e singularidade de sua sequência. Diferentemente de outros loci epimutáveis descritos em *A. thaliana* como *FWA* e *SDC* e até mesmo de outros epialelos descritos em plantas que normalmente assumem somente um estado epigenético metilado, *QQS* representa o primeiro caso de um gene bastante propenso a variar epigeneticamente e que pode assumir uma gama de estados epialélicos diferenciados com consequências diretas sobre seu padrão de expressão. Variação epialélica em *QQS* ocorre independentemente de variação genética e não parece ser influenciada pelo estado epigenético de transposons adjacentes, embora um possível efeito sob condições de estresse onde existe relaxamento da cromatina e aumento da expressão destes elementos de transposição não possa ser descartado. Acreditamos que esta tendência à variação epialélica seja resultante, em parte, do mecanismo de controle epigenético diferenciado que regula a expressão deste gene, provavelmente mais flexível e susceptível a variações estocásticas relacionadas a erros na maquinaria de manutenção ou até mesmo a variação em condições ambientais. Embora epialelos contrastantes de *QQS* (totalmente metilado vs. demetilado) sejam estavelmente herdados ao longo das gerações de autofecundação, a heritabilidade de estados epigenéticos intermediários precisa ser melhor investigada de maneira a elucidarmos a dinâmica de formação epialélica para este gene. É possível que os diversos estados epialélicos apresentados por *QQS* representem diferentes eventos de perda de metilação. Alternativamente, demetilação ocasional seguida de perda gradual de metilação e transição entre diferentes estados epialélicos também seria possível uma vez que diferentes epialelos de *QQS* em Col-0*, apesar de estavelmente herdados, apresentam uma leve tendência à perda de metilação e expressão. Independentemente do cenário em questão, *QQS* é um gene claramente sob um controle epigenético dinâmico, o que associado ao fato deste gene ter implicações sobre o metabolismo do amido, sugere que o surgimento de variantes epialélicas de *QQS* em populações naturais possa contribuir para a geração de

plasticidade fenotípica, aumentando a capacidade de adaptação a mudanças nas condições ambientais.

Em conclusão, em função de seu controle epigenético flexível e propensão à variação epialélica, sugerimos o uso *QQS* como um novo “sensor” molecular para identificação de condições ambientais capazes de influenciar a maquinaria de controle epigenético ou ainda, novos reguladores importantes ao silenciamento epigenético em plantas. O estudo de *QQS* também fornece informações valiosas a respeito das potenciais implicações ecológicas e evolutivas de processos epigenéticos e que provavelmente serão importantes para uma incorporação realista de variação epigenética na teoria evolucionária.

5. PERSPECTIVAS

- Análise detalhada das vias de controle epigenético que atuam sobre *QQS* para esclarecer a predisposição a variação epialélica neste locus.
- Investigação do provável processo de formação epialélica em *QQS* através da análise detalhada da herança transgeneracional de epialelos de *QQS* identificados em Col-0* e identificação de condições ambientais capazes de induzir variação epialélica neste locus.
- Entendimento de qual a relação entre controle epigenético de *QQS* seu padrão de expressão temporal e espacial ao longo do desenvolvimento.
- Fenotipagem de diferentes epialelos de *QQS* em diversas condições de crescimento e estresses.
- Análise da frequência de variação epialélica em *QQS* em um número maior de indivíduos coletados em populações naturais.
- Avaliação da possível relação entre controle epigenético e o processo de evolução de genes novos.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- Aceituno FF, Moseyko N, Rhee SY, and Gutiérrez RA. 2008. The rules of gene expression in plants: organ identity and gene body methylation are key factors for regulation of gene expression in *Arabidopsis thaliana*. *BMC Genomics* **9**: 438.
- Becker C, Hagmann J, Müller J, Koenig D, Stegle O, Borgwardt K, and Weigel D. 2011. Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. *Nature* **480**: 245–249.
- Bender J, and Fink GR. 1995. Epigenetic control of an endogenous gene family is revealed by a novel blue fluorescent mutant of *Arabidopsis*. *Cell* **83**: 725–734.
- Bird A. 2007. Perceptions of epigenetics. *Nature* **447**: 396–398.
- Bossdorf O, Richards CL, and Pigliucci M. 2008. Epigenetics for ecologists. *Ecology Letters* **11**: 106–115.
- Cai J, Zhao R, Jiang Huifeng, and Wang W. 2008. *De novo* origination of a new protein-coding gene in *Saccharomyces cerevisiae*. *Genetics* **179**: 487–496.
- Carvunis A-R, Rolland T, Wapinski I, Calderwood MA, Yildirim MA, Simonis N, Charlotiaux B, Hidalgo CA, Barbette J, Santhanam B, et al. 2012. Proto-genes and *de novo* gene birth. *Nature* **487**: 370–374.
- Chen S, Zhang YE, and Long M. 2010. New genes in *Drosophila* quickly become essential. *Science* **330**: 1682–1685.
- Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, and Jacobsen SE. 2008. Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* **452**: 215–219.
- Cubas P, Vincent C, and Coen E. 1999. An epigenetic mutation responsible for natural variation in floral symmetry. *Nature* **401**: 157–161.
- Dolinoy DC. 2008. The *agouti* model: an epigenetic biosensor for nutritional and environmental alterations on the fetal epigenome. *Nutrition Reviews* **66**: S7–S11.
- Donoghue MTA, Keshavaiah C, Swamidatta SH, and Spillane C. 2011. Evolutionary origins of Brassicaceae specific genes in *Arabidopsis thaliana*. *BMC Evol Biol* **11**: 47.

- Dorweiler JE, Carey CC, Kubo KM, Hollick JB, Kermicle JL, and Chandler VL. 2000. *Mediator of paramutation1* is required for establishment and maintenance of paramutation at multiple maize loci. *Plant Cell* **12**: 2101–2118.
- Dunoyer P, Brosnan CA, Schott G, Wang Y, Jay F, Alioua A, Himber C, and Voinnet O. 2010. An endogenous, systemic RNAi pathway in plants. *EMBO J* **29**: 1699–1712.
- Durand S, Bouché N, Strand EP, Loudet O, and Camilleri C. 2012. Rapid establishment of genetic incompatibility through natural epigenetic variation. *Curr Biol* **22**: 326–331.
- Feinberg AP, and Irizarry RA. 2010. Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proc Natl Acad Sci USA* **107**: 1757–1764.
- Feng S, Cokus SJ, Zhang X, Chen P, Bostick M, Goll Mary G, Hetzel J, Jain J, Strauss SH, Halpern ME, et al. 2010. Conservation and divergence of methylation patterning in plants and animals. *Proc Natl Acad Sci USA* **107**: 8689–8694.
- Feng S, and Jacobsen SE. 2011. Epigenetic modifications in plants: an evolutionary perspective. *Curr Opin Plant Biol* **14**: 1–8.
- Gao Z, Liu Hai-liang, Daxinger L, Pontes O, He X, Qian W, Lin H, Xie M, Lorkovic ZJ, Zhang S, et al. 2010. An *RNA polymerase II*- and *AGO4*-associated protein acts in RNA-directed DNA methylation. *Nature* **465**: 106–109.
- Gascioli V, Mallory AC, Bartel DP, and Vaucheret H. 2005. Partially redundant functions of Arabidopsis DICER-like enzymes and a role for *DCL4* in producing trans-acting siRNAs. *Curr Biol* **15**: 1–7.
- Goll MG, and Bestor TH. 2005. Eukaryotic cytosine methyltransferases. *Annu Rev Biochem* **74**: 481–514.
- Gollery M, Harper J, Cushman J, Mittler T, Girke T, Zhu J, Bailey-serres J, and Mittler R. 2006. What makes species unique? The contribution of proteins with obscure features. *Genome Biol* **7**: R57.
- Grewal SI. 2000. Transcriptional silencing in fission yeast. *J Cell Physiol* **184**: 311–318.
- He X-J, Chen T, and Zhu J-K. 2011. Regulation and function of DNA methylation in plants and animals. *Cell Res* **21**: 442–465.
- Heinen TJAJ, Staubach F, Haming D, and Tautz D. 2009. Emergence of a new gene from an intergenic region. *Curr Biol* **19**: 1527–1531.

- Henderson IR, Zhang X, Lu C, Johnson L, Meyers BC, Green PJ, and Jacobsen SE. 2006. Dissecting *Arabidopsis thaliana* DICER function in small RNA processing , gene silencing and DNA methylation patterning. *Nat Genet* **38**: 721–725.
- Henderson IR, and Jacobsen SE. 2007. Epigenetic inheritance in plants. *Nature* **447**: 418–424.
- Henderson IR, and Jacobsen SE. 2008. Tandem repeats upstream of the *Arabidopsis* endogene *SDC* recruit non-CG DNA methylation and initiate siRNA spreading. *Genes Dev* **22**: 1597–1606.
- Ito H, Gaubert H, Bucher E, Mirouze M, Vaillant I, and Paszkowski J. 2011. An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature* **472**: 115–119.
- Jacobsen SE, and Meyerowitz EM. 1997. Hypermethylated *SUPERMAN* epigenetic alleles in *Arabidopsis*. *Science* **277**: 1100–1103.
- Jamalkandi SA, and Masoudi-Nejad A. 2009. Reconstruction of *Arabidopsis thaliana* fully integrated small RNA pathway. *Funct Integr Genomics* **9**: 419–432.
- Jia Y, Lisch DR, Ohtsu K, Scanlon MJ, Nettleton Daniel, and Schnable PS. 2009. Loss of *RNA-dependent RNA polymerase 2 (RDR2)* function causes widespread and unexpected changes in the expression of transposons, genes, and 24-nt small RNAs. *PLos Genet* **5**: e1000737.
- Johannes F, Porcher E, Teixeira FK, Saliba-Colombani V, Simon M, Agier N, Bulski A, Albuissou J, Heredia F, Audigier P, et al. 2009. Assessing the impact of transgenerational epigenetic variation on complex traits. *Plos Genet* **5**: e1000530.
- Kakutani T. 2002. Epi-alleles in plants : inheritance of epigenetic information over generations. *Plant Cell Physiol* **43**: 1106–1111.
- Kalisz S, and Puruggana MD. 2004. Epialleles via DNA methylation: consequences for plant evolution. *Trends Ecol Evol* **19**: 309–314.
- Khalturin K, Hemmrich G, Fraune S, and Bosch TCG. 2009. More than just orphans : are taxonomically-restricted genes important in evolution ? *Trends in Genetics* **25**: 404–413.
- Kinoshita T, and Jacobsen SE. 2012. Opening the door to epigenetics in PCP. *Plant Cell Physiol* **53**: 763–765.
- Kinoshita Y, Saze H, Kinoshita T, Miura A, Soppe WJJ, Koornneef M, and Kakutani T. 2006. Control of *FWA* gene silencing in *Arabidopsis thaliana* by SINE-related direct repeats. *Plant J* **49**: 38–45.

- Kurihara Y, Matsui A, Kawashima M, Kaminuma E, Ishida J, Morosawa T, Mochizuki Y, Kobayashi N, Toyoda T, Shinozaki K, et al. 2008. Identification of the candidate genes regulated by RNA-directed DNA methylation in Arabidopsis. *Biochem Biophys Res Commun* **376**: 553–557.
- Law JA, and Jacobsen SE. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Net Rev Gent* **11**: 204–220.
- Li L, Foster C, Gan Q, Nettleton Dan, James MG, Myers AM, and Wurtele ES. 2009. Identification of the novel protein *QQS* as a component of the starch metabolic network in Arabidopsis leaves. *Plant J* **58**: 485–498.
- Lira-Medeiros CF, Parisod C, Fernandes RA, Mata CS, Cardoso MA, and Cavalcanti PGF. 2010. Epigenetic variation in mangrove plants occurring in contrasting natural environment. *Plos One* **5**: e10326.
- Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, and Ecker JR. 2008. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* **133**: 523–536.
- Manning K, Tör M, Poole M, Hong Y, Thompson AJ, King GJ, Giovannoni JJ, and Seymour GB. 2006. A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nat Genet* **38**: 948–952.
- Martin A, Troadec C, Boualem A, Rajab M, Fernandez R, Morin H, Pitrat M, Dogimont C, and Bendahmane A. 2009. A transposon-induced epigenetic change leads to sex determination in melon. *Nature* **461**: 1135–1138.
- Melquist S, and Bender J. 2003. Transcription from an upstream promoter controls methylation signaling from an inverted repeat of endogenous genes in Arabidopsis. *Genes Dev* **17**: 2036–2047.
- Meyer P. 2011. DNA methylation systems and targets in plants. *FEBS Letters* **585**: 2008–2015.
- Miura K, Agetsuma M, Kitano H, Yoshimura A, Matsuoka M, Jacobsen SE, and Ashikari M. 2009. A metastable *DWARF1* epigenetic mutant affecting plant stature in rice. *Proc Natl Acad Sci USA* **106**: 11218–11223.
- Okano M, Bell DW, Haber DA, and Li E. 1999. DNA methyltransferases *Dnmt3a* and *Dnmt3b* are essential for *de novo* methylation and mammalian development. *Cell* **99**: 247–257.
- Paszkowski J, and Grossniklaus U. 2011. Selected aspects of transgenerational epigenetic inheritance and resetting in plants. *Curr Opin Plant Biol* **14**: 195–203.

- Payer B, Lee JT, and Namekawa SH. 2011. X-inactivation and X-reactivation: epigenetic hallmarks of mammalian reproduction and pluripotent stem cells. *Hum Genet* **130**: 265–280.
- Queitsch C, Sangster T a, and Lindquist S. 2002. *Hsp90* as a capacitor of phenotypic variation. *Nature* **417**: 618–624.
- Reinders J, and Paszkowski J. 2009. Unlocking the Arabidopsis epigenome. *Epigenetics* **4**: 1–7.
- Reinders J, Wulff BBH, Mirouze M, Marí-Ordóñez A, Dapp M, Rozhon W, Bucher E, Theiler G, Paskowski J, and Richards Eric J. 2009. Compromised stability of DNA methylation and transposon immobilization in mosaic Arabidopsis epigenomes. *Genes Dev* **23**: 939–950.
- Richards E J. 1997. DNA methylation and plant development. *Trends Genet* **13**: 319–323.
- Richards E J. 2008. Population epigenetics. *Curr Opin Genet Dev* **18**: 221–226.
- Roux F, Colomé-tatché M, Edelist C, Wardenaar R, Guerche P, Hospital F, Colot V, Jansen RC, and Johannes F. 2011. Genome-wide epigenetic perturbation jump-starts patterns of heritable variation found in nature. *Genetics* **188**: 1015–1017.
- Russo VEA, Martienssen RA, and Riggs AD. 1996. (eds) Epigenetic mechanisms of gene regulation. *Cold Spring Harbour Laboratory Press, Woodbury*.
- Schmitz RJ, Schultz MD, Lewsey MG, O'Malley RC, Urich MA, Libiger O, Schork NJ, and Ecker JR. 2011. Transgenerational epigenetic instability is a source of novel methylation variants. *Science* **334**: 369–373.
- Schmitz RJ, and Ecker JR. 2012. Epigenetic and epigenomic variation in *Arabidopsis thaliana*. *Trends Plant Sci* **17**: 149–154.
- Schotta G, Ebert A, Dorn R, and Reuter G. 2003. Position-effect variegation and the genetic dissection of chromatin regulation in *Drosophila*. *Semin Cell Dev Biol* **14**: 67–75.
- Scott RJ, and Spielman M. 2004. Epigenetics: imprinting in plants and mammals - the same but different? *Curr Biol* **14**: R201– R203.
- Seo PJ, Kim MJ, Ryu J-Y, Jeong E-Y, and Park C-M. 2011. Two splice variants of the *IDD14* transcription factor competitively form nonfunctional heterodimers which may regulate starch metabolism. *Nat Commun* **2**: 303.
- Sharma S, Kelly TK, and Jones P a. 2010. Epigenetics in cancer. *Carcinogenesis* **31**: 27–36.

- Silveira AB, Gauer L, Tomaz JP, Cardoso PR, Carmello-Guerreiro S, and Vincentz M. 2007. The Arabidopsis *AtbZIP9* protein fused to the VP16 transcriptional activation domain alters leaf and vascular development. *Plant Science* **172**: 1148–1156.
- Slotkin RK, and Martienssen R. 2007. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* **8**: 272–285.
- Soppe WJJ, Jacobsen SE, Alonso-blanco C, Jackson JP, Kakutani T, Koornneef M, and Peeters AJM. 2000. The late flowering phenotype of *fwa* mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Mol Cell* **6**: 791–802.
- Tautz D, Domazet-Lošo T. 2011. The evolutionary origin of orphan genes. *Nat Rev Genet* **12**: 692–702.
- Teixeira FK, Heredia F, Sarazin A, Roudier F, Boccara M, Ciaudo C, Cruaud C, Poulain J, Berdasco M, Fraga MF, et al. 2009. A role for RNAi in the selective correction of DNA methylation defects. *Science* **323**: 1600–1604.
- Teixeira FK, and Colot V. 2009. Gene body DNA methylation in plants: a means to an end or an end to a means? *EMBO J* **28**: 977–998.
- Teixeira FK, and Colot V. 2010. Repeat elements and the Arabidopsis DNA methylation landscape. *Heredity* **105**: 14–23.
- Toll-Riera M, Bosch N, Bellora N, Castelo R, Armengol L, Estivill X, and Albà MM. 2009. Origin of primate orphan genes: a comparative genomics approach. *Mol Biol Evol* **26**: 603–612.
- Vaughn MW, Tanurdzić M, Lippman Z, Jiang Hongmei, Carrasquillo R, Rabinowicz PD, Dedhia N, McCombie WR, Agier N, Bulski A, et al. 2007. Epigenetic natural variation in *Arabidopsis thaliana*. *PLoS Biol* **5**: e174.
- Waddington C. 1942. The epigenotype. *Endeavor* **1**: 18–20.
- Xiao W, Liu Hongbo, Li Y, Li X, Xu C, Long M, and Wang S. 2009. A rice gene of *de novo* origin negatively regulates pathogen-induced defense response. *Plos One* **4**: e4603.
- Zemach A, Mcdaniel IE, Silva P, and Zilberman D. 2010. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* **328**: 916–919.
- Zhang J. 2003. Evolution by gene duplication: an update. *Trends Eco Evol* **18**: 292–298.

- Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, et al. 2006. Genome-wide high-resolution mapping and functional analysis of DNA methylation in Arabidopsis. *Cell* **126**: 1189–1201.
- Zhang X. 2008. The epigenetic landscape of plants. *Science* **320**: 489–492.
- Zhao C, Craig JC, Petzold HE, Dickerman AW, and Beers EP. 2005. The xylem and phloem transcriptomes from secondary tissues of the Arabidopsis root-Hypocotyl 1. *Plant Physiol* **138**: 803–818.
- Zheng B, Wang Z, Li S, Yu B, Liu J, and Chen X. 2009. Intergenic transcription by *RNA Polymerase II* coordinates *Pol IV* and *Pol V* in siRNA-directed transcriptional gene silencing in Arabidopsis. *Genes Dev.*
- Zilberman D, Gehring M, Tran RK, Ballinger T, and Henikoff S. 2007. Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nat Genet* **39**: 61–69.

DECLARAÇÃO

Declaro para os devidos fins que o conteúdo de minha tese de Doutorado intitulada "Qua-Quine Starch de *Arabidopsis thaliana*, um gene novo regulado por metilação de DNA e propenso a variação epialélica"

() não se enquadra no § 3º do Artigo 1º da Informação CCPG 01/08, referente a bioética e biossegurança.

Tem autorização da(s) seguinte(s) Comissão(ões):

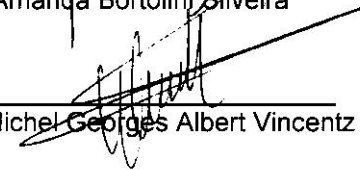
(x) CIBio – Comissão Interna de Biossegurança, projeto No. 2003/01, Instituição: IB.

() CEUA – Comissão de Ética no Uso de Animais, projeto No. _____, Instituição: _____

() CEP - Comissão de Ética em Pesquisa, protocolo No. _____, Instituição: _____

** Caso a Comissão seja externa ao IB/UNICAMP, anexar o comprovante de autorização dada ao trabalho. Se a autorização não tiver sido dada diretamente ao trabalho de tese ou dissertação, deverá ser anexado também um comprovante do vínculo do trabalho do aluno com o que constar no documento de autorização apresentado.*

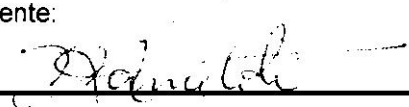

Aluno: Amanda Bortolini Silveira


Orientador: Michel Georges Albert Vincentz

Para uso da Comissão ou Comitê pertinente:

(x) Deferido () Indeferido

Carimbo e assinatura


Prof. Dr. MARCELO LANCELOTTI
Presidente da Comissão Interna de Biossegurança
Instituto de Biologia - UNICAMP

Para uso da Comissão ou Comitê pertinente:

() Deferido () Indeferido

Carimbo e assinatura



DECLARAÇÃO

Declaro, para os devidos fins, que o projeto CIBio/IB-2003/01, intitulado “Análise funcional de fatores de regulação da transcrição do tipo bZIP de angiospermas”, cujo pesquisador responsável é o Prof. Dr. MICHEL GEORGES ALBERT VINCENTZ, encontra-se devidamente aprovado e regularizado junto a CIBio/IB-UNICAMP e a CTNBio, conforme legislação vigente.

Cidade Universitária “Zeferino Vaz”,
28 de junho de 2012.


Prof. Dr. MARCELO LANCELLOTTI
Presidente da CIBio
Instituto de Biologia - UNICAMP