UNIVERSIDADE FEDERAL DE PELOTAS

Programa de Pós-Graduação em Biotecnologia

Tese

ANÁLISE EVOLUTIVA DE GENES DE HOMEOSTASE DE FERRO E DE ELEMENTOS REPETITIVOS EM ESPÉCIES MODELO.

Filipe de Carvalho Victoria

Pelotas, 2011

FILIPE DE CARVALHO VICTORIA

ANÁLISE EVOLUTIVA DE GENES DE HOMEOSTASE DE FERRO E DE ELEMENTOS REPETITIVOS EM ESPÉCIES MODELO.

Tese apresentada ao Programa de Pós-Graduação em Biotecnologia da Universidade Federal de Pelotas, como requisito parcial à obtenção do título de Doutor em Ciências (área do conhecimento: Melhoramento Vegetal).

Orientador: Antônio Costa de Oliveira

Pelotas, 2011

Dados de catalogação na fonte:
Ubirajara Buddin Cruz – CRB 10/901
Biblioteca de Ciência & Tecnologia - UFPel

V645a Victoria, Filipe de Carvalho
Análise evolutiva de genes de homeostase de ferro e de elementos repetitivos em espécies modelo / Filipe de Carvalho Victoria. – 154f. : il. color. – Tese (Doutorado). Programa de Pós-Graduação em Biotecnologia. Universidade Federal de Pelotas. Centro de Desenvolvimento Tecnológico, 2011. – Orientador Antonio Costa de Oliveira.

1.Biotecnologia. 2. *Physcomitrella patens* Brid. 3. Gene families. 4.*Polytrichum juniperinum* Hedw.. 5.Microssatélites. 6.Famílias gênicas. 7.Inferência bayesiana I.Oliveira, Antonio Costa de. II.Título.

CDD: 581.151

Banca examinadora:

Prof. Dra. Adriana Silva Hemmerly, Universidade Federal do Rio de Janeiro

Prof. Dra. Rosa Lia Barbiéri, Empresa Brasileira de Pesquisa Agropecuária

Prof. Dr. Luciano da Silva Pinto, Universidade Federal de Pelotas

Prof. Dr. Antônio Costa de Oliveira, Universidade Federal de Pelotas

Dedicatória

À minha esposa Margéli, pelo suporte e alegria mesmo naqueles momentos que deveriam ser seus.

Agradecimentos

A meu orientador Antônio Costa de Oliveira, por ter acolhido o projeto desta tese e oportunizar o seu pleno desenvolvimento. Pelos conselhos científicos e por abraçar com otimismo os meus interesses de pesquisa.

Ao Dr. Luciano Carlos da Maia, pelo apoio, amizade e incansável ajuda durante a realização do presente trabalho. Aos professores José Antônio Peters, Luciano do Amarante e Rogério Oliveira de Sousa pelo auxílio na concepção de experimentos, ensinamentos e críticas essenciais no amadurecimento do projeto de pesquisa.

A todos os professores do Programa de Pós-graduação em Biotecnologia, em especial ao Odir Antônio Dellagostin por ser sempre solícito, bem com pela convivência harmoniosa. À funcionária Fabiane Perez pela competência a prontidão em atender todas minhas dúvidas e necessidades junto a secretária do PPGB/UFPEL.

Aos colegas de laboratório do Centro de Genômica e Fitomelhoramento, Daniel Rosa Farias, Sydney Antonio Frehner Kavalco e Naciele Marini pela colaboração na concepção e elaboração dos experimentos. Aos demais colegas pela receptividade, convívio e apoio durante minha passagem.

A meus pais Jorge Victoria (*In memorium*) e Helena de Carvalho Victoria, pelo amor, mesmo à distância e acima de tudo pela dedicação em ofertar condições para seguir meus sonhos acima dos seus próprios. Ao meu irmão Alexandre Félix de Carvalho Victoria pela amizada e compreensão, mesmo quando a ausência fez-se necessária. A minha segunda mãe Aidê de Freitas Machado e meu segundo pai Daltro Osório de Albuquerque pelo carinho e incentivo, seja nos momentos sérios aos mais descontraídos, pelo exemplo de caráter, por estarem do meu lado tanto na alegria quanto na tristeza, vibrando com minhas conquistas ou me consolando em minhas derrotas. A Ayasha Magaskawe pela companhia e incodicional afeto.

A minha esposa, Margéli Pereira de Albuquerque pelo exemplo de pessoa e profissional, pela compreensão e amor, meu exemplo e insipiração de vida.

Muito Obrigado

The most exciting phrase to hear in science, the one that heralds new discoveries, is not 'Eureka!' (I found it!) but 'That's funny ...' **Isaac Asimov**

RESUMO

VICTORIA Filipe de Carvalho. **Análise evolutiva de genes de homeostase de ferro e de elementos repetitivos em espécies modelo.** 2011. 152 f. Tese (Doutorado) – Programa de Pós-Graduação em Biotecnologia. Universidade Federal de Pelotas, Pelotas.

O ferro é um elemento essencial para o crescimento e desenvolvimento das plantas, envolvido em processos metabólicos essenciais, como fotossíntese e respiração. Porém, são poucos os dados relacionando a interação entre diferentes genótipos e ambientes. Análises comparativas entre plantas inferiores e plantas cultivadas podem possibilitar o melhor entendimento destes processos. O uso de briófitas como modelo para estudos de processos biológicos em plantas surge como uma estratégia promissora devido ao padrão relativamente simples de desenvolvimento destas plantas. O presente trabalho objetivou identificar padrões de ocorrência de marcadores moleculares em plantas modelo, bem como inferir acerca da filogenia das famílias gênicas envolvidas na homoestase do ferro em plantas, possibilitando a criação de estratégias de transferência de informação genômica entre espécies modelo e espécies órfãs. Utilizando ferramentas de bioinformática foram realizadas análises exploratórias para detectar as ocorrências de elementos repetitivos em bancos de ESTs de onze espécies de plantas. Para a validação destes marcadores moleculares foram desenvolvidos 100 conjuntos de iniciadores a partir das sequências contendo microssatélites obtidas para *Physcomitrella patens* Brid. e testadas contra o DNA genômico de *Polytrichum juniperinum* Hedw. Foram realizadas análises filogenéticas e de divergência das famílias gênicas *Iron Regulated Transporter* (*IRT*), *Ferric Redectase Oxidase* (*FRO*), *Nicotinamide synthase* (*NAS*), *Yellow Stripe-Like* (*YSL*) e *Natural Resistance-Associated Macrophage Protein* (*NRAMP*), envolvidas na homoestase de ferro por meio de inferência bayesiana, utilizando genes de arroz, *Arabidopsis* e *Physcomitrella patens* Brid. na busca de homólogos em diferentes espécies de plantas terrestres, com o auxílio da ferramenta *Blast* (NCBI). Também foram desenvolvidos iniciadores para elementos transponíveis reconhecidamente associados a genes *Ysl* de milho e utilizados

conjuntamente com os iniciadores EST-SSR por meio da técnica *IRAP/REMAP* buscando encontrar marcadores microssatélites associados a cópias desta familia gênica. Como resultados foram identificados 13.133 marcadores microssatélites em bancos de dados não redundantes de regiões expressas (EST) de onze espécies de plantas. Os motivos dinucleotídeos foram mais frequentes em espécies basais, enquanto os motivos trinucleotídeos foram mais frequentes em espécies derivadas. Em 30% dos conjuntos de iniciadores EST–SSR testados contra o DNA de *P. juniperinum*, foi obtido bandas polimórficas promissoras para estudos de mapeamento comparativo e de diversidade genética. Foram encontrados 243 homólogos de genes relacionados as famílias gênicas envolvidas com a homoestase de ferro em trinta espécies de plantas. A análise de *fingerprinting* realizada sugere que a maioria destes genes estão submetidos a seleção positiva, indicando acúmulo de mutações adaptativas, essencial para a manutenção e otimização da resposta gênica. A análise de tempo de divergência indica que os genes *IRT* são mais basais e os genes *FRO* os mais recentes entre as familias gênicas estudadas. As famílias *NRAMP* e *YSL* são evolutivamente próximas. A análise bayesiana das sequências e de regiões promotoras dos genes *NRAMP* não indica duplicações recentes em gramíneas, sendo as duplicações provenientes de divergência ancestral a origem do grupo. Parálogos foram identificados somente em dicotiledôneas. Por meio da transferência de marcadores *IRAP/REMAP* é observado que genes *YSL* de *P. patens* estão cercados por retroelementos do tipo cópia, a exemplo do que ocorre com o gene *ZmYSL1* em milho. Também foi estabelecido o cultivo, em condições axênicas, de *Polytrichum juniperinum* Hedw. utilizando esporos como explantes, onde foi observado que protonemas são obtidos utilizando meio de cultura livre de fitorreguladores, regenerando gametófitos em cultivo *in vitro*.

Palavras-chaves: *Physcomitrella patens* Brid. *Polytrichum juniperinum* Hedw. Famílias gênicas. Microssatélites. Inferência bayesiana.

ABSTRACT

VICTORIA Filipe de Carvalho. **Evolutive analysis of iron uptake genes and repetitive elements in model plant species.** 2011. 152 f. Tese (Doutorado) – Programa de Pós-Graduação em Biotecnologia. Universidade Federal de Pelotas, Pelotas.

Iron is an essential element for plant development, involved in metabolic processes, such respiration and photosynthesis. However, data regarding the genotype by environment interaction are lacking. Comparative analysis with lower plant groups and crop plants can increase the understanding about these processes. The use of bryophytes as model plants rise as a promising strategy since they present simpler patterns of development. The present work aimed to identify the occurrence patterns of molecular markers in model plant species, as well as to infer about the phylogenetical relationships of gene families related with iron homoestasis in plants, allowing the development of tranfer strategies of genomic data across model and orphan species. Using bioinformatics tools, a survey analysis was performed to detect repetitive elements in EST banks of eleven plant species. To validate the SSR markers found, 100 primer pairs were developed on the microsatelite sequences obtained for *Physcomitrella patens* Brid. and tested against genomic DNA of *Polytrichum juniperinum* Hedw. Phylogenetic and divergence time analysis was performed for the gene families *Iron Regulated Transporter* (*IRT*), *Ferric Redectase Oxidase* (*FRO*), *Nicotinamide synthase* (*NAS*), *Yellow Stripe-Like* (*YSL*) and *Natural Resistance-Associated Macrophage Protein* (*NRAMP*), related to the iron homoestasis, with help of the Bayesian inference and using the rice, *Arabidopsis* and *P. patens* genes for the *Blast* search in distinct land plants species. Also, primers for transposable elements recognizably related to *Ysl* genes were developed and applied jointly with the SSR primers by the IRAP/REMAP technic searching to find microsatellite markers associated to copies of this gene family. A total of 13,133 SSR markers were discovered in non- redundant EST databases made for all eleven species chosen for this study. The dimer motifs are more frequent in

lower plant species, such as green algae and mosses, and the trimer motifs are more frequent for the majority of higher plant groups, such as monocots and dicots. Thirty percent of EST-SSE were successfully transferred with a relative polimorphism information across *Physcomitrella patens* Brid. and *P. juniperinum*, being promising for mapping and comparative genome analyses in plants. A total of 243 iron uptake gene sequences for 30 plant species were found using rice and *Arabidopsis thaliana* (L.) Heynh. homologues as queries. The evolutionary fingerprinting analyses suggested a positive selective pressure on iron uptake genes for most of the plant homologues analyzed, enabling an optimization and maintenance of gene function. The divergence time analysis indicates *IRT* as the most ancient gene family and *FRO* as the most recent. *NRAMP* and *YSL* genes appear as a close branch in the evolution of iron uptake gene families. No recent duplication in grasses were found based in the bayesian inference, and paralogue copies were only observed for dicot species. The *Nramp* cis-acting homology search indicated an ancestral duplication hypothesis for this gene family in grasses. Using *IRAP/REMAP* techniques, it was observed that *YSL* homologues in *Physcomitrella* are surrounded by *copia-like* retrotransposons as occurs in the maize *ZmYSL1* copy. Also *Polytrchum juniperinum* Hedw. *in vitro* cultures were estabilished using spores as explants. Protonemal and gametophyte development were obtained using a growth regulator free culture medium.

Key-words: *Physcomitrella patens* Brid. *Polytrichum juniperinum* Hedw. Gene families. Microsatelites. Bayesian inference.

SUMÁRIO

1. INTRODUÇÃO GERAL

As condições ambientais adversas, como seca, salinidade e temperaturas extremas, encontradas pela planta durante o seu ciclo de vida, impõem limitações severas ao crescimento e reprodução vegetal, restringindo o seu potencial genético e, em última análise, refletindo na produtividade das culturas agrícolas (*DHALIWHAL; ARORA*, 1999). Em solos ácidos a anoxia decorrente do alagamento pode levar à solubilização de grandes quantidades de ferro, antes precipitado formando quelatos e ligado à matéria orgânica presente no solo (PONNAMPERUMA, 1972). As plantas necessitam de mecanismos para absorver ferro do solo e atender a demanda para o crescimento e desenvolvimento, e também evitar a toxicidade pelo excesso de ferro, pois ambas as situações são responsáveis por várias disfunções nutricionais que afetam a fisiologia das plantas (PONNAMPERUMA, 1972; CHANEY et al. 1972).

O metabolismo do ferro é um mecanismo complexo sob um balanço homeostático, representando dois problemas principais para as plantas: i) deficiência como conseqüência de problemas de solubilidade e ii) toxidez devido ao excesso de solubilidade em condições anaeróbicas (SANTOS e COSTA de OLIVEIRA, 2007). Em condições de suficiência de ferro, as plantas reduzem Fe^{3+} e transportam o Fe^{2+}, resultante através da membrana plasmática via um transportador de baixa afinidade, ainda não caracterizado em nível molecular (CURIE e BRIAT, 2003). Em condições de deficiência de ferro, as plantas desenvolveram diferentes estratégias para aumentar a captação deste nutriente. Respondem a deficiência de ferro pela indução de respostas direcionadas a aquisição do elemento da rizosfera, estando divididos em plantas de Estratégia I (redução do Fe^{3+} a Fe^{2+}) e Estratégia II (quelação do Fe^{3+}) segundo RÖMHELD (1987) e MARSCHNER e RÖMHELD (1994).

Cultivares de arroz com diferentes níveis de tolerância à toxidez por excesso de ferro foram desenvolvidas através de melhoramento genético (FAGERIA; RABELO, 1987; SAHRAWAT *et al.,* 1996), e práticas agronômicas como plantio alternativo, manejo hídrico adequado, e a

aplicação de fertilizantes (WINSLOW *et al.,* 1989) vem sendo desenvolvidas e utilizadas. Neste panorama, a prática mais eficiente tem sido o uso de genótipos resistentes (SAHRAWAT *et al.,* 1996). Entretanto, devido à diversidade de ambientes em que a toxidez por excesso de ferro pode ocorrer, nenhuma destas opções é universalmente aplicável ou eficiente (BECKER; ASCH, 2005). Assim, o conhecimento sobre o impacto do excesso de ferro na fisiologia de plantas de arroz se torna necessário para a cultura.

De acordo com ROMHELD, (1987), existem dois sistemas ou estratégias para a absorção do ferro solúvel do solo, em condições de deficiência de ferro na planta, sendo divididos em plantas de Estratégia I (redução do Fe^{3+} a Fe^{2+}) e Estratégia II (quelação do Fe^{3+}). Existem várias famílias gênicas envolvidas na homeostase de ferro em plantas. Estes genes têm sido estudados em várias espécies de plantas e importantes avanços foram feitos, buscando a elucidação dos mecanismos. Entre as principais famílias gênicas envolvidas na homeostase de ferro estão: os genes *ZIP* (Zrt/Irt-related Proteins), que codificam para proteínas que atuam no transporte de Zn (*ZRT – Zinc Regulated Transporter*) e Fe^{2+} (*IRT – Iron Regulated Transporter*), sendo considerado o mais importante de raízes (CONNOLLY et al., 2002) e o principal transportador de ferro em *Arabidopsis* (VERT et al., 2002, VAROTTO et al., 2002). Os genes *FRO* (*Ferric Reductase Oxidase*) que codificam para enzima Fe^{3+} quelato redutase presente nas células epidérmicas radiculares, é responsável pela redução do Fe^{3+} a Fe^{2+}, etapa importante no processo de absorção desse micronutriente, em plantas da estratégia I (YI e GUERINOT, 1996); os genes *YSL* (*Yellow Stripe Like*) utilizados por gramíneas (estratégia II) em situação de deficiência de ferro e consiste na liberação de fitosideróforos (PS) no solo pelas raízes das plantas, onde são complexados com Fe^{+3} e absorvidos pelas células epidérmicas radiculares da planta na forma Fe^{+3}-PS via um transportador específico do tipo *YSL1*, sem necessidade de redução extracelular (CURIE et al., 2001); os genes *NRAMP* (*Natural Resistance-Associated Macrophage Protein*) que constituem uma família altamente conservada de proteínas integrais de membrana que estão envolvidos no transporte de ferro em vários organismos, incluindo bactérias, fungos, plantas e animais (CELLIER et al., 2001); e os genes *NAS* (*Nicotianamina sintase*), que codificam Nicotinamina

(NA), um quelante de metais que é ubíqua em plantas superiores. Nenhuma destas famílias gênicas foram estudadas amplamente quanto as relações filogenéticas dentro das principais linhagens de plantas terrestes.

À parte do significativo progresso no conhecimento das condições que levam à ocorrência da toxidez por excesso de ferro em plantas, são pouco conhecidas as interações entre excesso de ferro e diferentes genótipos (SAHRAWAT, 2004). Diversos fatores envolvendo a tolerância ao excesso de ferro vêm sendo identificados, bem como a identificação de QTLs (*quantitative trait loci*) em distintas populações (WU *et al.,* 1997; WU *et al.,* 1998; WAN *et al.,* 2003; SHIMIZU *et al.,* 2005). Análises comparativas em grupos primitivos e grupos irmãos das espécies cultivadas podem possibilitar o melhor entendimento destes processos.

As briófitas ocupam uma posição filogenética crítica para nossa compreensão sobre a origem das plantas terrestres. Suas relações filogenéticas tem atraído o interesse de sistematas desde a segunda metade do século XX (BOWE, 1935; CAMPBELL, 1905; CAMPBELL, 1971; HASKELL, 1949; KHANNA, 1965; MILLER, 1974, 1982; SCHOFIELD, 1985). Este interesse reside no fato de que este grupo de plantas aparantemente incluir os principais componentes da linhagem basal das plantas terrestres. Bryophyta *sensu strictu* é um dos grupos mais antigos de plantas ainda presentes na flora terrestre. Estas se originaram cerca de 450 milhões de anos atrás (HECKMAN et al., 2001; MAGALLÓN et al., 2009) e estão representadas atualmente por aproximadamente 10.000 espécies que colonizaram os mais diversos ambientes. Dentro da Divisão Bryophyta as quatro linhagens principais são: Sphagnopsida, Andreopsida, Polytrichopsida e Bryopsida (SHAW, GOFFINET, 2000). Bryopsida é o maior grupo em número de espécies, representando cerca de 90% das espécies de musgos conhecidas. Polytrichopsida apresenta número bem menor de espécies, comparada a linhagem anterior, mas ainda é considerado o segundo maior grupo em espécies e variabilidade ecológica, com plantas tipicamente pioneiras. Esta linhagem é considerada como grupo basal dentro de Bryophyta s. s., com espécies caracteristicamente rústicas, bem distribuídas em ambos os hemisférios. Ecologicamente, as polytrycáceas apresentam tanto o hábito xerofítico (*Polytrichum piliferum* Hedw.) como também

mesofíticos, podendo ocorrer em ambientes com alto grau de umidade (*Polytrichum commune* Hedw). Estudos evolutivos confirmam a origem monofilética das plantas terrestres com indicativos de que as briófitas formem um ramo irmão com as traqueófitas, entretanto as relações específicas entre estes grupos ainda estão sendo investigadas (KENRICK, CRANE, 1997; NICKERENT *et al*. 2000; SCHAEFER, ZRŸD, 2001).

O uso potencial dos musgos como modelo para estudos de processos biológicos em plantas é devido ao padrão relativamente simples de desenvolvimento destas plantas, conveniência para análises de linhagens celulares, resposta similar do grupo aos fatores de crescimentos e estímulos ambientais como os encontrados em outras plantas terrestres, e a facilidade para aplicação de abordagens genéticas, resultante da dominância da geração gametofítica no ciclo de vida destas plantas.

A espécie de musgo mais extensivamente estudada é *Physcomitrella patens* (Hedw.) Bruch. & Schimp, sendo primeira espécie basal com o genoma completamente sequenciado (RENSING *et al*., 2008) A partir dos dados genômicos de *P. patens* é possível inferir acerca dos eventos que possibilitaram a evolução dos genomas das plantas terrestres e fazer comparações com os genomas de muitas angiospermas já sequeciados [p. ex. *Arabidopsis thaliana* (L.) Heynh., *Oryza sativa* L. e *Populus trichocarpa* Torr. & A. Gray]. Como propriedades gerais do genoma de *P. patens* destacam-se a ocorrência de mais de 14 mil elementos transponíveis, sendo quase cinco mil compostos por *full-length Long terminal repeat retrotransposons* (*LTR-Rs)*, onde 46% são do tipo *gypsy* e 2% do tipo *copia*. Estes dados indicam que ocorrem três vezes mais *LTR-Rs* em *P. patens* do que no genoma de *A. thaliana* e cerca de um terço do que ocorre no genoma de *O. sativa* (RENSING *et al.,* 2008). Este tipo de elementos repetitivos também foram identificados associação com cópias de genes envolvidos na homeostase de ferro em milho (CURIE *et al.,* 2001) sendo promissores como marcadores em análises comparativas em genomas vegetais. Outra classe de marcadores moleculares muito promissores são os microssatélites. Devido a natureza multi-alélica, caracter co-dominante, abundância em genomas e sua reprodutibilidade são considerados uma ferramenta poderosa com

diversas aplicações na genética e melhoramento de plantas (VARSHNEY *et al.*, 2005). Segundo Oliveira et al. (2006), o alto nível de transferibilidade de marcadores microssatélites, grande quantidade e dispersão nos genomas são os principais atributos, que tornam uma potente ferramenta. Estudos comparativos em genomas de distintas espécies de plantas são necessários para encontrar padrões de distribuição de microssatélites possibilitam a transferência deste marcadores para espécies ainda pouco estudadas.

Portanto, os musgos representam uma ótima oportunidade para estudos genéticos de plantas, combinando simplicidade com conveniência técnica, buscando comparar o genoma de uma planta simples àqueles das plantas mais complexas.

2. OBJETIVOS

Por se tratar de um grupo basal dentro da filogenia das plantas terrestres, espera-se contribuir com estudos evolutivos e de genômica comparativa a partir das análises realizadas com a espécie de musgo *Polytrichum juniperinum* Hedw., buscando compreender como ocorrem os processos biológicos em briófitas comparando-os com os processos em plantas cultivadas.

A partir dos bancos de dados públicos de regiões expressas (*EST-Expressed Sequence Tags*) e de amostras de *Polytrichum juniperinum,* coletadas no Rio Grande do Sul, na Patagônia chilena e na Antártica Marítima pretende-se: (1) Avaliar a composição dos elementos repetitivos nos principais grupos de plantas; (2) Por meio do reconhecimento dos padrões de ocorrência de elementos repetitivos, selecionar e construir *primers* para os *loci* mais frequentes em briófitas e grupos afins, buscando identificar a ocorrência destes no genoma de *Polytrichum juniperinum*, validando-os como marcadores moleculares passíveis de transferência para espécies filogeneticamente relacionadas; (3) Comparar as sequências de famílas gênicas responsáveis pela homeostase do ferro em plantas disponíveis no *Genbank,* inferindo acerca de história evolutiva destes genes; (4) Avaliar a similaridade genética de amostras de *Polytrichum* juniperinum em três áreas geográficas distintas (Rio Grande do Sul, Patagônia chilena e Antártica Marítima), por meio de marcadores IRAP/REMAP e, a partir do

sequenciamento de regiões homólogas entre plantas inferiores e superiores, comparar as respostas obtidas com o conhecido para plantas de interesse econômico.

3. ARTIGO 1. Establishment of the moss *Polytrichum juniperinum* Hedw. under axenic conditions.

(Aceito para publicaçao no periódico Bioscience Journal)

ESTABLISHMENT OF THE MOSS *Polytrichum juniperinum* HEDW. UNDER AXENIC CONDITIONS.

Estabelecimento e Desenvolvimento do musgo *Polytrichum juniperinum* Hedw. sob condições de cultivo axênico.

Filipe de Carvalho VICTORIA[1]; Antônio Costa DE OLIVEIRA[2]; José Antônio PETERS[3]

[1] Biólogist, MSc in Botany, Graduate Student in Biotecnology-UFPEL; [2]PhD in Genetics, Plants Genomics Center, UFPEL; [3]PhD in Botany, Plants Tissues Cultive Laboratory, UFPEL.

Resumo *Polytrichum juniperinum* Hedw. (Polytrichaceae) é uma espécie de musgo de ampla distribuição mundial. Culturas *in vitro* foram estabelecidas a partir de esporos de espécimes coletados na natureza. O desenvolvimento, tanto de protonema quanto de gametófitos, foi observado utilizando o meio básico MS em três tratamentos, livre de fitorreguladores, suplementados com uma fonte de auxina (AIA), suplementados com uma fonte de citocinina (BAP) e suplementado com ambos reguladores. Nos cultivos resultantes de meio livre de reguladores e de meios contendo auxina, foi observado o desenvolvimento total dos gametófitos, enquanto nos meios contendo citocinina não foram observados desenvolvimento e regeneração de gametófitos. Estes resultados sugerem a utilização do meio livre de reguladores para cultivo de *Polytrichum juniperinum* em cultivos axênicos.

Palavras-chaves: Desenvolvimento *in vitro*; *Polytrichum juniperinum*; Meio MS; cultura de tecidos de musgos.

Summary *Polytrichum juniperinum* Hedw. (Polytrichaceae) is a moss with a worldwide distribution. *In vitro* culture was established from *P. juniperinum* spores collected in nature. Both protonema and gametophore stages of gametophyte development were obtained. The Murashige-Skoog regulator-free nutrient medium or

supplemented with AIA and BAP conferred a fully development and regeneration of gametophytes. Tissues grown on cytokinin did not produce any gametophytes. These results indicate the possibility to use a medium without growth regulators to obtain gametophytes for this species in axenic conditions.

Key words *In vitro* development; *Polytrichum juniperinum*; MS medium; mosses tissue culture.

INTRODUCTION

Micropropagation or *in vitro* vegetative propagation of plants constitutes a way of perpetuating healthy and aseptic explants, facilitating the application of regeneration techniques and genetic transformation. Aseptic culturing is necessary for certain experimental procedures (Sabovljević et al. 2006) and it is highly convenient for the maintainance of plant genotype collections free of different pathogens. The obtaintion of bryophytes isolated cultures has been reported as a complex task by many investigators (Gang et al. 2003; Bijelović & Sabovljević 2003; Cvetić et al. 2007, Silva et al. 2009, Silva et al. 2010), due to a possible interaction of these plants with other organisms in non-axenic conditions. Nevertheless, bryophytes have great advantages over vascular plants as models for plant biology investigations: (1) relatively simple structure compared to other higher plants, (2) haploid gametophyte as the dominant vegetative phase, and (3) lower chromosome numbers (Gang et al. 2003). The culture of bryophyte cells in suspension media, as well as the dominant gametophyte phase of mosses, have been reported as favorable model systems for genetic, biochemical, metabolic, and developmental studies (Cove et al. 2006; Ono et al. 1988).

The introduction of new species into axenic conditions and maintenance of stable cell is therefore essential as a start for in-depth investigation of the physiology and potential uses of bryophytes. *Polytrichum juniperinum* Hedw. (Polytrichaceae) is a common moss species with a worldwide distribution, adapted to open, dry and sandy environments, growing on a variety of peatlands, especially on drained habitats (van der Velde and Bijlsma 2003). The ancestral position of Bryophyta for land plants relatioship being that group as a target to understand the envolved processes to conquest the land enviroments by the plants. Early iniciatives aimed to verify the *in vitro* development for several moss species, such as *Physcomitrella patens* Brid. (Cove et al. 2006), *Ceratodon purpureus* (Hedw.) Brid. (Sabovljević et al. 2003), *Pogonatum urnigerum* (Hedw.) P. Beauv. (Cvetić et al. 2007) e *Atrichum* spp. (Ono et al. 1987; Gang et al. 2003; Sabovljević et al. 2006). The present study aimed to establish a *in vitro* culture for *P. juniperinum* and examine its development under axenic conditions.

MATERIALS AND METHODS

Fully developed *Polytrichum juniperinum* plants were indentified and collected by the first author in the autumm 2008 at two sites in Southern Brazil, Gramado (29° 23’ S; 50° 52’W) and Canela (29° 21’S; 50° 50’W) in the highlands named Serra Gaúcha. Fresh, unopened sporophytes were surface sterilized as described by Cvetić *et al.* (2007) by dipping in 25% commercial bleach (8% active NaOCl) for 3 minutes, and thoroughly rinsed in sterile distilled water. The cap was then removed and the spores released on the nutrient medium.

As basal medium for establishment of *in vitro* culture, the Murashige and Skoog (1962) basic medium containing 100 mg L^{-1} of inositol an 15 g L^{-1} of sucrose, solidified with 7 g L^{-1} of agar was used. In order to observe the influence of growth

regulators on the *in vitro* development of this species the follow media composition were used: MS 1 (MS regulator-free); MS 2 (MS + 1.0mg L^{-1} AIA, 0.05 mg L^{-1} Kinetin); MS 3 (MS + 1.0 mg L^{-1} AIA, 0.1 mg L^{-1} Kinetin); MS 4 (MS + 1.0 mg L^{-1} AIA, 1.0 mg L^{-1} Kinetin); MS 5 (MS + 1.0 mg L^{-1} AIA, 1.5 mg L^{-1} Kinetin); MS 6 (MS+ 1.0 mg L^{-1} AIA); MS 7 (MS + 1.0 mg L^{-1} AIA, 1.0 mg L^{-1} BAP) and MS 8 (MS + 1.0 mg L^{-1} BAP). All media above were shed in 90x60 mm Petri dishes. Prior to the sterilization the pH was adjusted to 5.8.

Culture were grown at 25±1ºC under long-day conditions (16 h light/8 h dark) supplied by cool-white fluorescent tubes at a photon flow rate of 48 μmol m^2s^{-1}. When a protonemal mass formation was observed, it was subcultured monthly in the same medium until the rise of the first shoot. One unopened sporophyte per dish was used, in a total of four replicates for each treatment (completely randomized design). Calli development and regenerating gamethophyte amounts where evaluated in each medium proposed, and a test of comparison of means was performed using the Tukey test (5% of probability) with the aid of Statistix 9.0 *for Windows* software.

RESULTS AND DISCUSSION

The inoculated spores of *P. juniperinum* took 15 days to germinate and 20-40 days for completely protonemata formation (Figure 1A). After tree subcultures, the first gametophyte shoots rose in the MS regulator-free medium and in the MS6, MS7 and MS8 medium (Figure 1B). No differences were found for spore germination and protonemata development, when all media used were compared. Gametophyte regeneration was not observed in media containing Kinetin as a cytokinin source. In the medium containing AIA as auxin and BAP as an alternative cytokinin source, the regeneration was successful. The same was observed with regulator-free MS, with no

significantly differences observed within other tested media (Table 1). On the other hand, calli formation were only found in media with 0.1 and 1.0 mg L^{-1} Kinetin.

The results demonstrate that the fully in vitro development for *Polytrichum juniperinum* can be reach using regulator-free medium. Early studies reports the importance of hormonal requirements for shoot regeneration and multiplication (Bopp & Atzorn 1992). Cytokinins have been shown to induce bud formation in protonemata cultures of some moss species (Speiss 1976; Bijelović & Sabovljević 2003). In the present study these growth regulator not demostrate a diferencial shoot formation response. The cytokinins dosages used were not sufficient for buds development, suggests a diferencial requirement for bud induction in *P. juniperinum* where compare with other mosses species.

In studies of cytokinin action on different moss species (Speiss 1976), calli were obtained with most of the species, except for the polytrichaceous species studied. However, Gang et al. (2003) demonstrate for *Atrichum undulatum* (Hedw.) P. Beauv. that calli was obtained when the growth medium contained Benzyladenine (BA). The callus formation was obtained to *Pogonatum urnigerum* (Hedw.) P. Beauv., when a medium with low sugar values was used (Cvetić et al. 2007). For *P. juniperinum* calli was observed only in the media when the Kinetin contents was increase (MS4 and MS5). In our study is the new report for calli occurrence in Polytrichaceae species in axenic conditions.

Cvetić et al. (2007) reported a fast calli senescence formation and relatively fast senescence of protonemata, probably due to the fact that the protonemata of moss species are not persistent in nature. The positive effect found on the AIA and Kinetin ratio tested in MS4 and MS5 media probably due to an increase in protonemata

growth, delaying the senescence and keeping the calli viable, which was not observed for the other tested media in the present study.

The spore germination and protonemata development was observed on all media used. Similar results were found in other initiatives to establish an *in vitro* protocol for moss development and gametophyte regeneration (Cvetić et al. 2005, 2007; Sabovljević et al. 2003, 2006), when the use of growth regulators were not necessary. These results indicate that it is possible to use simple media for spore germination and gametophyte regeneration of *P. juniperinum* in axenic conditions.

ACKNOWLEDGEMENTS

The authors thank the IBAMA for the collection permision given for the first author and the Comissão de Aperfeiçoamento de Pessoal de Nivel Superior (CAPES) for the financial support.

REFERENCES

BOPP, M.; ATZORN, R. Hormonelle Regulation der Moosentwicklung. **Naturwissenschaften,** Berlin, v. 79, p.337–346, 1992.

BIJELOVIĆ, A.; SABOVLJEVIĆ, M. Callus induction and plant regeneration in the moss Aloina aloides (Schultz.) Kindb. (Pottiaceae, Bryopsida). **Archives of Biological Sciences**, Belgrade, v. 55, n. 3-4, p. 77-80, 2003.

COVE, D.J.; BEZANILLA, M.; HARRIES, P.; QUATRANO, R. Mosses as model systems for the study of metabolism and development. **Annual Review of Plant Biology**, v. 57, p. 497-520, 2006.

CVETIĆ, T.; SABOVLJEVIĆ, A.; SABOVLJEVIĆ, M.; GRUBIŠIĆ, D. *In vitro* cultura and apogamy-alternative pathway in the life cycle of the moss *Amblystegium serpens* (Amblystegiaceae). **Archives of Biological Science,** Belgrade, v. 57, n. 4, p. 267-272, 2005.

CVETIĆ, T.; SABOVLJEVIĆ, A.; SABOVLJEVIĆ, M.; GRUBIŠIĆ, D. Development of the moss *Pogonatum urnigerum* (Hedw.) P. Beauv. under *in vitro* culture conditions. **Archives of Biological Science**, Belgrade, v. 59, n. 1, p. 57-61, 2007.

GANG, Y.Y.; DU, G.S.; SHI, D.J.; WANG, M. Z.; LI, X. D.; HUA, Z. L. Establishment of in vitroregeneration systems of the Atrichum mosses. **Acta Botanica Sinica**, Beijing, v. 45, n. 12, p. 1475-1480, 2003.

MURASHIGE, T.; SKOOG, F. A revised medium for rapid growth and bioassays with tobacco tissue cultures. **Plant Physiology**, Danvers, v. 15, n. 3, p. 473-497, 1962.

ONO, K; MURASAKI, Y; KAWAUCHI, K. Establishment and characteristics of a cell suspension culture from a moss, *Atrichum undulatum*. **The Botanical Magazine**, Tokyo, v. 100, n. 2, p. 217–221, 1987.

ONO, K.; MURASAKI, Y.; TAKAMIYA, M. Induction and morphogenesis of cultured cells of bryophytes. **Journal of Hattori Botanical Laboratory**, v. 65, p. 391-401, 1998.

SABOVLJEVIĆ C, M.; BIJELOVIĆ, A.; DRAGICEVIĆ, I. *In vitro* Culture of Mosses: *Aloina aloides* (K.F.Schultz) Kindb., *Brachythecium velutinum* (Hedw.) B.S. & G., *Ceratodon purpureus* (Hedw.) Brid., *Eurhynchium praelongum*(Hedw.) B.S. & G. and *Grimmia pulvinata* (Hedw.) Sm. **Turkey Journal of Botany**, Ankara, v. 27, p. 441-446, 2003.

SABOVLJEVIĆ, A.; CVETIĆ, T.; SABOVLJEVIĆ, M. Establishment and development of the Chaterine's moss *Atrichum undulatum* (Hedw.) P. Beauv. (Polytrichaceae), in *in vitro* conditions. **Archives of Biological Science***,* Belgrade, v. 58, n. 2, p. 87-93, 2006.

SILVA, A. S. M.; PORTO, K. C.; SIMABUKURO, E. A. Effect of lighth and water availability on spore germination and protonemal growth of the Neotropical moss

Thamniopsis incurva (Pilotrichaceae). **Cryptogamie, Algologie,** Paris, v. 30, n. 2, p. 243-257, 2009.

SILVA, A. S. M.; PORTO, K. C.; SIMABUKURO, E. A. Effects of light and nutrients on different germination phases of the cosmopolitan moss *Bryum argenteum* Hedw. (Bryaceae). **Brazilian Archives of Biology and Technology,** Curitiba, v. 53, n. 4, p. 763-769. 2010

SPEISS, L.D. Developmental effects of zeatin, ribosyl-zeatin, and *Agrobacterium tumefaciens* B6 on certain mosses. **Plant Physiology,** Danvers, v. 58, n. 1, p. 107-109, 1976.

VAN DER VELDE, M.; BIJLSMA, R. Phylogeography of five *Polytrichum* species within Europe, **Biological Journal of Linnean Society**, London, v. 78, n. 2, p. 203-213, 2003.

Table 1. *In vitro* micropropagation results for *Polytrichum juniperinum* Hedw. Positive (+) and negative (-) signs indicate a positive or negative response to a given medium, respectivel, (N/A) indicates a unsuccesfull gametophyte regeneration.

	MS1	*MS2*	*MS3*	*MS4*	*MS5*	*MS6*	*MS7*	*MS8*
Germination	+	+	+	+	+	+	+	+
Callus formation	-	-	+	+	-	-	-	-
Gametophyte formation	+	-	-	-	-	+	+	+
Nr of regenerating gametophyte	66 a	N/A	N/A	N/A	N/A	66 a	60 a	68 a

Means followed by the same letter do not differ by the Tukey test ($p<0,05$).

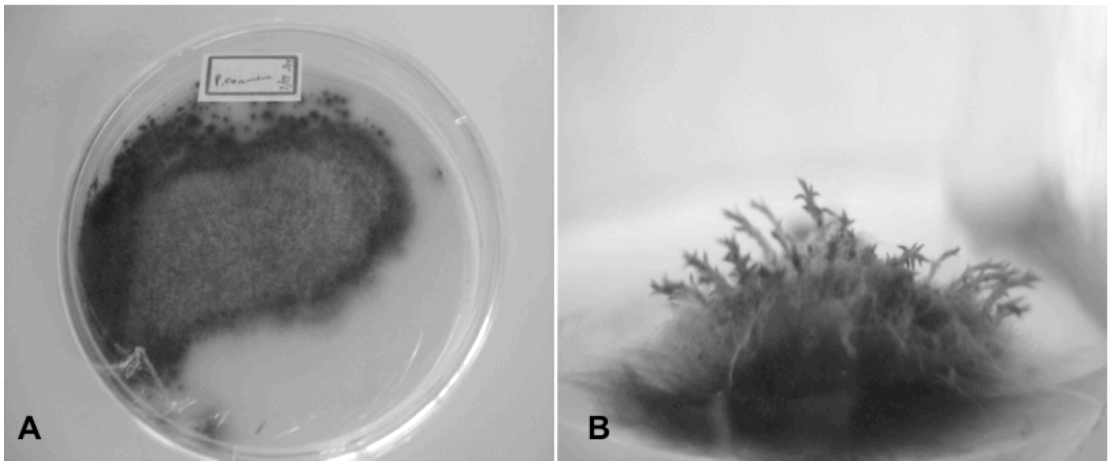


Fig. 1. A. *Polytrichum juniperinum* Hedw. protonemata obtained after 40 days *in vitro* cultive. **B.** *Polytrichum juniperinum* Hedw. regenerate gametophytes by pure MS medium.

4. ARTIGO 2. *In silico* comparative analysis of SSR markers in plants

(Publicado na BMC Plant Biology)

Victoria et al. BMC Plant Biology 2011, 11:15
http://www.biomedcentral.com/1471-2229/11/15

BMC Plant Biology

RESEARCH ARTICLE Open Access

In silico comparative analysis of SSR markers in plants

Filipe C Victoria[1,2], Luciano C da Maia[1], Antonio Costa de Oliveira[1*]

Abstract

Background: The adverse environmental conditions impose extreme limitation to growth and plant development, restricting the genetic potential and reflecting on plant yield losses. The progress obtained by classic plant breeding methods aiming at increasing abiotic stress tolerances have not been enough to cope with increasing food demands. New target genes need to be identified to reach this goal, which requires extensive studies of the related biological mechanisms. Comparative analyses in ancestral plant groups can help to elucidate yet unclear biological processes.

Results: In this study, we surveyed the occurrence patterns of expressed sequence tag-derived microsatellite markers for model plants. A total of 13,133 SSR markers were discovered using the *SSRLocator* software in non-redundant EST databases made for all eleven species chosen for this study. The dimer motifs are more frequent in lower plant species, such as green algae and mosses, and the trimer motifs are more frequent for the majority of higher plant groups, such as monocots and dicots. With this *in silico* study we confirm several microsatellite plant survey results made with available bioinformatics tools.

Conclusions: The comparative studies of EST-SSR markers among all plant lineages is well suited for plant evolution studies as well as for future studies of transferability of molecular markers.

Background

In agriculture, productivity is affected by environmental conditions such as drought, salinity, high radiation and extreme temperatures faced by plants during their life cycle, that impose severe limitations to the growth and propagation, restricting their genetic potential and, ultimately, reflecting yield losses of agricultural crops. Although, advances have been achieved through classical breeding, further progress is needed to increase abiotic stress tolerance in cultivated plants. New gene targets need to be identified in order to reach these goals, requiring extensive studies concerning the biological processes related to abiotic stresses. Comparative analysis between primitive and related groups of cultivated species may shed some light on the understanding of these processes.

Microsatellites or SSRs (Simple Sequence Repeats) are sequences in which one or few bases are tandemly repeated, ranging from 1-6 base pair (bp) long units. They are ubiquitous in prokaryotes and eukaryotes, present even in the smallest bacterial genomes [1-3]. Variations in SSR regions originate mostly from errors during the replication process, frequently DNA Polymerase slippage. These errors generate base pair insertions or deletions, resulting, respectively, in larger or smaller regions [4]. SSR assessments in the human genome have shown that many diseases are caused by mutation in these sequences [5]. The genomic abundance of microsatellites, and their ability to associate with many phenotypes, make this class of molecular markers a powerful tool for diverse application in plant genetics. The identification of microsatellite markers derived from EST (or cDNAs), and described as functional markers, represents an even more useful possibility for these markers when compared to those based on assessing anonymous regions [6-8]. EST-SSRs offer some advantages over other genomic DNA-based markers, such as detecting the variation in the expressed portion of the genome, giving a ''perfect''marker-trait association; they can be developed from EST databases

* Correspondence: acostol@terra.com.br
[1]Plant Genomics and Breeding Center, Faculdade de Agronomia Eliseu Maciel, Universidade Federal de Pelotas, RS, Brasil
Full list of author information is available at the end of the article

at no cost and unlike genomic SSRs, they may be used across a number of related species [9].

Many studies indicate UTRs as being more abundant in microsatellites than CDS regions [10]. In a study of micro- and minisatellite distribution in UTR and CDS regions using the Unigene database for several higher plants groups, higher occurrence of these elements in coding regions were found for all the studied species [11]. Disagreements between earlier reports and the later, reflect a deficiency in annotation when translated and non-translated fractions are separated in the Unigene transcript database. Dimer repeats were also frequent in CDS regions, which could be due to the fact that the Unigene database contains predominantly EST clusters. Therefore, there is a tendency for under-representing the UTR regions in the annotated sequences [11].

The characterization of tandem repeats and their variation within and between different plant families, could facilitate their use as genetic markers and consequently allow plant-breeding strategies that focus on the transfer of markers from model to orphan species to be applied. EST-SSR also have a higher probability of being in linkage disequilibrium with genes/QTLs controlling economic traits, making them more useful in studies involving marker-trait association, QTL mapping and genetic diversity analysis [9].

On model organisms, microsatellites have been reported to correspond to 0.85% of *Arabidopsis thaliana* (L.) Heynh, 0.37% of maize (*Zea mays* L.), 3.21% of tiger puffer (*Takifugu rubripes* Temminck & Schlegel), 0.21% of the nematode *Caenorhabditis elegans* Maupas and 0.30% of yeast (*Saccharomyces cerevisiae* Meyer ex. E.C. Hansen) genomes [10]. Moreover, they constitute 3.00% of the human genome [12]. All kinds of repeated element motifs, excluding trimers and hexamers, are significantly less frequent in the coding sequences when compared to intergenic DNA streches of *A. thaliana*, *Z. mays, Oryza sativa* subsp *japonica* S. Kato (rice), *Glycine max* (L.) Merr. (soybean) and *Triticum aestivum* L. (wheat) [10].

Close to 48.67% of repeat elements found in many species are formed by dimer motifs. In *Picea abies* (L.) H. Karst. (Norway spruce), for example, the dimer occurrence is 20 times more frequent in clones originating from intergenic regions vs. transcript regions [13]. Approximately 14% of protein translated sequences (CDS - coding sequences) contain repetitive DNA regions, and this phenomenon is 3 folds more frequent in eukaryotes than prokaryotes [14]. Clustering studies showing microsatellite occurrence in distinct protein families (non-homologous) from either prokaryotic or eukaryotic genomes, indicate that the origins of these loci occurred after eukaryotic evolution [14-16]. The highest and lowest repeat counts were found in rodents and *C. elegans*, respectively [3].

In plant species, some reports have described the levels of occurrence of microsatellites associated to transcribed regions [7,8,10,11,17-22]. However, some comparative and/or descriptive approaches, still can offer new perspectives on the features of these markers. Furthermore, frequently new groups of plant species have their genome sequenced, enabling the reassessment of databases using new sequences, representing divergent evolutionary groups and/or with different genetic models.

The online platforms for nucleotide, protein and transcript (ESTs) databases available for the majority of species are relatively small when compared with model species, eg *Physcomitrella patens* (Hedw.) Bruch & Schimp., *O. sativa* and *A. thaliana*. Since the protocols for the isolation of repetitive element loci, such as microsatellites, require intensive labour and can be expensive, the exploitation of these elements *in silico* on databases of model plants and their respective transfer to orphan species, is a potentially fruitful strategy.

In this study we present our results on the SSR survey for the development of plant SSR markers. The survey was based on clustered non-redundant EST data, their classification, characterization and comparative analysis in eleven phylogenetically distant plant species including two green algae, a hepatic, two mosses, two fern, two gymnosperms, a monocot and a dicot.

Results and Discussion

We analysed 560,360 virtual transcripts with the *SSRLocator* software (Table 1). The species with most abundant records in Genbank was *Arabidopsis thaliana* with 224,496 virtual transcripts (40%), followed by *Oryza sativa* with 121,635 (21.7%), *Physcomitrela patens* with 79,537 (14.19%), *Pinus taeda* with 58,522 (10.44%) and *Chlamydomonas reinhardtii* with 40,525 (7.2%). The remaining species added up to 11.7% of virtual transcripts analysed. When total genome sizes are compared for the model plants included in this analysis, the virtual transcripts of *P. patens* (511 Mb) represent 0.01% of genome size. For *O. sativa* (389 Mb) and *A. thaliana* (109.2 Mb) the ESTs analysed represent 0.02% and 0.18%, respectively, of the genome. The highest average bp count per EST sequence was found for *Selaginella* spp. (924 bp) followed by *M. polymorpha* (777 bp), *C. reinhardtii* (775 bp) and *P. taeda* (760 bp). The lower average bp per sequence was found for *G. gnemon* (563 bp) and *A. capillus-veneris* (580 bp). For the model plants, *A. thaliana* showed the lowest average bp count (321 bp), with *P. patens* and *O. sativa* presenting similar bp counts (737 and 755 bp, respectively). Shorter observed sequences could be an indication of

Table 1 EST database size and Overall occurrence of SSR, percentages and average length motifs per specie

Species	*EST database count*	*pb*	*Average pg count per EST*	*GC Content %*
Chlamydomonas reinhardtii	40,525	31,388,333	775	57.22
Mesostigma viride	6,401	4,273,634	668	51.36
Marchantia polymorpha	10,086	7,836,025	777	54.75
Syntrichia ruralis	7,114	4,764,692	670	49.20
Physcomitrella patens	79,537	58,636,814	737	47.60
Selaginella spp.	19,830	18,318,250	924	51.38
Adiantum capillus-veneris	16,138	9,363,530	580	45.97
Gnetum gnemon	6,076	3,420,021	563	44.33
Pinus taeda	58,522	44,467,932	760	43.64
Oryza sativa	121,635	91,859,132	755	47.52
Arabidopsis thaliana	224,496	72,013,660	321	41.10

incomplete representation of genes, but one must keep in mind that average gene sizes could vary among species, i.e., rice fl-cDNAs (1,747 bp) are 14% longer than *Arabidopsis* fl-cDNAs (1,532 bp) (TAIR 9 and RIKEN, accessed in 12.2.2010). The overall bp counts are very similar to those found by other authors [23].

The frequency of SSR per EST database was higher (4.66%) in *Selaginella* spp virtual transcripts (Table 2). For model plants, 3.57% and 0.84% SSRs/EST were found for *O. sativa* and *A. thaliana*, respectively.

The average motif length, excluding compound SSRs, was 27.03 bp. *Mesostigma* EST database shows the longest SSR average size with 34.13 bp, and the shortest size was found for *Marchantia polymorpha* with 22.56 bp mean size. The SSR size for model plants was similar. For *P. patens*, *O. sativa* and *A. thaliana*, average sizes of 24.2, 23.4 and 26.5 bp were found, respectively. A total 1,106 EST sequences contained more than one SSR. Among the species, *O. sativa* and *P. patens* are on the extremes of the distribution with 37.34% and 3.46% of virtual transcripts containing one or more microsatellites. However, *Adiantum capillus-veneris* EST database contained the highest percentage of transcripts displaying more than one SSR (20.86%) based on the database size. Similar results were found in our group [11], using the Unigene database for grasses and other allies. In the same study, rice was shown to have the highest frequency of ESTs containing more than one SSR (11.28%). In the present study, a similar value was found for rice (10.20%). These small differences could be due to different redundancy reduction parameters used in Unigene species database and CAP3 default settings. Other reports for higher plants [19,20,24-26], showed different ranges, but never higher than 2-3 fold. The variations encountered in different reports are related to the strategy employed by investigators (software, repeat number and motif type) [11]. The results for each species, regarding the percentage of SSRs found per EST database size are shown on Table 2.

Table 2 EST database size and Overall occurrences of SSRs, percentages and average length motifs per species

Species	*Number of SSR loci*	*SSR/EST database (%)*	*Average motif length (bp)*	*EST sequences with SSRs (%)*	*N. of seq. containing more than one SSR (%)*	*Single SSRs*	*Compound SSRs*
Chlamydomonas reinhardtii	980	2.41	33.21	886 (2.19)	94 (9.78)	899	81
Mesostigma viride	81	1.26	34.12	73 (1.14)	8 (9.87)	73	8
Marchantia polymorpha	437	4.33	22.56	436 (4.32)	1 (0.52)	425	12
Syntrichia ruralis	190	2.67	23.84	149 (2.09)	41 (10.09)	189	1
Physcomitrella patens	2753	3.46	24.20	2577 (3.24)	176 (6.6)	2670	83
Selaginella spp.	968	4.66	23.71	868 (4.38)	100 (11.13)	927	41
Adiantum capillus-veneris	749	4.64	31.14	599 (3.71)	150 (20.86)	624	125
Gnetum gnemon	212	3.48	23.62	195 (3.21)	17 (8.45)	203	9
Pinus taeda	568	0.97	30.89	530 (0.91)	38 (6.85)	539	29
Oryza sativa	4347	3.57	23.44	3934 (3.23)	413 (10.19)	4199	148
Arabidopsis thaliana	1890	0.84	26.52	1822 (0.81)	68 (3.62)	1837	53

The microsatellite survey using *SSRLocator* showed that 13,133 SSRs were available as potential marker loci. From those, 12,585 loci were found in single formation and only 590 were found in compound formation. The fern *A. capillus-veneris* showed the highest percentage (20%) of compound SSR loci. When compared with other available SSR marker search tools, similar results were found. Using MISA software, a total of 13,861 SSRs were available as potential marker loci, being 13,172 SSRs single and 689 compound SSRs for all studied species. *Adiantum* EST database showed the highest percentage of SSR in compound formation (15.55%). This trend does not hold for the majority of lower plants. *P. patens*, for example, presented few EST-SSRs in compound formation (3.57%) and possibly the fern lower database size is masking the results. When it is compared with the majority of plant groups, *P. taeda* is the only species showing a high percentage of compound SSRs (5.81%), corroborating other studies which report that compound and imperfect tandem repeats are most common in pines [27-29].

A total of 3,723 EST-SSRs were found in *P. patens* database using the MISA software [23]. The *SSRLocator* analysis resulted in 2,839 SSR for this species. When the same non-redundant databases were run in other bioformatics tools, the results were similar to MISA. Using the SciKoco package [30] combined with MISA, Sputinik and Modified scripts, it was possible to narrow SSR results to a 2-fold range variation.

The search for repetitive elements in EST databases of the eleven taxa listed above enabled the comparison of patterns of occurrence of these elements in lower and higher plants (Figure 1). In some species such as *C.* reinhardtii, *Mesostigma viride* and bryophytes, we found that dimer (NN) microsatellites are more common when compared to higher plants (Figure 2). The trimer (NNN) microsatellites are predominant in higher plants (See additional files), in agreement with other SSR survey studies [6,10,11,21] supporting the relative distribution of motifs in these plant groups. However, gymnosperm species showed the lowest SSR occurrence within the derived plant groups. *Pinus* and *Gnetum* results indicate low SSR frequencies as intrinsic characteristics of gymnosperms, such as suggested by other results obtained with distinct methods [10,23,28,29]. The patterns of occurrence of dimers and trimers found in the EST databases of the selected species are shown on Additional files 1 and 2, respectively.

The average GC-content in the 11 datasets was 48.55%. Significantly increased GC-contents were detected for the green algae *Chlamydomonas* (57.22%) and *Mesostigma* (51.36%), for the moss *Syntrichia ruralis* (54.75%) and the fern moss *Sellaginella* spp. (51.38%). These results are in agreement with other genomic comparative analyses of a wide range of plant groups, where the lower groups presented the higher contents [23,31,32]. The remaining species showed similar results (Table 1).

Dimer and Trimer most frequent motifs

For algae species, the most frequent dimer motifs were AC/GT and CA/TG (Figure 2). For example, in *C. reinhardtii*, from 548 dimer occurrences, 199 AC/GT and 233 CA/TG motifs were found. The predominant trimer motifs found were GCA/TGC, CAG/CTG and GCC/GGC (Additional file 3) with 55, 46 and 39 occurrences in 263 trimers found for algae species. For nonvascular plants, the predominant dimer motifs were AG/CT (239/1,049), AT/AT (226/1,049) and GA/TC (340/1,049), as found for *P. patens*. For mosses, the most

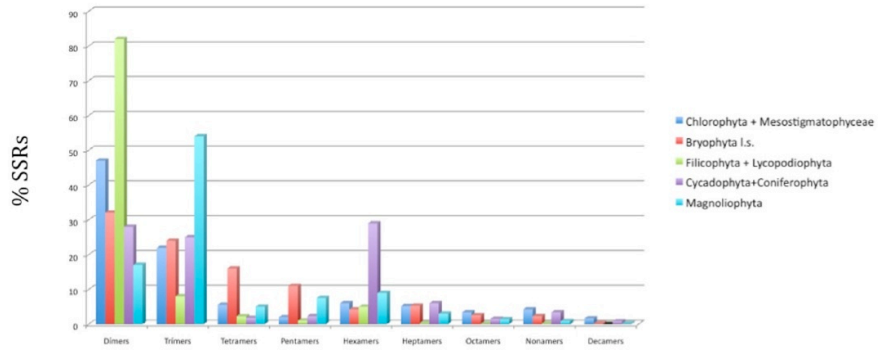


Figure 1 SSR motifs occurrences by plant group studied. SSR motifs (%) in all plant groups studied (Chlorophyta+Mesostigmatophyceae = unicelullar green algae; Bryophyta l.s. = hornworts, liverworts and mosses; Filicophyta+Lycopodiophyta = ferns; Cycadophyta+Coniferophyta = Gimnosperms; Magnoliophyta = flowering plants)

Victoria *et al. BMC Plant Biology* 2011, **11**:15
http://www.biomedcentral.com/1471-2229/11/15

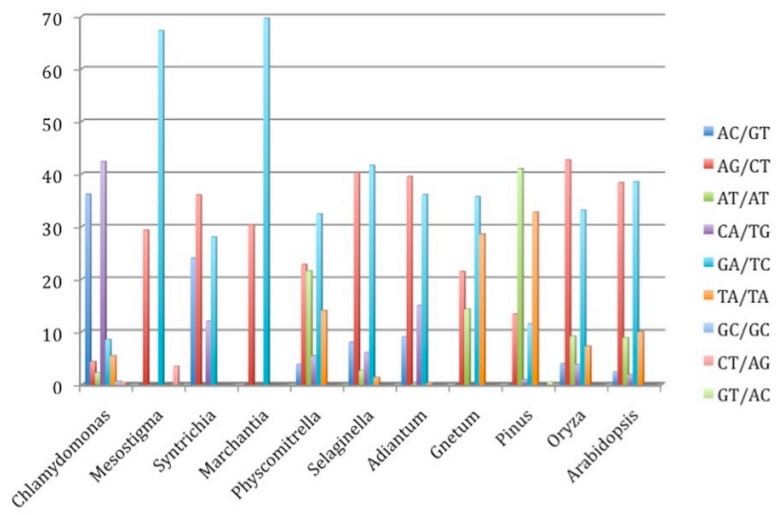


Figure 2 Predominant loci containing dinucleotide microsatellites motifs per species.

frequent trimers found within the studied species were GCA/TGC, AAG/CTT and AGC/GCT. For vascular plants, the most frequent motifs were AG/CT and GA/TC. In *O. sativa*, 246 (43%) and 191(33%) occurrences for these motifs were found, respectively, in a total of 578 dimer occurrences. The GC/GC was only detected in *C. reinhardtii*. There has been a report on the abundance of GC elements in *Chlamydomonas* genome libraries [33].

For the other species this motif has not been reported in high frequencies [10,11,23,28,34].

Among trimer motifs, there was a predominance of AAG/CTT, AGA/TCT, GGA/TCC and GAA/TTC in higher plants. In lower plants, the motifs GCA/TGC and CAG/CTG were predominant. The trimer motif CCG/CGG is predominant in the algae *C. reinhardtii* and the model moss *P. patens*, and could reflect the high GC content in these two species. However, this relationship does not hold for the other cryptogams analysed. The increased CCG/CGG frequency has been described earlier for grasses and has been related to a high GC-content [10]. In this context, the CCG/CGG increase in *Chlamydomonas* and *P. patens* was consistent, but, a previous study reported that it can not be taken as a rule, since higher GC values were found for other lower groups with low CCG/CGG contents [23]. For rice CCG/CGG is the predominant motif and its content appears to be high in the members of the grass family [11,21].

Comparing all plant groups selected for this *in silico* study, the most frequent dimer motifs found were AG/CT and GA/TC, occurring for all plant species. The most frequent trimers were AAG/CTT and GCA/TGC occurring in the 11 studied species.

Tetramers, Pentamers and Hexamers

Tetramer and pentamer motifs were rare for all studied species except for *M. viride*. This algae showed the higher frequencies in loci formed by motifs longer than three nucleotides with 36.95% of tetramer and 19.56% of pentamer motifs. Although these results are in agreement with other study [23], it is difficult to state that this is a rule for this species, since the EST database size for *Mesostigma* is the smallest one available among the studied databases. In general, tetramer and pentamer motifs predominantly found for *Oryza, Physcomitrella* and *Selaginela* where CATC/GATG, CTCC/GGAG, GATC/GATC, TGCT/AGCA (Additional file 4) and CTTCT/AGAAG, GGAGA/TCTCC, GGCAG/CTGCC, TCTCG/CGAGA and TGCTG/CAGCA (Additional file 5) and these were the most frequent motifs, at least for two out of three of these species.

Hexamer motifs were predominant in novel taxa such as gymnosperms and flowering plants [3,21,35]. *P. taeda* and *G. gnemom* showed the highest frequency (26.95%) of these motifs, but none of the hexamer motifs found in *Gnetum* and *Pinus* were found in common with other plant EST databases. However, one can not state the absence of hexamer motif patterns in plant groups, since in Bryophytes there is a possibility of patterns occurring within closely related groups. For *P. patens* and *M. polymorpha* the AGCAGG/AGCAGG, AGCTGG/CCAGGT, CAGCAA/TTGCTG and TGGTGC/GCA

Table 3 Distribution of Blast hits for *Physcomitrella patens* SSR loci sequences against several taxa with GO assignment

Taxa	Best Hits (%)
Physcomitrella patens	26.90
Oryza sativa	10.89
Vitis vinifera	10.80
Arabidopsis thaliana	9.00
Populus trichocarpa	8.60
Zea mays	7.18
Picea sitchensis	5.60
Ricinus communis	4.80
Glycine max	3.90
Sorghum bicolor	3.90
Medicago truncatula	1.48
Nicotiana tabacum	0.75
Solanum tuberosum	0.63
Micromonas pusilla	0.56
Micromonas sp.	0.55
Chlamydomonas reinhardtii	0.48
Triticum aestivum	0.47
Solanum lycopersicum	0.46
Elaeis guineensis	0.41
Hordeum vulgare	0.40
Ostreococcus lucimarinus	0.39
Ostreococcus tauri	0.35
Cyanothece sp.	0.29
Psium sativum	0.28
Brassica rapa	0.28
Spinacia oleraceae	0.25
Gossypium hirsutum	0.21
Pinus contorta	0.21

CCA motifs occur in both species (Additional file 6). Based on plastid molecular data, Marchantiophyta and Bryophyta originated about 450 Mya [36] and its possible that some repeats are conserved for recently formed groups, but it would be necessary to include others species in further analyses to confirm this hypothesis. For the other SSR types (7, 8, 9 and 10 repeats) frequencies were very low (less than 2 occurrences per motif) and were not further characterized.

Physcomitrella patens SSR loci *versus* Gene Ontology assignments

For the 4,909 SSR loci found for *P. patens* EST sequences, 1,750 had GO assignments. More than 25% of these hits were exclusive to *P. patens*. However, up to 70% of SSR loci were found as conserved across the moss and the higher plant species *O. sativa*, *Vitis vinifera* L. and *A. thaliana*. On Table 3, the distribution of the best Blast hits is presented.

Regarding biological processes, the majority of SSR loci found were involved with metabolic (32.17%) and cellular (31.02%) processes (Figure 3). Comparing all *P. patens* genome sequences with Gene Ontology assignment and those containing SSRs (Figure 4), there was a concentration of SSRs in metabolic process genes. Biological adhesion, rhythmic processes, growth and cell killing processes had the lowest SSR contents among the *P. patens* transcripts. Similar results were found comparing *P. patens* and *A. thaliana* EST libraries [37]. This author suggested that genes that are involved in protein metabolism and biosynthesis are well conserved between mosses and vascular plants. These patterns were confirmed for mosses using *Syntrichia ruralis* and *P. patens* transcript databases, respectively [38,39]. For cellular components (Figure 5) the majority of SSRs found are related to intracellular component gene sequences (52.52%) and membrane elements (12.15%). This ontology levels were reported as the majority of GO assignments in for *P. patens* annotated sequences [39]. Currently, more than half of cellular component GO annotations for *P. patens* genome [32] are related with membrane structure (Figure 6). Our results show the enrichment of SSR occurrence mainly for genes related to this structural level. The whole genome molecular function assignment level in Gene Ontology revealed a predominance of binding genes (80.51%), suggesting these are representatively higher in *P. patens* genome (Figure 7). However, when EST sequences containing SSRs are assessed with the Gene Ontology assigned molecular function (Figure 8), a relative increase of other functions is revealed. Sequences associated with binding decrease (42.81%), and those related to catalytic activity (33.76%), and structural molecule activity (10.80%) increase. These findings agree to the expectations concerning the cellular function and are consistent with ratios observed for rice, *Arabidopsis*, and for the bryophytes *Syntrichia ruralis* and *P. patens* [32,38-41]. The higher occurrence of SSR loci in this ontology level indicate a good potential for using these molecular markers to saturate pathways associated to those functions described above.

Predicted coding for SSR loci

The predicted amino acid content for the SSR loci detected in the eleven species studied is shown in Figure 9. The amino acids arginine (Arg), alanine (Ala) and Serine (Ser) were predominant for all species. Alanine was predominant for the majority of cryptogams, ranging from 14.85% to 29.7%. Exceptions were observed for *Adiantum*, *Mesostigma* and *Physcomitrella*, in which serine (Ser), glutamic acid (Glu) and leucine (Leu) were the predominant amino acid (up to 17%). Serine (up to 11%) was predominant for fern species and for *Gnetum*

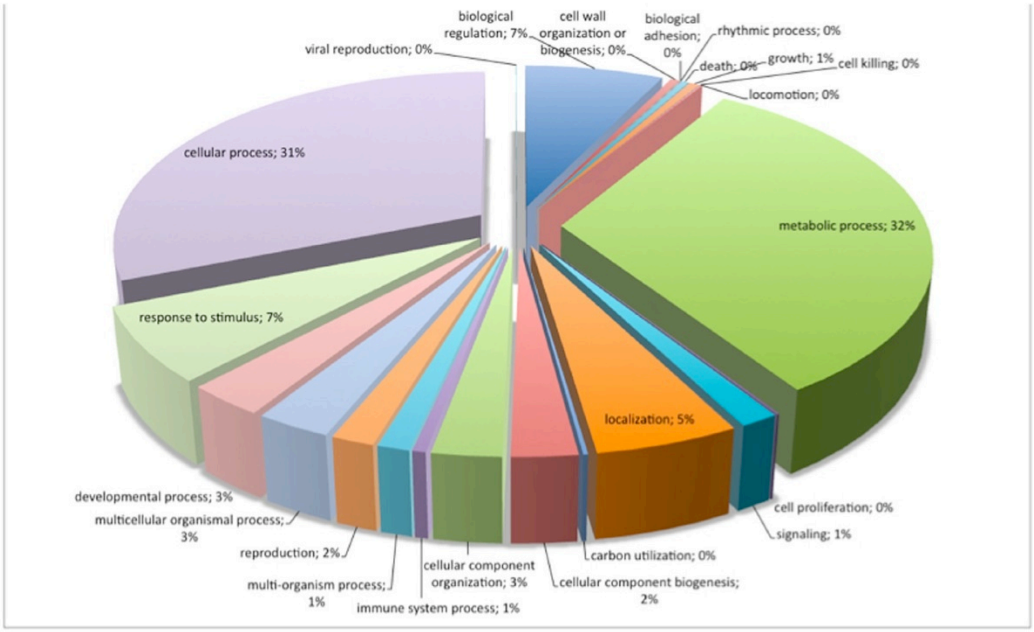


Figure 3 Distribuition of *Physcomitrella patens* SSR loci within sequences of known biological processes in Gene Ontology.

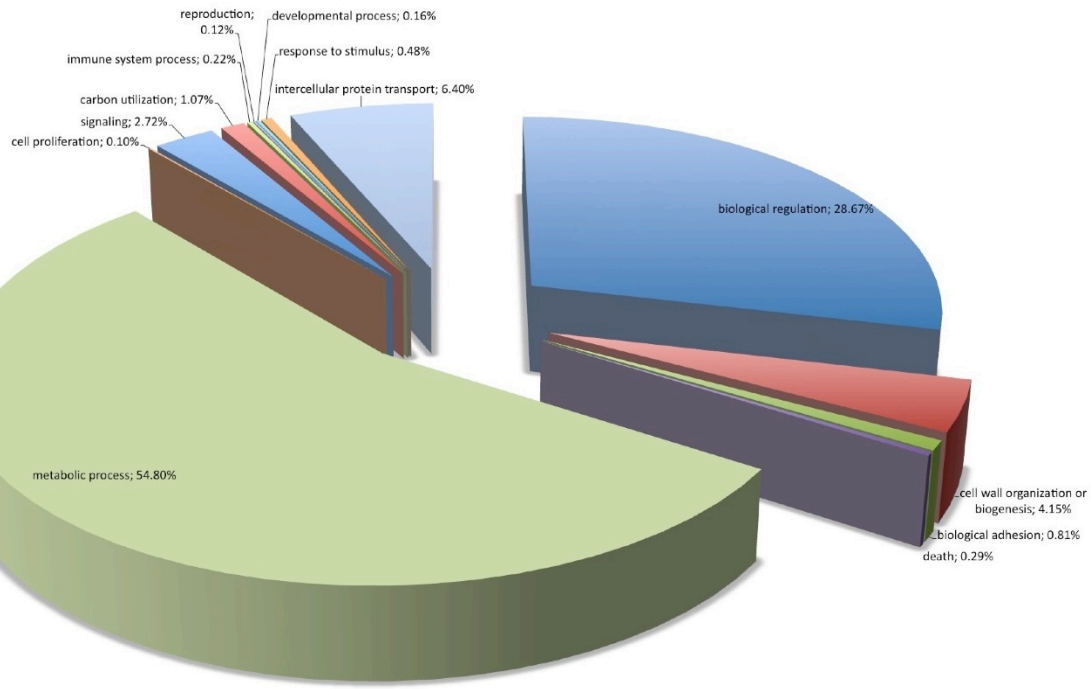


Figure 4 Distribuition of *Physcomitrella patens* genome sequences with Gene Ontology assignments into biological processes. (Data: Rensing et al., 2008).

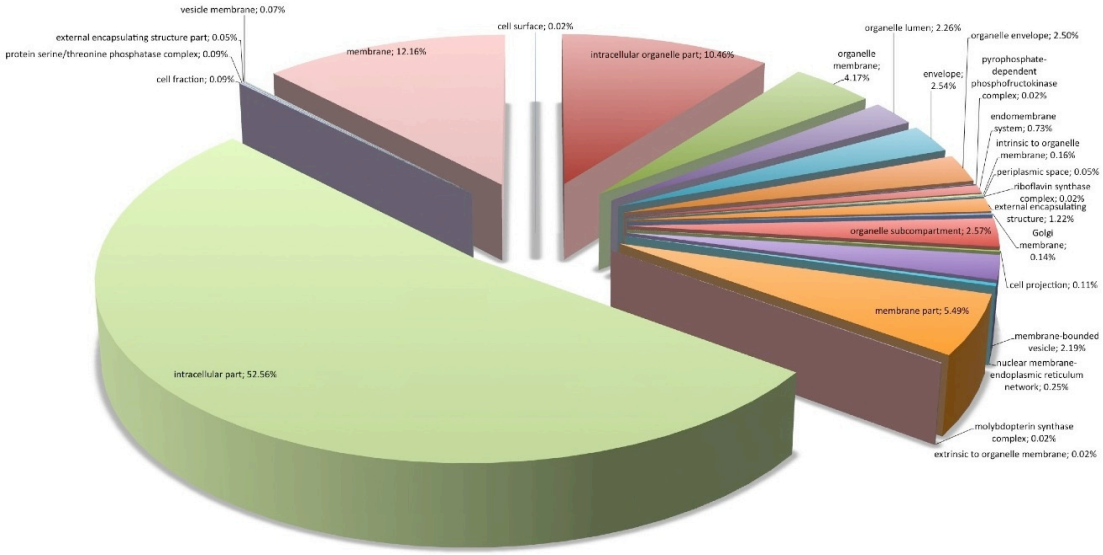


Figure 5 Distribuition of *Physcomitrella patens* SSR loci within sequences of known cellular component in Gene Ontology.

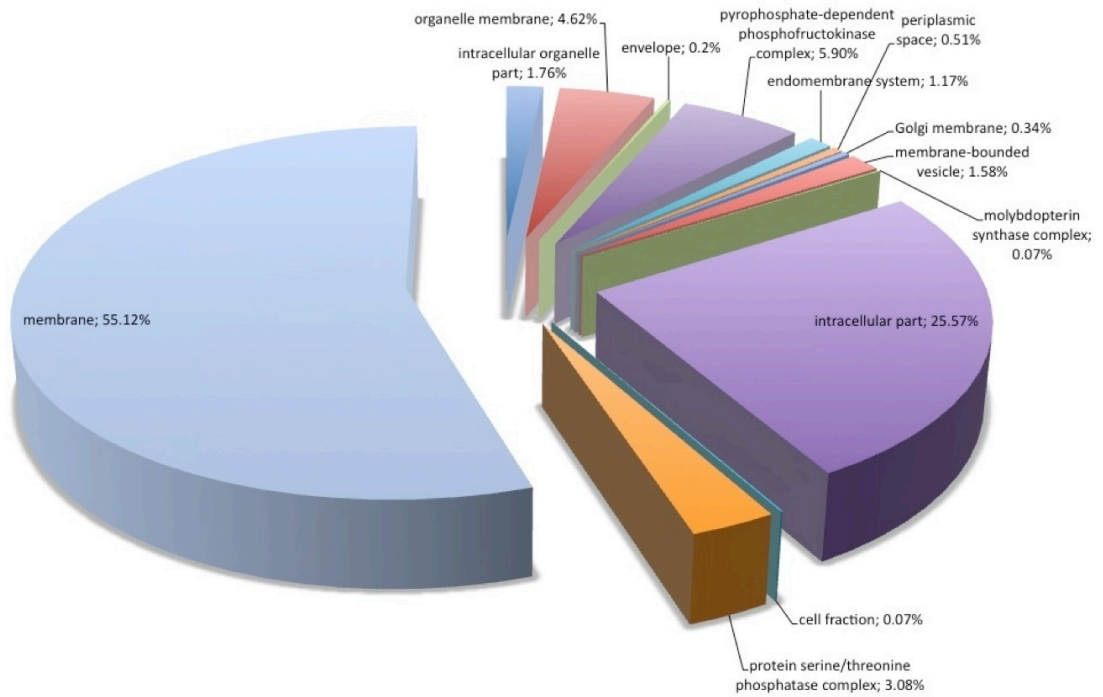


Figure 6 Distribuition of *Physcomitrella patens* genome sequences with Gene Ontology assignments into cellular component. (Data: Rensing et al., 2008).

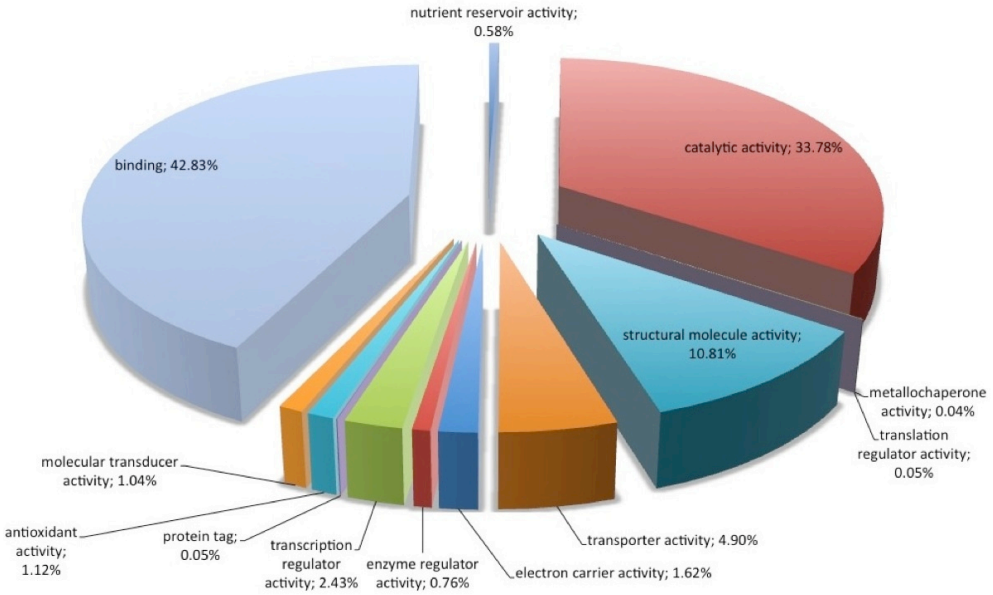


Figure 7 Distribuition of *Physcomitrella patens* SSR loci within sequences of known molecular function in Gene Onthology.

and *Arabidopsis*, *Pinus* and *Oryza* showed arginine as the predominant amino acid (10.46% and 23.31%, respectively). Tyrosine (Tyr), asparagine (Asp), aspartic acid (Asn) were the amino acids found at lower frequencies among SSR loci for all species and were practically absent in the algae species surveyed. In bryophytes, methionine was only found in *Physcomitrella*, but at a small frequency (1.7%). For all higher plant species databases used in this survey, arginine, alanine, serine, glutamic acid, proline (Pro) and leucine were among the

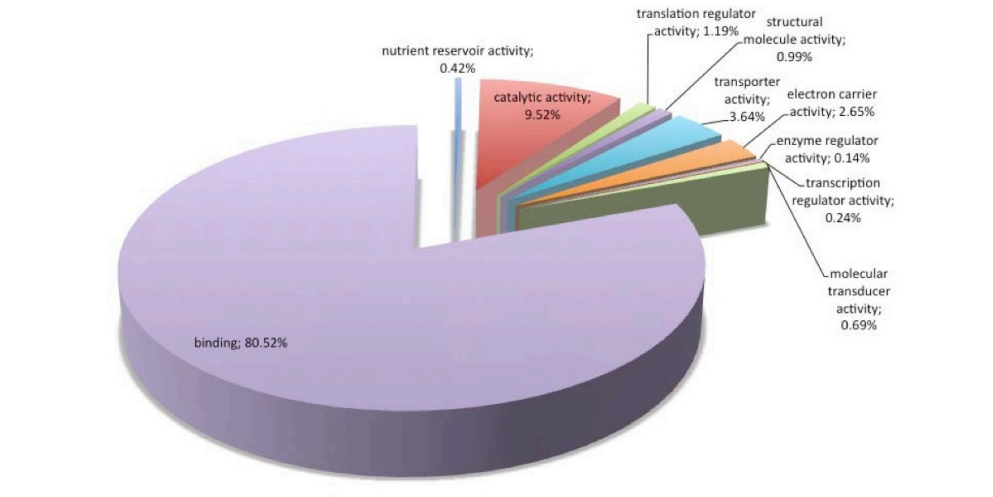


Figure 8 Distribuition of *Physcomitrella patens* genome sequences with Gene Ontology assignments into molecular function. (Data: Rensing et al., 2008).

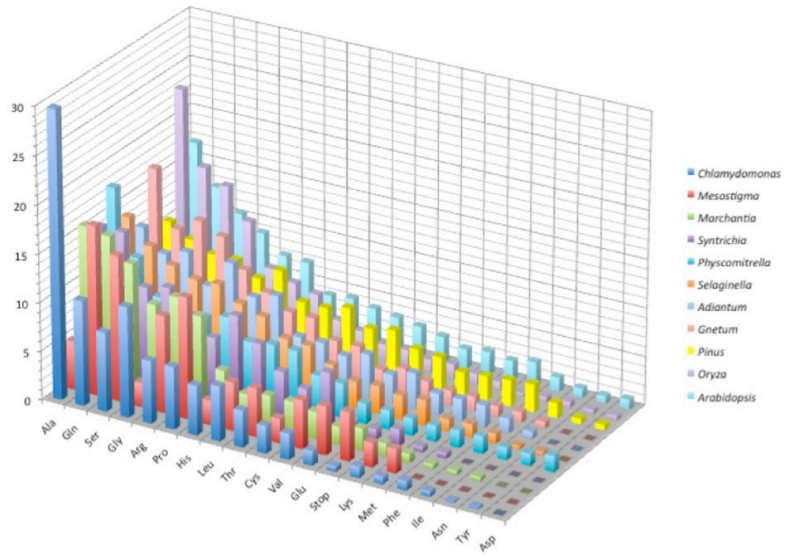


Figure 9 Predicted amino acid occurrences in SSR loci within plant groups studied.

predominant amino acids, agreeing with previous reports for flowering plants [11,3,22,42-45]. No reports were found for amino acid distribution in SSR loci in lower plants.

The small EST databases available for some species did not seem to have hampered the results, since the predicted loci distribution found were consistent within the taxonomic groups. The absence of a relationship between genome size and tandem repeat loci content were reported based in grass genome studies [11], where large genomes such as sugarcane (*Saccharum officinarum* L.), maize and wheat did not present higher frequencies of SSR loci.

Relationship of Codon-bias with EST-SSR motif occurrences

The high GC-content in some EST-SSR motifs found in the present study can be a result of a codon usage preference by plant species. When we compare the codon usage for the model species included in this study (*Chlamydomonas reinhardtii*, *Physcomitrella patens*, *Oryza sativa* and *Arabidopsis thaliana*) the occurrence of some repeat motifs are reflected in codon-bias known for each species. Higher frequencies of GC were found in the first and third codon position for all four species. However, for the basal plant (*C. reinhardtii*), the preference for GC3 was much higher than the other three species. The first (GC1) and the third (GC3) codon position reached 64.8% and 86.21% of the occurrences, respectively. For rice, GC1 and GC3 frequencies were 58.19% and 61.6%, respectively. For the other model plants, the occurrences at GC3 were lower than the occurrences in GC1, i.e., for *Physcomitrella patens* and *Arabidopsis thaliana*, GC1 (55.49% and 50.84%, respectively) and GC3 (54.6% and 42.4%, respectively) values were found. When one associates these codon usage values with the SSR motif frequencies found, a striking result is obtained for *C. reinhardtii* and rice. In the first, the most frequent motifs were GCA/TGC, CAG/CTG and GCC/GGC and could be explained by the GC1s and GC3s codon preference. In rice the CCG/CGG predominant motif could also be a reflection of GC3s codon preference. For *Arabidopsis*, the most frequent motif found in this study (GAA/TTC) is also the most preferred codon used by this species (GAA) with 34.3% of the occurrences. It also reflects the GC1 preference in the codon usage in this species. In the model moss species the most frequent motifs do not show a relationship with the GC codon usage (Figure 10). Despite the similarities in average codon bias between *P. patens* and *Arabidopsis thaliana*, the distribution pattern is different, with 15% of moss genes being unbiased [46]. An association between the frequency of microsatellite motifs and codon usage could explain the occurrences found in *P. patens*. For example, the most representative motifs GCA/TGC, AAG/CTT and AGC/GCT are also found among the most used codons GCA, AAG and AGC (20.7%, 33.6% and 15%, respectively).

The width of the GC3 distribution in flowering plants was found to be a result of variation in the levels of

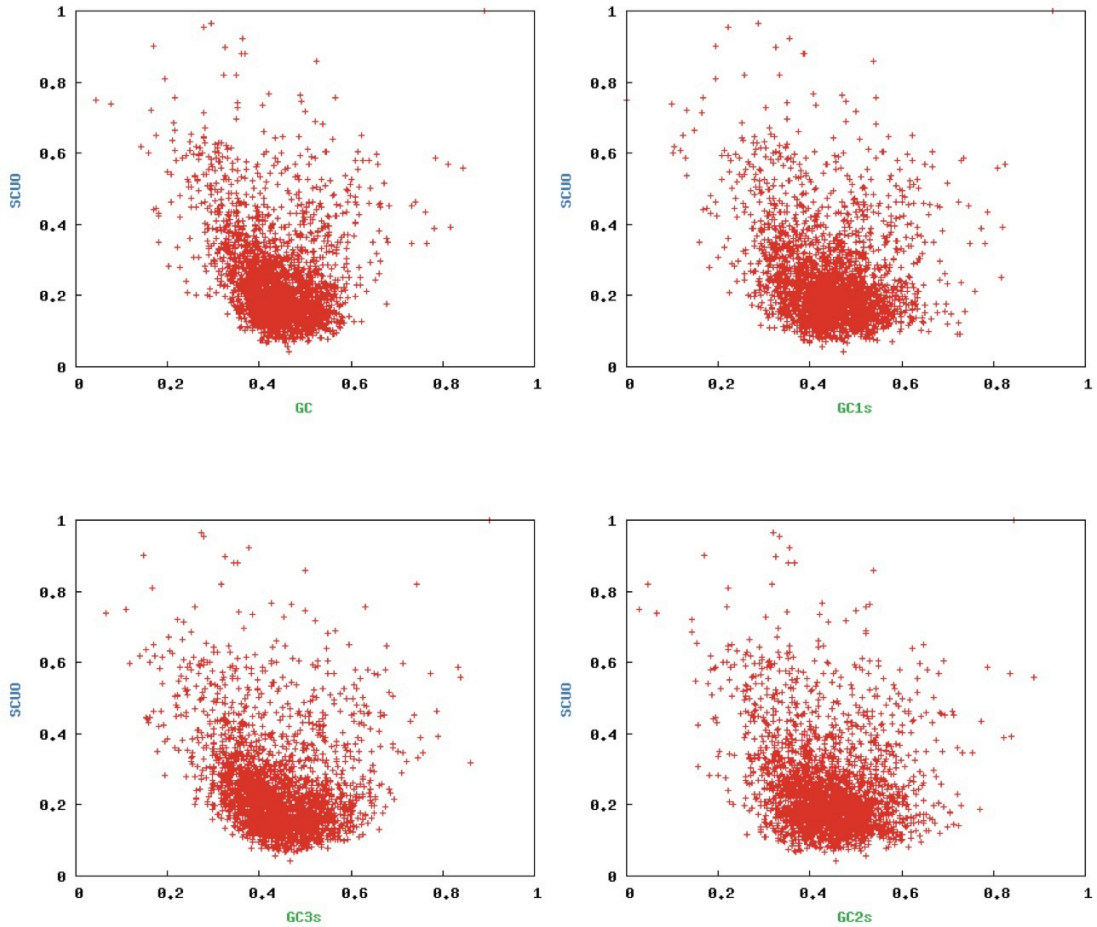


Figure 10 Correlation between synonymous codon usage bias and GC composition in *Physcomitrella patens* EST-SSR sequences.

directional mutation pressure or selection against mutational biases. Likewise, the low frequency of GC2 occurrences is a result of a strong selective pressure against peptide substitution. The balance between these forces could be shaping the distribution of EST-SSR by means of codon usage preference [47].

Positive and negative selection sites in EST-SSR across species

SSRs represent hyper mutable loci subject to reversible changes in their length [8]. Significant differences in SSR representations exist even among closely related species, suggesting that SSR abundance may change relatively rapidly during evolution [48]. To infer about the selection pressures (dN/dS ratio) on EST-SSR found for the 11 species chosen for this work, we used the common most frequent motif in all species (AAG/CTT and GCA/TGC). The dN-dS test revealed few negatively selected sites in the triplets for each EST-SSR (Additional file 7). The positive selection in SSR based sequence was reported in other studies [8,49-51]. More than 50% of sites for both motifs analyzed across species were under a positive selection (dN/dS > 1), suggesting a weak selection pressure on these EST-SSR motifs, as was reported for other species [52,53]. The occurrence of selective sweeps or background selection in ancestral lineages [54] cannot be discarded, however it could not be tested with the present data.

In silico transferability of EST-SSR across species

Across-species transferability of EST-SSRs is greater than genomic SSRs, as they originate from expressed regions and therefore they are more conserved across a number of related species [6].

The virtual PCR shows a lower transferability of *Chlamydomonas reinhardtii* EST-SSR for most of the

Victoria *et al. BMC Plant Biology* 2011, **11**:15
http://www.biomedcentral.com/1471-2229/11/15

plant species tested. The best results were found for *Adiantum* and *Arabidopsis*, where successful rates of positive EST-SSR amplicons derived from algae were 26% and 9%, respectively. When EST-SSR primers designed from *Arabidopsis* were used against other species, again low transferability rates were found, being the best positive cases found in *Physcomitrella*, *Pinus* and rice with amplification rates of 1.04%, 1.20% and 1.90%. The summary of *in silico* PCR results can be accessed in the Additional files section of this article. Some reports suggest that SSR markers have higher transferability rates when used between closely related species [6,22,55]. In this work virtual PCR amplification did follow the same trend.

For the positive EST-SSRs found for the *in silico* transfer, ten sets of *Physcomitrella* EST-SSR primers were used to illustrate the transferability results using an electronic tool [56] to simulate gel electrophoresis (Figure 11). For the three tested EST-databases only two primers amplified a single locus in each species (SSR9 and SSR10). In the other sets 2, 3 and even 4 virtual amplicons were observed (Additional file 8). For *Chlamydomonas*, 70% of the tested primers resulted in one amplicon and 10% each resulted in 2, 3 or 4 amplifications. However, only 20% of amplicons obtained in this algae species are related to the EST-SSR sequence, suggesting that the majority of designed EST-SSR primers act as degenerate when applied to *Chlamydomonas*. For rice, 30%, 40% and 10% of tested primers resulted in one, two or three amplifications, respectively. In *Arabidopsis* 40%, 40% and 20% of tested primers results in one, two or three amplifications, respectively. For both flowering plants, 50% of tested primers amplified moss EST-SSR homologue sequences, showing a high rate of success for transferability across species. These results agree with other studies where the transfer success rates decrease with the increasing evolutionary distance [55,57-60]. The use of this molecular marker across distant taxonomical groups are not impossible, however our findings confirm that only a few retain their EST-SSR homologue sequences, making this effort hardly worthwhile [61].

Conclusions

These results make it possible to create strategies for transferring molecular markers based on microsatellites from model to orphan species.

Microsatellites were found in all species studied and variable transfer rates were found as a function of genetic distance among taxa. The motifs found are influenced by species codon usage preference. The two most common motifs among the eleven species are under a positive selection pressure. Primers generating one amplicon in the genome of origin may generate multiple amplicons in other taxa and only a few retain their original targeting sequence. The similarities between the

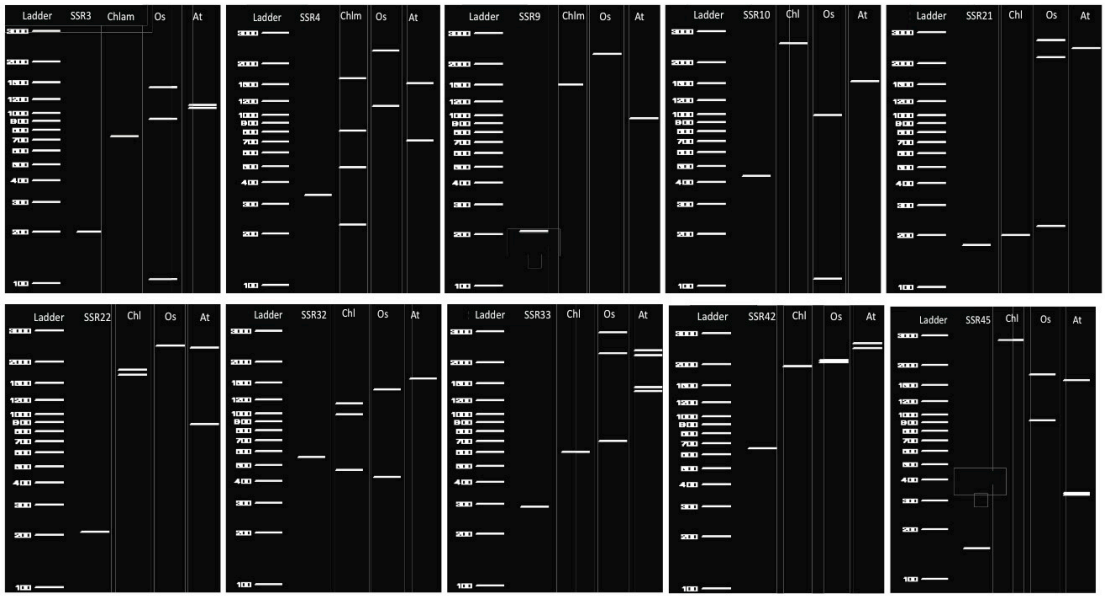


Figure 11 Eletronical eletrophoresis gel for 10 primers set design for *Physcomitrella patens* EST-SSR (SSRn) across *Chlamydomonas reinhardtii* (Chml) *Oryza sativa* (Os) and *Arabidopsis thaliana* (At) EST databases.

results here presented and other initiatives using similar bioinformatics Perl scripts, such as MISA [23], support *SSRLocator* as a useful tool for SSR survey analyses.

Methods

An exploratory *in silico* analysis of SSRs was made in ESTs databases of 11 taxa, as follows: two unicellular green algae (*Chlamydomonas reinhardtii* Dang, *Mesostigma viride* Lauterborn.), three bryophytes s. l. [*Marchantia polymorpha* L., *Physcomitrella patens* and *Syntricha ruralis* (Hedw.) Weber & Mohr], two ferns (*Selaginella* spp. and *Adiantum capillus-veneris* L.), two gymnosperms (*Gnetum gnemon* L. and *Pinus taeda* L.) and two flowering plants, a monocot (*Oryza sativa*) and a dicot (*Arabidopsis thaliana*). These species were chosen because the amount of available ESTs data in Genbank (NCBI). As these databases may have redundancy, we used the program CAP3 [62] for MacOX, to construct contigs with the sequences and get non-redundant sequences for each database following the default settings.

Taxa data were loaded into the software *SSRLocator* [63], to investigate the presence of tandem repetitive elements (SSRs). The analysis was performed following the search parameters for repetitive elements in class I (≥ 20 bp) described as more efficient molecular markers [17]. Data resulting from *in silico* analyses were assessed for occurrence patterns in chosen taxa databases. The same analysis was performed using MISA script http://pgrc.ipk-gatersleben.de/misa/ software to search for SSR occurrences per contig. Several instructions in the algorithm used in *SSRLocator* resemble those from MISA [19] and SSRIT [17]. However, additional instructions have been inserted in *SSRLocator's* code. Instead of allowing the overlap of a few nucleotides when two SSRs are adjacent to each other and one of them is shorter than the minimum size for a given class as found in MISA and SSRIT, a module written in Delphi language records the data and eliminates such overlaps. For GC content, Perl scripts were used and the results were stored in text files (.txt) for later comparative analyses.

For the predicted amino acid contents in the SSR loci, an additional routine script was written in the *SSRLocator* software. This script determined which amino acids were coded by trimer, hexamer and nonamer motifs found in the EST database analysed [63].

To validate the frequencies obtained using the *SSRLocator* software, the *Physcomitrella patens* EST database was chosen.

This database was run with other SSR search scripts and softwares, such as MISA [19] and SPUTINIK [64], running in SCIROKO package [30], MINE SSR http://www.genome.clemson.edu/resources/online_tools/ssr, SSRIT following the SSR categories defined above [17]. The results were exported into Microsoft Excel spreadsheets (MacOSX-Oficce 2008) and respectively grouped by taxon.

A codon-bias for the model plants included in this research (*Chlamydomonas reinhardtii*, *Physcomitrella patens*, *Oryza sativa* and *Arabidopsis thaliana*) was made comparing with the preferencial codon table for each species available at http://www.kazusa.or.jp/codon/. The sequences containing EST-SSR for *Physcomitrella patens* was submitted to CodonO server [65] to confirm the preferencial codon usage compared with the know codon table for this species. To investigate the selective pressure on the triplets on the EST-SSR which occurs in all studied species a dN-dS statistics [66] was used to verify the synonymous and noun-synonymous substitutions in the preferential codons nearby the repeats chosen using the molecular phylogenetics package MEGA4 [67].

The *Physcomitrella patens* SSR results were run through a Gene Ontology (GO) assignment database in order to assess associations between SSR loci and biological processes, cellular components and molecular function of known genes. A fasta file with all EST-SSRs found in *P. patens* was subjected to Blast2GO software and ran against the GO annotated sequences, and the obtained hits were compiled.

To verify the potential transferability of this molecular markers we have tested *in silico* all EST-SSR found for the plant ancestral lineage, and for the derivative plant group, represented here by the green algae *Chlamydomonas reinhardtii* and *Arabidopsis thaliana*, across the others species EST database used for the present SSR survey. Electronic PCR [68] was used to verify the transferability of EST-SSRs across studied species. The positive results found were used to simulate a gel electrophoresis with aid of SIMGEL.exe included in the SPCR package [56] using the *Physcomitrella patens* EST-SSR sequences to design primers and *Chlamydomonas*, rice and *Arabidopsis* as templates. The virtual amplicons resulted for each primer set tested across species were aligned to verify the homology between the amplicons.

Additional material

Additional file 1: Patterns of occurrence for dimer SSR motifs in percentage.

Additional file 2: Patterns of occurrence for trimer SSR motifs in percentage.

Additional file 3: Predominant trinucleotide microsatelites motifs loci occurrences per species.

Additional file 4: Predominant tetramers microsatelites motifs loci occurrences per species.

Additional file 5: Predominant pentamers microsatelites motifs loci occurrences per species.

Additional file 6: Predominant hexamers microsatelites motifs loci occurrences per species.

Additional file 7: dN/dS table for the common most frequent motifs for 11 species tested EST databases.

Additional file 8: Eletronical PCR results table.

Acknowledgements
We would like to thank the Developmental Center of Technology (CDTec/UFPEL) for the support to the first author. This work was supported by the National Council for Scientific and Technological Development CNPq (process # 480938/2009-1 and 475122/2007-0).

Author details
[1]Plant Genomics and Breeding Center, Faculdade de Agronomia Eliseu Maciel, Universidade Federal de Pelotas, RS, Brasil. [2]Graduate Program in Biotechnology, Universidade Federal de Pelotas, RS, Brasil.

Authors' contributions
FCV carried out all *in silico* studies, including the SSR survey, the electronic PCR and the sequence alignment for selective sites mining and drafted the manuscript. LCM created the SSR script used and participated in the design of the study. ACO conceived the study, and participated in its design and coordination. All authors read and approved the final manuscript.

Received: 10 July 2010 Accepted: 19 January 2011
Published: 19 January 2011

References

1. Morgante M, Olivieri AM: **PCR-amplified microsatellites as markers in plant genetics.** *The Plant Journal* 1993, **3(1)**:175-182.
2. Jurka J, Pethiyagoda C: **Simple repetitive DNA sequences from Primates: Compilation and analysis.** *Journal of Molecular Evolution* 1994, **40**:120-126.
3. Tóth G, Gáspári Z, Jurka J: **Microsatellites in different eukaryotic genomes: survey and analysis.** *Genome Research* 2000, **10**:967-981.
4. Iyer RR, Pluciennik A, Rosche WA, Sinder RR, Wells RD: **DNA polymerase III proofreading mutants enhance the expansion and deletion of triplet repeat sequence in Escherichia coli.** *Journal of Biological Chemistry* 2000, **275(3)**:2174-2184.
5. Mirkin SM: **DNA structures, repeat expansions and human hereditary disorders.** *Current Opinion in Structural Biology* 2006, **16(3)**:351-358.
6. Varshney RK, Graner A, Sorrells ME: **Genic microsatellite markers in plants: features and applications.** *Trends in Biotechnology* 2005, **23(1)**:48-55.
7. Varshney RK, Hoisington DA, Tyagy AK: **Advances in cereal genomics and applications in crop breeding.** *Trends in Biotechnology* 2006, **24(11)**:490-499.
8. Kashi Y, King DG: **Simple sequence repeats as advantageous mutators in evolution.** *Trends Genet* 2006, **22**:253-259.
9. Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS: **Transferable EST-SSR markers for the study of polymorphism and diversity in bread wheat.** *Molecular Genetics and Genomics* 2003, **270**:315-323.
10. Morgante M, Hanafey M, Powell W: **Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes.** *Nature Genetics* 2002, **3(2)**:194-200.
11. Maia LC, Souza VQ, Kopp MM, Carvalho FIF, Oliveira AC: **Tandem repeat distribution of gene transcripts in three plant families.** *Genetics and Molecular Biology* 2009, **32(4)**:1-12.
12. Subramanian S, Mishra RK, Singh L: **Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions.** *Genome Biology* 2003, **4(2)**:R13.
13. Li YC, Korol AB, Fahima T, Beiles A, Nevo E: **Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review.** *Molecular Ecology* 2002, **11**:2453-2465.
14. Marcotte EM, Pellegrini M, Yeates TO, Eisenberg D: **A census of protein repeats.** *Journal of Molecular Biology* 1999, **293**:151.
15. Kashi Y, King D, Soller M: **Simple sequence repeats as a source of quantitative genetic variation.** *Trends in genetics* 1997, **13**:74-78.
16. Wren JD, Forgacs E, Fondon JW III, Pertsemlidis A, Cheng SY, Gallardo T, Williams RS, Shohet RV, Minna JD, Garner HR: **Repeat polymorphisms within gene regions: phenotypic and evolutionary implications.** *American Journal of Human Genetics* 2000, **67**:345-356.
17. Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S: **Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential.** *Genome Research* 2001, **11(8)**:1441-52.
18. McCouch SR, Teytelman L, Xu Y, *et al*: **Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.).** *DNA research* 2002, **9(6)**:199-207.
19. Thiel T, Michalek W, Varshney RK, Graner A: **Exploiting EST databases for the development of cDNA derived microsatellite markers in barley (*Hordeum vulgare* L.).** *Theoretical and Applied Genetics* 2003, **1-6**:411-422.
20. Nicot N, Chiquet V, Gandon B, Amilhat L, Legeai F, Leroy P, Bernard M, Sourdille P: **Study of simple sequence repeat (SSR) markers from wheat expressed sequence tags (ESTs).** *Theoretical and Applied Genetics* 2004, **1-9(4)**:8008-5.
21. Lawson MJ, Zhang L: **Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes.** *Genome Biology* 2006, **7**:R14, 3.
22. Zhang L, Yuan D, Yu S, Li Z, Cao Y, Miao Z, Qian H, Tang K: **Preference of simple sequence repeats in coding and non coding regions of *Arabidopsis thaliana*.** *Bioinformatics* 2004, **20**:1081-1086.
23. von Stackelberg MV, Rensing SA, Reski R: **Identification of genic moss SSR markers and a comparative analysis of twnty-four algal and plant gene indices reveal species-specific rather than group-specific characteristics of microsatellites.** *BMC Plant Biology* 2006, **6**:9.
24. Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ: **Microsatellite markers from sugarcane (Saccharum spp.) ESTs cross transferable to erianthus and sorghum.** *Plant science* 2001, **16(6)**:1115-1123.
25. Kantety RV, La Rota M, Matthews DE, Sorrells ME: **Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat.** *Plant molecular biology* 2002, **48(5-6)**:5-1-11.
26. Asp T, Frei UK, Didion T, Nielsen KK, Lübberstedt T: **Frequency, type, and distribution of EST-SSRs from three genotypes of *Lolium perenne*, and their conservation across orthologous sequences of *Festuca arundinacea*, *Brachypodium distachyon*, and *Oryza sativa*.** *BMC plant biology* 2007, **12(7)**:36.
27. Echt CS, May-Marquardt P, Hseih M, Zahorchak R: **Characterization of microsatellire markers in eastern white pine.** *Genome* 1996, **39**:1102-1108.
28. Echt CS, May-Marquardt P: **Survey of microsatellite DNA in pine.** *Genome* 1997, **40**:9-17.
29. Fisher PJ, Gardner RC, Richardson TE: **Single locus microsatellites isolated using 5'anchored PCR.** *Nucleic Acids Research* 1996, **24**:4369-4372.
30. Kofler R, Schlotterer C, Lelley T: **SciRoKo: A new tool for whole genome microsatellite search and investigation.** *Bioinformatics* 2007, **23**:1683-1685.
31. Qiu Y-L, Lee J, Bernasconi-Quadroni B, Soltis DE, *et al*: **The earliest Angiosperms: Evidence from mitochondrial, palstid and nuclear genomes.** *Nature* 1999, **402**:404-407.
32. Rensing SA, Lang D, Zimmer AD, *et al*: **The *Physcomitrella* genome reveals insights into the conquest of land by plants.** *Science* 2008, **319**:64-69.
33. Wakarchuk WW, Müller FW, Beck C: **F. Two GC-rich elements of *Chlamydomonas reinhardtii* with complex arrangements of directly repeated sequences motifs.** *Plant Molecular Biology* 1992, **18**:143-146.
34. Yashoda R, Sumathi R, Chezhian P, Kavitha S, Ghosh M: ***Eucalyptus* microsatellites mined *in silico*: survey and evaluation.** *Journal of Genetics* 2008, **87(1)**:21-25.
35. Jiang D, Zhong GY, Hong QB: **Analysis of microsatellites in citrus unigenes.** *Acta genetica Sinica* 2006, **33(4)**:345-53.
36. Magallón S, Hilu KW: **Land plants (Embryophyta).** In *The Timetree of Life.* Edited by: S. B. Hedges, S. Kumar. Oxford, University Press; 2009:133-137.
37. Nishiyama T, Fujita T, Shin-I T, Seki M, Nishide H, Uchiyama I, Kamiya A, Carninci P, Hayashizaki Y, Shinozaki K, Kohara Y, Hasebe M: **Comparative genomics of Physcomitrella patens gametophytic transcriptome and Arabidopsis thaliana: Implication for land plant evolution.** *PNAS* 2003, **100(13)**:8007-8012.
38. Oliver MJ, Dowd SE, Zaragoza J, Mauget SA, Payton PR: **The rehydration transcriptome of the desiccation-tolerant bryophyte Tortula ruralis: Transcript classification and analysis.** *BMC Genomics* 2004, **5**:89.
39. Lang D, Eisinger J, Reski R, Resing SA: **Representation and High-Quality Annotation of the Physcomitrella patens Transcriptome Demonstrates a High Proportion of Proteins Involved in Metabolism in Mosses.** *Plant Biology* 2005, **7**:238-250.

40. Ware D, Jaiswal P, Ni J, Pan X, Chang K, Clark K, Teytelman L, Schmidt S, Zhao W, Cartinhour S, McCouch S, Stein L: **Gramene: a resource for comparative grass genomics.** *Nucleic Acids Research* 2002, **30**:103-105.
41. Rhee SY, Beavis W, Berardini TZ, *et al*: **The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community.** *Nucleic Acids Research* 2003, **31**:224-228.
42. Jung S, Abbott A, Jesudurai C, Tomkins J, Main D: **Frequency, type, distribution and annotation of simple sequence repeats in Rosaceae ESTs.** *Functional & integrative genomics* 2005, **5(3)**:136-43.
43. La Rota M, Kantety RV, Yu JK, Sorrells ME: **Nonrandom distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat, and barley.** *BMC Genomics* 2007, **18(1)**:23, 6.
44. Varshney RK, Thiel T, Stein N, Langridge P, Graner A: **In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species.** *Cell Mol Biol Lett* 2002, **7**:537-546.
45. Parida SK, Anand Raj Kumar K, Dalal V, Singh NK, Mohapatra T: **Unigene derived microsatellite markers for the cereal genomes.** *Theor Appl Genet* 2006, **112**:808-817.
46. Rensing SA, Fritzomsky D, Lang D, Reski R: **Protein encoding genes in an ancient plant: analysis of codon usage, retained genes and splice sites in a moss, *Physcomitrella patens*.** *BMC genomics* 2005, **6**:43.
47. Kawabe A, Miyashita NT: **Patterns of codon usage bias in three dicot an four monocot plant species.** *Genes and Genetic System* 2003, **78**:343-352.
48. Mrázek J: **Analysis of distribuition indicates diverse functions of simple sequence repeats in *Mycoplasma* genomes.** *Molecular Biology and Evolution* 2006, **23**:1370-1385.
49. King DG, Kashi Y: **Indirect selection for mutuability.** *Heredity* 2007, **99**:123-124.
50. King DG, Soller M: **Variation and fidelity: The evolution of simple sequence repeats as functional elements in adjustable genes.** In *Evolutionary Theory and Processes: Modern Perspectives*. Edited by: Wasser SP. Kluwer Academic Publisher, the Netherlands; 1999:65-82.
51. Vigouroux Y, Matsuoka Y, Doebley J: **Directional evolution for microstellites size in maize.** *Molecular Biology and Evolution* 2003, **20**:1480-1483.
52. Ellis JR, Burke JM: **EST-SSRs as a resource for population genetic analyses.** *Heredity* 2007, **99**:125-132.
53. Yatabe Y, Kane NC, Scotti-Saintagne C, Rieseberg LH: **Rampant gene exchange across a strong reproductive barrier between the annual sunflowers, *Helianthus annuus* and *H petiolaris*.** *Genetics* 2007, **175**:1883-1893.
54. Wrigth SI, Gaut BS: **Molecular population genetics and the search for adaptative evolution in plants.** *Molecular Biology and Evolution* 2005, **22(3)**:506-519.
55. Chapman MA, Hvala J, Strever J, *et al*: **Development, polymorphism, and cross-taxon utility of EST-SSR markers from safflower (*Carthamus tinctorius* L.).** *Theoretical and Applied Genetics* 2009, **120**:85-91.
56. Cao Y, Wang L, XU K, Kou C, Zhang Y, Wei G, He J, Wang Y, Zhao L: **Information theory-based algorithm for *in silico* prediction of PCR products with whoke genomic sequences as templates.** *BMC bioinformatics* 2005, **6**:190.
57. Brondani C, Rangel PHN, Borba TCO, Brondani RPV: **Transferability of microsatellite and sequence tagged site markers in *Oryza* species.** *Hereditas* 2003, **138**:187-192.
58. Castillo A, Budak H, Varshney RK, Dorado G, Graner A, Hernandez P: **Tranferability and polimorphism of barley EST-SSR markersused for phylogenetic analysus in *Hordeum chilense*.** *BMC plant biology* 2008, **8**:97.
59. Yodav OP, Mitchell SE, Fulton TM, Kresovich S: **Tranferring molecular markers from sorghum, rice and other cereals to pearl millet and identifying polumorphic markers.** *Journal of SAT Agricultural Research* 2008, **6**:1-4.
60. Zeid M, Yu JK, Goldowitz I, Denton ME, *et al*: **Cross-amplification of EST-derived markers among 16 grass species.** *Field Crops Research* 2010, **118**:28-35.
61. Barbará T, Palma-Silva C, Paggi GM, Bered F, Fay MF, Lexer C: **Cross-species transfer of nuclear microsatellites markers: potential and limitations.** *Molecular Ecology* 2007, **16**:3759-3767.
62. Huang X, Madan A: **CAP3: A DNA sequence assembly program.** *Genome Research* 1999, **9**:868-877.
63. Maia LC, Palmieri DA, Souza VQ, Kopp MM, Carvalho FIF, Oliveira AC: **SSR Locator: Tool for Simple Sequence Repeat Discovery Integrated with Primer Design and PCR Simulation.** *International Journal of Plant Genomics* 2008, Article ID 412696, 9 pages.
64. Abajan C, SPUTINIK: 1994 [http://espressosoftware.com/sputnik/index.html].
65. Angellotti MC, Bhuiyan SB, Chen G, Wan X-F: **CodonO: codon usage bias analysis within and across genomes.** *Nucleic Acids Research* 2007, **35**: W132-W136.
66. Yang Z, Bielawski JP: **Statistical methods for detecting molecular adaptation.** *Trends in Ecology and Evolution* 2000, **12**:496-503.
67. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.** *Molecular Biology and Evolution* 2007, **24**:1596-1599.
68. Schuler GD: **Sequence mapping by eletronic PCR.** *Genome Research* 1997, **7(5)**:541-550.

doi:10.1186/1471-2229-11-15
Cite this article as: Victoria *et al.*: *In silico* comparative analysis of SSR markers in plants. *BMC Plant Biology* 2011 **11**:15.

5. ARTIGO 3. Phylogenetic relationships and selective pressure on iron uptake genes in plants

(A ser submetido para publicação no periódico Molecular Biology and Evolution)

Phylogenetic relationships and selective pressure on iron uptake genes in plants

Filipe de Carvalho Victoria[1,2], Clauber Mateus Priebe Bervald[3], Luciano Carlos da Maia[2] Rogerio Oliveira de Sousa[4] , Olivier Panaud [5] and Antônio Costa de Oliveira[2].

[1]*Graduate Student in Biotechnology, Centro de Biotecnologia, Universidade Federal de Pelotas, RS, Brasil.*

[2]*Plant Genomics and Breeding Center, Faculdade de Agronomia Eliseu Maciel, Universidade Federal de Pelotas, RS, Brasil.*

[3]*Ricetec Seeds* Inc*., Rua 18 de novembro, 341, Navegantes, 90240-040 - Porto Alegre, RS – Brasil.*

[4]*Soil Science Departament, Faculdade de Agronomia Eliseu Maciel, Universidade Federal de Pelotas, RS, Brasil.*

[5]*Plant Genome and Development Laboratory, University of Perpignan, France.*

ABSTRACT

Iron is an essential element for plant development, involved in metabolic processes, such respiration and photosynthesis. Comparative analysis with lower plant groups and the closed clades of crop plants can increase the knowledge about these processes. The use of bryophytes as model plants rise as a promising strategy since they present simpler patterns of development. Using a phylogenetic analysis of five iron uptake gene families (*NAS, NRAMP, YSL, FRO* and *IRT*) selected in monocots, dicots, Gymnospermae and bryophytes, it was possible to infer about homology diversification in these genes for plants species. The homologue genes were found using known iron uptake gene sequences of *Oryza sativa*, *Arabidopsis thaliana* and *Physcomitrella patens* as queries. The phylogeny was constructed using the most common bioinfomatics tools available. A total of 243 gene sequences for 30 plant species were found. The evolutionary fingerprint analysis suggested a negative selective pressure of iron uptake genes for most of the plant homologues analyzed. The Nicotinamide synthase and Yellow Stripe genes appear to accumulate more negative selection sites, suggesting a strong purifying pressure in these two gene families. The divergence time analysis indicates *IRT* as the most ancient gene family and *FRO* as the most recent. *NRAMP* and *YSL* genes appear as a closed branch in the evolution of iron uptake gene families.

Key-Words: *Ferric reductase oxidase, IRT, Nicotinamide synthase, NRAMP, Yellow Stripe-Like, bayesian inference, multi-copy gene family evolution*.

INTRODUCTION

The mineral nutrition of plants is a major factor related to growth and development, consequently affecting the productivity. Iron is a key micronutrient for plants, taking part in redox centers of proteins essential for photosynthesis and respiration (Taiz and Zeiger 2004). However, iron deficiency causes a metabolic imbalance which is deleterious to plant development (Briat and Lebrun 1997). In iron-rich environments, such as calcareous soils, the excess of this element induces the production of hydroxyl radicals involved in the oxidative stress, and can be cause several damages to cellular structures, eventually leading to death (Guerinot and Yi 1994; Briat *et al.* 1995). Therefore plants must balance iron concentration in a homeostatic way, providing the necessary amounts of the micronutrient and preventing the condition of internal cation excess (Briat and Lobréaux 1997; Grusak *et al.* 1999).

The mechanisms for iron uptake from the soil have been classified in two distinct strategies (Marschner and Römheld 1994). The strategy I is shared between dicots and non-graminaceous monocot species. Iron molecules in their ferric form are more soluble at low pH, therefore plants acidify the rhizosphere and release chelators or reductants, such as organic acids and phenolics in order to improve their uptake (Römheld and Marschner, 1986; Römheld et al. 1984). Plants increase iron absorption, under iron deficiency conditions, through three reactions (Hell and Stephan 2003): i) proton excretion by a specific H^+-ATPase pump action in the cellular membrane, acidifying the soil pH solution and increasing the ferric solubility; ii) Fe^{3+} reduction by Fe^{3+} chelate-reductase enzyme (*FRO*, Ferric Reductase Oxidase) to a more soluble form (Fe^{2+}); and iii) Fe^{2+} transport to root through a specific transporter (*Irt1*, Iron Regulated Transporter) (Römheld and Marschner 1986; Marschner and Röhmeld 1994; Eide et al. 1996; Santos and Costa de Oliveira 2007).

To surpass the lower iron availability, an independent strategy evolved in grasses. In strategy II, plants use the chelation for iron acquisition from the soil. These species release a lower molecular weight compound (phytosiderophore-PS), a non-proteic amino acid synthetized from methionine (Curie and Briat 2003). These compounds show higth afinity and mineral chelating properties, thus efficiently binding to Fe^{3+} in rhizosphere, creating a stable complex Fe^{3+}-PS (Takagi et al. 1984; Roberts et al. 2004). This is recognized and transported to inner of root cells by action of specific transporters (*Ys1*, Yellow Stripe 1) (Mori 1991) without the extracellular

reduction dependence (Römheld and Marschner 1986; von Wirén et al. 1995).

Some homologues of genes encoding the Fe^{+2} transporter related to strategy I have already been characterized in *Arabidopsis*, tomato and soybean (Grotz et al. 1998; Eckhardt et al. 2001; Vert et al. 2001; Moreau et al. 2002) and the ferric-chelate reductase component was identified in *Arabidopsis* (*AtFro*2) (Robinson *et al.*, 1999) and pea (*PsFro*1) (Waters et al. 2002). The NAS (Nicotinamide synthase), which is related to the nicotinamide (NA) synthesis precursor, is essential in the mugineic acid pathway. NA can bind several metals, including Fe^{2+} and Fe^{3+}, but the complex formed is not secreted, suggesting a probable participation at intra and intercellular metal transport processes in both strategy I and II species (Kim et al. 2006). Other specific genes are related with regulation of iron transport, such *IRT* (*Iron Regulated Transporter*), considering the most important in the root system.

Additional components related to iron homeostasis have been characterized. Although not related to mineral uptake, these components are probably involved in iron intracellular targeting and storage. Members of the widely distributed *NRAMP* (Natural Resistance-Associated Macrophage Protein) family of cation transporters have also been characterized in *Arabidopsis* and rice (Belouchi et al. 1997; Thomine et al. 2000). Experimental evidence for a role of *AtNramp1*, *AtNramp3* and *AtNramp4* and *OsNramp1* in iron homeostasis already exists. *ATNRAMP1* may be related to iron subcellular transport and its targeting to storage compartments such as vacuoles or plastids (Curie et al. 2000). Ferritin has an well established role in iron storage and buffering of the mineral availability in the cytoplasm of mammalian cells. In plants, Ferritin is located in the plastid stroma, where it performs a similar function (Ragland et al. 1990). Ferritin cDNAs have been isolated from pea, soybean, maize and *Arabidopsis* (Briat and Lobréaux 1997).

These genes can vary in their copy number among different plant genomes. The identification of orthologue and paralogue copies becomes necessary for the understanding of the evolutionary relationships among these gene families. A phylogenetic approach is then useful to indicate how these families evolved. These approaches can allow a comparison that will enable to infer which copies have diverged by speciation or duplication based on their similarities and differences, since a proposed evolution model was indicated. In this study we used phylogenetic tools to infer which copies for each gene in a few iron homeostasis gene families can be a result of speciation or duplication events, contributing for the comparative genomic

analysis in plant species.

METHODS

Search for iron uptake homologues in plants species

A local alignment between *Oryza sativa* L. (*query*), *Arabidopsis thaliana* (L.) Heynh. and *Physcomitrella patens* Brid. was performed using the *BLAST* tool (Altschul et al., 1990) accessed from web http://blast.ncbi.nlm.nih.gov/Blast.cgi. The sequences of gene related with iron homeostasis in plants were selected using the *Genbank* access number provided by Gross et al. (2003). The gene homology was defined by the maximum alignment score associated with the higher query coverage with the lower *E-value*.

The selected sequencies were subjected to global alignment in ClustalW (Larkin *et al*., 2007). An initial tree was built using the *Neighbor-Joining* method (Saitou and Nei, 1987), with a 1000 *boostrap* replicates, with help of MEGA 4 software (Tamura *et al*., 2007), to verify the phylogenetic signal. The monophyletic groups were subjected to the modelling to find the best substitution model in the *Jmodeltest* (Posada, 2008). The appropiate model was selected to run a Bayesian analysis in the *BEAST* software (Drummond and Rambaut, 2007), being the resulting trees used to build the consensus tree for the phylogeny. The phylogenetic analysis was performed with help of EPOS for MacOSX 10.5.1. software (Griebel *et al.*, 2008).

Selection analysis of iron uptake gene copies in plants

Rates of synonymous and non-synonymous substitutions (dS and dN) were estimated by dividing the observed number of synonymous and non-synonymous substitutions per gene (DS and DN) by the number of synonymous and non-synonymous sites (S and N), i.e. dS=DS/S and dN=DN/N. Using the HyPhy software (Pond, Frost and Muse 2004), we estimated pairwise substitution rates using the most appropriated substitution computed before. We performed PAML ML estimates using the following parameters: pairwise comparison (i.e. runmode=-1) and the transition-transversion ratio, κ, estimated from sequence data using codon frequencies estimated from the nucleotide frequencies at the three codon positions (F3x4 model). In order to obtain the evolutionary fingerprint and the synonymous and non-synonymous individual rate for each copy in all studied sequences we used the algorithms implemented in http://www.datamonkey.org/, and the results were included in the

phylogenetic trees obtained.

Divergence time of iron uptake genes in plants

The divergence time estimates were computed using the Bayesian inference using the *Physcomitrella patens* copies for iron uptake genes and the *Adh* as an outgroup, with 100,000 replicates. The dates were estimated using the probably date for bryophyte origin (Megallón; Hilu, 2009). The resulting tree was edited with help of *FigTree* software (Vlad et al., 2008).

RESULTS AND DISCUSSION

Using the data available for the model plant genome projects in *GenBank*, such as *BGI-RIS* and the *IRGSP* for rice, *TAIR* for *Arabidopsis thaliana* and *COSSMOSS* for *Physcomitrella patens*, it was possible to construct an overview of iron homeostasis in the main plant lineages. As a starting point, five protein families: *YSL*, *FRO*, *NAS*, *IRT*, and *NRAMP*, which have positive molecular implications on iron uptake, intracellular targeting and storage in other plants, were chosen for the analysis. Taking specific members of these families as query sequences, searches for homologous sequences in the non-redundant (NR), expressed sequence tag (EST) and high throughput genomic sequences (HTG) in *GenBank* current databases were carried out. A total of 243 sequences were analysed and 112 putative genes were predicted for 30 plant species, based on the homology with model plant annotated genes. Together with rice, *Arabidopsis* and *P. patens* previously reported sequences, we analyzed in this study a total of forty-three genes possibly involved with iron homeostasis in rice: eighteen genes related to the Yellow Stripe Fe^{3+}-phytosiderophore transporter, two genes homologous to the *FRO* family, thirteen sequences related to the *ZIP* family (including two *IRT* genes) and eight putative *NRAMP* genes. The location of five putative new genes which were predicted from the indica rice data and whose scaffolds were not anchored to genetic maps still remains unclear.

Ferric Reductase Oxidase gene family relationship

The homologues for *FRO* genes (Ferric Reductase Oxidase), which code for a Fe^{3+} chelate reductase enzyme (responsable for Fe^{3+} to Fe^{2+} reduction), are represented by eight homologues in *Arabidopsis thaliana* (*AtFro1* to *AtFro8*) and only two in rice (*OsFro1* and *OsFro2*). The *Arabidopsis* genes *Fro6*, *Fro7* and *FRO8* were reported to be often expressed in photosynthetic tissues (Mukherjee *et al.*, 2006).

BLAST searches revealed six similar sequences of *FRO* genes in *Physcomitrella patens* (Supplementary File). Apparently only two of these moss sequences found are closely related with the *FRO* gene family, forming a distinct clade with *OsFro1*, *AtFro6 e AtFro7* (*FRO1* clade), the other rice and *Arabidopsis* homologues are clustered in a sister clade, which by means of comparison was named *FRO2* clade (Figure 1A). However, the topology observed when grouping the *FRO* homologues of these model plants, suggests the existence of two orthologue groups that probably diverged before the diversification of land plants. The Bryophyta *l. s.* are the ancestral plants clade and their origin is about 450 mya (Megallón and Hilu, 2009). Therefore, the *FRO1* lineage has been closely related with the ancestral of this gene family, since *P. patens* homologues are closer to this clade. Gross et al. (2003) found a similar cluster pattern when compared *FRO* homologues of rice, *Arabidopsis* and pea (*Pisum sativum* L.).

FRO gene homologues were found for another four species of dicots (*Malus xiaojinensis* Chen & Jiang, *Populus trichocarpa* Torr. & A. Gray, *Ricinus communis* L. and *Vitis vinifera* L.) and two monocots (*Sorghum bicolor* [L.] Moench and *Zea mays* L.), but no independent relationship for monocots or dicots are demonstrated in the bayesian tree (Figure 1A). Other reports for *FRO* genes in plants (Groom et al. 1996; Torres et al. 1998) detected additional homologues in *Arabidopsis* and rice. These studies report that several homologues for *FRO* genes in plants could presumably act as burst oxidases generating reactive oxygen species as a defense against pathogens (Groom et al. 1996; Torres et al. 1998). The. major difference would be based on the fact that in the *FRO* members Fe^{3+} act as the final acceptor for the transferred electron instead of molecular oxygen, as in the case of burst oxidases (Robinson et al. 1999). These two different functions in related lineages and the absence of independent branches for the major flowering plant groups agree with our hypothesis about the early divergence of *FRO* genes in plants.

***Nicotinamide synthase* gene family relationship**

The first occurrence of *NAS* genes in plants was found in barley (*Hordeum vulgare* L.) (Higuchi et al. 1999) and later these genes were found in tomato (Ling et al., 1999), *Arabidopsis* (Suzuki et al. 1999), rice (Higuchi et al. 2001) and maize (Mizuno et al. 2003). The transcripts *OsNas1* and *OsNas2*, as well as the corresponding coded proteins, were accumulated in response to lower iron source (Inoue et al. 2003). The expression of *OsNas1* is induced in response to an iron

deficiency both in roots and leaves, especially in chlorotic leaves (Higuchi et al. 1999). Three orthologue copies of *NAS* genes (*OsNas1*, *OsNas2* and *OsNas3*) were expressed in rice under iron deficiency (Inoue et al. 2003)

In our search, three rice *NAS* homologues (*OsNas1, OsNas2* and *OsNas3*), four homologues for *Arabidopsis* (*AtNas1*, *AtNas2, AtNas3* and *AtNas4*) and two for *P. patens* (*PHY150995* and *PHY215944*) were found. Homologues for other flowering plants species were also found (Table 3). When the model plant homologues were aligned with other plant species putative *NAS* genes, the moss copies are found as outgroupers. The flowering plant *NAS* copies are grouped in two distinct clades, reflecting the divergence between monocots and dicots (Figure 1B). This topology indicates the existence of two major and independent lineages for this gene family, one for mocots and other for dicots, both related directly with the *PHY150995* lineage. In *Thlaspi caerulescens*, the same arrangement for *NAS* genes was found, suggesting that it reconstructs the different evolutionary rates and/or selection pressure between this two plant groups (Figure 5B) (Mari et al. 2006). These dicotomy was found for *NAS* genes when *Neurospora crassa* Shear & B. O. Dodge and *Magnoporthe grisea* (T. T. Hebert) M. E. Barr homologues were compared, the dicot species showing a closer relationship with the filamentous fungi sequences (Trampczynska et al. 2006). The rice copies *OsNas1/OsNas2* and the *Arabidopsis AtNas3/AtNas4* branches observed suggest a paralogy divergence event in rice and *Arabidopsis*, respectively. The same response to iron deficiency of *OsNas1* and *OsNas2* may reflect that the duplication event in rice is recent, and is consistent with a tight clustering of the two copies.

***Natural Resistence-Associated Macrophage Protein gene* family relationship**

The *NRAMP* (*Natural Resistance-Associated Macrophage Protein*) constitutes a highly conserved integral membrane protein family involved with iron transport in several organisms, including bacteria, fungi, plants and animals (Cellier et al. 1995). These genes are widely distributed in all plants families, acting mainly in divalent cation transport (Curie and Briat 2003). The first occurrence of *NRAMP* genes were identified in mammals (*Nramp1*), coding a macrophage membrane protein responsible for cation concentration in phagosome, regulating the phagocited bacterial regulation (Williams et al. 2000).

Four homologue sequences were found for *P. patens* using the eight *O. sativa NRAMP* copies as queries. All moss homologues were grouped in *NRAMP*

phylogenetic relationship (Figure 2B) and no evidence for the divergence of the major flowering plants groups was observed, since monocots and dicots homologues are grouped together in different branches. The homologues for *P. patens* were grouped at the root of the phylogeny, which indicates the occurrence of a single process for the divergence of *NRAMP* copies for non-vascular and vascular plants. Thomine and Schroeder (2004), based on *NRAMP* genes analysis in recent plant species such as rice, tomato, *Arabidopsis, Medicago truncatula* Gaertn. and soybean (*Glycine max* [L.] Merr.), also evidenced the early divergence of *NRAMP* copies, reporting the existence of two subfamilies for these genes, one closer to animal homologues and the other related with green plants. The narrowest interspecific relationship was found by Gross et al (2003) for *OsNramp2* and *OsNramp*8 genes in rice and the *AtNramp2, AtNramp3, AtNramp4* and *AtNramp5* genes in *Arabidopsis*, forming a comparted branch with the root of the phylogeny (Figure 2A), suggesting that the copy relationships in rice is an evidence for a recent duplication in this species. When other plant homologues are added to the analysis (Supplementary file), a significant detachment of rice copies occurs. Several studies have reported the multi-copy nature of *NRAMP* gene family in higher plants (Marshner and Romheld 1994; Eide 1998; Mori 1999; Bennetzen 2002). This is also present in lower groups, i.e., when we include the moss homologue occurrences found in the *BLAST* search using as a query the known genes for the main model plants, rice (Gross *et al.*, 2003) and *Arabidopsis* (Thomine and Schreoder 2004). These results strengthen the ancestral duplication hypothesis of this gene family.

IRT genes

The *ZIP* genes coding for a Zn (*ZRT* – Zinc Regulated Transporter) and Fe_{2+} (*IRT* – Iron Regulated Transporter) transporter proteins were studied. *Irt1* is an specific Fe^{2+} transporter (Eide *et al.*, 1996), being the major transporter in plant roots (Connolly *et al.*, 2002), and the main iron uptake protein in *Arabidopsis* (Vert *et al.*, 2002, Varotto *et al*., 2002). However, *Irt1* is also responsible for the uptake of Mn^{+2}, Zn^{+2}, and Co^{+2} (Curie and Briat, 2003). Another member of the *ZIP* family is *Irt2*, an *Irt1* homologue found in epidermic cells of *A. thaliana* roots (Curie and Briat, 2003). Orthologues of *Irt1* have been characterized in tomato and rice, both mRNAs were accumulated in roots under iron deficiency conditions (Eckhardt *et al.*, 2001, Bughio *et al.*, 2002, Ishimaru *et al.*, 2006). Although *Irt1* is capable of mediating the transport of several metals in bacteria (Eide *et al.*, 1996), iron must be the most important metal

for this transporter, since plants can survive with decreasing contents of Mn, Zn and Cu, but will die if the iron level does not increase (Vert *et al.*, 2001).

For *P. patens*, four *IRT* homologues were found using rice sequences as a query. Three *Arabidopsis IRT* genes and ten other *ZIP* (*ZRT/IRT-related proteins*) homologues were included to infer this gene family topology in plants. For other nine flowering plant species, multicopies for *IRT/ZIP* genes were found (Additional file 1). In the resulting phylogenetic tree (Figure 2A) the moss homologues constitute the root of topology. These genes are associated with a branch composed exclusively by *Arabidopsis* copies, forming the other plant copies an internal group. In this restricted group the monocot and dicot clades are separated. In the monocot branch, only rice copies are placed together, such as reported for other evolutionary studies with these plant genes (Gross et al, 2003; Chen et al., 2008; Migeon, 2010). For the dicots *Arabidopsis* and *Malus xiaojinensis*, *IRT* copies are grouped close in a single branch, within the larger cluster of *Arabidopsis ZIP* copies. Other studies reported the close relationship between dicot and monocot *IRT* genes, suggesting an incipient divergence history in this two plant groups (Guerinot, 2000; Mäser *et al*., 2001; Vert *et al.*, 2002; Gross *et al.* 2003;), supported by the occurrence of several conserved domains in this gene family (Eng *et al*., 1998). Therefore, the results obtained in this study, enriched by ancestral plant lineage homologues, the monocot were more distant to the ancestral lineage of *IRT* genes than dicot copies, suggesting an occurrence of a paralogue lineage for this gene. The paralogy event was observed in *Arabidopsis*, rice and *Malus*, at least three events for *Arabidopsis* and a single event for the other species.

Yellow Stripe Like gene family relationship

Previous homology studies of iron uptake gene families had reported 18 putative genes for the *YSL* family for rice, expressed both in roots and in shoots of this plant species (Gross et al. 2003; Koike et al. 2004). For *Arabidopsis*, the presence of *YSL* genes are surprising, since this species is known as a strategy I plant for iron absorption, not synthetizing or using phytosiderophores (Waters et al. 2006). Early studies had shown that occurrence of *YSL* genes are related with the transport of metals complexed with nicotinamide (NA). This relation was determined for *AtYsl1* (Le Jean et al. 2005; Waters et al. 2006), *AtYSL2* (Didonato et al. 2004) and *AtYsl3* (Waters et al. 2006). The same function was observed in *Thlaspi caerulescens* J. Presl. & C. Presl., a metal accumulating species for the gene *TcYsl3* (Mari et al. 2006;

Gendre et al. 2007).

The *BLAST* alignment using *O. sativa YSL* (*yellow stripe-like*) as a query returns eight genes in *Arabidopsis* (*AtYsl1* to *AtYsl*8), and only two homologues for *P. patens* in Genbank (Table 1). When comparing these three plant homologues with other flowering plant homologues for *YSL* genes (Figure 3A) the basal branch is constituted by the moss homologues, which was expected. The other homologue arrangements demonstrated the possibility of two different duplication times for *YSL* gene copies in flowering plant species. These duplication events probably occurred before the monocot/dicot divergence since both clades observed are consisted of copies of the major groups of higher plants. The ancestral divergence of *OPT* gene copies was considered as an ancient event, based of the proximity of *Arabidopsis* orthologues with the *Saccharomyces* species, suggesting the high conservation of these gene family among fungi, metazoan and plant (Koh et al. 2002; Stacey et al. 2008). However, our analysis suggests an early duplication for these gene copies, at least for higher plants, since no homologue copies of lower plants (represented by *P. patens* homologues) was found inside the major monocots/dicots *YSL* clades.

The comparison of *Arabidopsis* and *O. sativa YSL* homologues indicated that several arrangements within rice *YSL* copies are reflecting a recent duplication and expansion of this gene family (Gross *et al*. 2003). These authors observed at least four paralogous cases in rice *YSL* gene copies, suggesting also four recent duplication events in this species. In our analysis, the clustering of rice homologues *OsYsl3* and *OsYsl4* agrees with a recent duplication hypothesis, but in a single event, suggesting that these are paralogous copies in the rice genome. It is also true for *Arabidopsis AtYsl4* and *AtYsl6* orthologues, perhaps after the first duplication event for these genes in higher plants (Figure 3A). The other rice *YSL,* such *OsYsl5* and *OsYsl6*, homologues are strongly related to other plant homologues, suggesting an early divergence of these copies, before the monocot/dicot divergence. The observed relationship between the genes *OsYsl2*, *OsYsl9*, *OsYsl15* and *OsYsl16* suggests a recent duplication event, since the internal branchs are composed only by monocot genes, and does not demonstrate a paralogy divergence event in rice, but can be true for all monocots. A differential function for the *Arabidopsis* homologues can be reflected in this phylogenetic analysis, since no monocot copy of *YSL* genes was found clustering close to dicot homologues. Probably the higher selection pressure on the iron absorption mechanism has been the main factor for the detachment of rice

and *Arabidopsis* copies, as it is known for other iron homeostasis genes in procaryotes (Osorio et al. 2008) and eukaryotes (Roualt and Tong 2005; Blázquez et al. 2007).

Selective pressure on iron uptake gene copies in plants

When one compares the selective sites for the iron uptake genes for the studied species, four out of seven were under negative selection. These results suggest the strong action due to the Darwinian selection. For *IRT*, *FRO* and *NAS* genes, the synonymous and non-synonymous substitutions have been equivalent (Figures 5 A-F). For the other genes the non-synonymous substitutions have been in higher frequencies agreeing with the hypothesis of a strong selection on these gene families. All sites analyzed demonstrate a fast evolving rate (Table 1). When one compares the fingerprint of iron uptake genes with *ADH* genes (added as a non-related family) in plants, the selection found was similar for all genes except for *NRAMP* genes (Figure 6A-F), suggesting a higher pressure on this gene family during plant evolution.

Using *Adh* sequences from several plant species, it was possible to compare the evolving rates within this gene family and the iron uptake gene families. For this gene family, the substitution occurs slightly faster in plants (Yokoyama and Harry, 1993), and was confirmed in our analysis (Figure 3B). The substitution rates demonstrated by the dN/dS ratio analysis indicate that the *NRAMP* genes did evolve faster than other genes related with iron acquisition. The non-synonymous substitutions prevail against the synonymous substitution indicating a weak diversifying selective pressure acting on these genes in plants. In some rice copies (*OsNramp2*, *OsNramp3*, *OsNramp4*, *OsNramp5* and *OsNramp7*) this could be an indication of the higher influence of these orthologues in the iron acquisition mechanisms. Perhaps in rice these genes evolve faster, probably because of the prevalence of the strategy II in grass plants, enabling the strategy I homologues to accumulate mutations. However, the fast evolution of *NRAMP* genes suggested by the substitution rates is not an exclusivity of grass species, since it occurs both in early land and higher plant species.

For *FRO* genes, a higher substitution rate was observed in one *Arabidopsis thaliana* orthologue (*AtFro8*) and in one rice orthologue (*OsFro2*) suggesting that these orthologues are under a strong selection, and probably are functionally more requested. *AtFro8* was detected in mitochondria (Heazlewood et al. 2004), and that

localization implies that this ferric chelate reductase must be involved in mitochondrial iron homeostasis (Jeong and Connolly 2009). Considering the evolutionary story of mitochondrial genes, a higher substitution rate would be expected. The higher substitution rates found in *Physcomitrella patens* homologues (Figure 1A) agrees with that hypothesis.

In the monocot clade for *NAS* genes, the maize copies *ZmNas1* and *ZmNas2* appear accumulating more substitutions when compared to others grass orthologues. No reference was found in the literature to explain these results. However, in other grass species such as rice and barley the *Nas1* and *Nas2* encode the major *NAS* enzyme (Inoue et al 2003).

Much has been learned about the genes and proteins necessary for primary Fe and Zn uptake in plants (Curie and Briat, 2003; Schmidt, 2003). Genes for phytosiderophore synthesis have been identified (Higuchi et al., 1999, 2001; Kobayashi et al., 2001; Inoue et al., 2003), and a gene for Fe (III)-phytosiderophore uptake, *ZmYsl1*, has been identified in maize (Curie et al., 2001). In our analysis the KA/KS rates for *YSL* gene orthologues demonstrate similarities for all plants species analyzed. The synonymous substitutions appear in higher frequencies than the non-synonymous, suggesting a selective pressure to not modify the aa loci for these genes. The higher substitution rates are found to *Sorghum bicolor* (Sorghumbicolor5), *Populus trichoicarpa* (Populustrichocarpa1), *Ricinus communis* (Ricinuscommunis4) and *Zea mays* (LOC100280197). None of the above are described in the literature as a true functional *YSL* gene. Within the *Arabidopsis YSL* family, the *Ysl1* and *Ysl3* proteins are most closely related to *Ysl2* (Didonato et al., 2004). However, the gene structure of these three family members is dissimilar: *Ysl1* has only four exons, whereas *Ysl2* has six exons and *Ysl3* has seven exons (Waters et al. 2006). Thus, the gene structure suggests that these genes are evolutionarily distant, but conservation of the protein sequences suggests that the proteins perform similar functions. Probably, the higher substitution rates found in *Sorghum, Ricinnus, Populus* and *Zea* copies reflects a conservation of the proteins with similar structure of a membrane protein or a phytosiderophore, but some expression essays are necessary to confirm this hypothesis.

Divergence between iron uptake gene families in plants

Using an ancient plant group it was possible to infer on the origins of each gene family related with the iron homeostasis used in this approach. For *P. patens* only four of the iron uptake gene families can be taken as of monophyletic origins (Figure 4). One of *NAS* moss copies was grouped as a basal orthologue for the metal uptake genes in plants. The other *NAS* copy for *P. patens* was closer to *FRO*, *NRAMP* and *YSL* copies. Both moss *NAS* copies have diverged at c.a. 350 Mya, reconstructing the Bryophyte origins, suggesting the paraphyletic origins of this gene family. Perhaps the functional *NAS* copy is the most recent, the same copy which is closer with the other *NAS* genes in higher plant groups, suggesting that as the basal copy of *NAS* genes in plants and excluding a duplication event in mosses for these gene family.

The earliest gene family appears to be the *IRT*, which diverged from the other iron related genes around 250 Mya (Figure 4). The most derivative genes are the *FRO* genes that have diverged ca. 90 Mya. *NRAMP* and *YSL* are sharing a same clade and diverged ca. 115 Mya, suggesting a close relationship between both families. Therefore, siderophores are found in early groups such *Archae* and *Bacteria* (Williams, 1979; Sriyosachati and Cox, 1986; Wolf and Crosa, 1986; Tai and Holmes, 1988; Zimmermann et al., 1989) and the distribution profiles of these genes in Eukaryota can be explained by the presence in a universal common ancestor and the posterior loss of certain domains that lack in recent groups, or the horizontal transfer of these genes after the split of those domains (Almeida et al., 2008).

REFERENCES

Almeida FC, Leszczyniecka M, Fisher PB, DeSalle R. 2008. Examining Ancient Inter-domain Horizontal Gene Transfer Evolutionary. Bioinformatics. 4: 109–119.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. Journal of Molecular Biology. 215: 403-410.

Belouchi A, Kwan T, Gros P. 1997. Cloning and characterization of OsNramp family from *Oryza sativa*, a new family of membrane proteins possibly implicated in the transport of metal ions. Plant Molecular Biology. 33: 1085-1092.

Bennetzen J. 2002. The rice genome. Opening the door to comparative plant biology. Science. 296: 60–63.

Blázquez L, De Juan J, Ruiz-Martínez et al. 2007. Genes related to iron metabolism and susceptibility to Alzheimer's disease in Basque population. Neurobiology of Aging. 28(12): 1941-1943.

Briat JF, Lobreaux S. 1997. Iron transport and storage in plants. Trends in Plant Science. 2: 187-193.

Briat JF, Fobis-Loisy I, Grignon N, Lobréaux S, Pascal N, Savino G, Thoiron S, von Wirén N, van Wuytswinkel O. 1995. Cellular and molecular aspects of iron metabolism in plants. Biol Cell. 84:69-81.

Briat JF, Lebrun M. 1999. Plant responses to metal toxicity. C. R. Acad. Sci. Paris 322: 43–54.

Bughio N, Yamaguchi H, Nishizawa NK, Nakanishi H, Mori S. 2002. Cloning an iron regulated metal transporter from rice. Journal of Experimental Botany. 53:1677-1682.

Cellier M, Privé G, Belouchi A, Kwan T, Rodriques V, Chia W, Gros P. 1995. Nramp

defines a family of membrane proteins. Proc Natl Acad Sci USA. 92:10089-10093.

Chen WR, Feng Y, Chao YE. 2008. Genomic analysis and expression pattern of OsZIP1, OsZIP3 , and OsZIP4 in two rice (*Oryza sativa* L.) genotypes with different zinc efficiency. Russian Journal Of Plant Physiology. 55 (3): 400-409.

Connolly EL, Fett JP, Guerinot ML. 2002. Expression of the IRT1 metal transporter is controlled by metals at the levels of transcript and protein accumulation. *Plant Cell*. 14: 1347- 1357.

Curie C., Briat JF.2003. Iron transport and signaling in plants. Annu. Rev. Plant Biol. 54:183–206.

Curie C, Alonso JM, Jean ML, Ecker JR, Briat JF. 2000. Involvement of NRAMP1 from *Arabidopsis thaliana* in iron transport. Biochemical Journal. 347:749-755.

Curie C, Panaviene Z, Loulergue C, Dellaporta SL, Briat JF, Walker EL. 2001. Maize yellow stripe1 encodes a membrane protein directly involved in Fe(III) uptake. Nature. 409: 346-349.

DiDonato RJJr, Roberts LA, Sanderson T, Eisley RB, Walker EL. 2004. Arabidopsis Yellow Stripe-Like2 (YSL2): a metal-regulated gene encoding a plasma membrane transporter of nicotianamine-metal complexes.The PlantJournal. 39:403–414.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology. 7: 214.

Eckhardt U, Mas Marques, Buckhout TJ. 2001. Two-iron regulated cation transporters from tomato complement metal uptake-deficient yeast mutants. Plant Mol Biol. 45:437–448.

Eide DJ, Broderius M, Fett JP, Guerinot ML. 1996. A novel iron-regulated metal transporter from plants identified by functional expression in yeast. PNAS. 93: 5624-

5628.

Eide DJ, Broderius M, Fett JP, Guerinot ML. 1996. A novel iron-regulated metal transporter from plants identified by functional expression in yeast. *PNAS*. v. 93, p. 5624-5628.

Eide DJ. 1998. The molecular biology of metal ion transport in *Saccharomyces cerevisiae*. *Ann Rev Nutr.* 18: 441–69.

Eng BH, Guerinot ML, Eide D, Saier MHJr. 1998. Sequence analyses and phylogenetic characterization of the ZIP family of metal ion transport proteins. Journal Membr Biology. 1(166-1), 1-7.

Gendre D, Czernic P, Conejero G, Pianelli K, Briat JF, Lebrun M, et al. 2007. TcYSL3, a member of the YSL gene family from the hyperaccumulator *Thlaspi caerulescens*, encodes a nicotianamine-Ni/Fe transporter. The Plant Journal 49: 1 – 15.

Griebel T, Brinkmeyer M, Böcker S. 2008. EPoS: a modular software framework for phylogenetic analysis Bioinformatics Applications Note. 24(20)*:* 2399–2400.

Groom QJ, Torres MA, Fordham-Skelton AP, Hammond-Kosack KE, Robinson NJ, Jones JD. 1996. RbohA a rice homologue of the mammalian gp91phox respiratory burst oxidase gene. Plant J. 10:515-522.

Gross J, Stein RJ, Fett-Neto AG, Fett JP. 2003. Iron homeostasis related genes in rice. Genetics and Molecular Biology. 26(4): 477-497.

Grusak MA, Pearson JN, Marentes E. 1999. The physiology of micronutrient homeostasis in field crops. Field Crops Res. 60:41-56.

Guerinot ML. 2000. The ZIP family of metal transporters. Biochim Biophys Acta. 1465(1-2):190-198.

Guerinot ML, YI Y. 1994. Iron: nutritious, noxious, and not readily available. Plant Physiol. 104:815-820.

Heazlewood JL, Tonti-Filippini JS, Gout AM, Day DA, Whelan J, Millar AH. 2004. Experimental analysis of the Arabidopsis mitochondrial proteome highlightsignaling and regulatory components, provides assessment of targeting prediction programs, and indicates plant-specific mitochondrial proteins, The Plant Cel.16: 241–256.

Hell R, Stephan UW. 2003. Iron uptake, trafficking and homeostasis in plants. Planta. 216:541-551.

Higuchi K, Watanabe S, Takahashi M, Kawasaki S, Nakanishi H, Nishizawa NK, Mori S. 2001. Nicotianamine synthase gene expression differs in barley and rice under Fe-deficient conditions The Plant Journal. 25(2): 159-167

Higuchi K, Suzuki K, Nakanishi H, Yamaguchi H, Nishizawa NK, Mori S. 1999. Cloning of nicotianamine synthase genes, novel genes involved in the biosynthesis of phytosiderophores. Plant Physiol. 119:471–480.

Inoue H, Higuchi K, Takahashi M, Nakanishi H, Mori S, Nishizawa NK. 2003. Three rice nicotianamine synthase genes, OsNAS1, OsNAS2, and OsNAS3 are expressed in cells involved in long- distance transport of iron and differentially regulated by iron. Plant J. 36: 366–381.

Ishimaru Y, Suzuki M, Tsukamoto T, et al. 2006. Rice plants take up iron as an Fe3+ -phytosiderophore and as Fe2+. Plant Journal. 45: 335-346.

Jeong J, Connolly EL. 2009. Iron uptake mechanisms in plants: Functions of the FRO family of ferric reductases Plant Science. 176: 709–714.

Kim SA, Punshon T, Lanzirotti A, LI L, Alonso JM, Ecker JR, Kaplan J, Guerinot ML. 2006. Localization of iron in *Arabidopsis* seed require the vacuolar membrane transporter VIT1. Science. 314:1295-1298.

Kobayashi T, Nakanishi H, Takahashi M, Kawasaki S, Nishizawa NK, Mori S. 2001. In vivo evidence that Ids3 from *Hordeum vulgare* encodes a dioxygenase that converts 2#-deoxymugineic acid to mugineic acid in transgenic rice. Planta. 212: 864–871

Koh S, Wiles AM, Sharp JS, Naider, FR, Becker JM, Stacey G. 2002. An Oligopeptide Transporter Gene Family in *Arabidopsis*. Plant Physiology. 128: 21–29.

Larkin MA, Blackshields G, Brown NP et al. 2007. Clustal W and Clustal X version 2.0. Bioinformatics. 23: 2947–2948.

Le Jean M, Schikora A, Mari S, Briat JF, Curie C. 2005. A loss-of-function mutation in AtYSL1 reveals its role in iron and nicotianamine seed loading. The Plant Journal. 44: 769 – 782.

Ling HQ, Koch G, Baümlein H, Ganal MW. 1999. Map-based cloning of chloronerva, a gene involved in iron uptake of higher plants encoding nicotianamine synthases. Proc. Natl Acad. Sci. USA. 96: 7098-7103.

Magallón S, Hilu KW. 2009. Land plants (Embryophyta). *In:* Hedges SB, Kumar S. editors. The Timetree of Life. Oxford. University Press. p. 133–137.

Mari S, Gendre D, Pianelli K, et al. 2006. Root-to-shoot long-distance circulation of nicotianamine and nicotianamine-nickel chelates in the metal hyperaccumulator Thlaspicaerulescens. Journal of Experimental Botany 57: 4111–4122.

Marschner H, Röhmeld V. 1994. Strategies of plants for acquisition of iron. Plant Soil 165:375–88.

Mäser P, Thomine S, Schroeder JI, et al. 2001. Phylogenetic relationships within cation transporter families of *Arabidopsis*. Plant Physiol 126:1646-1667.

Migeon A, Blaudez D, Wilkins O, Montanini B, Campbell MM, Richaud P, Thomine S, Chalot M. 2010. Genome-wide analysis of plant metal transporters, with an

emphasis on poplar Cell. Mol. Life Sci. 67:3763–3784

Mizuno D, Higuchi K, Sakamoto T, Nakanishi H, Mori S, Nishizawa NK. 2003. Three nicotianamine synthase genes isolated from maize are differentially regulated by iron nutritional status. Plant Physiol. 132: 1989–1997.

Morariu VI, Srinivasan BV, Raykar VC, Duraiswami R, Davis LS. 2008. Automatic online tuning for fast gaussian summation. Advances in Neural Information Processing Systems (NIPS). 8.

Moreau S, Thomson RM, Kaiser BN, Trevaskis B, Guerinot ML, Udvardi MK, Puppo A, Day DA. 2002 GmZIP1 encodes a symbiosis-specific zinc transporter in soybean. J Biol Chem. 277:4738-4746.

Mori S. 1999. Iron acquisition by plants. Curr Opin Plant Biol. 2: 250–253.

Mori S, Nishizawa NK, Hayashi H, Chino M, Yoshimura E, Ishihara J. 1991 Why are young rice plants highly susceptible to iron deficiency? Plant Soil. 130: 143–156.

Mukherjee I, Campbell NH, Ash JS, Connolly EL. 2006. Expression profiling of the Arabidopsis ferric chelate reductase (FRO) gene family reveals differential regulation by iron and copper, Planta. 223: 1178–1190.

Grotz N, Fox T, Connolly EL, Park W, Guerinot ML, Eide D. 1998. Identification of a family of zinc transporter genes from Arabidopsis that respond to zinc deficiency, Proc. Natl. Acad. Sci. U. S. A. 95: 7220–7224.

Osorio H, Martinez V, Nieto PA, Holmes DS, Quatrini R. 2008. Microbial iron management mechanisms in extremely acidic environments: Comparative genomics evidence for diversity and versatility. *BMC Microbiol.* 8: 203.

Posada D. 2008. jModelTest: Phylogenetic Model Averaging. *Molecular Biology and Evolution*. 25: 1253-1256.

Pond SLK, Frost SDW, Muse SV. 2005. HyPhy: Hypothesis testing using phylogenies. Bioinformatics. 21(5):676–9.

Ragland M, Briat JF, Gagnon J, Laulhere JP, Massenet O, Theil EC. 1990. Evidence for conservation of ferritin sequences among plants and animals and for a transit peptide in soybean. J Biol Chem 265:18339-18344.

Roberts LA, Pierson AJ, Panaviene Z, Walker EL. 2004. *Yellow stripe1*: expanded roles for the maize iron-phytosiderophore transporter. Plant Physiol 135: 112–120.

Robinson NJ, Proctor CM, Connolly EL, Guerinot ML. 1999. A ferric chelate reductase for iron uptake from soils. Nature. 397:694-697.

Röhmeld V, Marschner H. 1986. Mobilization of iron in the rhizosphere of different plant species. Adv. Plant Nutr. 2: 155-204.

Römheld V, Müller C, Marschner H. 1984. Localization and capacity of proton pumps in roots of intact sunflower plants. Plant Physiol. 76:603-606.

Römheld V. 1987. Different strategies for iron acquistion in hgher plants. Physiologia Plantarum. 70: 321–234.

Rouault TA, Tong WH. 2005. Opinion: Iron-sulphur cluster biogenesis and mitochondrial iron homeostasis. Nat Rev Mol Cell Biol. 6:345-51.

Saitou N, Nei M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution. 4: 406-425.

Santos LS, Costa de Oliveira A. 2007. Rice iron metabolism:from source to solution. Journal of Crop Science and Biotechnology. 10: 64-72.

Schmidt W. 2003. Iron solutions: acquisition strategies and signaling pathways in

plants, Trends Plant Sci. 8: 188–193.

Sriyosachati S, Cox CD. 1986. Siderophore-mediatediron acquisition from transferrin by *Pseudomonas aeruginosa*. Infect. Immun. 52:885-891.

Stacey MG, Patel A, McClain WE, Mathieu M, Remley M, Rogers EE, Gassmann W, Blevins DG, Stacey G. 2008. The *Arabidopsis* AtOPT3 protein functions in metal homeostasis and movement of iron to developing seeds. Plant Physiol. 146:589–601

Suzuki K, Higuchi K, Nakanishi H, Nishizawa NK, Mori S. 1999. Cloning of nicotianamine synthase gene from *Arabidopsis*. Soil Sci. Plant Nutr. 45: 993-1002.

Tai SS, Holmes RK. 1988. Iron regulation of the cloned diphtheria promoter in Escherichia coli. Infect. Immun. 56:2430-2436.

Taiz L, Zeiger E. 2004. Fisiologia Vegetal. 3ª ed. Porto Alegre: Artmed.

Takagi S, Nomoto K, Takemoto T. 1984. Physiological aspect of mugineic acid, a possible phytosiderophore of graminaceous plant. J. Plant Nutr. 7: 469-477.

Tamura K, Dudley J, Nei M, Kumar, S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution. 24: 1596-1599.

Thomine S, Schroeder JI. 2004. Plant Metal Transporters with Homology to Proteins of the NRAMP Family. *In:* Cellier M, Gros P. *The NRAMP Family*. Andes/Kluwer series: Molecular Biology Intelligence Unit. p113-121.

Thomine S, Wang R, Ward JM, Crawford NM, Schroeder JI. 2000. Cadmium and iron transport by members of a plant metal transporter family in *Arabidopsis* with homology to Nramp genes. Proceedings of the National Academy of Sciences. 97: 4991-4996.

Torres MA, Onouchi H, Hamada S, Machida C, Hammond-Kosack KE, Jones JD.

1998. Six A*rabidopsis thaliana* homologues of the human respiratory burst oxidase (gp91phox). Plant J. 14:365-370.

Trampczynska A, Bottcher C, Clemens S. 2006. The transition metal chelator nicotianamine is synthesized by filamentous fungi. FEBS Letters. 580: 3173–3178.

Varotto C, Maiwald D, Pesaresi P, Jahns P, Salamini F, Leister D. 2002. The metal ion transporter IRT1 is necessary for iron homeostasis and efficient photosynthesis in *Arabidopsis thaliana*. *Plant Journal,* 31: 589–599.

Vert G, Briat JF, Curie C. 2001. *Arabidopsis* IRT2 gene encodes a root-periphery iron transporter. Plant J. 26:181–189.

Vert G, Grotz N, Dédaldéchamp FG, Guerinot ML, Briat JF, Curie C. 2002. IRT1, an *Arabidopsis* Transporter Essential for Iron Uptake from the Soil and for Plant Growth. Plant Cell. 14(6): 1223–1233.

von Wirén N, Marschner H, Römheld V. 1995. Uptake kinetics of iron-phytosiderophores in two maize genotypes differing in iron efficiency. Plant Physiol. 93:611-616.

Waters BM, Blevins DG, EIDE DJ. 2002. Characterization of FRO1, a Pea Ferric-Chelate Reductase involved in root iron acquisition. Plant Physiology. 129: 85-94.

Waters BM, Chu HH, Didonato RJ, Roberts LA, Eisley RB, Lahner B, et al. 2006. Mutations in *Arabidopsis* yellow stripe-like1 and yellow stripe-like3 reveal their roles in metal ion homeostasis and loading of metal ions in seeds. Plant Physiology 141: 1446 – 1458.

Williams LE, Pittman JK, Hall JL. 2000. Emerging mechanisms for heavy metal transport in plants. Biochim Biophys Acta. 1465:104-126.

Williams PH. 1979. Novel iron uptake system specified by ColV plasmids: an important component in the virulence of invasive strains of *Escherichia coli*. Infect.

Immun. 26:925- 932.

Wolf MK, Crosa JH. 1986. Evidence for the role of a siderophore in promoting *Vibrio angiuillaruln* infections. J.Gen. Microbiol. 132:2949-2952.

Yokoyama S, HARRY DE. 1993. Molecular phylogeny and evolutionary rates of alcohol dehydrogenase in vertebrates and plants. Mol. Biol. Evol. 10:1215-1226.

Zimmermann L, Angerer A, Braun V. 1989. Mechanistically novel iron(III) transport system in *Serratia marcescens*. J. B acteriol. 171:238-243.

Figure 1. *FRO* (A) and *NAS* (B) phylogenetic trees. The node value means the posterior significance of bayesian analysis. The branch value means the non-synonymous substitutions (dN) per synonymous substitutions (dS).

Figure 2. *IRT/ZIP* (A) and *NRAMP* (B) phylogenetic tree. The node value means the posterior significance of bayesian analysis. The branch values means the non-synonymous substitutions (dN) per synonymous substitutions (dS).

Figure 3. *YSL* and *ADH* phylogenetic trees. The node value means the posterior significance of bayesian analysis. The branch values means the non-synonymous substitutions (dN) per synonymous substitutions (dS).

Figure 4. Divergence within Iron uptake gene families in *Physcomitrela patens* using *ADH* genes as an outgroup. The scale bar indicates the divergence time in Mya.

Figure 5. dN/dS substitutions sites in the alignment of all plants studied. A. IRT genes. B. NAS genes. C. NRAMP genes. D. Yellow Stripe-like genes. E. FRO genes. F. ADH genes.

Figure 6. Evolutionary fingerprint of dN/dS substitutions sites in alignment length for all plants studied. A. *FRO* genes. B. *IRT* genes. C. *NAS* genes. D. *NRAMP* genes. E. *YSL* genes. F. *ADH* genes.

Figure 1

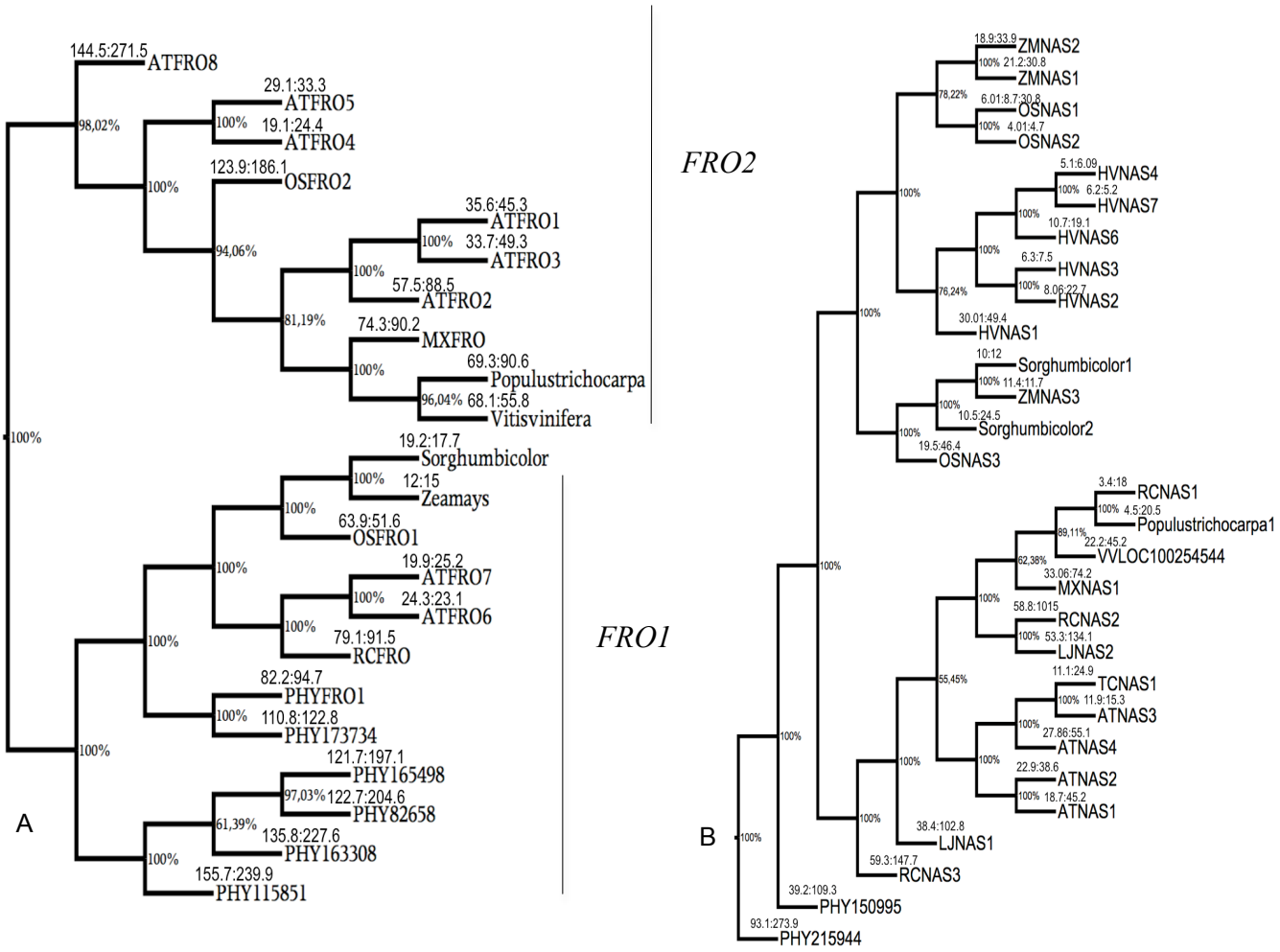

144.5:271.5
ATFRO8
29.1:33.3
ATFRO5
19.1:24.4
ATFRO4
123.9:186.1
OSFRO2
35.6:45.3
ATFRO1
33.7:49.3
ATFRO3
57.5:88.5
ATFRO2
74.3:90.2
MXFRO
69.3:90.6
Populustrichocarpa
68.1:55.8
Vitisvinifera
19.2:17.7
Sorghumbicolor
12:15
Zeamays
63.9:51.6
OSFRO1
19.9:25.2
ATFRO7
24.3:23.1
ATFRO6
79.1:91.5
RCFRO
82.2:94.7
PHYFRO1
110.8:122.8
PHY173734
121.7:197.1
PHY165498
122.7:204.6
PHY82658
135.8:227.6
PHY163308
155.7:239.9
PHY115851
A
FRO1
FRO2
18.9:33.9
ZMNAS2
21.2:30.8
ZMNAS1
6.01:8.7:30.8
OSNAS1
4.01:4.7
OSNAS2
5.1:6.09
HVNAS4
6.2:5.2
HVNAS7
10.7:19.1
HVNAS6
6.3:7.5
HVNAS3
8.06:22.7
HVNAS2
30.01:49.4
HVNAS1
10:12
Sorghumbicolor1
11.4:11.7
ZMNAS3
10.5:24.5
Sorghumbicolor2
19.5:46.4
OSNAS3
3.4:18
RCNAS1
4.5:20.5
Populustrichocarpa1
22.2:45.2
VVLOC100254544
33.06:74.2
MXNAS1
58.8:1015
RCNAS2
53.3:134.1
LJNAS2
11.1:24.9
TCNAS1
11.9:15.3
ATNAS3
27.86:55.1
ATNAS4
22.9:38.6
ATNAS2
18.7:45.2
ATNAS1
38.4:102.8
LJNAS1
59.3:147.7
RCNAS3
39.2:109.3
PHY150995
93.1:273.9
PHY215944
B

Figure 2

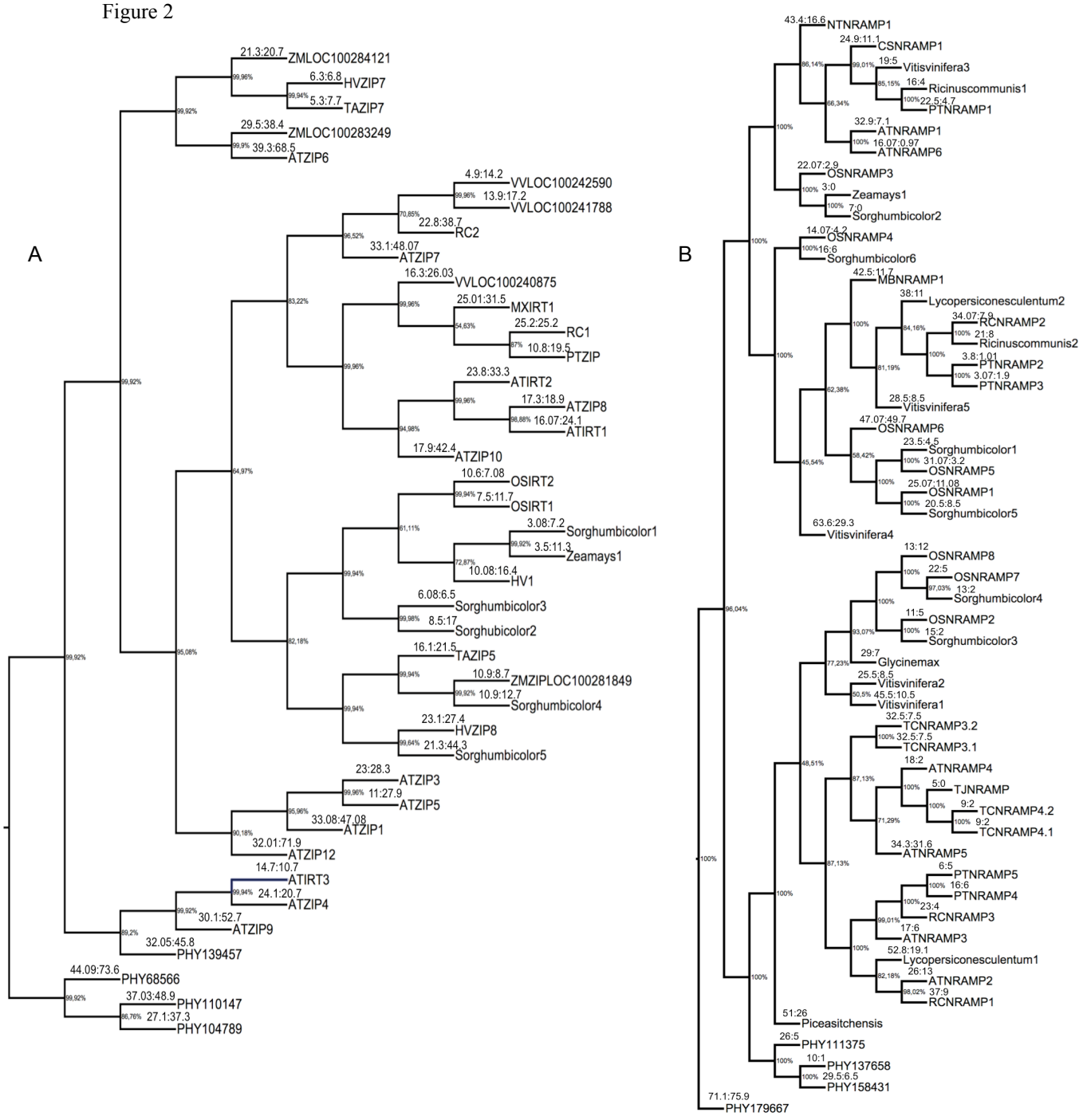

A
21.3:20.7
ZMLOC100284121
6.3:6.8
HVZIP7
5.3:7.7
TAZIP7
29.5:38.4
ZMLOC100283249
39.3:68.5
ATZIP6
4.9:14.2
VVLOC100242590
13.9:17.2
VVLOC100241788
22.8:38.7
RC2
33.1:48.07
ATZIP7
16.3:26.03
VVLOC100240875
25.01:31.5
MXIRT1
25.2:25.2
RC1
10.8:19.5
PTZIP
23.8:33.3
ATIRT2
17.3:18.9
ATZIP8
16.07:24.1
ATIRT1
17.9:42.4
ATZIP10
10.6:7.08
OSIRT2
7.5:11.7
OSIRT1
3.08:7.2
Sorghumbicolor1
3.5:11.3
Zeamays1
10.08:16.4
HV1
6.08:6.5
Sorghumbicolor3
8.5:17
Sorghubicolor2
16.1:21.5
TAZIP5
10.9:8.7
ZMZIPLOC100281849
10.9:12.7
Sorghumbicolor4
23.1:27.4
HVZIP8
21.3:44.3
Sorghumbicolor5
23:28.3
ATZIP3
11:27.9
ATZIP5
33.08:47.08
ATZIP1
32.01:71.9
ATZIP12
14.7:10.7
ATIRT3
24.1:20.7
ATZIP4
30.1:52.7
ATZIP9
32.05:45.8
PHY139457
44.09:73.6
PHY68566
37.03:48.9
PHY110147
27.1:37.3
PHY104789
B
43.4:16.6
NTNRAMP1
24.9:11.1
CSNRAMP1
19:5
Vitisvinifera3
16:4
Ricinuscommunis1
22.5:4.7
PTNRAMP1
32.9:7.1
ATNRAMP1
16.07:0.97
ATNRAMP6
22.07:2.9
OSNRAMP3
3:0
Zeamays1
7:0
Sorghumbicolor2
14.07:4.2
OSNRAMP4
16:6
Sorghumbicolor6
42.5:11.7
MBNRAMP1
38:11
Lycopersiconesculentum2
34.07:7.9
RCNRAMP2
21:8
Ricinuscommunis2
3.8:1.01
PTNRAMP2
3.07:1.9
PTNRAMP3
28.5:8.5
Vitisvinifera5
47.07:49.7
OSNRAMP6
23.5:4.5
Sorghumbicolor1
31.07:3.2
OSNRAMP5
25.07:11.08
OSNRAMP1
20.5:8.5
Sorghumbicolor5
63.6:29.3
Vitisvinifera4
13:12
OSNRAMP8
22:5
OSNRAMP7
13:2
Sorghumbicolor4
11:5
OSNRAMP2
15:2
Sorghumbicolor3
29:7
Glycinemax
25.5:8.5
Vitisvinifera2
45.5:10.5
Vitisvinifera1
32.5:7.5
TCNRAMP3.2
32.5:7.5
TCNRAMP3.1
18:2
ATNRAMP4
5:0
TJNRAMP
9:2
TCNRAMP4.2
9:2
TCNRAMP4.1
34.3:31.6
ATNRAMP5
6:5
PTNRAMP5
16:6
PTNRAMP4
23:4
RCNRAMP3
17:6
ATNRAMP3
52.8:19.1
Lycopersiconesculentum1
26:13
ATNRAMP2
37:9
RCNRAMP1
51:26
Piceasitchensis
26:5
PHY111375
10:1
PHY137658
29.5:6.5
PHY158431
71.1:75.9
PHY179667

Figure 3

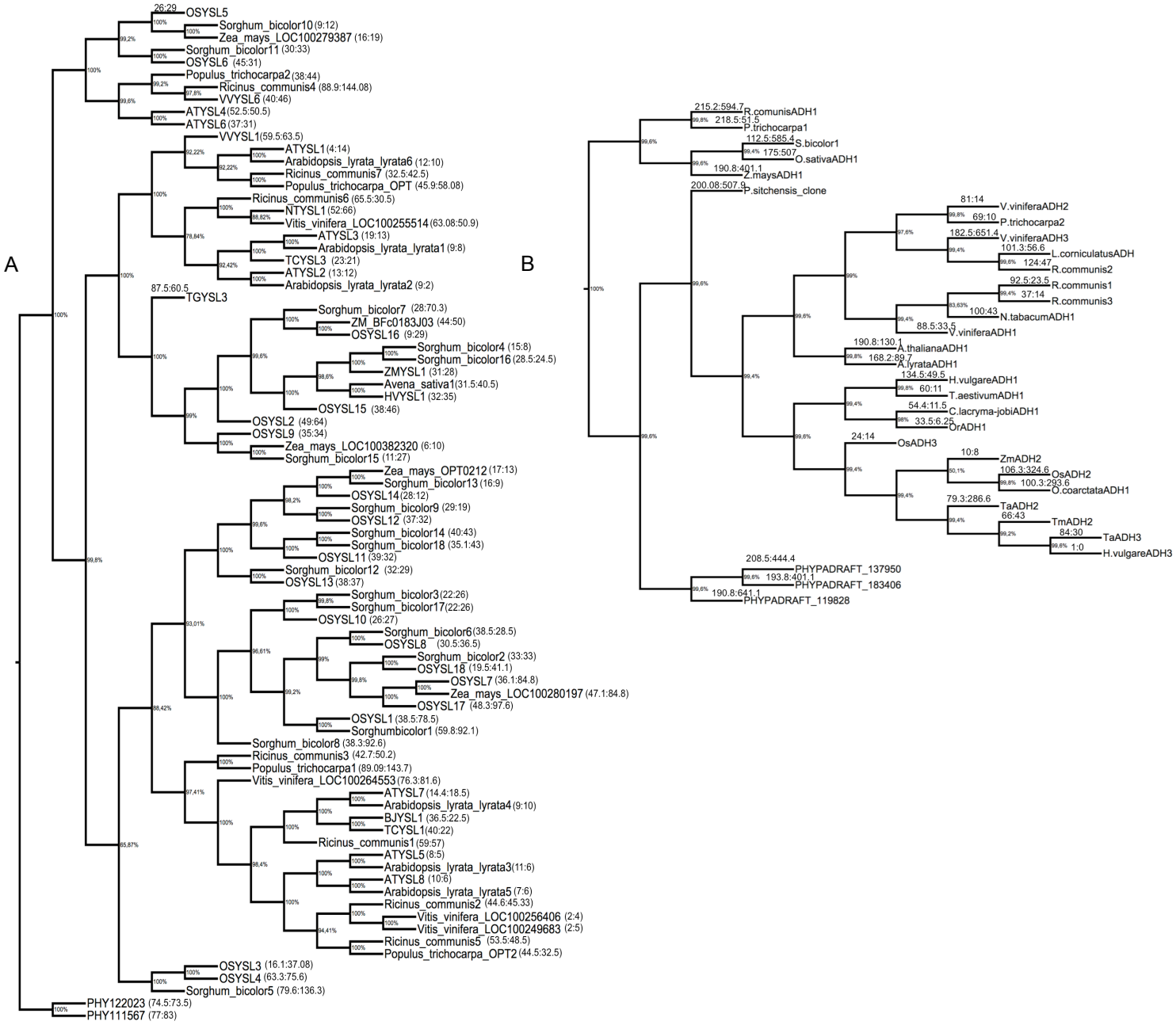

A
B
OSYSL5
Sorghum_bicolor10 (9:12)
Zea_mays_LOC100279387 (16:19)
Sorghum_bicolor11 (30:33)
OSYSL6 (45:31)
Populus_trichocarpa2 (38:44)
Ricinus_communis4 (88.9:144.08)
VVYSL6 (40:46)
ATYSL4 (52.5:50.5)
ATYSL6 (37:31)
VVYSL1 (59.5:63.5)
ATYSL1 (4:14)
Arabidopsis_lyrata_lyrata6 (12:10)
Ricinus_communis7 (32.5:42.5)
Populus_trichocarpa_OPT (45.9:58.08)
Ricinus_communis6 (65.5:30.5)
NTYSL1 (52:66)
Vitis_vinifera_LOC100255514 (63.08:50.9)
ATYSL3 (19:13)
Arabidopsis_lyrata_lyrata1 (9:8)
TCYSL3 (23:21)
ATYSL2 (13:12)
Arabidopsis_lyrata_lyrata2 (9:2)
TGYSL3
Sorghum_bicolor7 (28:70.3)
ZM_BFc0183J03 (44:50)
OSYSL16 (9:29)
Sorghum_bicolor4 (15:8)
Sorghum_bicolor16 (28.5:24.5)
ZMYSL1 (31:28)
Avena_sativa1 (31.5:40.5)
HVYSL1 (32:35)
OSYSL15 (38:46)
OSYSL2 (49:64)
OSYSL9 (35:34)
Zea_mays_LOC100382320 (6:10)
Sorghum_bicolor15 (11:27)
Zea_mays_OPT0212 (17:13)
Sorghum_bicolor13 (16:9)
OSYSL14 (28:12)
Sorghum_bicolor9 (29:19)
OSYSL12 (37:32)
Sorghum_bicolor14 (40:43)
Sorghum_bicolor18 (35.1:43)
OSYSL11 (39:32)
Sorghum_bicolor12 (32:29)
OSYSL13 (38:37)
Sorghum_bicolor3 (22:26)
Sorghum_bicolor17 (22:26)
OSYSL10 (26:27)
Sorghum_bicolor6 (38.5:28.5)
OSYSL8 (30.5:36.5)
Sorghum_bicolor2 (33:33)
OSYSL18 (19.5:41.1)
OSYSL7 (36.1:84.8)
Zea_mays_LOC100280197 (47.1:84.8)
OSYSL17 (48.3:97.6)
OSYSL1 (38.5:78.5)
Sorghumbicolor1 (59.8:92.1)
Sorghum_bicolor8 (38.3:92.6)
Ricinus_communis3 (42.7:50.2)
Populus_trichocarpa1 (89.09:143.7)
Vitis_vinifera_LOC100264553 (76.3:81.6)
ATYSL7 (14.4:18.5)
Arabidopsis_lyrata_lyrata4 (9:10)
BJYSL1 (36.5:22.5)
TCYSL1 (40:22)
Ricinus_communis1 (59:57)
ATYSL5 (8:5)
Arabidopsis_lyrata_lyrata3 (11:6)
ATYSL8 (10:6)
Arabidopsis_lyrata_lyrata5 (7:6)
Ricinus_communis2 (44.6:45.33)
Vitis_vinifera_LOC100256406 (2:4)
Vitis_vinifera_LOC100249683 (2:5)
Ricinus_communis5 (53.5:48.5)
Populus_trichocarpa_OPT2 (44.5:32.5)
OSYSL3 (16.1:37.08)
OSYSL4 (63.3:75.6)
Sorghum_bicolor5 (79.6:136.3)
PHY122023 (74.5:73.5)
PHY111567 (77:83)
R.comunisADH1
P.trichocarpa1
S.bicolor1
O.sativaADH1
Z.maysADH1
P.sitchensis_clone
V.viniferaADH2
P.trichocarpa2
V.viniferaADH3
L.corniculatusADH
R.communis2
R.communis1
R.communis3
N.tabacumADH1
V.viniferaADH1
A.thalianaADH1
A.lyrataADH1
H.vulgareADH1
T.aestivumADH1
C.lacryma-jobiADH1
OrADH1
OsADH3
ZmADH2
OsADH2
O.coarctataADH1
TaADH2
TmADH2
TaADH3
H.vulgareADH3
PHYPADRAFT_137950
PHYPADRAFT_183406
PHYPADRAFT_119828

Figure 4

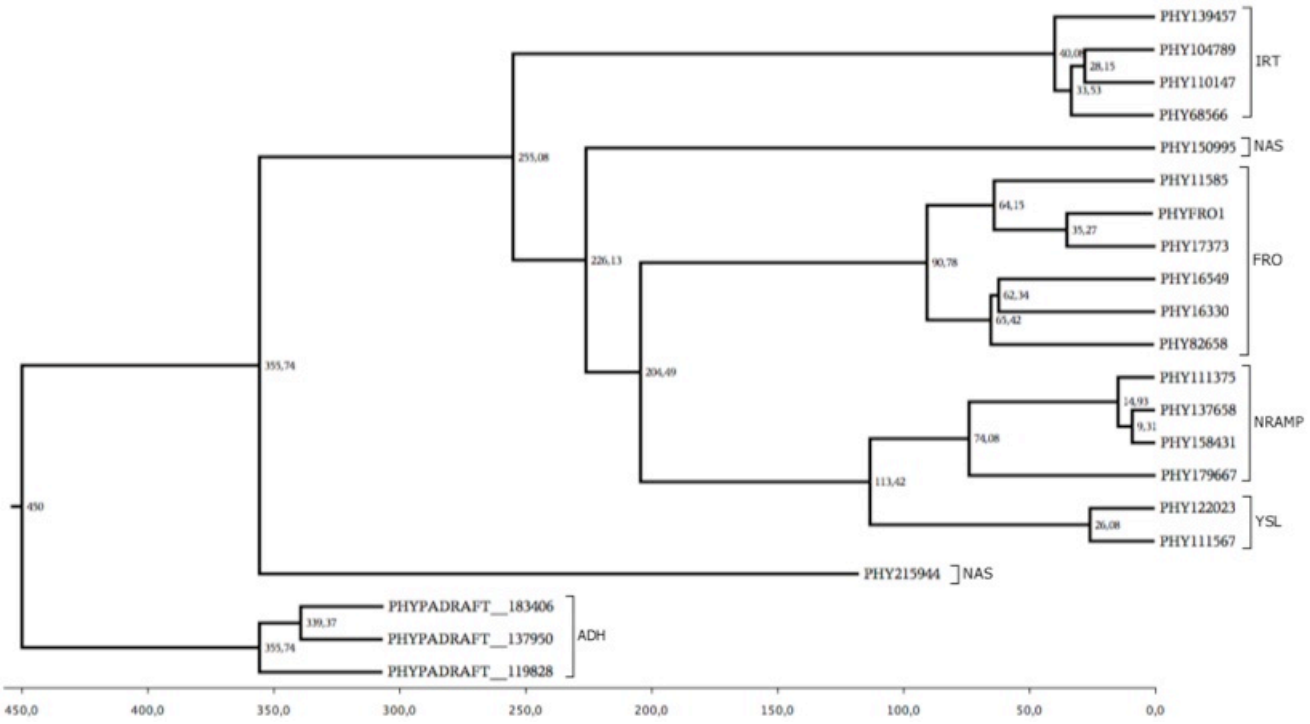

PHY139457
PHY104789
PHY110147
PHY68566
IRT
PHY150995
NAS
PHY11585
PHYFRO1
PHY17373
PHY16549
PHY16330
PHY82658
FRO
PHY111375
PHY137658
PHY158431
PHY179667
NRAMP
PHY122023
PHY111567
YSL
PHY215944
NAS
PHYPADRAFT__183406
PHYPADRAFT__137950
PHYPADRAFT__119828
ADH
450
355,74
255,08
226,13
204,49
90,78
74,08
113,42
64,15
35,27
62,34
65,42
14,93
9,31
26,08
93,08
28,15
33,53
339,37
355,74
450,0
400,0
350,0
300,0
250,0
200,0
150,0
100,0
50,0
0,0

Figure 5

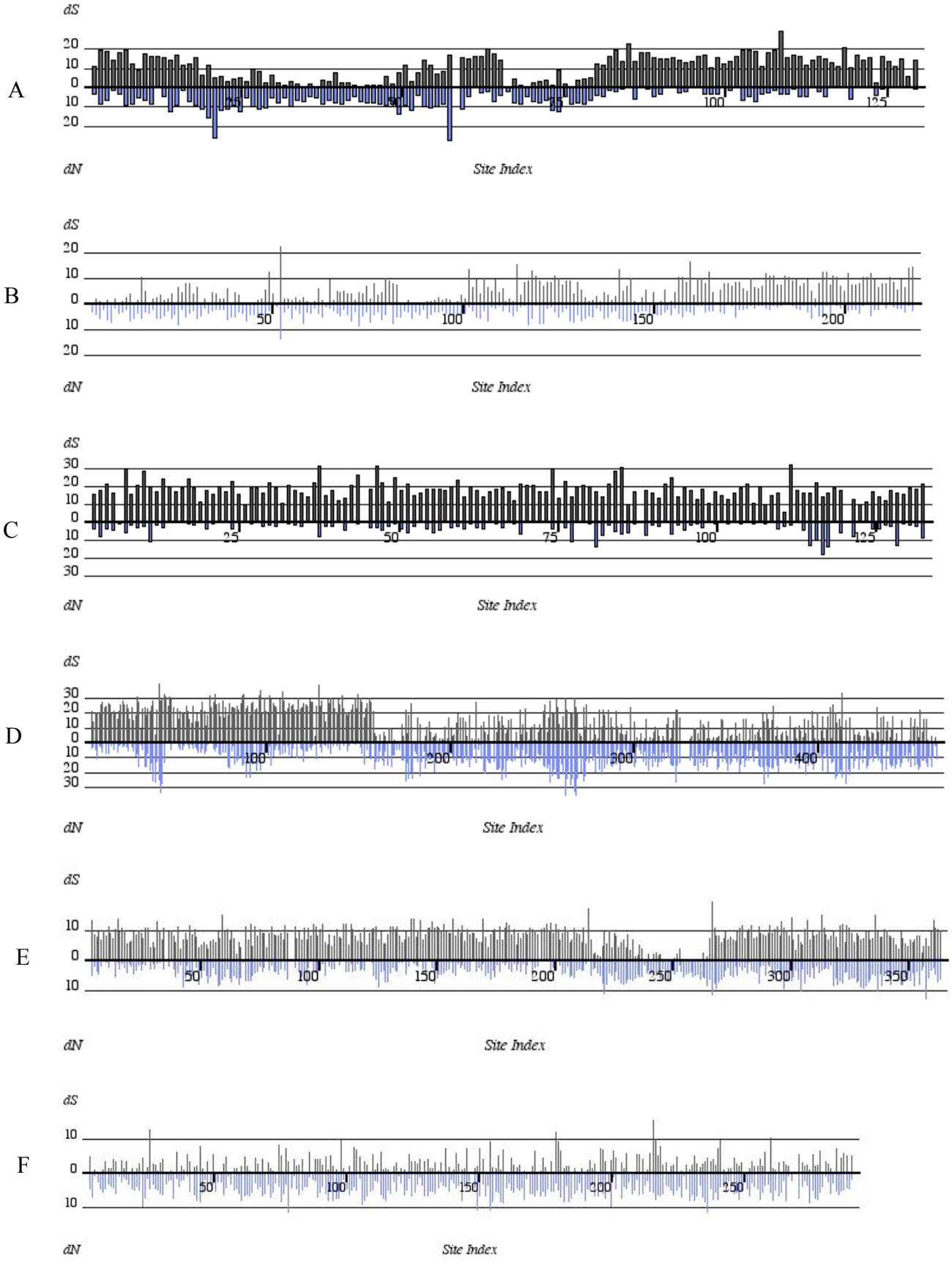

A
dS
20
10
0
10
20
25
50
75
100
125
dN
Site Index
B
dS
20
10
0
10
20
50
100
150
200
dN
Site Index
C
dS
30
20
10
0
10
20
30
25
50
75
100
125
dN
Site Index
D
dS
30
20
10
0
10
20
30
100
200
300
400
dN
Site Index
E
dS
10
0
10
50
100
150
200
250
300
350
dN
Site Index
F
dS
10
0
10
50
100
150
200
250
dN
Site Index

Figure 6

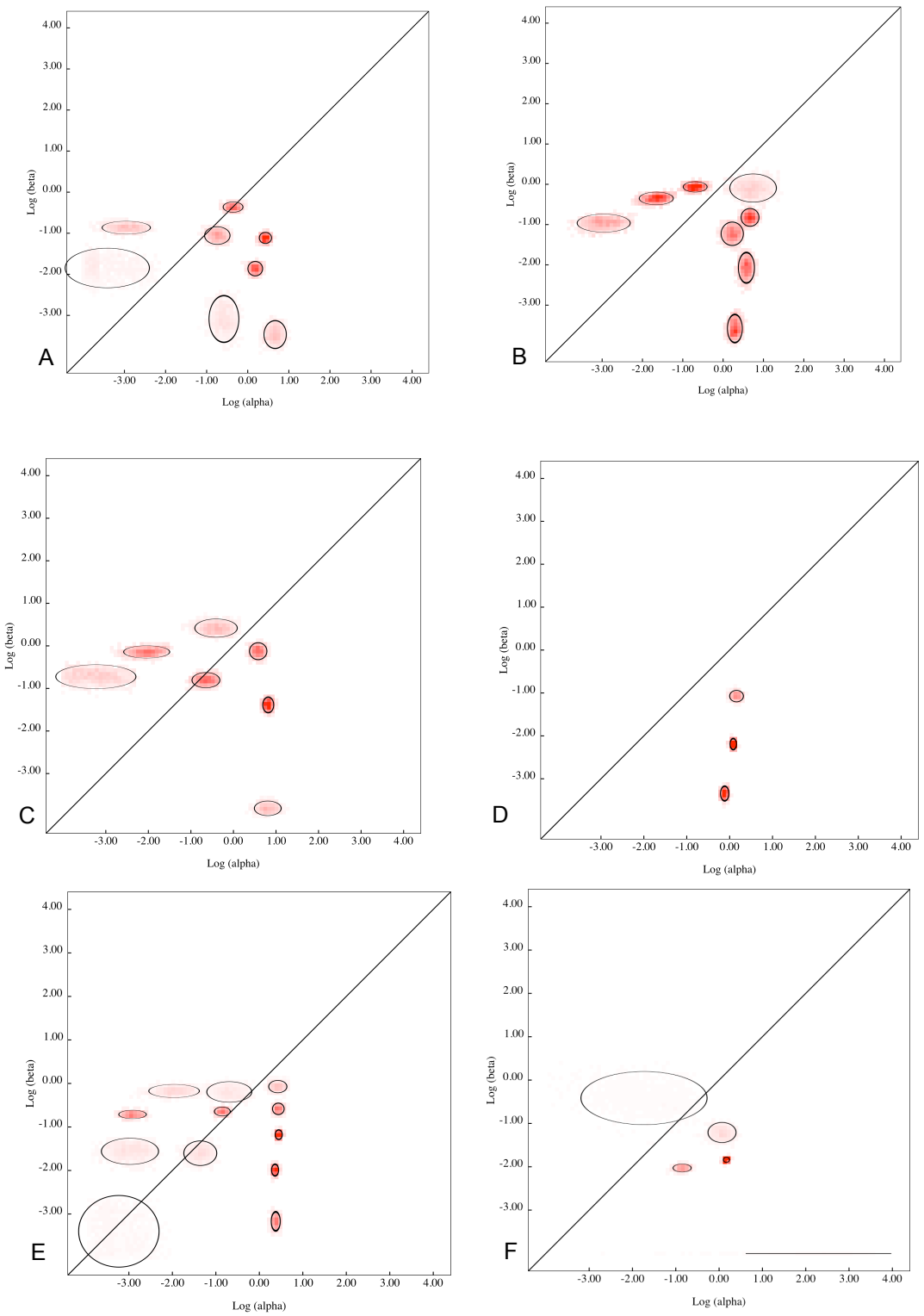


Table 1. Selection pressure analysis in iron uptaken gene families. Slac. Single Likelihood Ancestor Counting. * The analysis does not resulting an rate class for dN/dS.

Gene Family	Sequences	Codons	Tree length (subs/site)	Directionally selected residues	Directionally selected sites	Negatively selected sites (SLAC)	Positively selected sites (SLAC)	Sites with dN/dS<=1	Sites with dN/dS>1	Global dN/dS ratio
NAS	29	264	14.22	6	18	51	29	250	14	0.725
NRAMP	49	134	6.85	1	1	123	-	*	*	0.102
YSL	81	351	19.21	20	132	201	58	309	42	0.453
FRO	22	403	21.21	2	4	180	25	*	*	0.381
IRT	40	144	15.36	5	16	60	17	115	29	0.476

6. ARTIGO 4. Divergence time of NRAMP gene familes in plants

(A ser submetido para publicação no periódico Molecular Biology and Evolution)

Evolution, divergence time and *cis-acting* analysis in *Nramp* gene family across plants evolutionary lineages.

Filipe de Carvalho Victoria[1,2], Daniel da Rosa Farias[2,3], Clauber Mateus Priebe Bervald[2,4], Luciano Carlos da Maia[2] Rogerio Oliveira de Sousa[5] and Antônio Costa de Oliveira[2].

[1]*Graduate Student in Biotechnology, Centro de Biotecnologia, Universidade Federal de Pelotas, RS, Brazil.*

[2]*Plant Genomics and Breeding Center, Faculdade de Agronomia Eliseu Maciel, Universidade Federal de Pelotas, RS, Brazil.*

[3]*Gradute Student in Agronomy, FAEM, Universidade Federal de Pelotas, RS, Brazil.*
[4]*Ricetec Sementes, Rua 18 de novembro, 341, Navegantes, 90240-040 - Porto Alegre, RS – Brazil.*
[5]*Soil Science Departament, Faculdade de Agronomia Eliseu Maciel, Universidade Federal de Pelotas, RS, Brazil.*

ABSTRACT

The *NRAMP* genes represent a large family of metal transporters that are well conserved from bacteria to mammals. This gene family occurs in several copies in different plant genomes. The identification of orthologue and paralogue copies become necessary for understanding the relationships within gene families. In this study we used phylogenetics tools to infer on the divergence time of *NRAMP* gene members in rice. Searches for homologues of NRAMP genes in plants were made using the rice (*Oryza sativa* L.) copies as queries. The obtained sequences were subjected to a global alignment and the conserved regions were selected to evaluate the relationship of *NRAMP* homologues in plants. The bayesian phylogenetic arrangements were tested for the best substitution model using the *Physcomitrella patens* (Hedw.) Bruch & Schimp as outgroup. In the resulting phylogenetic tree, no evidence of different lineages for the major flowering plant groups was observed, since monocot and dicot homologues were grouped together in different branches. The homologues for *P. patens* were grouped at the root of the phylogeny, indicating the occurrence of a single process for the divergence of *NRAMP* copies in non-vascular and vascular plants. No recent duplication in rice was found, paralogue copies were only observed for dicot species, confirmed by the cis-acting homology search in the promoter regions of *NRAMP* genes. These results indicate an ancestral duplication hypothesis for this gene family in grasses. In grass genomes, no paralogue copies for *NRAMP* genes were found, suggesting that all ortologue copies can be contributing for metal homeostasis in *Oryza sativa* and allies.

Key-words: *multi-copy gene families; plant molecular evolution; Cis-acting elements; relative codon usage; Bayesian inference.*

INTRODUCTION

To fulfill many essential functions ranging from metal absorption to metal sequestration and storage plants need metal transporters (Hell and Stephan, 2003). Because the abundance and the bioavailability of micronutrient metals can be very limiting in some soils, plants have developed efficient absorption strategies. In addition to metal absorption, plants also need to be able to transport transition metals to the growing organs and to the cell compartments where they are necessary. A fine control of metal concentrations is required in chloroplasts in photosynthetic tissues, where metals play essential roles in photosynthesis but can cause serious oxidative damage. In some cases, plants also have to deal with toxic heavy metals such as cadmium, lead and mercury or toxic excess of essential metals. In this case, transporters can function either in excluding metals at the root or sequestering metals in some cell compartments such as the vacuole (Kim et al., 2006). The *NRAMP* (*Natural Resistance-Associated Macrophage Protein*) constitutes a highly conserved integral membrane protein family involved with iron transport in several organisms, including bacteria, fungi, plants and animals (Cellier et al. 1995). These genes are widely distributed in all plant families, acting mainly in divalent cations transports (Curie and Briat, 2003). The first occurrence of *NRAMP* genes were identified in mammals (*Nramp1*), coding a macrophage membrane protein responsible for cation concentration in the phagosoma, regulating the phagocited bacterial activity (Williams et al*.,* 2000).

Plants also harbor *NRAMP* gene families involved in metal uptake from the environment. Conservation of *NRAMP* protein function in eukaryotes underlies the fundamental importance of redox metal homeostasis, and phylogenetic analyses suggest that Nramp functional homologs from fungi, plants and animals derived successively from a common ancestor (Cellier et al., 2001).

In higher eukaryotes, gene transcription is controlled by a variety of mechanisms such as chromatin modifications or degradation via complementary miRNAs. Gene promoters and their *cis*-acting regulatory element composition, however, are the initial checkpoints for transcriptional gene activities and define the potential spatiotemporal expression of a gene (Wang et al., 2009).

Cell signaling is one aspect of the complex system of communication that coordinates basic cellular activities and interactions of a cell with its environment. Transcriptional regulatory networks that drive organ-specific and cell-specific patterns of gene expression and mediate interactions with the environment represent one aspect of plant cell signaling (Priest et al., 2009).

In plants, transcriptional regulation is mediated by a large number of transcription factors (TFs) controlling the expression of tens or hundreds of target genes in various, sometimes intertwined, signal transduction cascades (Wellmer and Riechmann, 2005).

Cis-acting regulatory elements are important molecular switches involved in the transcriptional regulation of a dynamic network of gene activities controlling various biological processes, including abiotic stress responses, hormone responses and developmental processes (Yamaguchi-shonozaki and Shinozaki, 2005). Although often only 5 to 20 bp in length, they are critical for understanding gene regulation (Liu et al., 2004).

Interaction between transcription factors (TFs) and *cis*-acting sequences in the upstream region of genes is crucial for gene regulation at the transcription level, because this controls a number of target genes responsible for various downstream physiological responses (Kong and Yang, 2010). Various computational methods have been developed to predict promoters or other cis-elements upstream of target genes (Hahmuradov et al., 2005; Zhang et al., 2005; Zhou et al., 2007).

Since some cis-acting elements are known to be directly involved in gene transcription regulation under Fe deficiency, one will be able to use these *cis*-elements to identify putative genes of interest. A recent study demonstrated the genomic analysis of *cis*-elements (e.g. TATA-box) of rice microRNA (miRNAs) genes surrounding the regions of promoters and detected 249 promoters for 212 rice pre-miRNA sequences (Cui et al., 2009). Thus, the computational approach facilitates search for the genomic wide target genes and analysis of their functional annotation.

Comparative genomics has been proven to be a powerful tool for the discovery of a large variety of functional elements taking into account their conservation between related species. In particular, it has been shown that comparative genomics approaches are able to detect genetic elements that are often difficult to discover due to their small size and/or limited information content. (Haberer et al., 2006). This strategy has been proposed to aid regulatory element

identification by examining orthologous sequences from multiple species (Liu et al., 2004).

METHODS

A local alignment between *Oryza sativa* L. (*query*), *Arabidopsis thaliana* (L.) Heynh. and *Physcomitrella patens* Brid. was performed with the BLAST tool (Altschul et al., 1990) accessed from web http://blast.ncbi.nlm.nih.gov/Blast.cgi. The sequences of rice *NRAMP* genes were selected using the *Genbank* access number provided by Gross et al. (2003). The gene homology was defined by the maximum alignment score associated with the higher query coverage and the lower *E-value*.

The selected sequences were subjected to the global alignment in ClustalW (Larkin *et al*., 2007). An initial tree was built using the *Neighbor-Joining* method (Saitou and Nei, 1987), and 1,000 boostrap replicates, with the aid of MEGA 4 software (Tamura *et al*., 2007), to verify the phylogenetic signal. The monophyletic groups were subjected to modelling to find the best substitution model in the *Jmodeltest* (Posada, 2008). The appropriate model was selected to run a Bayesian analysis in the *BEAST* software (Drummond and Rambaut, 2007), being the resulting trees used to build the consensus tree for the phylogeny. The phylogenetic analysis was performed with help of EPOS for MacOSX 10.5.1. software (Griebel *et al.*, 2008).

Synonymous and non-synonymous substitutions rates (dS/dN) were estimated by dividing the observed number of synonymous and non-synonymous substitutions per gene (dS and dN, respectively) by the number of synonymous and non-synonymous sites, S and N, respectively. Using the HyPhy software (Pond et al., 2004), we estimated pairwise substitution rates using the most appropriated substitution computed before.

The divergence time estimates were computed using the Bayesian inference using all sequences selected for this study, with 100,000 replicates. The dates were estimated using the probably date for bryophyte origins (Megallón and Hilu, 2009) associated with the dS/dN rates for date calibration. The resulting tree was edited with help of *FigTree* software (Vlad et al., 2008).

For each gene, codon-usage bias was computed with General Codon Usage Analysis software *GCUA* (McInerney 1998) for the effective number of codons (Wright 1990). The relative synonymous codon usage (RSCU) of each codon in each

gene were also computed, which quantifies the abundance of each codon relative to that expected under equal usage of alternative codons of the same amino acid. Heat maps of RSCU were constructed with CIMMiner (http://discover.nci.nih.gov/cimminer) (Weinstein et al. 1997).

The putative promoter region (1,000 bp) upstream from a selected species with complete genome sequences was extracted manually. Different online tools were used, according to species, TAIR (http://www.arabidopsis.org/) for *Arabidopsis thaliana,* RAP-DB (http://rapdb.dna.affrc.go.jp/) for *Oryza sativa*, SoyBase (http://soybase.org/index.php) for *Glycine max* (L.) Merr.*,* and phytozone (http://www.phytozome.net/) for *Populus trichocarpa* Torr. & A. Gray and *Sorgum bicolor* (L) Moench. The selected promoters were used for the cis-acting regulatory element prediction was runned in the PRECISE software (Trindade et al. 2005) using the PlanProm DB (Ilahm et al. 2003) as a promoter reference database, downloaded from the Release 2009.02 (http://cub.comsats.edu.pk/plantpromdb.htm). The most representative motifs were selected based in the relative frequency ratio F_{Tr}/F_{Rr} with a significative *z-value* each motif found, where F_{Tr} is the relative frequency of a motif within the selected set of sequences, and F_{Rr} is the relative frequency of a motif within the reference set of sequences. For detailed analysis and phylogenetic relationship these promoters were subjected to the FootPrinter software (Blanchette and Tompa 2003).

RESULTS AND DISCUSSION

Phylogenetic relationship of Nramp genes copies in selected plant species

Four homologue sequences were found for *P. patens* using *O. sativa Nramp* copies as queries. All moss homologues were grouped in the root of topology, as expected for the land plants phylogeny (Soltis et al. 1999). In *Nramp* phylogenetics relationship (Figure 1) no evidence for the divergence of the major flowering plant groups was observed, since monocot and dicot homologues were grouped together in different branches. The homologues for *P. patens* were grouped at the root of the phylogeny, which indicates the occurrence of a single process for the divergence of *Nramp* copies for non-vascular and vascular plants. Thomine and Schreoder (2004), based on *NRAMP* gene analyses in derived plant species, such as rice, tomato, *Arabidopsis, Medicago truncatula* Gaertn. and soybean (*Glycine max*) also evidenced the early divergence of *Nramp* copies, reporting the existence of two subfamilies for

these genes, one closer to animal homologues and the other related with green plants copies. The narrow inter-specific relationship was found by Gross et al (2003) for *OsNramp2* and *OsNramp*8 genes in rice and the *AtNramp2, AtNramp3, AtNramp4* and *AtNramp5* genes in *Arabidopspis*, forming a comparted branch with the root of the phylogeny, suggesting the copies relationships in rice as an evidence of a recent duplication in these species.

When other plant species homologues were added to the analysis (Supplementary files) a significantly detachment of rice copies was observed. All rice *Nramp* copies are related with the other grass species analyzed, *Sorghum bicolor* (L.) Moench. Considering the divergence of the BEP and the PACCAD clades, representing rice and sorghum respectively, this occurs in c.a. 40 Mya (Bremer 2002), suggesting a duplication in the Poaceae root for Nramp genes. The exceptions are *OsNramp6* and *OsNramp8*, but both are in an outside branch of other rice and sorghum copies, probably reflecting an ancestrality of the grass copies. Therefore, no exclusive branch for monocots and dicots were observed (Figure 1). The two major branches showed a distinct relationship within plant *Nramp* homologues with one of them closely related with the lower plant copies included in the study. The divergence of these clades apparentely occurred at 340 Mya, reconstructing the divergence within conifers and flowering plants (Soltis et al 2002; Megallón and Sanderson 2005; Zimmer et al. 2007). The close relationship of *AtNramp2*, *AtNramp3, AtNramp4, AtNramp5,*and their dicot allies, such as tomato (*Lycopersicum esculetum* Mill.), castor oil plant (*Ricinus communis* L.), soybean, the rice copies *OsNramp2, OsNramp7, OsNramp8* and the two sorghum related copies with the Sitka spruce (*Picea sitchensis*) and moss homologues reinforces the ancestral hypothesis of these clade.

Several approaches reports the multi-copy nature of *NRAMP* gene family in higher plants (Marshner and Romheld 1994; Eide 1998; Moria 1999; Bennetzen 2002), also evidencied in lower groups now with the moss homologues occurrence found in the *BLAST* search using as a query the know genes for the main model plants, rice (Gross *et al.*, 2003) and *Arabidopsis* (Thomine and Schreoder 2004). These results strengthen the ancestral duplication hypothesis of this gene family.

Selective pressure and divergence time of Nramp genes in plant species

All homologues studied showed substitution rates lower than 1.0, suggesting a strong selective pressure on the *Nramp* genes in plants (See supplementary files). For the moss copies PHY111375 and PHY179667, the non-synonymous substitution appears in a higher rate, indicating a positive selection among these phylogenetical branch, problably these *Nramp* homologues are not essential for iron homoestasis or are pseudogenes, since both accumulated more mutations when compared with the other moss copy (PHY179667). The rice *Nramp* copies are expressed in distinct plant parts (Banerjee and Chandel 2011), except for *OsNramp8*. Apparently these copies are not essential for iron acquisition, since experiments have reported a low expression of these genes in rice plant subjected to iron stress (Sperotto et al. 2010). The selective pressure ratio on these copies does not differ from the other rice copies, turning it impossible to make inferences about their functionality. Purifying selection (dN/dS<1) is a common characteristic of multi-copy genes and was described for other iron uptake genes in plants (Perovic et al. 2007).

Codon-bias analysis

Codon usage is highly nonrandom for all plant species analyzed (Fig 2), perhaps because these species do not differ dramatically in overall base composition, ranging from 41 to 56% of GC content (Victoria et al. 2011). On the other hand, despite these similarities, the distribution pattern is different when comparing the species, since Arabidopsis and rice have a preference usage for GC1 and GC3 bases, respectively (Victoria et al. 2011) and 15% of the moss genes are unbiased (Rensing et al. 2005). It is clear that base compositional differences among species contributes, at least in part, to their different relative usage of synonymous codons, with alternative codons with more G or C bases being relatively more frequently incorporated in high G + C content genomes and vice versa.

Several codons were underrepresented across species and their copies. Comparing all species, a low occurrence of codons for leucine (TTA), isoleucine (ATA) and arginine (CGC) were found in the *Nramp* genes. However, arginine (AGA, AGG) and alanine (GCT) were highly represented in the *Nramp* copies in plants. Only for *OsNramp1, OsNramp2* and the *Sorghum* homologues *Sb2* these codons show a lower usage and probably reflect a functional similarity in these two grass genes copies. Differences in codon usage for several amino acids reflect an effect of phylogeny. For example, glycine GGG codon is used mainly for monocot

copies. The lower usage of this codon is probably related with the detrimental effect on mRNA tertiary structure (Kreitman and Antezana 1999). It is not clear why the other codons are so rare in both absolute terms and especially in highly expressed genes, probably reflecting functional differences. For example *OsNramp1* and *OsNramp2* were reported as having a high basal expression in culture medium and when the iron supply is increased their transcription response positively (Zhou and Yang 2004). The ATA codon is overrepresented when comparing with other rice *Nramp* copies. These behavior appears following the structure of *Arabidopsis* copies where the ATA codon is also underrepresented in at least two copies that demonstrate a high efficiency in iron uptake (Thomine et al. 2000).

Cis-acting analysis

Promoters for rice, *Arabidopsis, Sorghum bicolor, Populus trichocarpa* and soybean Nramp genes copies were used to compare the cis-acting frequencies and distribution. A total of 64 motifs were found in these promoter regions. The 20 most representative elements were used for a phylogenetic analysis (Table 1). The footprint results are shown in Figure 3. The distribution of the mostly motifs indeed reconstruct the phylogenetic relationships found when the gene regions were compared by the bayesian inference. Some motifs were found only in the grass clade, such as TTCTTG, CTAGTGC, CCGCCGC, CGCCGCC and CGCTGCT found in closely related rice and sorghum copies, strengthening the ancestral duplication of rice *Nramp* copies. Whereas all copies analyzed were functional iron uptake genes, no specific elements were found in the promoters, since some copies showed higher frequencies for individual motif and no motif was found conserved in all promoters analyzed. Only AAAAAT, GAAAAA, TTTTCT, TTTGTT and CAGGAA were widely distributed in monocot and dicot promoters, however not for all species. Some important regulators for iron uptake genes such as *OsIRO2* (Ogo et al. 2007) and *IDE-like* (Kobayashi et al. 2005) were not found with high frequency motifs in *Nramp* promoters used in this analysis. However, some cis-element related with the gene expression under stress conditions, such as GT1GMSCAM4 (Ma et al. 2009), represented by the GAAAAA motif, were found in *OsNramp4, OsNramp6, OsNramp3,* in two copies for sorghum, *PtNramp1, PtNramp2, PtNramp3, AtNramp1* and *AtNramp6.* This cis-acting element was reported as a patogen-induced response in plants of *Arabidopsis* and soybean (Parker et al. 2004), the same function is also

atributted to the *Nramp* genes in mammals (Goswami et al. 2001). Probably, these motif conservation as being a reminiscence of the ancestral *Nramp* copy, since this is an ancient function (Courville et al. 2006) lost in some of modern plants, or it is an unknown function of the metal homoestasis gene family in plant species.

REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. Journal of Molecular Biology. 215: 403-410.

Banerjee S, Chandel G. 2011. Understanding the role of metal homeostasis related candidate genes in Fe/Zn uptake, transport and redistribution in rice using semi-quantitative RT-PCR. Journal of Plant Molecular Biology and Biotechnology. 2(1): 33-46.

Bennetzen J. 2002. The rice genome. Opening the door to comparative plant biology. Science. 296: 60–63.

Blanchette M, Tompa M. 2003. FootPrinter: a program designed for phylogenetic footprinting. Nucleic Acids Research. 31(13): 3840-3842.

Bremer K. 2002. Gondwanan evolution of the grass alliance of families (Poales). Evolution. 56(7):1374-1387.

Cellier MFM, Bergevin I, Boyer E., Richer E.2001. Polyphyletic origins of bacterial Nramp transporters. TRENDS in Genetics. 17(7): 365-370

Cellier MFM, Privé G, Belouchi A, Kwan T, Rodriques V, Chia W, Gros P. 1995. Nramp defines a family of membrane proteins. Proceedings of Natural Academy of Science of USA. 92:10089-10093.

Courville P, Chaloupka R, Cellier MFM. 2006. Recent progress in structure–function analyses of Nramp proton-dependent metal-ion transporters. Biochemistry and Cell Biology. 84:960-978.

Cui X, Xu SM, Mu DS, Yang ZM. 2009. Genomic analysis of rice microRNA promoters and clusters. Gene. 431(1-2): 61-66.

Curie C., Briat JF.2003. Iron transport and signaling in plants. Annual Revue of Plant Biology. 54:183–206.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology. 7: 214.

Eide DJ. 1998. The molecular biology of metal ion transport in Saccharomyces cerevisiae. Ann Rev Nutr. 18: 441–69.

Goswami T, Bhattacharjee A, Babal P, Searle S, Moore E, Li M, Blackwell JM. 2001. Natural-resistance-associated macrophage protein 1 is an H^+/bivalent cation anti-porter. Biochemical Journal. 354: 511–519.

Griebel T, Brinkmeyer M, Böcker S. 2008. EPoS: a modular software framework for phylogenetic analysis Bioinformatics Applications Note. 24(20)*:* 2399–2400.

Gross J, Stein RJ, Fett-Neto AG, Fett JP. 2003. Iron homeostasis related genes in rice. Genetics and Molecular Biology. 26(4): 477-497.

Haberer G, Mader MT, Kosarev P, Spannagl M, Yang L, Mayer KF. 2006. Large-scale cis-element detection by analysis of correlated expression and sequence conservation between Arabidopsis and Brassica oleracea. Plant Physiology.142: 1589–1602.

Hahmuradov IA, Solovye VV, Gammerman AJ. 2005. Plant promoter prediction with confience estimation. Nucleic Acids Research. 33(3): 1069-1076.

Hell R, Stephan UW. 2003. Iron uptake, trafficking and homeostasis in plants. Planta. 216: 541–551.

Kim SA, Punshon T, Lanzirotti A, LI L, Alonso JM, Ecker JR, Kaplan J, Guerinot ML. 2006. Localization of iron in *Arabidopsis* seed require the vacuolar membrane transporter VIT1. Science. 314:1295-1298.

Kobayashi T, Suzuki M, Inoue H, Itai RN, Takahashi M, Nakanishi H, Mori S, Nishizawa NK. 2005. Expression of iron-acquisition-related genes in iron-deficient rice is co-ordinately induced by partially conserved iron-deficiency-responsive elements Journal of Experimental Botany. 56(415): 1305–1316.

Kong WW, Yang ZM. 2010. Identification of iron-deficiency responsible microRNA genes and *cis*-elements in *Arabidopsis.* Plant Physiology and Biochemistry. 48: 153-159.

Kreitman M, Antezana M. 1999. The population and evolutionary genetics of codon bias. In: Singh RS, Krimbas CB, editors. Evolutionary genetics: from molecules to morphology. New York: Cambridge University Press. p. 82–101.

Larkin MA, Blackshields G, Brown NP et al. 2007. Clustal W and Clustal X version 2.0. Bioinformatics. 23: 2947–2948.

Liu Y, Liu XS, Wei L, Altman RB, Batzoglou S. 2004. Eukaryotic regulatory element conservation analysis and identification using comparative genomics. Genome Research**.** 14: 451–458.

Ma K, Xiao J, Li X, Zhang Q, Lian X. 2009. Sequence and expression analysis of the C3HC4-type RING finger gene family in rice. Gene 444: 33–45

Magallón S, Hilu KW. 2009. Land plants (Embryophyta). *In:* Hedges SB, Kumar S. editors. The Timetree of Life. Oxford. University Press. p. 133–137.

Marschner H, Röhmeld V. 1994. Strategies of plants for acquisition of iron. Plant Soil. 165:375–88.

McInerney JO. 1998. GCUA: General Codon Usage Analysis. Bioinformatics Applications Note. 14(4): 372-373.

Morariu VI, Srinivasan BV, Raykar VC, Duraiswami R, Davis LS. 2008. Automatic online tuning for fast gaussian summation. Advances in Neural Information Processing Systems (NIPS). 8.

Mori S. 1999. Iron acquisition by plants. Current Opinion in Plant Biology. 2: 250–253

Ogo Y, Itai RN, Nakanishi H, Kobayashi T, Takahashi M, Mori S, Nishizawa NK. 2007. The rice bHLH protein OsIRO2 is an essential regulator of the genes involved in Fe uptake under Fe-deficient conditions. The Plant Journal. 51, 366–377.

Park HC, Kim ML, Kang YH et al. 2004. Pathogen- and NaCl-induced expression of the SCaM-4 promoter is mediated in part by a GT-1 box that interacts with a GT-1-like transcription factor. Plant Physiol. 135: 2150–2161

Perovic D, Tiffin P, Douchkov D, Bäumlein H, Graner A. 2007. An integrated approach for the comparative analysis of a multigene family: The nicotianamine synthase genes of barley. Functional and Integrative Genomics. 7(2): 169-179.

Pond SLK, Frost SDW, Muse SV. 2005. HyPhy: Hypothesis testing using phylogenies. Bioinformatics. 21(5):676–9.

Posada D. 2008. jModelTest: Phylogenetic Model Averaging. Molecular Biology and Evolution. 25: 1253-1256.

Priest HD, Filichkin SA, Mockler TC. 2009. *Cis*-regulatory elements in plant cell signaling. Current Opinion in Plant Biology. 12: 643–649.

Saitou N, Nei M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution. 4: 406-425.

Shahmuradov AI, Gammerman AJ, Hancock JM, Bramley PM, Solovyev VV. 2003. PlantProm: a database of plant promoter sequences. Nucleic Acids Research. 31: 114-117.

Soltis PS, Soltis DE, Savolained, Crane PR, Barraclough TG. 2002. Rate heterogeneity among lineages of tracheophytes: integration of molecular and fossil data and evidence for molecular living fossils. Proceedings of Natural Academy of Science of U S A. 99(7):4430-5.

Soltis PS, Soltis DE, Wolf PG, Nickrent DL, Chaw SM, Chapman RL. 1999. The

phylogeny of land plants inferred from 18S rDNA sequences: pushing the limits of rDNA signal? Molecular Biology and Evolution. 16 (12): 1774-1784.

Sperotto RA, Borr T, Duarte GL, Santos LS, Grusak MA, Fett JP. 2010. Identification of putative target genes to manipulate Fe and Zn concentrations in rice grains. Journal of Plant Physiology. 167(17): 1500-1506.

Tamura K, Dudley J, Nei M, Kumar, S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution. 24: 1596-1599.

Thomine S, Schroeder JI. 2004. Plant Metal Transporters with Homology to Proteins of the NRAMP Family. *In:* Cellier M, Gros P. *The NRAMP Family*. Andes/Kluwer series: Molecular Biology Intelligence Unit. p113-121.

Thomine S, Wang R, Ward JM, Crawford NM, Schroeder JI. 2000. Cadmium and iron transport by members of a plant metal transporter family in *Arabidopsis* with homology to Nramp genes. Proceedings of the National Academy of Sciences. 97: 4991-4996.

Trindade LM, van Berloo R, Fiers M, Visser RGF. 2005. PRECISE: Software for Prediction of cis-Acting Regulatory Elements. Journal of Heredity. 96(5): 618–622

Victoria FC, da Maia LC, de Oliveira AC. 2011. *In silico* comparative analysis in plants. BMC Plant Biology. 11:15.

Wang X, Haberer G, Mayer KFX. 2009. Discovery of cis-elements between *sorghum* and rice using co-expression and evolutionary conservation. BMC Genomics**.** 10:284.

Weinstein JN, Myers TG, O'Connor PM, et al. 1997. An information-intensive approach to the molecular pharmacology of cancer. Science. 275:343–349.

Wellmer F, Riechmann JL. 2005. Gene network analysis in plant development by genomic technologies. The International Journal of Developmental Biology. 49: 745–759.

Williams LE, Pittman JK, Hall JL. 2000. Emerging mechanisms for heavy metal

transport in plants. Biochim Biophys Acta. 1465:104-126.

Wright F. 1990. The effective number of codons used in a gene. Gene. 87:23–29.

Yamaguchi-Shinozaki K, Shinozaki K. 2005. Organization of cis-acting regulatory elements in osmotic- and cold-stress-responsive promoters. Trends in Plant Science. 10(2):.88–94.

Zhang W, Ruan J, Ho TD, You Y, Yu T, Quatrano RS. 2005. *Cis*-regulatory element based targeted gene finding: genome-wide identification of abscisic acid-and abiotic stress-responsive genes in *Arabidopsis thaliana*. Bioinformatics. 21(14): 3074-3081.

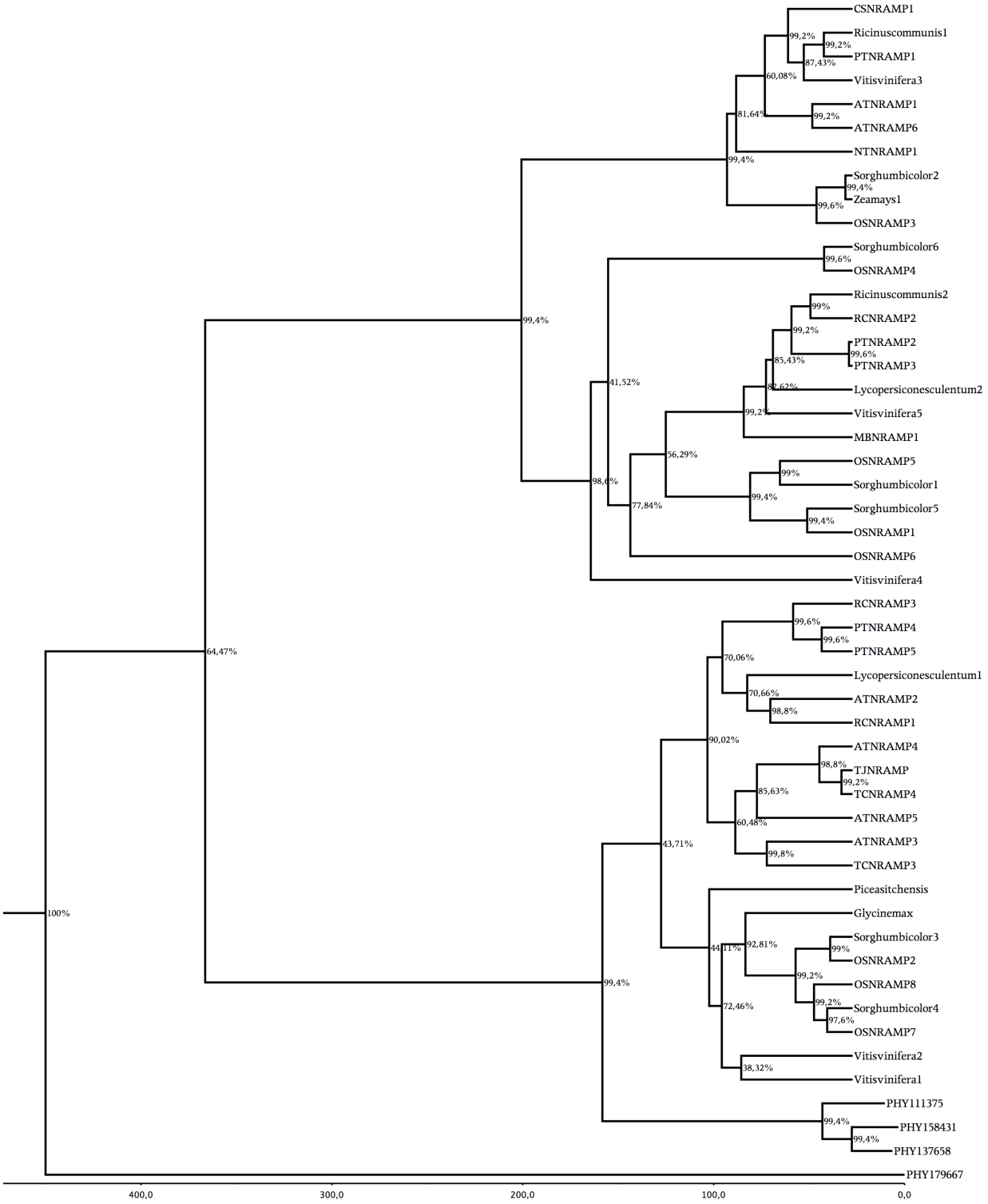
Zhou X, Yang Y. 2004. Differential expression of rice *Nramp* genes in response to pathogen infection, defense signal molecules and metal ions. Physiological and Molecular Plant Pathology. 65(5): 235-243.

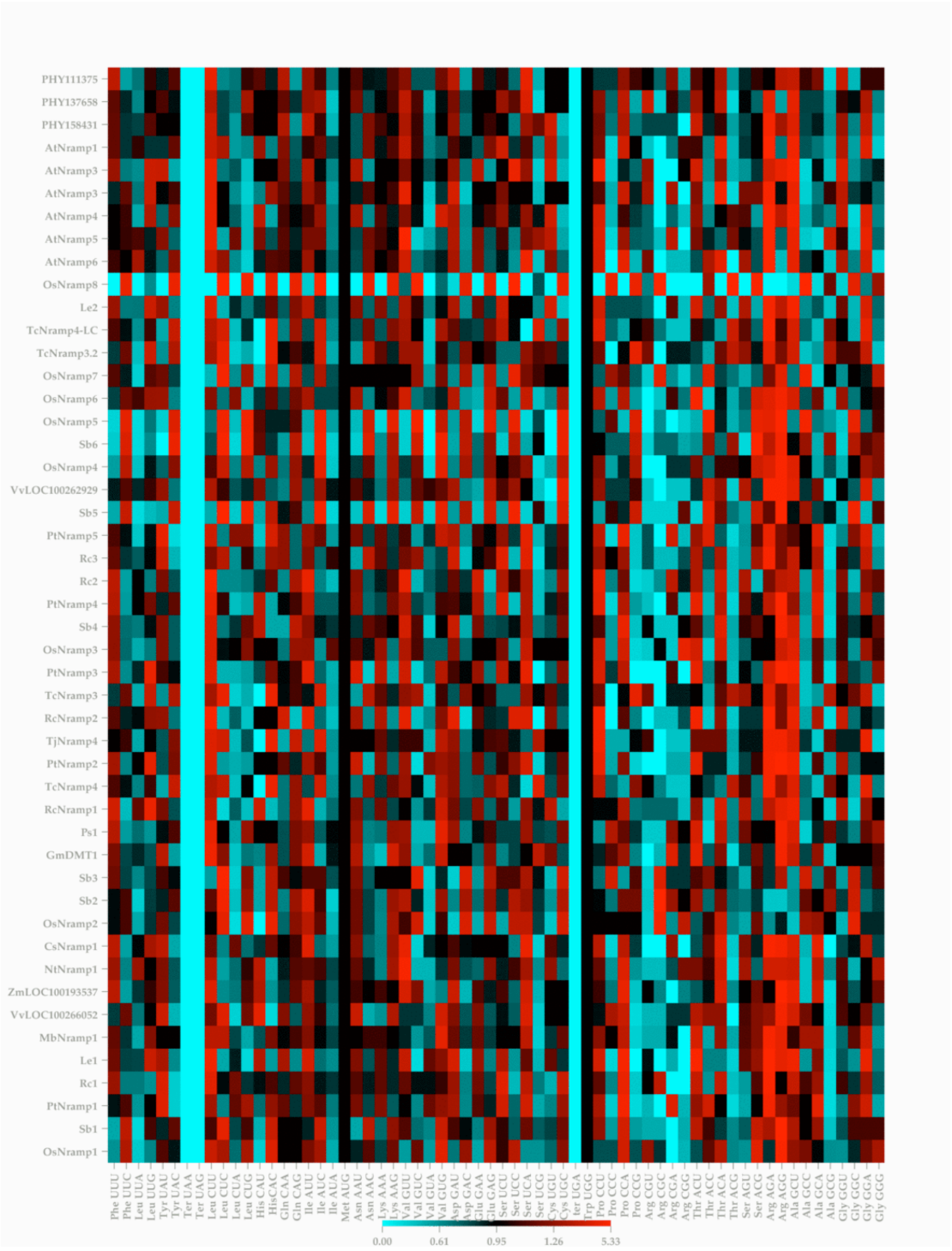
Zhou X, Wang G, Zhang W. 2007. UV-B responsive microRNA genes in *Arabidopsis thaliana*. Molecular Systems Biology. 3:103.

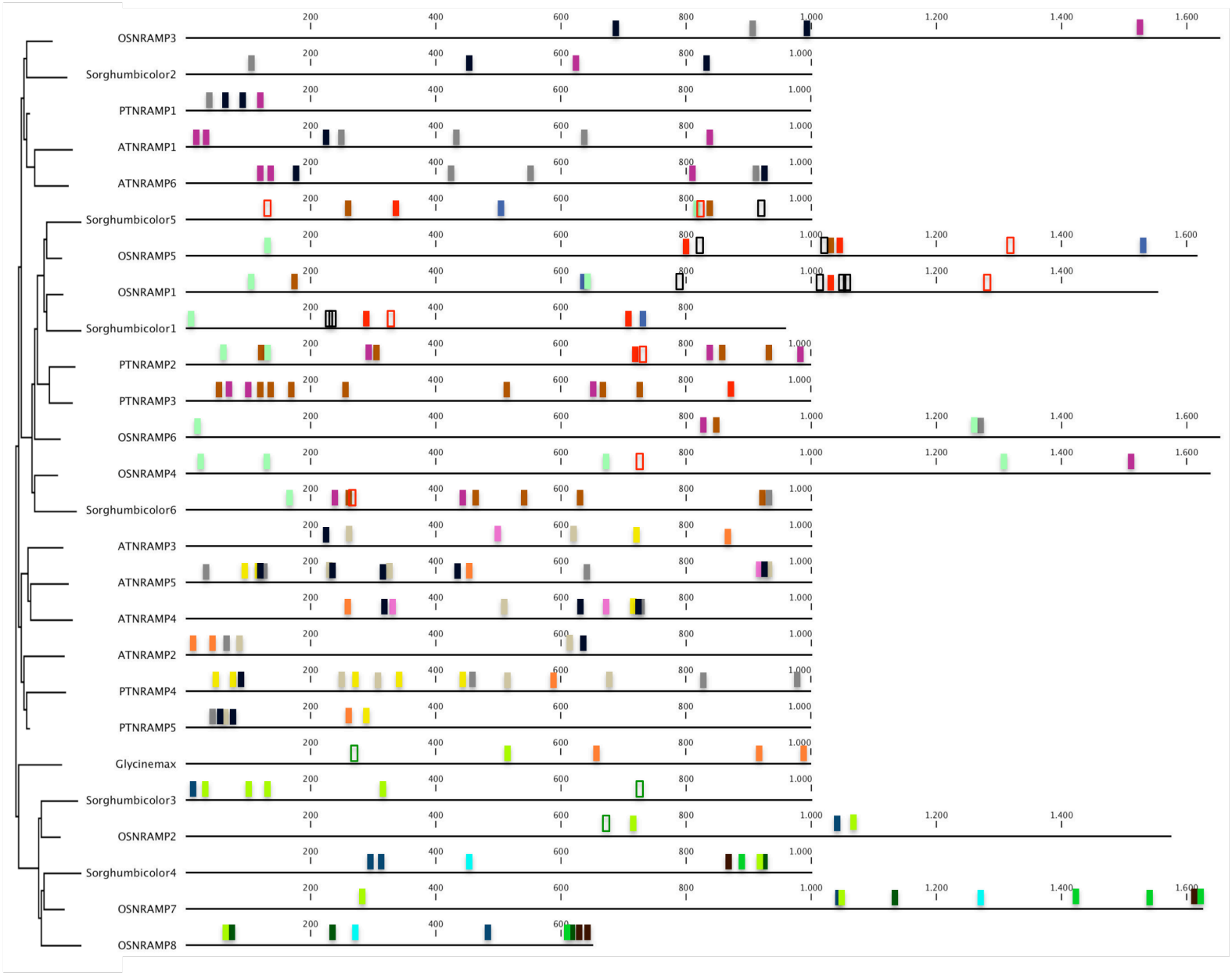
Figure 1. Phylogenetical tree based in the Bayesian inference analysis for Nramp genes copies in 16 plant species. The knot values indicates posterior significance. (Scale in Mya = million years ago)

Figure 2. Heat map of relative codon usage (RSCU) values for 47 *Nramp* plant homolgues. Each column represents a different codon, with the corresponding amino acid abbreviations and codon identity.

Figure 3. The most 20 representative cis-elements found in the phylogenetical footprint and their position in selected plant promoters plus strand. Os= *Oryza sativa*; At= *Arabidopsis thaliana*; Pt= *Populus trichocarpa.*


CSNRAMP1
Ricinuscommunis1
PTNRAMP1
Vitisvinifera3
ATNRAMP1
ATNRAMP6
NTNRAMP1
Sorghumbicolor2
Zeamays1
OSNRAMP3
Sorghumbicolor6
OSNRAMP4
Ricinuscommunis2
RCNRAMP2
PTNRAMP2
PTNRAMP3
Lycopersiconesculentum2
Vitisvinifera5
MBNRAMP1
OSNRAMP5
Sorghumbicolor1
Sorghumbicolor5
OSNRAMP1
OSNRAMP6
Vitisvinifera4
RCNRAMP3
PTNRAMP4
PTNRAMP5
Lycopersiconesculentum1
ATNRAMP2
RCNRAMP1
ATNRAMP4
TJNRAMP
TCNRAMP4
ATNRAMP5
ATNRAMP3
TCNRAMP3
Piceasitchensis
Glycinemax
Sorghumbicolor3
OSNRAMP2
OSNRAMP8
Sorghumbicolor4
OSNRAMP7
Vitisvinifera2
Vitisvinifera1
PHY111375
PHY158431
PHY137658
PHY179667
400,0
300,0
200,0
100,0
0,0


PHY111375
PHY137658
PHY158431
AtNramp1
AtNramp3
AtNramp3
AtNramp4
AtNramp5
AtNramp6
OsNramp8
Le2
TcNramp4-LC
TcNramp3.2
OsNramp7
OsNramp6
OsNramp5
Sb6
OsNramp4
VvLOC100262929
Sb5
PtNramp5
Rc3
Rc2
PtNramp4
Sb4
OsNramp3
PtNramp3
TcNramp3
RcNramp2
TjNramp4
PtNramp2
TcNramp4
RcNramp1
Ps1
GmDMT1
Sb3
Sb2
OsNramp2
CsNramp1
NtNramp1
ZmLOC100193537
VvLOC100266052
MbNramp1
Le1
Rc1
PtNramp1
Sb1
OsNramp1
Phe UUU
Phe UUC
Leu UUA
Leu UUG
Tyr UAU
Tyr UAC
Ter UAA
Ter UAG
Leu CUU
Leu CUC
Leu CUA
Leu CUG
His CAU
His CAC
Gln CAA
Gln CAG
Ile AUU
Ile AUC
Ile AUA
Met AUG
Asn AAU
Asn AAC
Lys AAA
Lys AAG
Val GUU
Val GUC
Val GUA
Val GUG
Asp GAU
Asp GAC
Glu GAA
Glu GAG
Ser UCU
Ser UCC
Ser UCA
Ser UCG
Cys UGU
Cys UGC
Ter UGA
Trp UGG
Pro CCU
Pro CCC
Pro CCA
Pro CCG
Arg CGU
Arg CGC
Arg CGA
Arg CGG
Thr ACU
Thr ACC
Thr ACA
Thr ACG
Ser AGU
Ser ACG
Arg AGA
Arg AGG
Ala GCU
Ala GCC
Ala GCA
Ala GCG
Gly GGU
Gly GGC
Gly GGA
Gly GGG
0.00
0.61
0.95
1.26
5.33



BOXES LEGEND

AATTTA	CAGGAA
AATCTT	CTAGTGC
AAAAAT	TATTTTT
TTTGTT	CGCCGCG
GAAAAA	CGTCGTC
CCGCCGC	AGATCC
CGCTGCT	TGTTCGT
TTCTTC	TCCATT
GCCGCCG	TTTTCT
TAATAA	TTTTTTA

Table 1. Most representative cis-acting motifs found in the promoter regions of *Oryza sativa, Arabidopsis thaliana, Populus trichiocarpa, Sorghum bicolor and Glycine max Nramp* gene copies.

Box	Length	n. of hits	z-value	Ft/Fr
TAATAA	6	38	0,0	126,26
ATTTTT	6	35	0,0	62,87
TTTTTT	6	91	0,0	37,45
CGCCGC	6	28	0,0	104,71
AAAAAT	6	40	0,0	51,23
GCCGCC	6	29	0,0	96,4
TTTATT	6	77	0,0	28,24
TTTTTTA	7	44	0,0	105,7
GAAAAA	6	54	0,0	41,72
AATTTA	6	35	1,53E-152	22,1
AATCTT	6	31	1,82E-020	4,92
CGCTGCT	7	5	3,78E-085	82,38
CTAGTGC	7	4	9,50E-038	47,52
CGTCGT	6	12	9,14E-010	6,33
CCGCCGC	6	8	1,11E-012	10,29
TTCTTG	6	42	2,93E-064	8,97
CAGGAA	6	22	4,53E-029	8,34
TTTTCT	6	52	8,09E-127	13,18
TTTGTT	6	45	5,83E-087	10,84
TATTTTT	7	40	4,37E-245	30,18

7. ARTIGO 5. Yellow-stripe like genes survey in Bryophytes based in the transferability of *IRAP/REMAP* markers

(A ser submetido para publicação no periódico BMC Evolutionary Biology)

Yellow-stripe like gene copies survey in Bryophytes based in the transferability of IRAP/REMAP markers.

Filipe de Carvalho Victoria[1,2], Nacieli Marini[2,3], Luciano Carlos da Maia[2], Rogério Oliveira de Sousa[4] and Antônio Costa de Oliveira[2§]

[1]*Graduate Student in Biotechnology, Centro de Biotecnologia, Universidade Federal de Pelotas, RS, Brazil.*

[2]*Plant Genomics and Breeding Center, Faculdade de Agronomia Eliseu Maciel, Universidade Federal de Pelotas, RS, Brazil.*

[3]*Gradute Student in Agronomy, FAEM, Universidade Federal de Pelotas, RS, Brazil.*

[4]*Soil Science Departament, Faculdade de Agronomia Eliseu Maciel, Universidade Federal de Pelotas, RS, Brasil.*

[§]Corresponding author

Email addresses:
FCV: filipevictoria@gmail.com
NM: nacymarini@gmail.com
LCM: lucianoc.maia@gmail.com
ROS: rosousa@ufpel.edu.br
ACO: acostol@terra.com.br

Abstract

Background

The sequencing of model plant species has provided an increase in the understanding of developmental and evolutionary processes occured during the formation of plants on Earth. However, basal plant species are less surveyed because the lack of genomic information available. The transference of molecular markers from close model species is well suited for plant evolution studies by the means of fullfilling the gaps and at the same time ellucidating the biological process through comparative genomics. The *Yellow Stripe Like* is a multi copy iron uptake gene family well studied in crop species. The presence and the copy number of these genes were not confirmed in basal species.

Results

In this study we test the transferability of 100 EST-SSR primers pairs from *Physcomitrella patens* Brid. a moss model species to six genotypes of *Polytrichum juniperinum* Hedw. Thirty percent of loci were sucesfully amplified with a relative polimorphism information. This markers were associated to IRAP/REMAP technics and similarities within southern american *Polytrichum* genotypes showing some distinction from the antarctic genotype. The *YSL* homologous in *Physcomitrella patens* appear being associated with a *copia-like* retrotransposon as occurs in maize *ZmYSL1*.

Conclusion

The transference of genomic information demonstrates a promissing approach to extensive gene survey across species as the *YSL* homologues searches in plants. The experimental results suggested the association of *YSL* genes with a copia-like element since the ancestral of land plants lineage.

Background

Wild species constitute a potential source of genetic variation for cultivated species. Besides, they can be analyzed to answer the long-lasting questions concerning the origins, evolution and spread of major agricultural crops of the world. Recently, there has been considerable progress in plant genomics, leading to novel molecular breeding tools to reduce the costs and to simplify the assays. Plant genome research has been focused on the major crops and model species and a vast amount of genomic information has been accumulated. This information will provide an opportunity to use it as sources of information for thousands of minor plants species, such as early land plants [1] and grass species [2].

Across-species transferability of SSRs derived from EST databases is greater than that of SSRs derived from enriched genomic DNA libraries, as they originate from expressed regions and therefore they are more conserved than non-coding regions across a number of related species [3]. They have been shown to be useful for comparative mapping across species, comparative genomics, and evolutionary studies and also to posses a higher potential for inter-specific transferability than genomic SSRs [4-9]. On the other hand, they are expected to be less polymorphic within the species due to its conserved nature [3]. In summary, EST-SSR has provided a valuable source of new PCR-based molecular markers in cereal crops.

In recent years, much has been learned about the genes and proteins necessary for primary Fe and Zn uptake from the soil [10]. Strategy II plants (the grasses) obtain Fe by secretion of Fe(III)-binding molecules, called phytosiderophores, and then by taking the Fe(III)-phytosiderophore complex into the root cells. Genes for phytosiderophore synthesis have been identified [11, 12, 13, 14], and a gene for Fe (III)-phytosiderophore uptake, *ZmYS1*, has been identified in maize [15]. Strategy I plants (non-grasses) obtain Fe by lowering the rhizosphere pH using H1-ATPase proteins, by reducing Fe(III) to Fe(II) with ferric reductase proteins, and by taking up the reduced Fe using Fe(II) transporter proteins [2, 16]. Ferric reductase genes have been identified in *Arabidopsis* and several other plant species [17, 18], as have Fe (II) transporter genes of the ZIP family [19, 20, 21, 22]. A homologous helitron was mapped on chromosome 6S in mayze. Thus was found closely to *YSL* related sequences and are unique to either B73 or Mo17 markers [23]. Analysis of the ends of the inserted sequence indicates that it is a long-terminal repeat retrotransposon [15].

The aims of this study were to transfer EST-SSR primers designed from *Physcomitrella patens* Brid. EST databank [24] to a basal moss species *Polytrichum juniperinum* Hedw. as well as to associate this molecular marker with IRAP/REMAP assesments using the transposable element related with maize *YSL* genes, to amplify an homologous region and infer about the *YSL* evolution in plants.

Results

EST-SSR transferability

From the 100 primer pairs based on the *Physcomitrella patens EST* database [24], 30 were functional and, when used with genomic DNAs from six *Polytrichum juniperinum* genotypes, gave 30% cross-species transferability across the six species. Among these cases,

43.3% of the EST-SSR primers yielded fragments of the expected size from one or more species (Fig. 1). Twenty-two (73.3%) of the thirty functional primers also showed transferability to one or more of the six *Polytrichum* genotypes tested. Thirteen (43.3%) of these 30 primers successfully amplified fragments in all genotypes. Polymorphism was detected with ten (33.33%) of the functional EST-SSRs. Six (20%) of these 30 primers also detected polymorphism between of one or more genotypes sampled almost all showing variation due to length of the SSRs, loci number (Fig. 1) or due to null alleles.

IRAP-REMAP transferability

To increase the probability of finding bands, one can combine primers from both 5' and 3' LTR ends or combine LTR primers with SSR primers to amplify intervening genomic DNA. As expected, the IRAP/REMAP analysis produced a high level of polymorphism (77% of bands appeared to be polymorphic), but with 50% of transferability (Fig. 2). The primer combinations used produced on average 1.5 and a maximum of five bands (Fig. 2). The average similarity of 0.71 and the cophenetic correlation coefficient of 0.83 (Fig. 3) suggested a relatively high similarity and a good adjustment between the original and the dendrogram-derived matrix, respectively [25]. *Polytrichum juniperinum* sampled in southern of Brazil (Gramado, Canela, Pelotas and Morro Redondo) appear as the most similar genotypes (0.89), suggesting a close relationship in this mosses subset, as expect due the geographic proximity of these populations. The inverse was also confirmed for the most distant populations, such as chilean and antarctic moss samples appear as the most distant similar pair. Using the *P. patens* K3 ecotype genotype as outgroup (Fig. 4), it was possible to infer the genotypes distances, where the antarctic moss is farther distant from the other moss sampled. The Gramado and Canela populations appear more closely related than Pelotas and Morro Redondo genotypes, as the Pelotas is a transitionary genotype to either Gramado/Canela or Morro Redondo genotype. The chilean genotype appears outlining the southern brazilian genotypes, making a linkage with the Morro Redondo population. However, a larger number of samples in Southern America would be needed to complete the geographic range of this species.

Retrotransposable element association with YSL genes in plants

Using the *ZmYSL1* as a query, two *Physcomitrella patens* putative *YSL* genes were recovered, based in their homology with the maize Strategy-II gene. These genes, named *PHY111567* and *PHY122023,* placed in the scaffold 2 and 38, respectively. The maize homologue was described in the chromosome 5, but other copies were found in chromosomes

6, 7 and 10 in the maize genome. The *Zmji* copies were found associated with the maize *YSL* homologues in each chromosome, being one copy in the chromosome 5 and two copies in other chromosomes. No amplicon related to *Zmji* primers appear associated with *YSL* moss homologues, but the LTR1 was found associated in the same scaffold containing *PHY111567* or *PHY122023*, with one and two copies respectively.

The higher distance from one LTR and a *YSL* copy was found in the chromosome 10 in maize (30 cM) placed downstream from the gene region (Fig. 5), a similar distance was found in the moss scaffold 2 (27.5 cM), but the LTR was placed in the upstream from the gene region. The lowest LTR distance from a *YSL* homologue was found for *PHY122023* placed 0.2 cM downstream from a LTR. In maize chromosomes two genes were placed closely of LTR regions. In chromosome 6 and 10, the LTR was less than 5 cM from the *YSL* copies.

Discussion

EST-SSR transferability

Comparative genetics revealed the conservation of gene and marker orders within several plants families [26, 27, 28,]. Comparative genomics facilitates the identification of putative orthologous loci, controlling agronomic traits within crop species [29, 30]. It also assists extending genetic information from model species to more complex species [31]. For example, the grass genomes are highly conserved, molecular markers developped in rice can be used for genetic analysis of other grasses [26, 30, 32]. Transferability of EST-SSRs in cereals was assessed in terms of successful amplification with at least 57% of them producing distintic amplicons (33). This result supports the findings in cereal crops [3, 34, 35, 36]. This is common with EST-SSR amplification and is probably due to the high rate of conservation of EST sequences, suggesting an amplification of either the orthologous or paralogous copies [37].

A high degree of polymorphism was detected between mapping parents in several crop species [3]. For mosses*, Physcomitrella patens* SSR markers were analyzed in two further species of the Funariaceae: *Physcomitrium sphaericum* (C.F. Ludw.) Fürnr. and *Funaria hygrometrica* Hedw.*,* whereas 79.7% of SSR PCRs performed well in the more closely related *P. sphaericum* and only 34 % in the more distantly related *F. hygrometrica* [38]. This results are in accordance with prior reports about interspecies transferability of EST-derived SSRs for numerous seed plants including *Triticum aestivum* L, *Hordeum vulgare* L., *Festuca arundinacea* Schreb., *Oryza sativa* L., *Medicago truncatula* Gaertn. and *Pinus taeda* L., where the transferability decreased with increasing phylogenetic distance and transfer success

rates differed from 96 % to 40 % [3, 39, 34, 37]. In the present work, the EST-SSR transfer from *P. patens* to *P. juniperinum* reach similar values as those found for less closely taxa.

IRAP-REMAP transferability

Genomic retrotransposons are the most abundant class of transposable elements and they outnumber the genes in the eukaryotic genomes [40, 41, 42]. Their copy number and genomic locations are plastic. Plant genomes contain hundreds of thousands of these elements, together forming the vast majority of the total DNA [43]. Because of their copy-and-paste mode of transposition, these elements tend to increase their copy number while they are active [44]. The differences in genome size observed in the plant kingdom are accompanied by variations in retrotransposon content, characterized by polymorphic insertion patterns within pools of many species, suggesting that retrotransposons might be important players in the evolution of genome size [45].

The use of retrotransposon-based markers can be a valuable tool for plant breeders [46]. IRAP and REMAP techniques can be used separately or combined for a more complete genome survey. The ubiquitous presence of LTR retrotransposons in plant genomes suggests that the use of these techniques would allow breeders to obtain markers close to virtually any important agronomical trait. Also, the hypervariable nature of these repeat elements should make them excellent sources of polymorphic markers [3, 36, 37, 47]. When comparing the results obtained with IRAP/REMAP, they proved to be as reliable molecular markers as AFLPs, but they also bring additional information, showing a great potential use in genome assessments for fingerprinting, mapping and diversity studies [47]. In this study, a phylogeography approach using molecular markers was applied to a non-model moss species by the transferability of genomic information from an allied species. *Polytrichum* species have been studied with molecular approaches, from isoenzymes to SSRs, but the number of markers used always were in a small number [48, 49]. The existence of such a strong phylogeographic signal suggests that long-distance dispersal is less important than mutation for generating patterns of variation on a global scale [50]. This pattern contrasts with the mainly founds for Polytrichaceous species [49], which commonly produce numerous, small spores where little or no intercontinental differentiation was detected, supporting the view that barriers to gene flow exist at the intercontinental scale for moss phylogeographies [51, 52, 53, 54]. Our results contribute to these patterns, since the antarctic *P. juniperinum* genotype showed an expected divergence from the other genotypes but in a lower degree suggesting a

sympatric diversification such as the one found in other moss phylogeographic studies [50]. The use of more informative markers, associated to several traits, brings the possibility of saturing the non-model species with molecular markers and with low costs and reproducible techniques for the most diverse molecular diagnoses and studies.

Retrotransposable element association with YSL genes in plants

The retrotransposon marker method also provides an efficient technique for evaluating retrotransposition history and behavior in natural and domesticated plant populations. It has been used to study the insertional polymorphism of the *Ty1-copia* group in plants [55, 56, 57, 58, 59], showing a highly polymorphism in barley, wheat, rye and oat, suggesting that the transpositional activity of this retroelement has persisted for millions of years in several cereal species [59]. A similar study has shown that a few retrotransposons have been transpositionally active in the recent past in several *Pisum* species [59, 60, 61] and in diploid *Avena* species [62]. For these reason the IRAP/REMAP should also be useful to evaluate the transposition history of retrotransposable elements in plants.

Transposon-tagging from a maize mutant defective in iron uptake allowed the cloning of *Yellow Stripe*1 (*YS1*), the only molecular component related to the transport of the Fe^{3+}-phytosiderophore complex characterized to date [15]. The LTR *copia-like* position in the moss scaffold associated to a *YSL* homolog was similar to the one found in the maize *Zmji/ZmYSL1* association, suggesting a possible divergency of this multi-copy gene family in plants was induced by a transposable element, since these elements are found in both early and recent plant groups.

Conclusion

These results make it possible to apply molecular markers transfer strategies based on microsatellites and transposable elements from model to orphan species.

The EST-SSR tranfer percentage found is similar to those obtained in most transfer studies, decreasing when one increases the distance between the two taxa sampled. Using the IRAP/REMAP markers it is possible to survey an unmapped or unknown gene target in orphan species.

Orthologues from an *YSL* multi copy gene family probably diverged in ancient species because of the nature of the *copia-like* retrotransposon association observed in moss and maize homologue regions.

Methods

Sampling the plant material

Full developed gametophytes of *Polytrichum juniperinum* were collected in six different regions, four in southern Brazil [Gramado (29° 23' 19" S; 50° 52' 23" W), Canela (29° 21' 57" S; 50° 50' 22" W), Morro Redondo (31° 54'22" S; 52° 56'70" W) and Pelotas (31° 58' 77" S; 52° 63' 27" W)], one in southern Chile [Punta Arenas(53° 09'16"S; 70° 59'59"W)] and one in the Maritime Antarctic [Admiral Bay, King George Island(62° 07'40" S; 58° 23'15" W)]. The plants were sampled with help of an sterile blade and a zip plastic bag with silica gel inside. The collection were carried to the laboratory for DNA extraction and PCR analyses. Samples of *Physcomitrella patens* K3 ecotype was used to validate the markers tranferability, follow the rapid DNA extraction protocol [63] for all moss genotypes sampled.

Molecular protocols

For the rapid extraction of DNA, small pieces (>10 mg) of shoot and leaf apices were placed into a 1.5 ml Eppendorf tube. Liquid nitrogen was added and immediately after the nitrogen had sublimed, the sample was ground with a micropestle, diluted in 50µl 100 mM TE-buffer, and incubated at 60°C for 15 min. The suspension was cleared by centrifugation at 13,200 rpm for 10 min, the supernatant transferred to a clean 1.5 ml Eppendorf tube, and 2 µl was immediately used for PCR.

The *EST-SSR* primers used following the *in silico* transferability results reported for moss genomes [24] using the *SSRLocator* software [25].

The IRAP/REMAP amplification reactions were performed according to the protocol described [46]. Two *LTR* primers from a retrotransposon element associated with *ZmYsl1* were obtained according to their described sequences [11]. Other six primers were obtained from *Physcomitrella patens* copia-like retrotranposon [1]. The combination of 10 oligos, referred to as Zmji-1F, Zmji-1R, Zmji-2F, Zmji-2R, LTR-1F, LTR-1R, LTR-2F, LTR2R, LTR-3F and LTR-3R, enabled the use associated to a *YSL* homologue in *Physcomitrela patens* resulted from a BLAST search using *ZmYSL1* as a query. The same *LTR* primer was used to combine with the EST-SSR primer transferred to *Polytrichum juniperinum* for the *REMAP* reaction.

The amplification program consisted of a Touchdown PCR with an initial denaturation at 94°C for 5 min, followed by 10 cycles composed of 94°C for 30 s, 62- 52°C for 60 s (decreasing 1°C per cycle) and 72°C for 30 s for denaturation, annealing, and extension,

respectively. After were nested 35 cycles in 52°C annealing time, maintaining the same denaturation and extension temperatures above. After amplification, a final extension step was performed at 72°C for 10 min. The amplification product was separated in agarose gel stained with GelRed^{TM} (Uniscience).

Amplica analysis

Amplification products were scored independently as 1 and 0 for presence and absence of bands, respectively, and the obtained binary data were used for the analyses. The genetic similarity between individual pairs of genotypes was analyzed by using the *NTSYS pc 2.1* software [25]. The average similarity for all genotype pairs was used as a cutoff value for defining the clusters. For the estimation of genetic similarity, the Dice coefficient was used (Dice 1945) and basing on the generated similarity matrices (*MSIRAP*, *MSREMAP* and *MSIRAP+MSREMAP*), 3 dendrograms were obtained through clustering analysis by the *Neighbor-joining* algorithm [64]. To verify the fitting between similarity matrices and the respective dendrogram-derived matrices, the cophenetic correlation coefficient (r) was used [65]. The statistical stability of the clusters and the final tree was estimated by a bootstrap analysis with 1000 replications, with the *DARwin* 5 software [66].

The resulting amplicon sizes were compared with the *in silico* PCR products, to verify the possible unespecific products. The *Zmji* and moss LTR region, and the maize and *Physcomitrella patens YSL* homologues were ploted in that respective plants genome using the ACT software [67] and the distances obtained from each marker to each YSL gene was mapped for each species in their respective chromosome (for *Zea mays* case) or scaffold (for moss case).

Authors' contributions

FCV carried out all *in silico* and molecular studies, including the mapping regions and sequences *Blast* surveys, the DNA extration and PCR protocols and the EST-SSR transferability analysis and drafted the manuscript. NM contributes in the molecular protocols, as DNA extration PCR techniques. LCM participate in the design of the study. ROS participate in the design of the study. ACO conceived of the study, and participated in its design and coordination. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank the Developmental Center of Technology (CDTec/UFPEL) for the support to the first author. This work was supported by the National Council for Scientific and Technological Development CNPq (process # 480938/2009-1 and 475122/2007-0).

References

1. Rensing SA, Lang D, Zimmer AD. *et al.*: **The *Physcomitrella* genome reveals insights into the conquest of land by plants**. *Science* 2008. **319**: 64–69.

2. Wang ML, Barkley NA, Yu JK, Dean RE, Newman ML, Sorrells ME, Pederson G: **Transfer of simple sequence repeat (SSR) markers from major cereal crops to minor grass species for germplasm characterization and evaluation.** *Plant Genetic Resources: characterization and utilization* 2005. **3**: 45-57.

3. Varshney RK, Graner A, Sorrells ME: **Genic microsatellite markers in plants: features and applications.** *Trends in Biotechnology* 2005. **23**: 48-55.

4. Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS: **Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat.** *Molecular Genetics and Genomics* 2003. **270**: 315-323.

5. Decroocq V, Favé MG, Hagen L, Bordenave L, Decroocq S: **Development and transferability of apricot and grape EST microsatellite markers across taxa.** *Theoretical and Applied Genetics* 2003. **106**: 912-922.

6. Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ: **Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum.** *Plant Science* 2001. **160**: 1115-1123.

7. Gao L, Tang J, Li H, Jia J: **Analysis of microsatellites in major crops assessed by computational and experimental approaches.** *Molecular Breeding* 2003. **12**: 245-261.

8.Yu JK, La Rota M, Kantety R, Sorrells M: **EST derived SSR markers for comparative mapping in wheat and rice.** *Molecular Genetics and Genomics* 2004. **271**: 742-751.

9.Zhang LY, Ravel C, Bernard M, Balfourier F, Leroy P, Feuillet C, Sourdille P: **Transferable bread wheat EST-SSRs can be useful for phylogenetic studies**

among the Triticeae species. *Theoretical and Applied Genetics* 2006. **113**: 407-418.

10. Curie C, Briat JF: **Iron transport and signaling in plants**. *Annual Review of Plant Biology* 2003. **54**: 183–206.

11 Higuchi K, Suzuki K, Nakanishi H, Yamaguchi H, Nishizawa NK, Mori S: **Cloning of nicotianamine synthase genes, novel genes involved in the biosynthesis of phytosiderophores.** *Plant Physiology* 1999. **119:** 471-479.

12 Higuchi K, Watanabe S, Takahashi M, Kawasaki S, Nakanishi H, Nishizawa NK, Mori S: **Nicotianamine synthase gene expression differs in barley and rice under Fe-deficient conditions.** *Plant Journal* 2001. **25:** 159-167.

13 Kobayashi T, Ogo Y, Itai RN, Nakanishi H, Takahashi M, Mori S, Nishizawa NK: **The transcription factor IDEF1 regulates the response to and tolerance of iron deficiency in plants**. *Proceeding of Nature Academy of Science USA* 2007. **104:** 19150-19155.

14 Inoue H, Higuchi K, Takahashi M, Nakanishi H, Mori S, Nishizawa NK: **Three rice nicotianamine synthase genes, *OsNAS1*, *OsNAS2*, and *OsNAS3* are expressed in cells involved in long-distance transport of iron and differentially regulated by iron**. *Plant Journal* 2003. **36**: 366-381.

15 Curie C, Panaviene Z, Loulergue C, Dellaporta SL, Briat JF, Walker EL: **Maize yellow stripe1 encodes a membrane protein directly involved in Fe(III) uptake.** *Nature* 2001**. 409**: 346-349.

16 Hell R, Stephan UW: **Iron uptake, trafficking and homeostasis in plants**. *Planta* 2003. **216**: 541-551.

17 Robinson NJ, Proctor CM, Connolly EL, Guerinot ML: **A ferric chelate reductase for iron uptake from soils**. *Nature* 1999. **397**: 694-697.

18 Waters BM, Blevins DG, Eide DJ: **Characterization of FRO1, a Pea Ferric-**

Chelate Reductase involved in root iron acquisition. *Plant Physiology* 2002**. 129**: 85-94.

19 Eide D, Broderius M, Fett JP, Guerinot ML: **A novel iron-regulated metal transporter from plants identified by functional expression in yeast**. *Proceeding of Nature Academy of Science USA* 1996. **93**: 5624-5628.

20 Eckhardt U, Mas Marques A, Buckhout TJ: **Two-iron regulated cation transporters from tomato complement metal uptake-deficient yeast mutants**. *Plant Molecular Biology* 2001**. 45**: 437–448.

21 Vert G, Briat JF, Curie C: **Arabidopsis IRT2 gene encodes a root-periphery iron transporter**. *Plant Journal* 2001. **26**: 81–189.

22 Bughio N, Yamaguchi H, Nishizawa NK, Nakanishi H, Mori S: **Cloning an iron regulated metal transporter from rice**. *Journal of Experimental Botany* 2002**. 53**: 1677-1682.

23 Brunner S, Pea G, Rafalski A: **Origins, genetic organization and transcription of a family of non-autonomous helitron elements in maize.** *The Plant Journa*l 2005. **43**: 799–810

24 Victoria FC, Maia LC, Oliveira AC: **In silico comparative analysis of SSR markers in plants.** *BMC Plant Biology* 2011. 11:15.

25 Rohlf FJ: ***NTSYS-pc: numerical taxonomy and multivariate analysis system, version 2.1.*** Exeter Software, New York, 2000. 98p.

26 vanDeynze AE, Nelson JC, O'Donoughue JS, Ahn SN, Siripoonwiwat W, Harrington SE, Ylesias ES, Braga DP, McCouch SR, Sorrells ME: **Comparative mapping in grasses. Oat relationships**. *Molecular Genomics and Genetics* 1995**. 249**: 349–356.

27 vanDeynze AE, Nelson JC, Yglesias ES, Harrington SE, Braga DP, McCouch SR, Sorrells

ME: **Comparative mapping in grasses. Wheat relationships**. *Molecular Genomics and Genetics* 1995**. 248**: 744–754.

28 Kuleung C, Baenziger PS, Dweikat I: **Transferability of SSR markers among wheat, rye, and triticale**. *Theoretical and Applied Genetics* 2004. **108**: 1147-1150.

29 Paterson AH: **Polyploidy, evolutionary opportunity and crop adaptation.** *Genetica* 2005. 123: 191–196

30 Devos KM, Gale MD: **Genome relationships: the grass model in current research.** *Plant Cell* 2000. **12**(5): 637–646.

31 Gale MD, Devos KM: **Comparative genetics in the grasses.** *Proceedings of Natural Academy of Science USA* 1998. **95**: 1971–1974.

32 Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS: **Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat.** *Molecular Genetics and Genomics* 2003. 270: 315–323

33 Sim S-C, Yu J-K, Jo Y, Sorrells ME, Jung G: **Transferability of cereal EST-SSR markers to ryegrass.** *Genome* 2009. **52**: 431–437 (2009)

34 Saha, MC, Mian MAR, Eujayl I, Zwonitzer JC, Wang L, May G: **Tall fescue EST-SSR markers with transferability across several grass species.** *Theoreticaland Applied Genetics* 2004. **109**: 783–791.

35 Yu JK, Dake TM, Singh S, Benscher D, Li WL, Gill B, Sorrells ME: **Development and mapping of EST-derived simple sequence repeat markers for hexaploid wheat.** *Genome* 2004. **47***:* 805–818.

36 Zhang LY, Bernard M, Leroy P, Feuillet C, Sourdille P: **High transferability of bread wheat EST-derived SSRs to other cereals.** *Theoreticaland Applied Genetics* 2005. **111**: 677–687.

37 Varshney RK, Sigmund R, Borner A, Korzun V, Stein N, Sorrells ME et al.: **Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice.** *Plant Science* 2004. **168**: 195–202.

38 von Stackelberg MV, Rensing SA, Reski R: **Identification of genic moss SSR markers and a comparative analysis of twnty-four algal and plant gene indices reveal species-specific rather than group-specific characteristics of microsatellites.** BMC Plant Biology 2006. **6**:9.

39 Thiel T, Michalek W, Varshney RK, Graner A: **Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.).**
Theoreticaland Applied Genetics 2003. **106**(3): 411-422.

40 Sabot,F, Schulman AH: **Parasitism and the retrotransposon life cycle in plants: a hitch-**
hiker's guide to the genome. *Heredity* 2006. **97**: 81–_88

41 Feschotte C, Jiang N, Wessler SR: **Plant transposable elements: where genetics meets genomics.** *Nature Reviews Genetics* 2002. **3**: 329-341.

42 IHGSC (International Human Genome Sequencing Consortium). **Initial sequencing and analysis of the human genome.** Nature. 2001. **409**: 860–921.

43 Schulman AH, Kalendar R: **A movable feast: diverse retrotransposons and their contribution to barley genome dynamics**, *Cytogenetic and Genome Research* 2005. **110**(1-4): 598-605.

44 **Mansour A: Utilization of genomic retrotransposons as cladistic markers.** *Journal of Cell and Molecular Biology* 2008. **7**(1**)**: 17-28.

45 Vitte C, Panaud O: **LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model.** *Cytogenet Genome Research* 2005.

110: 91-107.

46 Kalendar R, Grob T, Regina M, Suoniemi A, Schulman I: **IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques.** *Theoretical and Applied Genetics* 1999. **98**: 704–711.

47 BRANCO JSC, VIEIRA EA, MALONE G, KOPP MM, MALONE E, BERNARDES A, MISTURA CC, CARVALHO FIF, OLIVEIRA, AC: **IRAP and REMAP assessments of genetic similarity in rice**. *Journal of Applied Genetics* 2007. **48**: 107-113.

48 Hyvönen J, Hedderson TA, Smith Merill GL, Gibbings JG, Koskinen S: **On phylogeny of the Polytrichales.** *Bryologist* 1999. **101**: 489-504.

49 Van der Velde M, Bijlsma R: **Phylogeography of five *Polytrichum* species within Europe.** *Biological Journal of the Linnean Society* 2003. **78**: 203-213.

50 Huttunen S, Hedenäs L, Ignatov MS, Devos N, Vanderpoorten A: **Origin and evolution of the northern hemisphere disjunction in the moss genus *Homalothecium* (Brachytheciaceae).** *American Journal of Botany* 2008. **95**(6): 720–730.

51 Shaw AJ: **Molecular phylogeography and cryptic speciation in the mosses *Mielichhoferia elongata* and *M. mielichhoferiana* (Bryaceae).** *Molecular Ecology* 2000. **9**: 595–608

52 McDaniel SF, SHAW AJ: **Selective sweeps and intercontinental migration in the cosmopolitan moss *Ceratodon purpureus* (Hedw.) Brid.** *Molecular Ecology* 2005. **14**: 1121–1132.

53 Shaw AJ, Werner O, Ros RM: **Intercontinental Mediterranean disjunct mosses: Morphological and molecular patterns**. *American Journal of Botany* 2003. **90**: 540–550

54 Vanderpoorten A, Devos N, Hardy OJ, Goffinet B, Shaw AJ: **The barriers to oceanic island radiation in bryophytes: insights from the phylogeography of the**

moss *Grimmia montana*. *Journal of Biogeography* 2008. **35**(4): 654-663.

55 Kumar A, Pearce SR, McLean K, Harrison G, Heslop-Harrison JS, Waugh R, Flavell AJ: **The Ty1-*copia*group of retrotransposons in plants: genomic organization, evolution and use as molecular markers.** *Genetica* 1997. **100**: 205–217.

56 Waugh R, McLean K, Flavell AJ, Pearce SR, Kumar A, Thomas BB, Powell W: **Genetic distribution of *BARE-1*-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms** (S-SAP). *Molecular Genomics and Genetics* 1997. **253**: 687–694.

57 Purugganan M, Wessler SR: **Transposon signatures – species-specific molecular markers that utilise a class of multiple-copy nuclear-DNA.** *Molecular Ecology* 1995. **4**: 265–269

58 Gribbon BM, Pearce SR, Kalendar R, Schulman AH, Paulin L, Jack P, Kumar A, Flavell AJ: **Phylogeny and transpositional activity of Ty1-*copia*group retrotransposons in cereal genomes.** *Molecular Genomics and Genetics* 1999. **261**: 883–891.

59 Ellis THN, Poyser SJ, Know MR, Vershinin AV, Ambrose MJ: **Polymorphism of insertion sites of Ty1-*copia*class retrotransposon and its use for linkage and diversity analysis in pea**. *Molecular Genomics and Genetics* 1998. **260**: 9–19.

60 Pearce SR, Stuart-Rogers C, Knox MR, Kumar A, Ellis THN, Flavell AJ: **Rapid isolation of plant Ty1-*copia*group retrotransposon LTR sequences for molecular marker studies.** *Plant Journal* 1999. **19**: 711–717.

61 Pearce SR, Knox M, Ellis THN, Flavell AJ, Kumar A: **Pea Ty1-*copia*group retrotransposons: transpositional activity and use as markers to study genetic diversity in *Pisum*.** *Molecular Genomics and Genetics* 2000. **263**: 898–907.

62 Yu G-X, Wise RP: **An anchored AFLP-and retrotransposons-based map of diploid *Avena***. *Genome* 2000. **43**: 736–749.

63 Pedersen N, Russel SJ, Newton AE, Ansell SW**: A novel molecular protocol for the rapid extraction of DNA from bryophytes and the utility of direct amplification of DNA from a single dwarf male.** *Bryologist* 2006. **109**(2): 257–264.

64. Saitou N, Nei M: **The neighbor-joining method: A new method for reconstructing phylogenetic trees.** *Molecular Biology and Evolution* 1987. **4**: 406-425.

65 Sokal RR, Rohlf FJ: **The comparison of dendrograms by objective methods**. *Taxon* 1962. **11**: 30–40.

66. Perrier X, Jacquemoud-Collet JP. **DARwin software**. 2006. http://darwin.cirad.fr/darwin

67 Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG and Parkhill J: **ACT: the Artemis Comparison Tool.** *Bioinformatics* 2005. **21**(16): 3422-3423.

Figures

Figure 1- PCR amplification patterns of EST-derived microsatellites of *Physcomitrella patens* Brid. in three *Polytrichum juniperinum* Hedw. genotypes. Exemplary SSR marker A – E were amplified in the *P. patens* standard lab strain 'Villersexel-K3' (1) originating from France, and in Canela (2), Punta Arenas (3) and King George Island (4) accesses, respectively. The SSR marker A and D shows a double banding. The SSRs A reveal no size polymorphism between the Canela and Punta Arenas accessions, while the SSR D reveal the same condition for Punta Arenas and Antarctica accessions. The SSR B reveal distinct visible single size polymorphisms between the *Polytrichum* accessions. The SSR marker C shows probably an null allele amplified only in the antarctic genotype. DNA size markers (M) is a 50 bp ladder (SIGMA-Aldrich).

Figure 2 - IRAP markers transferability from *Physcomitrella patens* in four *Polytrichum juniperinum* genotypes. The genotypes Gramado (1), Pelotas (2), Punta Arenas (3) and King George Island (4) were tested with moss *LTR* and *Zmji* primers. IRAP primers LTR-1F/LTR-1R (A), LTR-2F/LTR-2R (B), LTR-3F/LTR-3R (C), LTR-4F/LTR-4R (D) were amplified in one or more tested genotypes. The combined primer LTR-2F/LTR-1R (E) was amplified only in one accession. LTR-2F/LTR-3F (F) combination results in the highest band obtained in the present study (5 bands). Zmji-1F/Zmji-1R (G) was found only in the southern genotypes sampled, the Pelotas fragment had a lower size compared with others. The primer Zmji-2F/Zmji-2R (H) resuts in a polymorphic band found only in Gramado and Punta Arenas accessions. DNA size markers (M) is a 100 bp ladder (SIGMA-Aldrich).

Figure 3 - Dendrogram of 6 *Polytrichum juniperinum* genotypes obtained from the combined marker analysis IRAP and REMAP, using the Dice similarity matrix and UPGMA clustering method. Cophenetic correlation coefficient for the matrix is 0.71 (r). *PjPA*=Punta Arenas (Chile); *PjKGI*= King George Island (Antarctic); *PjGRA*=Gramado (Brazil); *PjCAN*=Canela (Brazil); *PjPEL*=Pelotas (Brazil) and *PjMR*=Morro Redondo (Brazil) *P. juniperinum* genotypes acessed.

Figure 4 - Dendrogram of 6 *Polytrichum juniperinum* genotypes obtained from the combined marker analysis IRAP and REMAP, using *Physcomitrella patens* as outgroup and the Neighbor-joining clustering method. *PjPA*=Punta Arenas (Chile)*; PjKGI*= King George Island (Antarctic); *PjGRA*=Gramado (Brazil); *PjCAN*=Canela (Brazil); *PjPEL*=Pelotas (Brazil) and *PjMR*=Morro Redondo (Brazil) *P. juniperinum* genotypes acessed and *PpK3*=Villersexel K3 (France) *P. patens* ecotype acessed.

Figure 5 – LTR and *YSL* homologues positions in the *Zea mays* L. chromosomes and *Physcomitrella patens* scaffolds.

Figure 1.

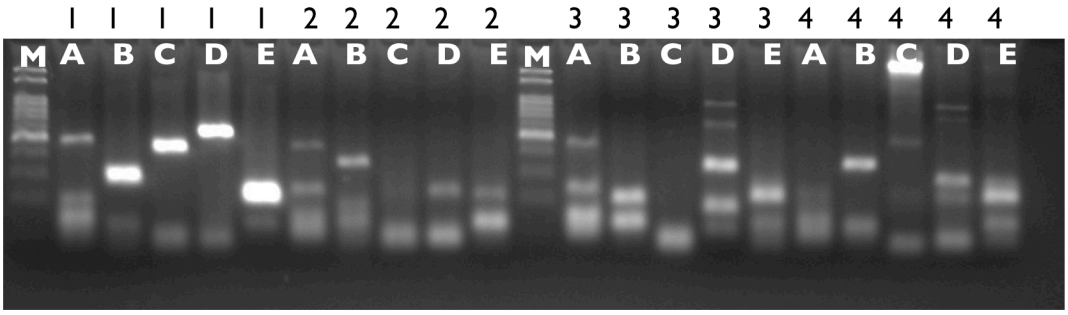


Figure 2.

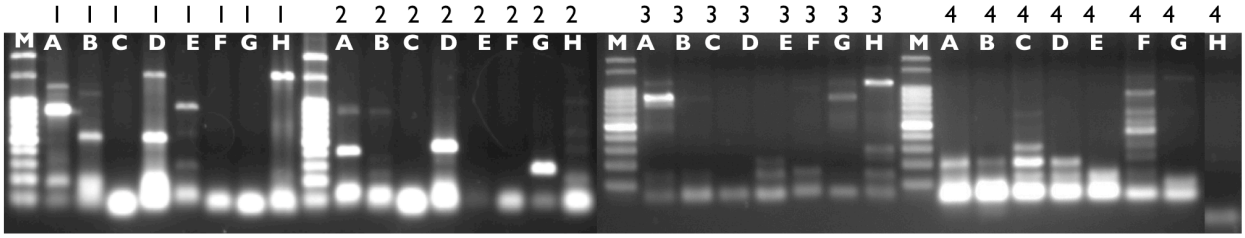


Figure 3

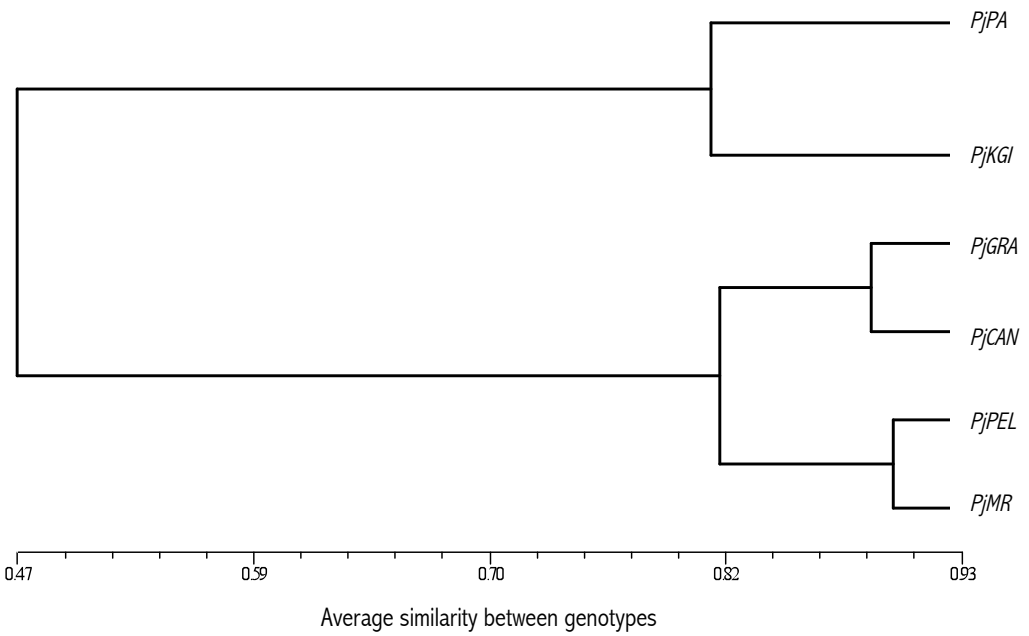


Figure 4

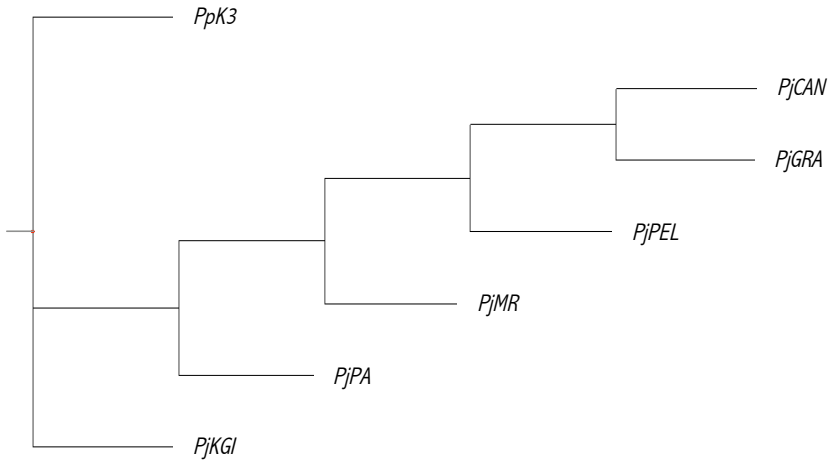
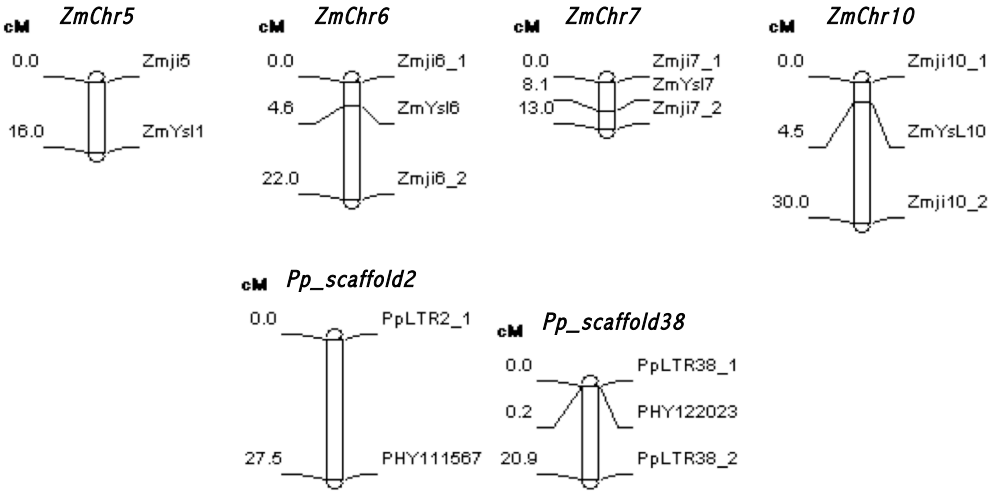


Figure 5



8. CONCLUSÕES GERAIS

Fazendo uso de briófitas como uma ferramenta experimental para a elucidação de processos biológicos complexos, no presente trabalho foi possível identificar que os marcadores microssatélites dinucletídeos (NN) foram mais frequentes em espécies basais e os motivos trinucleotídeos (NNN) foram mais frequentes em espécies derivadas na exploração *in silico* utilizando o *software SSRLocator*, corroborando resultados obtidos em estudos anteriores sobre frequencia e distribuição de elementos repetitivos em plantas. Cerca de 30% dos EST-SSR sintetizados a partir de sequencias expressas de *Physcomitrella patens* Brid. foram tranferidos com sucesso utilizando o DNA genômico de *P. juniperinum*, validando *SSRLocator* como ferramenta na exploração de marcadores microsatélites. A distribuição dos EST-SSR foi influenciada pelo uso preferencial de códon em cada grupo taxonômico, possibilitando assim estabelecer estratégias de transferência de marcadores entre espécies modelo e espécies órfas correlatas com maior precisão e reprodutibilidade

Foi observada uma natureza multi-cópia das famílias gênicas relacionadas com a homeostase de ferro nas plantas terrestes, sendo observada nas análises filogenéticas agrupamentos entre cópias de diferentes espécies. Não foram encontradas cópias parálogas dos genes *NRAMP* em gramíneas, sugerindo a diversificação no ancestral neste grupo, assim como a pressão positiva na potencial manutenção das cópias destes genes na homoestase do ferro ou funções correlatas, diminuindo a possibilidade de ocorrência de pseudogenes ou cópias não funcionais, corroborado pela ocorrência de elementos reguladores similares em regiões promotoras de cópias de arroz e sorgo

Retroelementos foram encontrados flanqueando regiões de ocorrência de homológos de genes *YSL* em briófitas assim como é reportado para *ZmYSL1* em milho, sugerindo estes elementos como responsáveis pela divergência ortóloga destes genes desde as linhagens basais de plantas terrestres.

Também foi possível estabelecer culturas *in vitro* de *Polytrichum juniperinum* Hedw. O desenvolvimento de protonema e a regeneração de gametófitos foram obtidos com sucesso em meios de cultura livres de reguladores de crescimento ou contendo somente auxina, podendo optar pelo meio de menor complexidade

reduzindo custos no cultivo *in vitro* para esta espécie de musgo. Somando estes resultados aos obtidos experimentos *in silico* e de laboratório para transferência de marcadores moleculares entre *P. patens* e *P. juniperinum*, há um forte indicativo do potencial uso desta última espécie como um modelo biológico em plantas.

9. REFERÊNCIAS

BECKER, M.; ASCH, F. Iron toxicity- Conditions and management concepts. **Journal of Plant Nutrition and Soil Science**, v. 168, p. 558-573. 2005.

BOWER, F.O. **Primitive Land Plants**. London. 1935. 658p.

CAMPBEL, E.O. Problems in the orogin and classification of bryophytes with particular reference to liverworts. **New Zealand Journal of Botany,** v. 9, p. 678-688. 1971.

CAMPBELL, D.H. **The Structure and Development of Mosses and Ferns.** London. 1905.

CELLIER, M.F.M; BERGEVIN, I.; BOYER, E.; RICHER. E. Polyphyletic origins of bacterial Nramp transporters. **TRENDS in Genetics**, v. 17, n. 7, p. 365-370, 2001.

CHANEY, R.L.; BROWN, J.C; TIFFIN, L.O. Obligatory reduction of ferric chelates in iron uptake by soybeans. **Plant Physiology***,* v. 50, p. 208-213, 1972.

CONNOLLY, E.L.; FETT J.P.; GUERINOT, M.L. Expression of the IRT1 metal transporter is controlled by metals at the levels of transcript and protein accumulation. **Plant Cell,** v. 14, p. 1347-1357, 2002.

COVE, D.J.; KNIGHT, C.D.; LAMPARTER, T. Mosses as model systems. **Trends in Plant Science**, v. 2, p. 99-105, 1997.

CURIE, C.; BRIAT, J.F. Iron transport and signaling in plants. **Annual Review of Plant Biology,** v.54, p. 183–206, 2003.

CURIE, C.; PANAVIENE, Z.; LOULERGUE, C.; DELLAPORTA, S.L.; BRIAT, J.F.; WALKER, E.L. Maize yellow stripe1 encodes a membrane protein directly involved in Fe(III) uptake. **Nature***,* v. 409, p. 346-349, 2001.

DHALIWHAL, G.S.; ARORA, R. Stresses in agroecosystems: concepts and approaches, in: G.S., DHALIWAL; R., ARORA, (Eds.), **Environmental Stress in crop plants**, Ajay Verma Commonwealth Publishers, 1999, pp. 1-18.

FAGERIA, N.K.; RABELO, N.A. Tolerance of rice cultivars to iron toxicity. **Journal of Plant Nutrition**, v. 10, p. 653-661. 1987.

HASKELL, G. Some evolutionary problems concerning the Bryophyta. **The Bryologist**, v. 52, p. 49-57, 1949.

HECKMAN, D.S.; GÊISER, D.M.; EIDELL, B.R.; STAUFFER, R.I.; KARDOS N.L.; HEDGES S.B. Molecular evidence for the early colonization of land by fungi and plants. **Science**, v. 293, p. 1129-1133, 2001.

KENRICK, P.; CRANE, P. The origin and early evolution of plants on land. **Nature**, v. 389, p. 33-39, 1997.

KHANNA, K. R. Differenctial evolutionary activity in bryophytes. **Evolution**, v. 18, p. 652-670, 1965.

MAGALLÓN, S.; HILU, K. W. Land plants (Embryophyta). *In:* **The Timetree of Life**, S. B. Hedges and S. Kumar, 2009. Eds. Oxford. University Press, pp. 133–137, 2009.

MARSCHNER H, RÖHMELD V. Strategies of plants for acquisition of iron. **Plant Soil,** v. 165, p. 375–88, 1994.

MILLER, H.A. Rhyniophyta, alternation of generations and the evolution of bryophytes. **Journal of the Hattory Botanical Laboratory***,* v. 38, p. 161-168, 1974.

MILLER, H.A. Bryophyte evolution and geography. **Biological Journal of the Linnean Society**, v. 18, p. 145-196, 1982.

NICKRENT, D.L.; PARKINSON, C.L.; PALMER, J.D.; DUFF, R.J. Multigene phylogeny of land plants with special references to bryophytes and the earliest land plants. **Molecular Biology and Evolution**, v. 17, p. 1885-1895, 2000.

OLIVEIRA, E. J.; PÁDUA, J.G.; ZUCCHI, M.I.; VENCOVSKY, R.; VIEIRA, M.L.C. Origin, evolution and genome distribution of microsatellites. **Genetics and Molecular Biology,** v. 29, n. 2, p. 294-307, 2006.

PONNAMPERUMA, F.N. The chemistry of submerged soils. **Advances in Agronomy**, v. 24, p. 29-96, 1972.

RENSING, S.A.; LANG, D.; ZIMMER, A.D.; *et al*. The *Physcomitrella* genome reveals insights into the conquest of land by plants. **Science***,* v. 319, p. 64-69, 2008.

RÖMHELD, V. Different strategies for iron acquistion in hgher plants. **Physiologia Plantarum,** v. 70, p. 321–234, 1987.

SAHRAWAT, K.L. Iron toxicity in wetland rice and the role of other nutrients. **Journal of Plant Nutrition**, v. 27, p. 1471-1504. 2004.

SAHRAWAT, K.L.; MULBAH, C.K., DIATTA, K., DELAUNE, R.D., PATRICK, W.H., SINGH, B.N., JONES M.P. The role of tolerant genotypes and plant nutrients in the management of iron toxicity in lowland rice. **Journal of Agricultural Science**, v. 126, p. 143-149. 1996.

SANTOS, L. S. ; COSTA DE OLIVEIRA, Antônio . Rice iron metabolism:from source to solution. **Journal of Crop Science and Biotechnology**, v. 10, p. 64-72, 2007.

SCHAEFER, D.G.; BISZTRAY, G.; ZRŸD, J.P. Genetic transformation of the moss *Physcomitrella patens*. *In*. BAJAJ, Y.P.S. **Biotecnology in Agriculture and Forestry**, v. 29, Plant Protroplasts and Genetic Engineering V, Berlin, Springer Verlag, Heidelberg, p. 349-364, 1991.

SCHAEFER, D.G.; ZRŸD, J.P. The moss *Physcomitrella patens*, now and then. **Plant Physiology**, v. 127, p. 1430-1438, 2001.

SCHOFIELD, W.B. **Introduction to Bryology.** New York. 1985. 481p.

SHAW, A. J.; GOFFINET, B. **Bryophyte Biology.** Cambridge University Press, England. 2000. 476p.

SHIMIZU, A.; GUERTA, C.Q.; GREGORIO, G.B.; KAWASAKI, S.; IKEHASHI, H. QTLs for nutritional contents of rice seedlings (*Oryza sativa* L.) in solution cultures and its implication to tolerance to iron toxicity. **Plant and Soil**, v. 275, p. 57-66. 2005.

VARSHNEY, R.K.; GRANER, A.; SORRELLS, M.E. Genic microsatellite markers in plants: features and applications. **TRENDS in Biotechnology**, v. 23, n. 1, p. 48-55, 2005.

VAROTTO, C.; MAIWALD, D.; PESARESI, P.; JAHNS, P.; SALAMINI, F.; LEISTER, D. The metal ion transporter IRT1 is necessary for iron homeostasis and efficient photosynthesis in Arabidopsis thaliana. **Plant Journal,** v. 31, p.589–599, 2002.

VERT, G.; GROTZ, N.; DÉDALDÉCHAMP, F. G.; GUERINOT, M. L.; BRIAT, J.F.; CURIE, C. IRT1, an *Arabidopsis* Transporter Essential for Iron Uptake from the Soil and for Plant Growth. **Plant Cell,** v. 14, n. 6, p. 1223–1233, 2002.

WAN, J.; ZHAI, H.; WAN, J.; IKEHASHI, H. Detection and analysis of QTLs for ferrous iron toxicity tolerance in rice, *Oryza sativa* L. **Euphytica**, v. 131, p. 201-206. 2003.

WINSLOW, M.D.; YAMAUCHI, M.; ALLURI, K.; MASAJO, T.M. Reducing iron toxicity in rice with resistant genotype and ridge planting. **Agronomy Journal**, v. 81, p. 458-460. 1989.

WU, P.; LUO, A.; ZHU, J.; YANG, J.; HUANG, N.; SENADHIRA, D. Molecular markers linked to genes underlying seedling tolerance for ferrous iron toxicity. **Plant and Soil**, v. 196, p. 317-320. 1997.

WU, P.; HU, B.; LIAO, C.Y.; ZHU, J.M.; WU, Y.R.; SENADHIRA, D.; PATERSON, A.H. Characterization of tissue tolerance to iron by molecular markers in different lines of rice. **Plant and Soil**, v. 203, p. 217-226. 1998.

YI, Y.; GUERINOT, M.L. Genetic evidence that induction of root Fe(III) chelate reductase activity is necessary for iron uptake under iron deficiency. **Plant Journal,** v. 10, p. 835–844, 1996.

10. ANEXOS

1. Arquivos suplementares do Artigo 2.

2. Arquivos suplementares do Artigo 3.

3. Arquivos suplementares do Artigo 4.

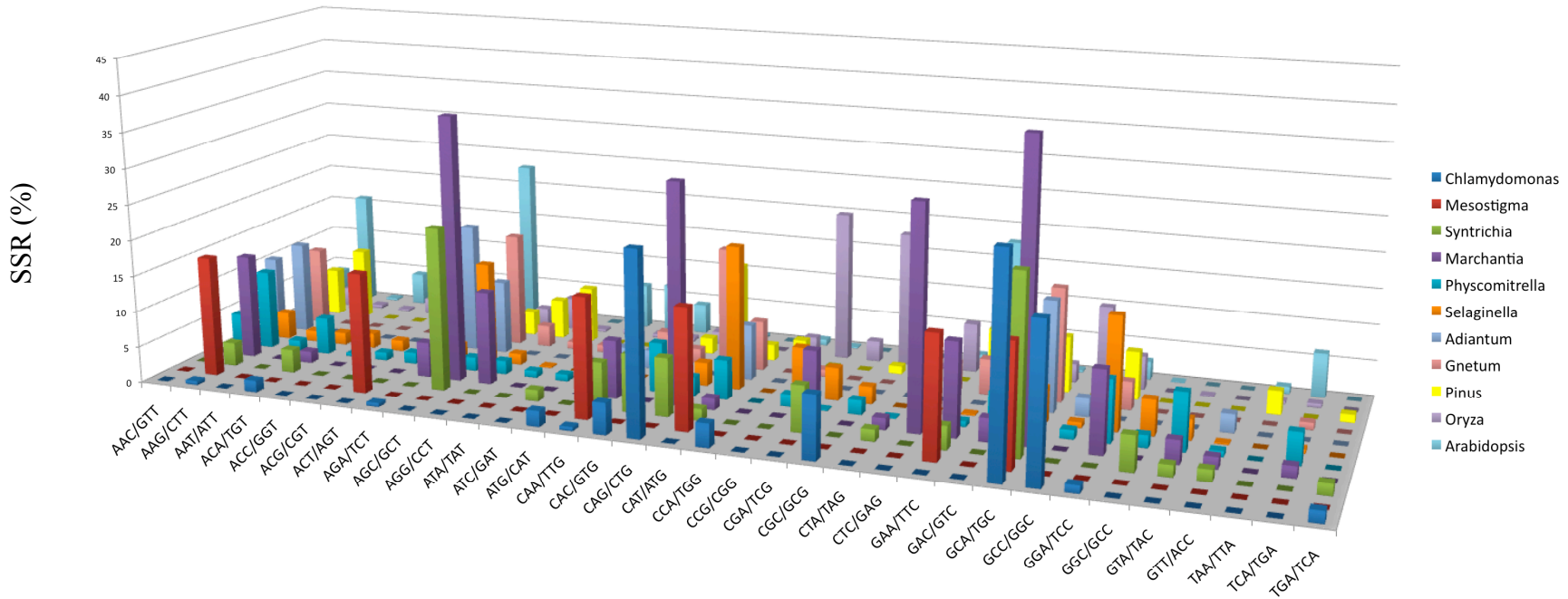
ANEXO 1

Additional file 1. Patterns of occurrence for dimer SSR motifs in percentage.

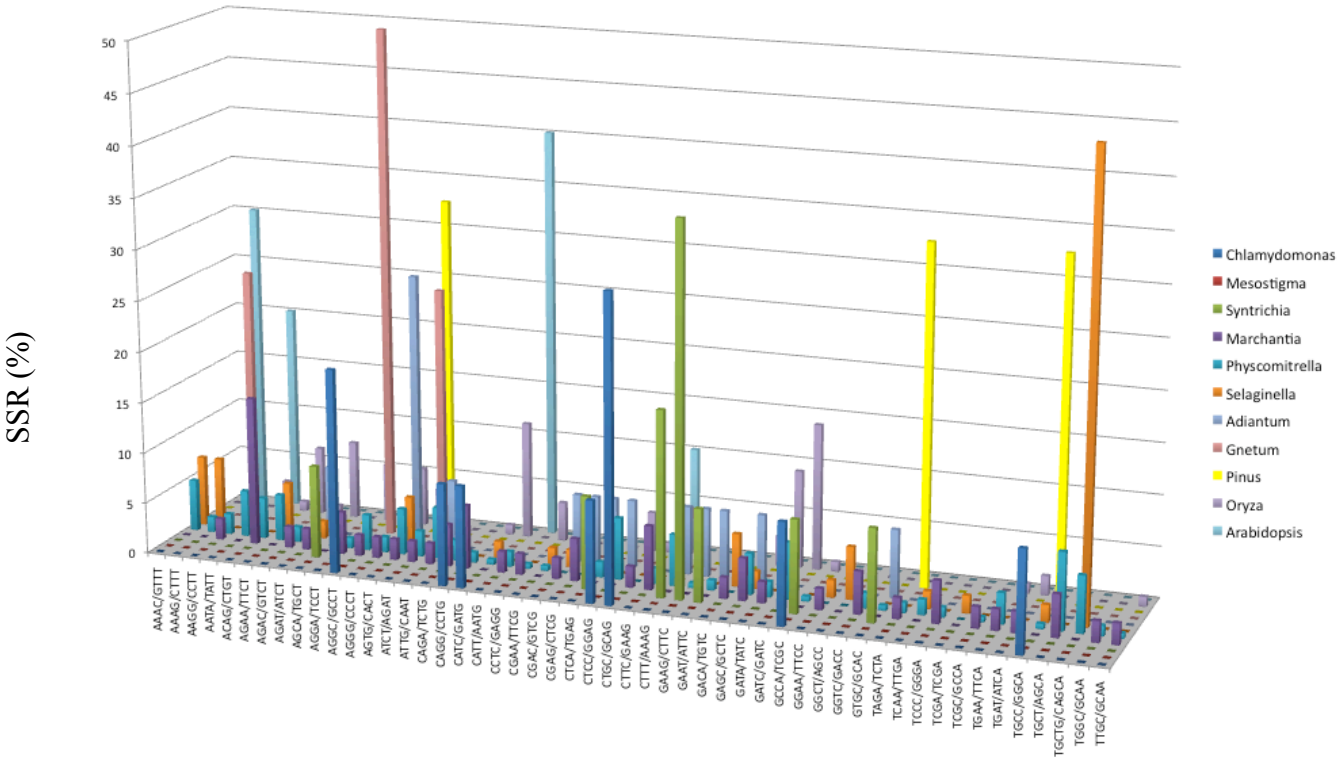
Repeat Motifs	*Algae*		*Bryophya l. s.*			*Vascular plants*					
						Ferns		*Gymnosperms*		*Flowering plants*	
Dimers	*C. reinhardtii*	*M. viride*	*M. polymorpha*	*S. ruralis*	*P. patens*	*Selaginella spp.*	*A. capillus-veneris*	*G. gnemon*	*P. taeda*	*O. sativa*	*A. thaliana*
AC/GT	36.2	-	24.0	-	3.8	8.1	9.0	-	-	4.0	2.4
AG/CT	4.4	29.3	36.0	30.4	22.8	40.3	39.5	21.4	13.4	42.6	38.3
AT/AT	2.2	-	-	-	21.5	2.7	0.1	14.3	41.0	9.2	8.9
CA/TG	42.4	-	12.0	-	5.4	6.0	15.0	-	0.9	3.8	1.9
GA/TC	8.5	67.2	28.0	69.6	32.4	41.6	36.1	35.7	11.5	33.1	38.5
TA/TA	5.5	-	-	-	14.0	1.3	0.3	28.6	32.7	7.3	10.0
GC/GC	0.5	-	-	-	-	-	-	-	-	-	-
CT/AG	0.4	3.4	-	-	-	-	-	-	-	-	-

Additional file 2. Patterns of occurrence for trimer SSR motifs in percentage.

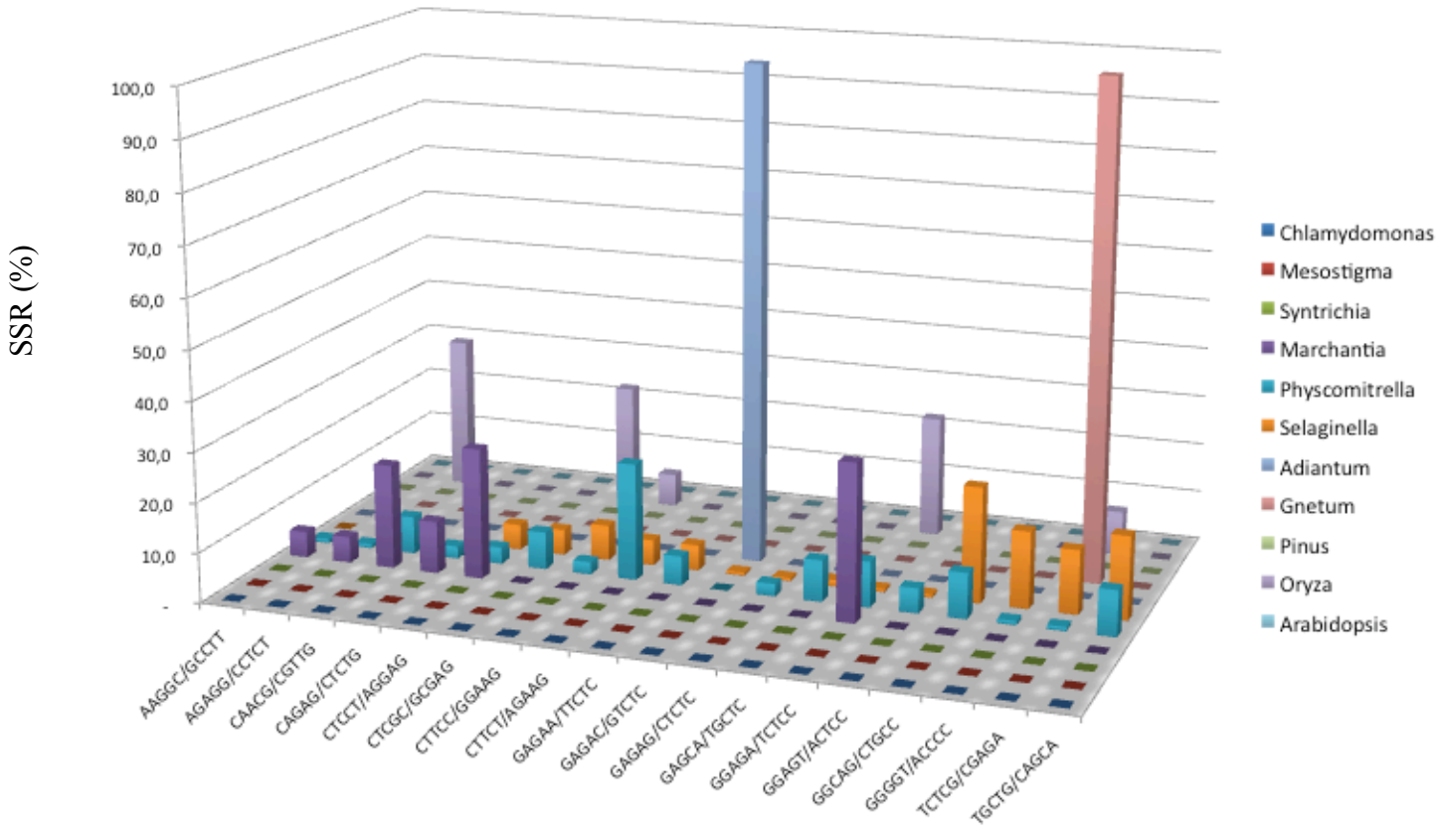
Repeat Motifs	*Algae*		*Bryophyta l. s.*			*Vascular plants*					
						Ferns		*Gymnosperms*		*Flowering plants*	
Dimers	*C. reinhardtii*	*M. viride*	*M. polymorpha*	*S. ruralis*	*P. patens*	*Selaginella spp.*	*A. capillus-veneris*	*G. gnemon*	*P. taeda*	*O. sativa*	*A. thalian*
AAC/GTT	-	-	-	-	4.4	0.8	10.3	-	3.3	0.3	3.9
AAG/CTT	0.5	16.7	3,2	14.5	11.1	3.8	12.8	10,9	6.7	2.6	15.9
AAT/ATT	-	-	-	-	1.4	1.5	-	-	10.0	0.5	0.3
ACA/TGT	1.6	-	3,2	1.6	5.2	1.7	-	-	-	0.2	4.5
ACC/GGT	-	-	-	-	0.5	2.1	-	-	-	2.3	1.3
ACG/CGT	-	-	-	-	1.1	1.5	-	-	-	1.4	-
ACT/AGT	-	16.7	-	-	1.6	0.4	-	-	1.1	0.2	1.3
AGA/TCT	0.5	-	-	4.8	8.3	1.9	17.9	7,9	5.6	2.6	22.9
AGC/GCT	-	-	22,6	37.1	1.9	14.1	10.3	15,8	3.3	2.6	0.5
AGG/CCT	-	-	-	12.9	1.9	1.5	-	3,0	5.6	4.6	1.1
ATA/TAT	-	-	-	-	0.9	0.2	-	1,0	7.8	0.3	-
ATC/GAT	-	-	1,6	-	0.9	-	-	1,0	-	0.6	6.3
ATG/CAT	2.2	-	-	-	1.1	-	2.6	-	-	0.5	6.7
CAA/TTG	0.5	16.7	6,5	8.1	1.1	-	2.6	4,0	3.3	0.5	4.2
CAC/GTG	4.3	-	8,1	-	6.9	0.8	2.6	2,0	2.2	1.8	0.6
CAG/CTG	25.0	-	8,1	30.6	2.7	3.2	-	16,8	13.3	3.3	0.4
CAT/ATG	-	16.7	1,6	1.6	5.5	20.0	7.7	6,9	2.2	-	-
CCA/TGG	3.3	-	-	-	-	-	-	-	3.3	2.5	0.8
CCG/CGG	-	-	-	-	1.7	6.8	-	1,0	-	20.9	0.2
CGA/TCG	-	-	6,5	9.7	0.2	4.6	-	-	-	2.9	-
CGC/GCG	8.7	-	-	-	2.0	2.5	-	-	1.1	18.7	-
CTA/TAG	-	-	1,6	1.6	0.2	-	-	-	-	0.2	-
CTC/GAG	-	-	-	30.6	-	-	-	4,0	-	6.8	1.0
GAA/TTC	-	16.7	3,2	12.9	0.8	0.2	-	5,0	7.8	2.6	17.8
GAC/GTC	-	-	-	3.2	5.3	2.7	12.8	-	4.4	1.9	-
GCA/TGC	29.3	16.7	24,2	40.3	8.9	4.9	15.4	15,8	7.8	2.6	0.3
GCC/GGC	21.2	-	-	-	1.3	0.8	2.6	-	-	11.1	-
GGA/TCC	1.1	-	-	11.3	8.6	15.6	-	4,0	6.7	4.7	2.6
GGC/GCC	-	-	4,8	-	1.7	4.9	-	-	-	-	0.1
GTA/TAC	-	-	1,6	3.2	8.0	2.9	-	-	-	0.1	-
GTT/ACC	-	-	1,6	1.6	0.8	0.2	2.6	-	-	-	-
TAA/TTA	-	-	-	-	-	-	-	-	3.3	0.4	1.0
TCA/TGA	-	-	-	1.6	4.4	0.6	-	1,0	-	0.6	6.2
TGA/TCA	1.6	-	1,6	-	-	-	-	-	1.1	-	-



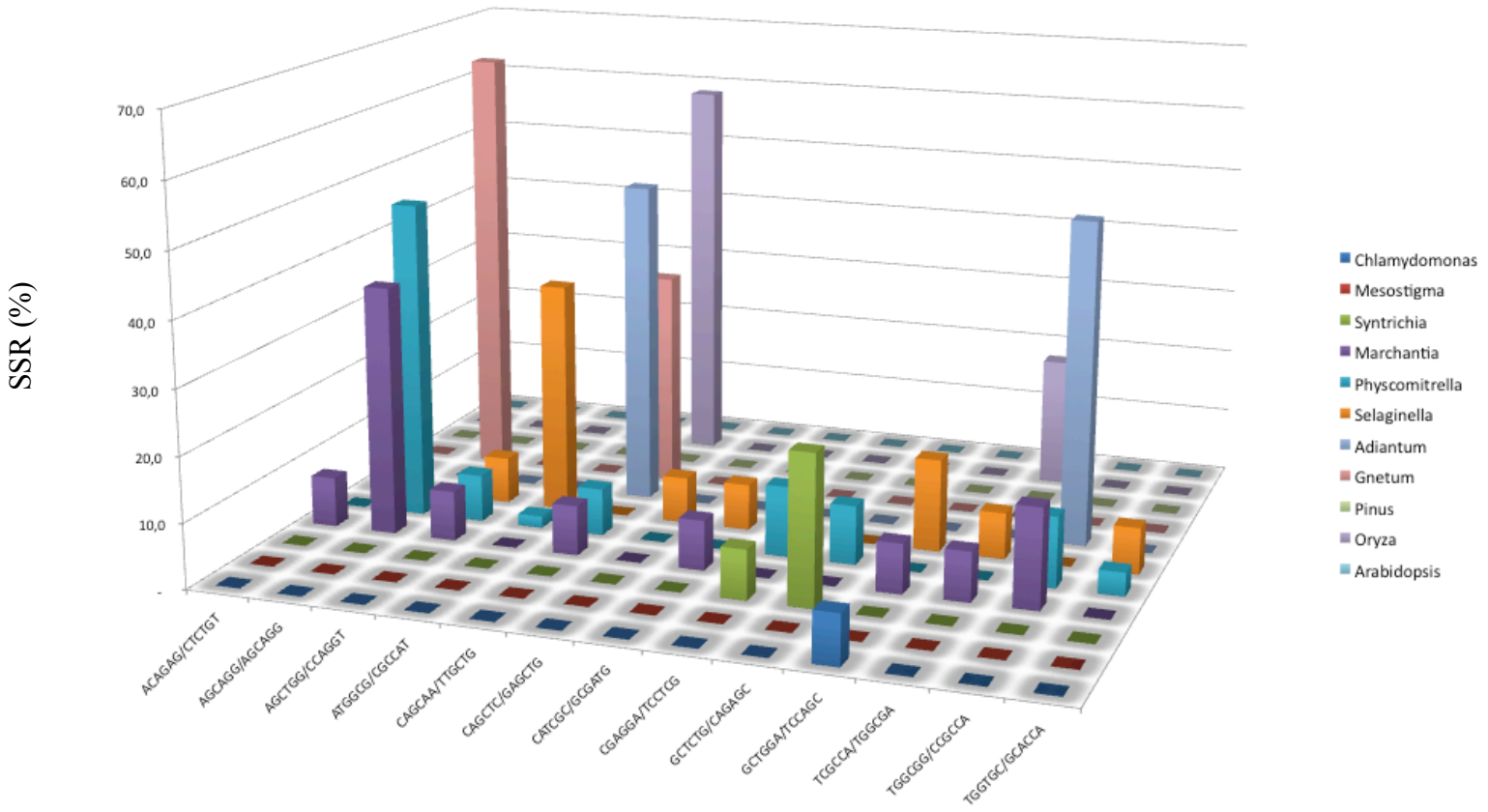
Additional file 3. Predominant loci containing trinucleotide microsatellites motifs per species.



Additional file 4. Predominant loci containing tetramers microsatellites motifs per species.



Additional file 5. Predominant loci containing pentamers microsatellites motifs per species.



Additional file 6. Predominant loci containing hexamers microsatellites motifs per species.

Additional file 7. dN/dS table for the common most frequent motifs for 11 species tested EST databases.

AGG/CCT motif					
Triplet	Syn sites (S)	Nonsyn sites (N)	dN-dS	Prob. (positive selection)	Normalized dN-dS
AAG	0.50454	2.23825	-1.73012	0.878433	-0.187245
CGG	0.557038	2.20646	1.08103	0.532335	0.116996
GAG	0.501536	2.23775	-0.206366	0.771498	-0.0223344
AGA	0.513401	2.23966	0.73118	0.614198	0.0791335
AAG	0.519388	2.22273	0.899794	0.657054	0.097382
AAG	0.514282	2.19635	-4.46748	0.985997	-0.483502
AAG	0.516528	2.18331	2.74813	0.279682	0.297422
AAG	0.731956	2.1026	-0.293794	0.690423	-0.0317965
AAG	0.547747	2.21259	1.69983	0.462505	0.183967
AAG	0.605001	2.23091	0.420291	0.597842	0.0454869
AAG	0.925331	2.05725	-0.578783	0.683643	-0.06264
AAG	0.578969	2.2423	1.53503	0.470588	0.166131
AGG	0.518176	2.32726	3.49594	0.276573	0.378355
AGA	0.918584	2.07523	0.160841	0.596693	0.0174073
GAG	0.688006	2.19088	2.07179	0.384344	0.224223
CAT	0.486148	2.45053	3.24799	0.299807	0.351521
GAA	0.865831	2.08804	-0.290612	0.63424	-0.0314521
ATA	0.910296	2.04134	-1.99686	0.815704	-0.216115
GAT	0.86488	2.09436	-1.2077	0.746108	-0.130706
GCA	0.799125	1.99794	1.50121	0.439544	0.162472
GCG	0.542355	2.2672	0.462816	0.5982	0.0500892
GCA	0.957027	1.96204	-1.68236	0.809953	-0.182077
GGC	0.720549	2.11616	-1.75598	0.805406	-0.190045
GAT	0.533094	2.41637	2.67645	0.333447	0.289664
GGA	0.590975	2.22215	4.62304	0.182718	0.500339
GAT	0.675461	2.08465	-0.676259	0.686507	-0.0731896
GAA	0.977162	1.96976	-2.07882	0.853884	-0.224985
GCA	0.621939	2.14384	1.60246	0.46101	0.17343
GCT	0.590266	2.13786	3.45118	0.231302	0.373511
GGT	0.739893	2.14588	-0.443171	0.685264	-0.0479631
GGT	0.607556	2.28413	-0.454509	0.677626	-0.0491902

AAT	0.695619	2.26213	0.438593	0.577958	0.0474677
GCA	0.958111	2.01635	2.05035	0.338642	0.221903
GCG	0.895376	2.07833	-0.97875	0.712776	-0.105927
TGC	0.623205	2.16294	3.95243	0.213227	0.42776
CCG	0.573093	2.37609	-0.666765	0.696335	-0.072162
CTC	0.962781	2.01998	-1.50465	0.79495	-0.162844
ATG	0.81405	2.01054	1.47434	0.441878	0.159563
ATG	0.93283	2.03281	-1.71463	0.813332	-0.185569
CTG	0.863353	2.09092	2.88078	0.270948	0.311778
CCT	1.0461	1.88006	0.647174	0.522228	0.0700417
CTT	0.837426	1.8965	-2.08193	0.810116	-0.225321
TTC	0.999669	1.78106	-7.41542	0.995243	-0.80255
CAG	0.650949	2.18821	2.24626	0.381008	0.243106
TCC	0.839913	1.93401	0.588486	0.54537	0.0636901
TTG	0.654223	2.30919	0.602481	0.561919	0.0652048
CCG	0.671886	2.25946	-2.77437	0.859282	-0.300262
TCG	0.690428	2.2716	3.77037	0.194335	0.408057
CGC	0.504898	2.45568	-0.647944	0.714364	-0.070125
GCC	0.805767	2.09712	0.374755	0.567291	0.0405587
GGC	0.484084	2.05504	3.56397	0.326727	0.385718
TTA	0.625928	2.1657	0.0739089	0.621917	0.00799894
TAC	1.18427	1.75413	1.61597	0.37398	0.174892
ACT	0.795257	2.02584	2.2628	0.351983	0.244896
CTC	0.597505	2.10293	1.39564	0.486205	0.151046
ACA	0.99943	1.97614	-1.71936	0.851214	-0.186081
CTT	0.750545	2.18723	1.04189	0.498332	0.112761
CCT	0.738113	1.95809	-0.979003	0.705654	-0.105955
CCC	0.831824	1.86786	-1.85934	0.805013	-0.201231
CAT	0.748215	2.0141	0.878961	0.523515	0.0951273
TTC	0.913514	2.04055	-0.47885	0.672306	-0.0518245
CTC	0.847362	2.03043	0.548342	0.55447	0.0593455
AGT	0.666666	2.32204	1.63807	0.442167	0.177283
CGT	0.983781	1.8522	-2.31964	0.875644	-0.251048
GCC	0.91735	2.01123	-0.674459	0.689912	-0.0729947
CTC	0.761539	2.08614	1.27782	0.471954	0.138295

ATC	0.756492	2.10574	-1.88623	0.82071	-0.204141
TTC	0.812806	2.00388	-0.429933	0.688247	-0.0465304
AAT	0.527396	2.38415	-2.45067	0.817206	-0.265229
ATC	0.77969	1.96889	-0.349656	0.667711	-0.0378422
TTG	0.871641	2.10883	-1.19322	0.745531	-0.129139
TTC	0.964533	1.89104	0.125074	0.592782	0.0135364
AAC	0.86428	2.07449	-2.26628	0.855083	-0.245273
AGT	0.813049	2.12136	1.67978	0.414022	0.181798
GCC	0.565472	2.41276	-2.10018	0.827655	-0.227296
AAT	0.930828	2.05557	-0.608088	0.683284	-0.0658115
GTG	0.704415	2.24775	-0.297315	0.667942	-0.0321776
CAA	0.707827	2.24365	1.55581	0.442373	0.168381
GTG	0.600711	2.07967	4.90052	0.16379	0.530369
TGG	0.931543	2.06846	1.72064	0.355232	0.18622
GAT	0.93811	2.05206	2.93919	0.225144	0.3181
CAG	0.976694	1.99653	-1.40334	0.762561	-0.151879
GGC	0.728668	2.00146	0.133063	0.602786	0.0144011
CCT	0.703238	2.28774	-0.697147	0.713933	-0.0754502
TTT	0.954248	2.01316	-2.58943	0.875632	-0.280246
CGC	0.616408	2.33745	4.88109	0.132056	0.528266
CCT	0.752321	2.19845	-0.750536	0.704652	-0.0812283
CAC	0.538658	2.27363	-0.73132	0.72696	-0.0791486
CGC	0.858961	2.10932	-2.58936	0.858316	-0.280239
CTC	0.683007	2.05798	1.92772	0.403855	0.208631
CTC	0.824259	2.00495	2.20401	0.334955	0.238533

GCA/TGC Motif

Triplet	Syn sites (S)	Nonsyn sites (N)	dN-dS	Prob. (positive selection)	Normalized dN-dS
TGA	0.964105	1.9946	-0.696999	0.696479	-0.0753993
CGG	0.975758	2.02424	-1.20895	0.760088	-0.13078
CTT	0.956179	1.99022	-0.734321	0.70043	-0.0794367
CGA	0.524856	1.94715	-3.36356	0.867423	-0.363861
ACT	0.778618	1.84346	-0.797655	0.73168	-0.086288
CTG	0.965429	1.92443	0.0138471	0.629397	0.00149794
TTT	0.598079	2.22956	-1.75438	0.804831	-0.189784

T-C	0.919839	1.95805	0.758547	0.519949	0.0820575
TGT	0.914656	1.8478	-1.13706	0.764196	-0.123004
TGC	0.870237	2.06078	-1.86354	0.840912	-0.201592
AAA	1.00098	1.98028	-1.46025	0.817387	-0.157965
CTG	1.01393	1.9579	-0.369786	0.691133	-0.0400024
GGT	0.889207	2.08062	1.26911	0.451958	0.137289
TTT	0.791561	2.03542	-0.660827	0.68159	-0.0714863
TGG	0.915063	1.84618	3.5097	0.139231	0.379669
GTG	0.923466	2.04116	-0.218732	0.648599	-0.0236618
TTT	0.645228	2.11009	-1.39617	0.785208	-0.151033
TAA	0.579224	2.39066	-2.30456	0.848339	-0.249301
GTG	1.05234	1.93305	-0.0954722	0.632961	-0.0103279
TTG	0.98962	1.83866	-0.39089	0.659854	-0.0422854
GTC	0.874017	2.0182	0.531494	0.563118	0.0574956
GTG	0.700909	1.96242	2.02161	0.396112	0.218692
GTG	0.919481	2.03356	-0.92235	0.705755	-0.0997772
ACG	1.05139	1.92144	1.31019	0.415418	0.141733
AGA	1.04619	1.91469	-0.256171	0.699221	-0.0277119
GCA	1.03101	1.96899	0.891654	0.488462	0.0964566
AAG	0.929835	2.04692	0.191718	0.584647	0.0207395
AGA	1.04672	1.94018	-0.238711	0.671239	-0.0258231
TGG	1,0	1.98193	1.02735	0.463264	0.111136
GAA	0.916272	2.04427	0.714288	0.52262	0.0772697
TGG	0.85646	2.09325	-0.246696	0.636585	-0.0266869
TTT	0.963607	2.03639	-1.76993	0.829636	-0.191466
TGT	0.97168	1.92199	-1.4884	0.776547	-0.161011
CTG	0.537339	2.00393	5.23092	0.150389	0.565866
TGA	0.96722	1.9158	-1.24627	0.761683	-0.134817
CGC	0.95481	2.01595	2.8842	0.20284	0.312004
GAC	1.00688	1.87989	0.400764	0.556512	0.0433535
GCC	0.635713	2.31333	-1.6752	0.781595	-0.181219
TGA	0.505971	2.2175	-3.39599	0.898789	-0.367369
GTT	0.708556	2.2249	-0.251788	0.660457	-0.0272377
GTT	0.886818	2.0313	-0.715176	0.710051	-0.0773657
GAT	0.929214	2.0636	0.883099	0.498664	0.0955312

GAT	0.927989	1.98006	-1.34771	0.792787	-0.145791
CCG	0.946245	2.04709	0.373589	0.566762	0.0404138
AGA	0.933718	2.0322	1.12892	0.465065	0.122123
TTG	1.17376	1.77383	0.52575	0.533089	0.0568742
TGG	0.710647	1.91497	4.25732	0.121616	0.460545
AGG	0.65716	2.2822	1.56934	0.452431	0.169767
TTT	0.849345	1.95905	4.28112	0.122747	0.463119
GTT	0.858695	1.98593	-2.95901	0.904002	-0.320097
TTC	0.840499	2.09835	3.81582	0.152434	0.412784
TTG	0.907337	2.06006	0.445719	0.565308	0.0482166
CTG	0.904673	2.09533	-0.113891	0.622613	-0.0123204
CTG	0.836133	1.97673	1.4095	0.440132	0.152475
TGA	0.867966	2.09405	-0.844905	0.707967	-0.0913994
AGT	0.959524	2.00338	0.595189	0.539524	0.0643858
TGC	1.03413	1.94442	1.02805	0.467732	0.111211
GTG	0.927491	2.03728	0.82515	0.506937	0.0892623
GGC	1.01648	1.86363	2.9511	0.213788	0.319242
CTA	0.619569	2.32578	4.47824	0.1214	0.484443
CGG	0.854097	2.07833	0.954245	0.497531	0.103228
GTA	1.0445	1.86508	-2.46023	0.855208	-0.26614
GGT	0.92804	1.90459	-0.109778	0.64778	-0.0118755
GCG	0.456573	2.05027	3.4234	0.337702	0.370334
TGG	0.609488	2.32016	-1.04676	0.725875	-0.113235
ATT	0.966533	1.99942	-1.70644	0.814532	-0.184598
GAA	0.697496	2.28822	0.380888	0.596604	0.0412034
AAT	0.959939	2.04006	-1.28721	0.790774	-0.139247
CCA	1.15799	1.74198	1.34791	0.423878	0.145813
CAG	0.955497	1.98314	-0.152298	0.654781	-0.0164752
TTG	0.80843	1.92657	0.988142	0.500227	0.106894
TCA	1.01052	1.95621	-1.3368	0.779021	-0.144612
TTT	0.527022	2.22695	4.35286	0.159572	0.47088
GGG	0.609576	2.30337	1.32563	0.488665	0.143403
CAG	0.877787	2.11017	-0.342059	0.649566	-0.037003
TTT	0.622635	2.25852	2.40493	0.368859	0.260159
GAT	0.666569	2.27578	-1.69776	0.78998	-0.183659

CGA	0.967111	2.0163	-1.01754	0.721065	-0.110074
AAT	0.916186	2.07558	-2.87149	0.914505	-0.31063
CGC	0.936391	1.99547	1.26265	0.446669	0.13659
AGC	0.957435	2.01966	-1.97683	0.872543	-0.213848
AAG	0.596464	1.99559	-1.52192	0.820359	-0.164637
GCT	0.60313	1.94089	-0.224675	0.71043	-0.0243047
GTA	0.571316	2.01791	-1.28653	0.790337	-0.139173
GCA	0.663421	1.90317	-4.84051	0.954666	-0.523633
GTC	0.63756	1.98423	-0.701934	0.708974	-0.0759333
GTG	0.875897	2.10261	0.601361	0.540245	0.0650535
GGC	0.565422	2.37087	-0.130901	0.647765	-0.0141605
CAG	0.693022	2.29678	-0.356401	0.672688	-0.0385544
GTT	0.573238	2.38855	-1.46546	0.811441	-0.158529
GTT	0.560504	2.34006	-0.651587	0.721229	-0.0704868

Additional file 8. Eletronical PCR results

Chlamydomonas EST-SSR				
Total rimers		Species	Positives Primers	Tranferability (%)
	319	*Mesostigma viride*	1	0,3
		Marchantia polymorpha	1	0,3
		Syntrichia ruralis	9	2,8
		Physcomitrella pates	7	2,2
		Selaginella ssp	2	0,6
		Adiantum capillus-veneris	86	26,9
		Gnetum gnemon	3	0,94
		Pinus taeda	4	1,2
		Oryza sativa	8	2,5
		Arabidopsis thaliana	29	9
Arabidopsis EST-SSR				
Total Primers Arabidopsis		Species	Primers Positivos	%
	1250	*Mesostigma viride*	1	0,08
		Marchantia polymorpha	1	0,08
		Syntrichia ruralis	4	0,32
		Physcomitrella pates	4	0,32
		Selaginella ssp	13	1,04
		Adiantum capillus-veneris	6	0,48
		Gnetum gnemon	9	0,72
		Pinus taeda	5	0,4
		Oryza sativa	15	1,2
		Arabidopsis thaliana	24	1,92

ANEXO 2.

Supplementary File. Homologues copies for Iron Uptaken gene families used in the present paper.

Sequence	Species	CDS Size	SEQUENCE ACESS
FRO like	*Sorghum bicolor*	2301	XM_002446494
ZmFRO	*Zea mays*	2283	NM_001154444
FRO Like	*Riccinus comunis*	2208	XM_002519344
FRO Like	*Vitis vinifera*	2145	XM_002272768
FRO like	*Populus trichocarpa*	2118	XM_002305210
OsFRO2	*Oryza Sativa*	1587	AB126085
MxFRO	*Malus xiaojinensis*	2166	EF577061
OsFRO1	*Oryza sativa*	2277	AB126084
PhyFRO1	*Physcomitrella patens*	2286	XM_001759823
PHY82658	*Physcomitrella patens*	1662	XM_001768649
PHY173734	*Physcomitrella patens*	2349	XM_001786041
PHY165498	*Physcomitrella patens*	2283	XM_001767084
PHY163308	*Physcomitrella patens*	2406	XM_001762353
PHY115851	*Physcomitrella patens*	2001	XM_001754526
OsIRT1	*Oryza sativa*	1105	AB070226
Predicted protein	*Sorghum bicolor*	1179	XM_002464063
IRT like	*Zea mays*	1146	BT064213
IRT like	*Zea mays*	1146	NM_001158638
Hypothetical protein	*Sorghum bicolor*	1125	XM_002464064
HvIRT1	*Hordeum vulgare*	1113	EU545802
IRT like	*Zea mays*	1194	NM_001154769
TaZIP5	*Triticum aestivum*	1131	DQ490132
Predicter protein	*Vitis vinifera*	1065	XM_002282389
Predicted protein	*Vitis vinifera*	1059	XM_002273361
Predicted protein	*Vitis vinifera*	1047	XM_002273143
Predicted protein	*Ricinus comunis*	1053	XM_002510913
TaZIP7	*Triticum aestivum*	1161	DQ490134
MxIRT1	*Malus xiaojinensis*	1095	AY605044
ZmZIP4	*Zea mays*	1164	EU967144
ZmZIP1	*Zea mays*	1191	EU964003
OsIRT2	*Oryza sativa*	1113	AB126086

Hypothetical protein	*Sorghum bicolor*	1137	XM_002466718
Hypothetical protein	*Sorghum bicolor*	1149	XM_002461670
Hypothetical protein	*Sorghum bicolor*	1131	XM_00243989
HvZIP8	*Hordeum vulgare*	1080	FJ208993
	Populus trichocarpa	1014	XM_002322319
Putative mRNA	*Ricinus communis*	1056	XM_002514508
HvZip7	*Hordeum vulgare*	1165	AM182059
PHYPADRAFT_139457	*Physcomitrella patens*	1128	XM_001772897
PHYPADRAFT_110147	*Physcomitrella patens*	993	XM_001771879
PHYPADRAFT_104780	*Physcomitrella patens*	1104	XM_001754540
PHYPADRAFT_68566	*Physcomitrella patens*	1056	XM_001755753
AtIRT1	*Arabidopsis thaliana*	1044	AT4G19690
AtIRT2	*Arabidopsis thaliana*	774	AT4G19680
AtZIP12	*Arabidopsis thaliana*	1068	AT5G62160
AtZIP10	*Arabidopsis thaliana*	1095	AT1G31260
AtZIP9	*Arabidopsis thaliana*	1035	AT4G33020
AtZIP8	*Arabidopsis thaliana*	948	AT5G4510
AtZIP7	*Arabidopsis thaliana*	1098	AT2G04032
AtZIP6	*Arabidopsis thaliana*	1025	AT2G30080
AtZIP5	*Arabidopsis thaliana*	1083	AT1G10970
AtZIP4	*Arabidopsis thaliana*	1227	AT1G10970
AtZIP3	*Arabidopsis thaliana*	1020	AT2G32270
ATZIP1	*Arabidopsis thaliana*	1227	AT3G12750
IRT3	*Arabidopsis thaliana*	1278	AT1G60960
OsNAS1	*Oryza sativa*	999	AB046401
HvNAS1	*Hordeum vulgare*	987	AB010086
ZmNAS2	*Zea mays*	1806	AB061271
HvNAS4	*Hordeum vulgare*	990	AB011266
HvNAS7	*Hordeum vulgare*	990	AB019525
HvNAS6	*Hordeum vulgare*	987	AB011269
HvNAS3	*Hordeum vulgare*	1008	AB011264
ZmNAS1	*Zea mays*	984	AB061270
HvNAS2	*Hordeum vulgare*	1008	AB011265
Hypothetical protein	*Sorghum bicolor*	1092	NC_012870
ZmNAS3	*Zea mays*	1080	AB042551

Hypothetical protein	*Sorghum bicolor*	1068	NC_012871
TcNAS1	*Thlaspi caerulescens*	966	AJ300446
Putative mRNA	*Ricinus communis*	963	XM_002512464
Predicted protein	*Populus trichocarpa*	966	XM_002330161
Predicted protein	*Vitis vinifera*	927	XM_002282139
LjNAS1	*Lotus japonicus*	957	AB480829
MxNAS1	*Malus xiaojinensis*	978	DQ403256
Putative mRNA	*Ricinus communis*	975	XM_002519163
LjNAS2	*Lotus japonicus*	939	AB480830
OsNAS2	*Oryza sativa*	978	AB046401
ZmNAS3	*Zea mays*	1113	EU971588
OsNAS3	*Oryza sativa*	1032	AB023819
PHYPADRAFT_150995	*Physcomitrella patens*	957	XM_001782185
PHYPADRAFT_215944	*Physcomitrella patens*	1449	XM_001770047
AtNAS2	*Arabidopsis thaliana*	963	AT5G56080
AtNAS4	*Arabidopsis thaliana*	975	AT1G5643
AtNAS3	*Arabidopsis thalinana*	963	AT1G09240
AtNAS1	*Arabidopsis thaliana*	963	AT5G0495
Putative mRNA	*Ricinus communis*	960	XM_002533321
OSNRAMP1	*Oryza sativa*	1554	DQ431468
OSNRAMP2	*Oryza sativa*	1395	L81152
OSNRAMP3	*Oryza sativa*	1653	U60767
OSNRAMP4	*Oryza sativa*	1638	NM_001052329
OSNRAMP5	*Oryza sativa*	1617	NM_001065847
OSNRAMP6	*Oryza sativa*	1527	NM_001049674
OSNRAMP7	*Oryza sativa*	1626	NM_001073667
OSNRAMP8	*Oryza sativa*	1608	NM_001057173
ATNRAMP1	*Arabidopsis thaliana*	1599	AF165125
ATNRAMP2	*Arabidopsis thaliana*	1593	NM_103618
ATNRAMP3	*Arabidopsis thaliana*	1530	NM_127879
ATNRAMP4	*Arabidopsis thaliana*	1539	NM_126133
ATNRAMP5	*Arabidopsis thaliana*	1593	NM_117995
ATNRAMP6	*Arabidopsis thaliana*	1584	NM_101464
CSNRAMP1	*Chengiopanax sciadophylloides (Franch. &*	1614	AB242564

	Sav.) C.B. Shang & J.Y. Huang		
GMDTM1	*Glycine max (L.) Merr.*	1551	AY169405
LE1	*Lycopersicon esculentum Mill.*	1593	AY196091
LE2	*Lycopersicon esculentum*	1530	AY196092
NRAMP1	*Nicotiana tabacum L.*	1862	AB505625
MBNRAMP1	*Malus baccata (L.) Borkh.*	1656	AY724413.1
PTNRAMP1	*Populus trichocarpa Torr. & A. Gray*	1557	XM_002307417
PTNRAMP2	*Populus trichocarpa*	1758	XM_002307418
PTNRAMP3	*Populus trichocarpa*	1626	XM_002332024
PTNRAMP4	*Populus trichocarpa*	1443	XM_002302388
PTNRAMP5	*Populus trichocarpa*	1503	XM_002332193
PHY158431	*Physcomitrella patens Bruch. & Schimp.*	1605	XM_001751394
PHY111375	*Physcomitrella patens*	1785	XM_001751703
PHY137658	*Physcomitrella patens*	1602	XM_001771546
PHY179667	*Physcomitrella patens*	1542	XM_001759309.1
Picesitchensis	*Picea sitchensis (Bong.) Carrière*	1557	EF676616
RCNRAMP1	*Ricinus communis L.*	1389	XM_002510083
RCNRAMP2	*Ricinus communis*	1530	XM_002524904
RC1	*Ricinus communis*	1587	XM_002520288
RC2	*Ricinus communis*	1497	XM_002520287.1
RC3	*Ricinus communis*	1641	XM_002527254
Sorghumbicolor1	*Sorghum bicolor (L.) Moench.*	1578	XM_002459595.1
Sorghumbicolor2	*Sorghum bicolor*	1608	XM_002461727.1
Sorghumbicolor3	*Sorghum bicolor*	1644	XM_002451435
Sorghumbicolor4	*Sorghum bicolor*	1653	XM_002438801
Sorghumbicolor5	*Sorghum bicolor*	1551	XM_002465622
Sorghumbicolor6	*Sorghum bicolor*	1635	XM_002443386
TCNRAMP3	*Thlaspi caerulescens J. Presl & C. Presl*	1539	EF639294

TCNRAMP4	*Thlaspi caerulescens*	1536	DQ418489
TJNRAMP4	*Thlaspi japonicum H. Boissieu*	1533	AB115423
VVLOC100266052	*Vitis vinifera L.*	1629	XM_002284520
VVLOC100262929	*Vitis vinifera*	1533	XM_002267036
ZMLOC100193537	*Zea mays L.*	1653	NM_001138648
AtYSL1	*Arabidopsis thaliana*	2022	AT4G24120
AtYSL2	*Arabidopsis thaliana*	1995	AT5G24380
AtYSL3	*Arabidopsis thaliana*	2028	AT5G53550
AtYSL4	*Arabidopsis thaliana*	2013	AT5G41000
AtYSL5	*Arabidopsis thaliana*	2145	AT3G17650
AtYSL6	*Arabidopsis thaliana*	2031	AT3G27020
AtYSL7	*Arabidopsis thalina*	2067	AT1G65730
AtYSL8	*Arabidopsis thaliana*	2175	AT1G48370
OsYSL1	*Oryza sativa*	2127	NM_001049074
Hypothetical protein	*Sorghum bicolor*	2115	XM_002457001
Hypothetical protein	*Sorghum bicolor*	2043	XM_002458662
Hypothetical protein	*Sorghum bicolor*	2067	XM_002448715
Hypothetical protein	*Zea mays*	912	BT040997
OsYSL2	*Oryza sativa*	2025	AB164646
Hypothetical protein	*Sorghum bicolor*	2037	XM_002452787
AsYSL1	*Avena sativa*	2031	FJ477297
HvYSL1	*Hordeum vulgare*	2037	AB214183
ZmYSL1	*Zea mays*	2049	AF186234
TgYSL3	*Tulipa gesneriana*	2013	AB477531
VvYSL1	*Vitis vinifera*	1986	AY538258
TcYSL3	*Thlaspi caerulescens*	2019	DQ268829
NtYSL1	*Nicotiana tabacum*	2028	AB263747
OsYSL3	*Oryza sativa*	1923	AB190913
Putative mRNA	*Ricinus communis*	2154	XM_002525837
Hypothetical protein	*Sorghum bicolor*	780	XM_002440808
Putativa mRNA	*Ricinus communis*	2085	XM_002520048
PtYSL1	*Populus trichocarpa*	1749	XM_002336761
Hypothetical protein	*Sorghum bicolor*	639	XM_002440817
Hypothetical protein	*Sorghum bicolor*	2034	XM_002439467

Hypothetical protein	*Sorghum bicolor*	1140	XM_002440816
BjYSL1	*Brassica juncea*	2070	EU557021
TcYSL1	*Thlaspi caerulescens*	2082	DQ268827
Hypothetical protein	*Vitis vinifera*	2115	XM_002269241
Hypothetical protein	*Vitis vinifera*	2127	XM_002266621
ZmOPT212	*Zea mays*	2151	HM021150
Putative mRNA	*Ricinus communis*	2085	XM_002532055
OsYSL4	*Oryza sativa*	2061	AB190914
Hypothetical protein	*Sorghum bicolor*	2115	XM_002457001
Hypothetical protein	*Sorghum bicolor*	2055	XM_002453135
OPT family	*Populus trichocarpa*	1998	XM_002306353
Hypothetical protein	*Sorghum bicolor*	2034	XM_002439467
Putative mRNA	*Ricinus communis*	2085	XM_002520048
Hypothetical protein	*Sorghum bicolor*	2064	XM_002452789
Putative mRNA	*Ricinus communis*	2154	XM_002525837
Hypothetical protein	*Sorghum bicolor*	780	XM_002440808
Hypothetical protein	*Sorghum bicolor*	2262	XM_002446767
OPT Family	*Populus trichocarpa*	1749	XM_002336761
Hypothetical protein	*Sorghum bicolor*	2169	XM_002446764
OsYSL5	*Oryza sativa*	2172	AB190915
Hypothetical protein	*Sorghum bicolor*	2046	XM_002446265
Hypothetical protein	*Zea mays*	2040	BT054164
Hypothetical protein	*Sorghum bicolor*	2013	XM_002446264
OPT Family	*Populus trichocarpa*	2010	XM_002298603
Hypothetical protein	*Zea mays*	2007	BT034471
VvYSL6	*Vitis vinifera*	1950	XM_002274523
Putative mRNA	*Ricinus communis*	1113	XM_002512861
Hypothetical protein	*Sorghum bicolor*	2169	XM_002449495
Hypothetical protein	*Sorghum bicolor*	2187	XM_002452447
Putative mRNA	*Ricinus communis*	2010	XM_002528326
Hypothetical protein	*Ricinus communis*	459	XM_002512860
OsYSL6	*Oryza sativa*	2037	AB190916
Putative mRNA	*Ricinus communis*	2130	XM_002510047
Hypothetical protein	*Vitis vinifera*	2142	XM_002277256
Hypothetical protein	*Vitis vinifera*	2142	XM_002279671

Unknow mRNA	*Picea sitchensis*	633	BT123731
OsYSL7	*Oryza sativa*	2052	AB190917
ZmYSL17	*Zea mays*	2124	NM_001153129
Hypothetical protein	*Sorghum bicolor*	1314	XM_002451338
Hypothetical protein	*Sorghum bicolor*	627	XM_002451337
OsYSL8	*Oryza sativa*	2085	AB190918
Zm_BFc0149E05	*Zea mays*	2151	BT067751
Hypothetical protein	*Sorghum bicolor*	2172	XM_002446765
OsYSL9	*Oryza sativa*	1974	AB190919
Hypothetical protein	*Zea mays*	2019	NM_001175069
Hypothetical protein	*Sorghum bicolor*	2007	XM_002448207
ZM_BFc0183J03	*Zea mays*	2019	BT086561
Hypothetical protein	*Vitis vinifera*	1998	XM_002274130
Putative mRNA	*Ricinus communis*	2016	XM_002515627
YSL1-like	*Vitis vinifera*	1986	XM_002280447
Predicted protein	*Arabidopsis lyrata subsp lyrata*	2028	XM_002864203
Putative protein	*Ricinus communis*	2004	XM_002518857
OPT Family	*Populus trichocarpa*	1959	XM_002299395
Hypothetical protein	*Sorghum bicolor*	1842	XM_002452788
Hypothetical protein	*Arabidopsis lyrata subsp lyrata*	1995	XM_002872066
Hypothetical protein	*Arabidopsis lyrata subsp lyrata*	2145	XM_002883029
Hypothetical protein	*Arabidopsis lyrata subsp lyrata*	2082	XM_002886916
OsYSL10	*Oryza sativa*	2061	AB190920
Hypothetical protein	*Sorghum bicolor*	2067	XM_002448715
TcYSL2	*Thlaspi caerulescens*	2400	DQ268828
OPT Family	*Populus trichocarpa*	2004	XM_002331883
Hyphotetical protein	*Vitis vinifera*	2127	XM_002269367
OsYSL11	*Oryza sativa*	2139	AB190921
Hypothetical protein	*Sorghum bicolor*	2235	XM_002446766
OPT Family	*Populus trichocarpa*	2070	XM_002305653
Hypothetical protein	*Sorghum bicolor*	918	XM_002448159

Hypothetical protein	*Arabidopsis lyrata subsp lyrata*	2166	XM_002891382
Hypothetical protein	*Vitis vinifera*	2172	XM_002269366
Hypothetical protein	*Vitis vinifera*	2184	XM_002266742
BjYSL4	*Brassica juncea*	1407	EU779810
OPT Family	*Populus trichocarpa*	1119	XM_002331882
OsYSL13	*Oryza sativa*	2175	AB164644
OsYsl12	*Oryza sativa*	2151	AB190922
OsYSL14	*Oryza sativa*	2184	AB164645
OsYSL15	*Oryza sativa*	2019	AB190923
TcYSL3	*Thlaspi caerulescens*	2019	DQ268829
Hypothetical protein	*Arabidopsis lyrata subsp lyrata*	2022	XM_002869691
OsYSL16	*Oryza sativa*	2028	AB190924
OsYSL17	*Oryza sativa*	1911	AB190925
OsYSL18	*Oryza sativa*	2040	AB190926
PHYPADRAFT_122023	*Physcomitrella patens*	1986	XM_001759493
PHYPADRAFT_111567	*Physcomitrella patens*	2004	XM_001751891
Os03g0189600	*Oryza sativa*	110	NM_001055758
Hypothetical protein	*Sorghum bicolor*	1155	XM_002465684
ZmADH1	*Zea mays*	1149	EU962922
HvADH1	*Hordeum vulgare*	1140	AF253472
TaADH1	*Triticum aestivum*	1140	EF122847
Putative mRNA	*Ricinus communis*	1161	XM_002510588
Predicted protein	*Populus trichocarpa*	1197	XM_002301895
AtADH1	*Arabidopsis thaliana*	1140	NM_106362
PHYDRAFT_183406	*Physcomitrella patens*	1143	XM_001763955
NtADH-Like	*Nicotiana tabacum*	1143	AY619947
VvADH1	*Vitis vinifera*	1143	AF194173
AlADH1	*Arabidopsis lyrata subsp lyrata*	1140	XM_002887624
Unknow mRNA	*Picea sitchensis*	1149	EF084688
LjADH1	*Lotus corniculatus*	1143	AJ717414
OsADH2	*Oryza sativa*	1128	X16297
OcADH2	*Oryza coarctata*	1134	EU371995

ZmADH2	*Zea mays*	1140	NM_001111940
TaADH2	*Triticum aestivum*	1140	EF122846
TmADH2	*Triticum monococcum*	1140	EF122854
C.lacryma-jobiADH1	*Coix lacryma-jobi*	1140	DQ455071
OrADH1	*Oryza rufipogon*	1140	GU798004
OnADH1	*Oryza nivara*	1106	GU798011
Hyptheical protein	*Sorghum bicolor*	990	XM_002449348
Predicted protein	*Populus trichocarpa*	1143	XM_002309864
HvADH3	*Hordeum vulgare*	1140	X12734
RcADH1	*Ricinus communis*	1143	XM_002534111
Ricinuscommunis2	*Ricinus communis*	1143	XM_002526121
Ricinuscommunis3	*Ricinus communis*	1143	XM_002526125
PHYDRAFT_137950	*Physcomitrella patens*	1185	XM_001771773
PHYDRAFT_119828	*Physcomitrella patens*	1155	XM_001757949
NtADH1	*Nicotiana tabacum*	1141	X81853
VvADH2	*Vitis vinifera*	1143	AF194174
VvADH3	*Vitis vinifera*	1149	AF194175
OsADH3	*Oryza sativa*	1143	AB267278
TaADH1-3	*Triticum aestivum*	1140	EF122841

ANEXO 3

Supplementary file.

Sequence	Species	GenBank Acess number	CDS lenghth	OSNRAMP Coverage	E-value
OSNRAMP1	*Oryza sativa* L.	DQ431468	1554	-	-
OSNRAMP2	-	L81152	1395	-	-
OSNRAMP3	-	U60767	1653	-	-
OSNRAMP4	-	NM_001052329	1638	-	-
OSNRAMP5	-	NM_001065847	1617	-	-
OSNRAMP6	-	NM_001049674	1527	-	-
OSNRAMP7	-	NM_001073667	1626	-	-
OSNRAMP8	-	NM_001057173	1608	-	-
ATNRAMP1	*Arabidopsis thaliana* (L.) Heynh.	AF165125	1599	69%	$2e^{-56}$
ATNRAMP2	-	NM_103618	1593	78%	$1e^{-117}$
ATNRAMP3	-	NM_127879	1530	80%	$1e^{-116}$
ATNRAMP4	-	NM_126133	1539	72%	$9e^{-138}$
ATNRAMP5	-	NM_117995	1593	76	$2e^{-152}$
ATNRAMP6	-	NM_101464	1584	73%	$5e^{-39}$
CSNRAMP1	*Chengiopanax sciadophylloides* (Franch. & Sav.) C.B. Shang & J.Y. Huang	AB242564	1614	73%	$1e^{-54}$

GMDTM1	*Glycine max* (L.) Merr.	AY169405	1551	72%	0.0
LE1	*Lycopersicon esculentum* Mill.	AY196091	1593	79%	$9e^{-157}$
LE2	-	AY196092	1530	71%	$7e^{-114}$
NRAMP1	*Nicotiana tabacum* L.	AB505625	1862	72%	$8e^{-56}$
MBNRAMP1	Malus baccata (L.) Borkh.	AY724413.1	1656	85%	$2e^{-165}$
PTNRAMP1	*Populus trichocarpa* Torr. & A. Gray	XM_002307417	1557	68%	$2e^{-108}$
PTNRAMP2	-	XM_002307418	1758	68%	$2e^{-96}$
PTNRAMP3	-	XM_002332024	1626	73%	$1e^{-54}$
PTNRAMP4	-	XM_002302388	1443	77%	$4e^{-136}$
PTNRAMP5	-	XM_002332193	1503	79%	$2e^{-127}$
PHY158431	*Physcomitrella patens* Bruch. & Schimp.	XM_001751394	1605	59%	$2e^{-82}$
PHY111375		XM_001751703	1785	65%	$2e^{-85}$
PHY137658		XM_001771546	1602	65%	$6e^{-85}$
PHY179667		XM_001759309.1	1542	72%	$3e^{-74}$
Picesitchensis	*Picea sitchensis* (Bong.) Carrière	EF676616	1557	72%	$4e^{-161}$
RCNRAMP1	*Ricinus communis* L.	XM_002510083	1389	74%	$3e^{-150}$
RCNRAMP2	-	XM_002524904	1530	73%	$5e^{-135}$
RC1	-	XM_002520288	1587	79%	$6e^{146}$
RC2	-	XM_002520287.1	1497	79%	$4e^{136}$

RC3	-	XM_002527254	1641	73%	$7e^{-57}$
Sorghumbicolor1	*Sorghum bicolor* (L.) Moench.	XM_002459595.1	1578	92%	0.0
Sorghumbicolor2	-	XM_002461727.1	1608	88%	0.0
Sorghumbicolor3	-	XM_002451435	1644	87%	$3e^{-111}$
Sorghumbicolor4	-	XM_002438801	1653	73%	$4e-^{53}$
Sorghumbicolor5	-	XM_002465622	1551	100%	0.0
Sorghumbicolor6	-	XM_002443386	1635	80%	0.0
TCNRAMP3	*Thlaspi caerulescens* J. Presl & C. Presl	EF639294	1539	75%	$2e^{-134}$
TCNRAMP4	-	DQ418489	1536	73%	$3e-^{143}$
TJNRAMP4	*Thlaspi japonicum* H. Boissieu	AB115423	1533	70%	$7e^{-139}$
VVLOC100266052	*Vitis vinifera* L.	XM_002284520	1629	87%	$2e^{-152}$
VVLOC100262929	-	XM_002267036	1533	70%	$5e^{-96}$
ZMLOC100193537	*Zea mays* L.	NM_001138648	1653	73%	$7e^{-63}$

Supplementary Files. dS/dN values for each node in the tree.

Node#	Sequences	dS/dN	dS Branch	dN Branch
1	*TjNramp/TjNramp4*	0.1438	0.05208407	0.3621
2	*TjNramp/TjNramp4/AtNramp4*	0.1426	0.03365858	0.2360
3	*Vitisvinifera2/TcNramp3*	0.2765	0.1125	0.4070
4	*GmDMT1/Vitisvinifera1*	0.5583	0.2724	0.4879
5	Nodes 4-5	0.2423	0.03635575	0.1501
6	Nodes 1-5	0.7509	0.4897	0.6521
7	*RcNramp1/AtNramp2*	0.2490	0.05583322	0.2243
8	*Piceasitchensis/AtNramp5*	0.1937	0.07281602	0.3759
9	Node 8/*Lycopersicumesculetum1*	0.1848	0.04810895	0.2603
10	Nodes 7-9	0.1476	0.02696615	0.1827
11	*PtNramp5/PtNramp4*	0.1406	0.03605660	0.2564
12	*RcNramp3/AtNramp3*	0.1224	0.04050035	0.3309
13	Nodes 11-12	0.04266466	0.00306767	0.07190200
14	Nodes 7-13	0.1154	0.03041214	0.2636
15	Nodes 1-14	0.1057	0.02401089	0.2272
16	*PHY137658/PHY158431*	0.1684	0.06521294	0.3873
17	*PHY111375/PHY179667*	0.2803	0.1153	0.4115
18	Nodes 17-18	1.0959	0.00168124	0.00153405
19	*OsNramp7/Sorghumbicolor4*	0.09992153	0.02821738	0.2824
20	*OsNramp2/Sorghumbicolor3*	0.1959	0.09256000	0.4725
21	Node 20/*OsNramp8*	0.2173	0.05309604	0.2443
22	Nodes 19-21	0.1587	0.06823205	0.4300
23	Nodes 16-22	0.1518	0.05858416	0.3858
24	Nodes 1-23	0.3798	0.02902556	0.07641390
25	*PtNramp1/Rc1*	0.1823	0.07399165	0.4058
26	*Vitisvinifera3/CsNramp1*	0.1194	0.03194067	0.2675
27	Nodes 25-26	0.1748	0.04044694	0.2314
28	*AtNramp1/NrNramp1*	0.8850	0.08333188	0.09415544
29	Node 28/*AtNramp6*	0.07255654	0.01163743	0.1604

30	Nodes 25-29	0.1170	0.03602797	0.3079
31	*Sorghumbicolor2/Zeamays1*	1.6418	0.00164181	1e-10
32	*OsNramp6/Osnramp3*	0.3119	0.02977769	0.09546819
33	Nodes 31-32	0.1009	0.02144023	0.2125
34	Nodes 25-33	0.1639	0.09590135	0.5851
35	*Sorghumbicolor6/OsNramp4*	0.1069	0.03851065	0.3602
36	Node 35/*Vitisvinifera4*	0.3010	0.08029322	0.2668
37	*Sorghumbicolor5/OsNramp1*	0.2594	0.04434316	0.1709
38	*OsNamp5/Sorghumbicolor1*	0.1925	0.06174375	0.3207
39	Nodes 37-38	0.1734	0.04847429	0.2796
40	Nodes 25-39	0.1665	0.05549075	0.3332
41	*MbNramp1/Lycopersicumesculetum2*	0.1400	0.04757590	0.3398
42	*Vitisvinifera5/Rc2*	0.1740	0.02936233	0.1688
43	Nodes 41-42	0.1545	0.00167751	0.01085477
44	*PtNRamp3/RcNramp2*	0.2030	0.03784224	0.1864
45	Node 44/*PtNramp2*	0.09005685	0.02292830	0.2546
46	Nodes 42-45	0.1297	0.06190546	0.4774
47	Nodes 35-46	0.5092	0.2211	0.4342
48	Nodes 1-47	0.2105	0.07101055	0.3373