UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
CENTRO DE BIOTECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA MOLECULAR E CELULAR

Ben Hur Neves de Oliveira

# A evolução do gene *MIR* no gênero *Oryza*

Dissertação de Mestrado

Porto Alegre, __/__/____

Ben Hur Neves de Oliveira

# A evolução do gene *MIR* no gênero *Oryza*

Dissertação submetida ao Programa de Pós-graduação em Biologia Molecular e Celular do Centro de Biotecnologia da Universidade Federal do Rio Grande do Sul como parte dos requisitos necessários para a obtenção do grau de Mestre em Biologia Molecular e Celular.

Orientador: Prof. Dr. Felipe Klein Ricachenevsky

Co-orientadora: Profª. Drª. Janette Palma Fett

Porto Alegre, __/__/____

Este trabalho foi desenvolvido no Laboratório de Fisiologia Vegetal, localizado no Departamento de Botânica da Universidade Federal do Rio Grande do Sul, com auxílio financeiro da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

**SUMÁRIO**

**Lista de abreviaturas**

ABI3/VP1: fator de transcrição *ABSCISIC ACID INSENSITIVE 3/VIVIPAROUS 1*
*AHA1*: gene codificador de uma próton ATPase de membrana
bHLH: fatores de transcrição da família *basic helix-loop-helix*
$CO_2$: dióxido de carbono
DNA: ácido desoxirribonucleico
Fe: ferro
$Fe^{2+}$: íon ferroso
$Fe^{3+}$: íon férrico
*FIT*: gene que codifica um *FER-LIKE IRON DEFICIENCY INDUCED TRANSCRIPTION FACTOR*
*FRO2*: gene que codifica uma redutase de membrana
Grupamentos Fe-S: grupamentos ferro-enxofre
*IDE1/2*: genes *IRON DEFICIENCY RESPONSIVE ELEMENT1/2*
IDEF1/2: fatores de transcrição *IDE-binding fator*
*IDS2*: gene *IRON DEFICIENCY SPECIFIC CLONE 2*
*IRT1*: gene *IRON REGULATED TRANSPORTER 1*
M: molar
Mb: megabases
*MIR:* gene *MITOCHONDRIAL IRON-RESPONSIVE*
NAC: fator de transcrição *NO APICAL MERISTEM, Arabidopsis transcription activation factor, and CUP - SHAPED COTYLEDON*
NAM-B1: fator de transcrição NAC, NAM-B1
*NRAMP1*: gene *NATURAL RESISTANCE-ASSOCIATED MACROPHAGE PROTEIN 1*
*YSL2:* gene *Yellow Stripe Like 2*
*YSL15:* gene *YELLOW STRIPE LIKE 15*
*ZIFL4:* gene *ZINC-INDUCED FACILITATOR-LIKE 4*
Zn: zinco

**Resumo**

O arroz (*Oryza sativa*) é tanto uma espécie modelo para a biologia vegetal, especialmente para as monocotiledôneas da família *Poaceae*, quanto uma espécie cultivada de grande importância comercial, sendo a base da alimentação de metade da população mundial. O arroz faz parte do gênero *Oryza*, composto por outras 23 espécies selvagens e uma espécie domesticada, que recentemente vem sendo exploradas como potenciais reservatórios de diversidade genética para melhoramento do arroz cultivado. Genomas das espécies diploides já foram sequenciados, tornando o gênero *Oryza* um gênero-modelo para estudos evolutivos. Neste trabalho, utilizamos os dados genômicos disponíveis para entender a evolução do gene *MIR* (*Mitochondrial Iron-Responsive Gene*), até então considerado órfão (i.e., presente apenas na espécie *Oryza sativa*). Genes órfãos e genes restritos taxonomicamente são novidades evolutivas, e auxiliam na compreensão de como genes são originados. Aqui, nós mostramos que o gene *MIR* está presente em outras espécies do gênero *Oryza*, sendo restrito à um subconjunto de espécies mais próximas de *Oryza sativa*. Também sugerimos que *MIR* tem sua origem a partir de sequências não codificantes, presentes em ancestrais comuns das espécies de genoma AA (i.e., com genoma organizado de maneira similar a *Oryza sativa*) e também em espécies mais distantes. Além disso, nossos dados indicam que estas sequências não codificantes são derivadas de um fragmento de éxon de genes de Rafinose Sintase, presente em diversos grupos de monocotiledôneas. Também mostramos que os genes *MIR* são regulados por deficiência de Fe, e que essa característica é conservada em quase todas as espécies que contém um gene *MIR* potencialmente funcional, exceto em *Oryza barthii*, em cujo genoma *MIR* foi translocado para outra posição e separado de sua região regulatória. Este trabalho mostra que *MIR* não é um gene órfão, mas restrito taxonomicamente, e que permanece sofrendo alterações que podem impactar na sua função sob deficiência de Fe.

**Abstract**

Rice (*Oryza sativa*) is both a model species for plant biology, especially for monocots from the *Poaceae* Family, and a cultivated species of great commercial relevance, being a basic food source for half of the world population. Rice is part of the *Oryza* genus, composed of another 23 wild species and another domesticated species, which recently have been explored as potential reservoirs of genetic diversity for cultivated rice breeding. Genomes of the diploid species have been already sequenced, making *Oryza* a model-genus for evolutionary studies. In this work, we used available genomic data to understand the evolution of the *MIR* (*Mitochondrial Iron-Responsive Gene*) gene, which was considered an orphan gene (i.e., found only in the species *Oryza sativa*). Orphan and taxonomically restricted genes are evolutionary novelties, and help to understand how genes are originated. Here we showed that *MIR* is present in other species of the *Oryza* genus, being restricted to a subclade of species closely related to *Oryza sativa*. We also suggest that *MIR* has its origin from non coding sequences also found in common ancestors to species that harbor AA genomes (i.e., species with a similar genomic organization to *Oryza sativa*) as well as more distantly related species. Moreover, our data indicate that these non coding sequences are derived from an exon fragment of Raffinose Synthase genes, which is present in several groups of monocots. We also showed that *MIR* genes are regulated by Fe deficiency, a characteristic that is conserved in all species that have a putative functional *MIR*, with the exception of *Oryza barthii*, in which *MIR* coding sequence was translocated to another position and separated from its regulatory region. This work shows that *MIR* is not an orphan gene, but rather taxonomically restricted, and that it is undergoing genomic changes that can impact its role in Fe deficiency.

# Introdução

## 1. Arroz: importância econômica e espécie modelo

O arroz é (*Oryza sativa*) uma das culturas mais importantes do mundo, sendo alimento básico para grande parte da população mundial. Estima-se que 19% do total de calorias consumidas pela população humana seja proveniente do arroz (Elert 2014). Embora a maior parte da produção seja derivada do continente asiático, o Brasil é o nono maior produtor mundial, e o maior fora da Ásia, com uma produção anual de aproximadamente 13 milhões de toneladas (FAO 2019). Também é destacado o papel do Brasil com um consumo médio 95 g/pessoa/dia, o que coloca o país como o 55º maior consumidor (FAO 2019). Dentre as regiões produtoras destaca-se o sul do país, com o estado do Rio Grande do Sul cultivando 8,750.774 toneladas do grão, sendo responsável por aproximadamente 61% de todo o arroz produzido no Brasil (IRGA 2017).

Além de sua importância econômica, o arroz também é um dos principais modelos para a biologia vegetal. Foi a segunda planta a ter seu genoma completamente sequenciado, tanto da subespécie *japônica* (Goff et al. 2002) quanto *indica* (Yu et al. 2002), e apresenta o menor genoma entre as plantas cultivadas, entre 398 a 466 Mb na região eucromática. Hoje conta com mais de 3000 genótipos com genoma disponível (Wang et al. 2018). A comunidade científica também dispõe de diversas ferramentas que tornam o arroz um excelente modelo, incluindo bancos de mutantes por inserção (Wang et al. 2013); (Li et al. 2017), bases de dados de expressão e co-expressão de transcritos (Jung et al. 2008); (Sato et al. 2013a); (Sato et al. 2013b), proteômica (Helmy et al. 2012) e interatômica (Gu et al. 2011), além de protocolos bem estabelecidos de transformação para super-expressão, silenciamento e edição de genes. A abundância de recursos faz com que o arroz seja a planta-modelo ideal para as monocotiledôneas, e a segunda mais utilizada depois de *Arabidopsis thaliana*.

## 2. O gênero *Oryza*

O arroz cultivado, *Oryza sativa*, foi domesticado há cerca de 8000 anos, provavelmente a partir de uma espécie selvagem, *Oryza rufipogon*, que passou a ser plantada intencionalmente por tribos de caçadores-coletores na Ásia (Callaway 2014). Neste processo, diversas características foram selecionadas, incluindo o pericarpo claro e a ausência de debulhe das sementes (Kovach et al. 2007); (Sang and Ge 2013). Além disso, é sabido que uma outra espécie, *Oryza glaberrima*, foi domesticada de maneira independente na África, a partir da espécie selvagem *Oryza barthii* (Wang et al. 2014). Além das duas espécies domesticadas, o gênero *Oryza* é composto por outras 22 espécies selvagens, totalizando 11 tipos de genomas diferentes e variação de 3,6 vezes em tamanho de genoma (Jacquemin et al. 2013); (Menguer et al. 2017). Apesar da diversificação recente (~15 milhões de anos), as espécies do gênero *Oryza* são adaptadas a ambientes bastante diversos, incluindo desde margens de rios e piscinas naturais até florestas e terrenos que passam por ciclos de inundação e seca (Vaughan et al. 2003); (Jacquemin et al. 2013). Estas espécies selvagens representam uma fonte de variação genética e alélica praticamente inexplorada, com grande potencial para o melhoramento de características de interesse econômico. Diversos exemplos de características que podem ser exploradas para introdução no arroz cultivado são conhecidas nestas espécies, e outras ainda devem ser identificadas (Menguer et al. 2017). Como exemplo do potencial de espécies selvagens, foi demonstrado que na variedade ancestral de trigo *Triticum turgidum* ssp. *diccocoides* o fator de transcrição NAM-B1 é expresso em folhas-bandeira acelerando a senescência e aumentando assim a translocação de Fe, Zn e proteína para os grãos. Este alelo, no entanto, não é funcional em variedades modernas, sendo necessária a introgressão do alelo da espécie selvagem para aumentar as concentrações destes nutrientes em grãos (Uauy et al. 2006). O gênero *Oryza* é especialmente atrativo na busca por genes e alelos de interesse em espécies selvagens, considerando-se a existência de marcadores moleculares espécie-específicos (Yamaki et al. 2013), de filogenia conhecida e de genomas completos das 13 espécies do gênero

(Stein et al. 2018), combinados com a grande quantidade de ferramentas moleculares disponíveis para estudos genéticos e genômicos em arroz. Da mesma forma, estudos evolutivos e filogenéticos são amplamente auxiliados pela existência de genomas das espécies próximas, permitindo a identificação de genes e famílias gênicas recentes surgidas durante a especiação dentro do gênero *Oryza* (Stein et al. 2018).

## 3. Ferro: importância em plantas

Embora o grão de arroz seja importante como fonte de calorias, é bastante pobre em nutrientes essenciais para a nutrição humana, como os minerais ferro (Fe) e zinco (Zn). Em comparação com outros cereais, como milho e trigo, o arroz apresenta as concentrações mais baixas de ambos os nutrientes, e a variabilidade mais baixa para a característica, restringindo o uso no melhoramento convencional para aumentar as concentrações de Fe e Zn em grãos (Kennedy and Burlingame 2003); (Pfeiffer and McClafferty 2008) Estima-se que cerca de 30% da população mundial esteja sob risco de deficiência de Fe e Zn, consideradas as duas deficiências nutricionais mais prevalentes em humanos (Graham et al. 2012). Dietas ricas em cereais, comuns em países e populações mais pobres, aumentam o risco de deficiência de ambos (Gómez-Galera et al. 2010). Por fim, é esperado que o aumento de $CO_2$ na atmosfera reduza ainda mais as concentrações de Fe e Zn em grãos de diversas culturas (Myers et al. 2014), tornando esforços para melhorar a qualidade nutricional de grãos essencial.

Fe (assim como Zn) é essencial para as plantas. Fe apresenta características químicas que permitem a formação de ligações com átomos doadores de elétrons como oxigênio, nitrogênio e enxofre, e a ligação a proteínas como grupos heme, grupamentos Fe-S e também como Fe não-heme (Römheld and Nikolic 2006). O Fe é encontrado em dois estados de oxidação sob pH fisiológico – ferroso ($Fe^{2+}$) e férrico ($Fe^{3+}$), sendo um eficiente doador / aceptor de elétrons. A facilidade em trocar de estado redox de $Fe^{2+}$ para $Fe^{3+}$ e vice-versa é importante para diversas funções exercidas pelo Fe. O Fe está envolvido, por exemplo, em reações redox da fotossíntese e respiração em cloroplastos e mitocôndrias (Briat et al. 2007), além da síntese da clorofila, fixação do nitrogênio e

síntese de DNA. Exemplos de proteínas que contêm Fe são citocromos, catalase, peroxidases, ferredoxina, ferro-superóxido dismutase, aconitase e lipoxigenases (Marschner 2012). Plantas sob deficiência de Fe também apresentam prejuízo no desenvolvimento, além de sintomas facilmente detectáveis como inibição do crescimento e clorose (deficiência de clorofila) nas folhas jovens (Römheld and Nikolic 2006).

## 4. Absorção de Fe em plantas

### 4.1 Absorção a partir do solo pelas raízes

O Fe é o quarto elemento mais abundante na crosta terrestre. No entanto, é comumente pouco disponível para absorção pelas plantas, devido à sua baixa disponibilidade em pH neutro a básico, condições nas quais se encontra como $Fe^{3+}$, formando óxidos-hidróxidos de baixa solubilidade. Estima-se que a concentração de Fe encontrada pela plantas nesse tipo de solo seja entre $10^{-14}$ a $10^{-17}$ M; porém, a concentração considerada ideal para o crescimento é na faixa de $10^{-4}$ e $10^{-9}$ M (Guerinot and Yi 1994). Assim, as plantas frequentemente se encontram sob deficiência de Fe, sendo necessário utilizar estratégias para solubilizar o metal na rizosfera e transportar através da membrana.

A maior parte das plantas utiliza a chamada estratégia I, ou estratégia de redução, descrita em detalhe em *Arabidopsis thaliana* (Fig. 1A). A estratégia I consiste no aumento de expressão de um conjunto de genes, incluindo (1) uma próton ATPase de membrana, codificada pelo gene *AtAHA1*, a qual reduz o pH da rizosfera, aumentando a solubilidade de $Fe^{3+}$; (2) uma redutase de membrana, codificada pelo gene *AtFRO2*, a qual reduz $Fe^{3+}$ para $Fe^{2+}$; e (3) um transportador de alta afinidade por $Fe^{2+}$, que transporta $Fe^{2+}$ para o citoplasma (Ricachenevsky and Sperotto 2014). Já as monocotiledôneas da família *Poaceae* (as "gramíneas") utilizam a estratégia II, ou estratégia de quelação (do inglês "chelation"; Fig. 1B). Esta estratégia consiste na síntese e secreção de fitossideróforos, aminoácidos modificados que são capazes de se ligar a Fe e outros metais, o que é feito pelo transportador de efluxo codificado pelo gene *OsZIFL4* (Nozoye et al. 2011); (Ricachenevsky et al. 2011); e no transporte do complexo

fitossideróforo-$Fe^{3+}$ para o citoplasma pelo transportador OsYSL15 (Lee et al. 2009); (Inoue et al. 2009). Portanto, as plantas da família *Poaceae*, que inclui diversas plantas de interesse econômico, como milho, sorgo e trigo, utilizam a estratégia II, a qual absorve diretamente o Fe na forma $Fe^{3+}$ ligado a fitossideróforo.

O arroz é parte da família *Poaceae*, e, portanto, utiliza a estratégia II. No entanto, foi observado que um transportador de $Fe^{2+}$, OsIRT1, é induzido sob deficiência de Fe, e capaz de absorver $Fe^{2+}$. Assim, foi sugerido que o arroz utilizaria uma estratégia combinada, contendo partes da estratégia I e uma estratégia II completa (Ricachenevsky and Sperotto 2014). Essa capacidade seria uma adaptação recente de plantas de arroz, a ambientes alagados, onde tanto o arroz cultivado quanto os seus progenitores selvagens crescem, e nos quais o íon ferroso é bastante abundante. Assim plantas de arroz apresentam algumas particularidades no mecanismo de absorção de Fe ainda não observadas em outras espécies.

Além disso, recentemente outros mecanismos acessórios para absorção de Fe têm sido descritos em plantas. Dentre eles, alguns trabalhos têm demonstrado que raízes de *A. thaliana* secretam compostos fenólicos em resposta à deficiência de Fe, o que aumenta a solubilidade de $Fe^{3+}$, auxiliando na sua absorção (Schmid et al. 2014); (Tsai and Schmidt 2017). Além disso, foi descrito que o transportador AtNRAMP1, também induzido sob baixas concentrações de Fe, é importante para a absorção de $Fe^{2+}$ em condições tanto de suficiência e deficiência de Fe, atuando conjuntamente com AtIRT1 (Castaings et al. 2016). Assim, fica evidente que as plantas apresentam múltiplos mecanismos para garantir o suprimento adequado de Fe. No entanto, genes homólogos àqueles envolvidos nesses processos estão presentes no genoma de arroz, mas até o momento não se sabe se as *Poaceae* também apresentam tais mecanismos acessórios.

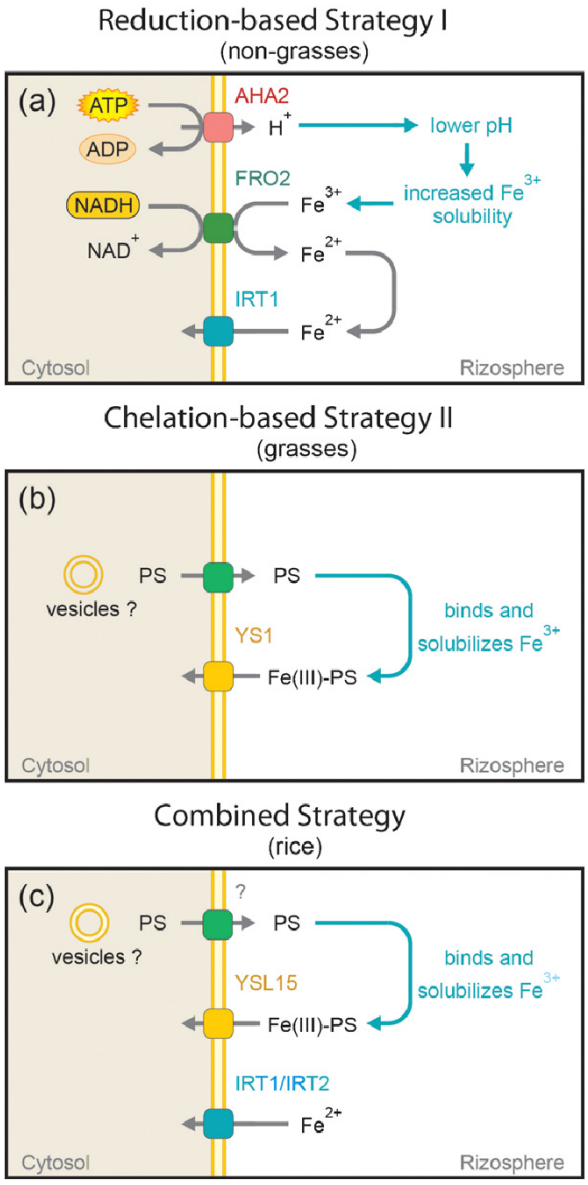


**Figura 1.** Estratégias para absorção de Fe a partir do solo. (a) Estratégia I, baseada em redução. Próton-ATPases (AHA2) secretam prótons para acidificar a rizosfera e aumentar a solubilidade de $Fe^{3+}$. $Fe^{3+}$ solubilizado é reduzido pela redutase férrica de membrana (FRO2), e o $Fe^{2+}$ é absorvido pelo transportador de metal específico (IRT1). (b) Estratégia II, baseada em quelantes. Fitossideróforos (PS – do inglês *phytosiderophore*) são sintetizados e secretados na rizosfera pelo transportador TOM1/OsZIFL4. Fitossideróforos quelam $Fe^{3+}$ na rizosfera, e complexos Fe(III)- fitossideróforo são absorvidos através do transportador OsYS1/OsYSL15. (c) Estratégia Combinada. Figura retirada de (Sperotto et al. 2012).

## 4.2 Regulação da resposta à deficiência de Fe

Os dois principais modelos nos quais a regulação transcricional da resposta à deficiência de Fe foi relativamente bem descrita são *A. thaliana* e arroz (Ivanov et al.

2012); (Kobayashi et al. 2012); (Brumbarova et al. 2015), ainda que muitos detalhes não sejam conhecidos. Em Arabidopsis, foi demonstrada a importância do fator de transcrição *FIT* para a indução dos genes-chave da resposta de deficiência de Fe, como *AtFRO2* e *AtIRT1*. *FIT* pertence à família de fatores de transcrição bHLH, é induzido transcricionalmente em raízes de plantas sob baixas concentrações de Fe, e mutantes com perda de função em *FIT* são cloróticos, a menos que grandes quantidades de Fe sejam suplementados no solo ou meio de cultivo (Colangelo and Guerinot 2004); (Jakoby et al. 2004); (Bauer et al. 2007). Interessantemente, FIT necessita formar heterodímeros com outros bHLH, como AtbHLH038, AtbHLH039, AtbHLH100 e AtbhHLH101, para então regular seus genes alvo (Wang et al. 2007); (Yuan et al. 2008). Todos estes bHLH, que fazem parte do subgrupo Ib desta superfamília em Arabidopsis, tem sua transcrição induzida sob deficiência de Fe. Trabalhos mais recentes têm se dedicado a compreender como cada um deles regula a resposta, quais tem funções similares e quais podem atuar de forma a compensar a perda de função de um ou mais membros deste subgrupo (Naranjo-Arcos et al. 2017) (Figura 2A). Por fim, alguns trabalhos têm explorado a regulação da degradação de FIT, e outros fatores que pode mediar a sua atividade (Brumbarova et al. 2015).

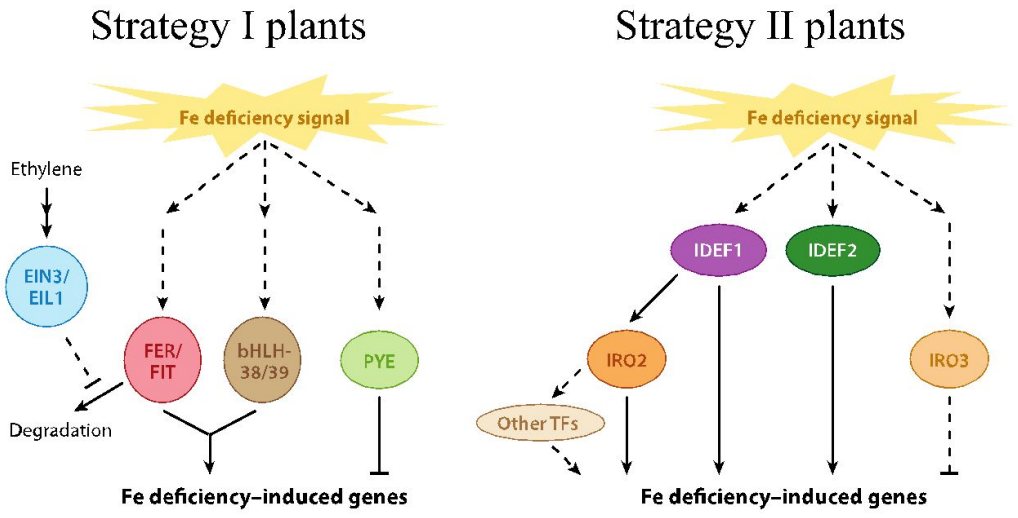


**Figura 2.** Regulação transcricional da resposta à deficiência de Fe em plantas de estratégia I e plantas de estratégia II (Figura de Kobayashi e Nishizawa et al., 2012).

Os principais fatores envolvidos na resposta transcricional de arroz e de plantas de estratégia II começaram a ser elucidados em um estudo utilizando o promotor do gene *IDS2* (*Iron Deficiency Specific clone 2*) de cevada, que codifica para uma enzima envolvida em modificações do ácido mugineico, um fitossideróforo (Kobayashi et al. 2003). Neste trabalho, os autores deletaram regiões do promotor de *IDS2* para identificar quais eram necessárias para a indução sob baixas concentrações de Fe, e identificaram duas sequências distintas, IDE1 e IDE2 (*Iron Deficiency responsive Element*), as quais atuam de forma sinérgica. Posteriormente, estudos em arroz identificaram fatores de transcrição que reconhecem IDE1 e IDE2, nomeados IDEF1 e IDEF2 (*IDE-binding factor*) (Kobayashi et al. 2007); (Ogo et al. 2008). IDEF1 e IDEF2 são parte das famílias ABI3/VP1 (*ABSCISIC ACID INSENSITIVE 3/VIVIPAROUS 1*) e NAC (*NO APICAL MERISTEM, Arabidopsis transcription activation factor, and CUP - SHAPED COTYLEDON*) de fatores de transcrição, respectivamente.

Enquanto IDEF2 controla a expressão de OsYSL2, um transportador importante para o carregamento de Fe no floema, IDEF1 parece ser o responsável pela regulação da maioria dos genes conhecidos da resposta à baixos níveis de Fe em arroz (Kobayashi et al. 2007); (Ogo et al. 2008), e por isso é considerado o principal fator de transcrição envolvido na resposta. Tanto *IDEF1* quanto *IDEF2* são expressos de maneira constitutiva (i.e., não são regulados por deficiência de Fe). IDEF1 reconhece a sequência alvo CATGC, e parece controlar a resposta de deficiência de Fe de maneira diferente nos estágios iniciais e tardios da resposta, mudando parcialmente seus alvos (Kobayashi et al. 2009). Diversos alvos de IDEF1 apresentam enriquecimento para a sequência CATGC em seus promotores. Interessantemente, IDEF1 é capaz de ligar Fe e outros metais divalentes por meio de repetições de histidina e asparagina, e regiões ricas em prolina, o que faz com que ele seja o mais provável candidato a sensor do status de Fe nas células (Kobayashi et al. 2012). Curiosamente, não foram descritos ortólogos de *IDEF1* ou *IDEF2* em plantas que não são da família *Poaceae*, e, portanto, a regulação da resposta nesse grupo de plantas parece ser distinta das não-*Poaceae*. Ainda assim, fatores de

transcrição ortólogos entre arroz e *A. thaliana* já foram descritos, com funções similares (Kobayashi and Nishizawa 2012).

## 5. Genes órfãos e o gene *MIR*

Genes órfãos são definidos como aqueles que são linhagem-específicos, não apresentando genes homólogos em espécies distantes (Tautz and Domazet-Lošo 2011). Genes órfãos podem ser derivados de diferentes processos, incluindo divergência extensiva de um gene homólogo, repetição de peptídeos de baixa complexidade, transferência lateral, perda de genes homólogos em espécies relacionadas, e origem *de novo* a partir de sequências não codificantes (Zhang et al. 2019). A geração de novos genes a partir de sequências não codificantes (por exemplo, não a partir de modificações em genes pré-existentes, por mutações e recombinações) tem sido debatida, sendo considerada como um mecanismo de novidade evolutiva menor, ou até mesmo irrelevante, quando comparado com as alternativas de duplicação e recombinação (Tautz and Domazet-Lošo 2011). No entanto, diversos trabalhos em plantas e em outros organismos já demonstraram funções importantes para genes linhagem-específicos (Xiao et al. 2009); (Cui et al. 2015). E, dentre as principais funções desempenhadas por esses genes órfãos, destaca-se principalmente a expressão desses em resposta à ocorrência de estresses bióticos e abióticos. Estudos em arabidopsis (Donoghue et al., 2011) e em arroz (GUO et al., 2007) já demonstraram isso. No caso de Arabidopsis, o gene *QQS* (do inglês, *QUA-QUINE STARCH*, AT3G30720), cuja função bioquímica já foi caracterizada, é um regulador da alocação de carbono e nitrogênio, atuando principalmente na partição do carbono e nitrogênio para amido, lipídios e proteínas das folhas se sementes (Li et al., 2009). No caso do arroz, o gene órfão, *OsDR10* é um supressor da resposta de defesa induzida por patógenos (Xiao et al., 2009). Portanto, a caracterização de genes órfãos e/ou linhagem-específicos é crucial para compreender tanto os mecanismos de evolução dos mesmos quanto determinar sua importância para a diversificação de espécies.

Em arroz, foi identificado e caracterizado um gene órfão denominado *MIR* (*Mitochondrial Iron-Regulated Gene*; (Ishimaru et al. 2009)(Bashir et al. 2011)). Neste trabalho, os autores demonstrarem que *MIR* é induzido drasticamente sob condições de deficiência de Fe, tanto em raízes quanto em folhas. Também foi demonstrado que *MIR* tem localização mitocondrial. Plantas mutantes com perda de função de *MIR* apresentam menor crescimento, acumulam maiores concentrações de Fe tanto em raízes quanto em parte aérea, porém apresentam aumento de expressão de diversos genes de resposta a deficiência de Fe, como *OsIRT1*, *OsYSL15*, e *OsNRAMP1* (Ishimaru et al. 2009). Os resultados indicam que *MIR* é importante para o controle da homeostase de Fe, e que na ausência desta proteína, plantas tem percepção do mesmo alterada (Ishimaru et al. 2009). Os autores destacam que *MIR* é um gene que não apresenta similaridade com qualquer outro em genomas de plantas, e, portanto, é um gene arroz-específico. O aumento de expressão de *MIR* sob deficiência de Fe, e também diminuição sob excesso de Fe, foram observados em trabalhos independentes (Bashir et al. 2014). No entanto, a função de *MIR* ainda é desconhecida, e a literatura segue considerando-o um gene órfão, mesmo após buscas extensivas por genes que sejam específicos de espécies de *Oryza* (Zhang et al. 2019).

## Objetivos

### Objetivo Geral

Identificar possíveis ortólogos de *MIR* em genomas de espécies selvagens de *Oryza*, e compreender os mecanismos de evolução que levaram ao surgimento deste gene funcional.

### Objetivos Específicos

- Identificar ortólogos de *MIR* nos genomas disponíveis de espécies do gênero *Oryza*;
- Reconstruir a filogenia dos genes *MIR-like* identificados;
- Determinar a presença de elementos de resposta à deficiência de Fe nos promotores de genes *MIR-like*;
- Analisar a expressão de genes *MIR-like* em condições de deficiência de Fe;
- Propor um modelo para a evolução do gene *MIR*.

**Manuscrito a ser submetido**

# The *Mitochondrial Iron Regulated* (*MIR*) gene is *Oryza* genus-specific and evolved before speciation of major AA-genome lineages

Ben-Hur Neves de Oliveira[1], Andriele Wairich[1], Andreia Carina Turchetto-Zolet[2], Janette Palma Fett[1,3], Felipe Klein Ricachenevsky[1,4*]

[1]Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, RS, Brazil.
[2]Departamento de Genética, Universidade Federal do Rio Grande do Sul, RS, Brazil.
[3]Departamento de Botânica, Universidade Federal do Rio Grande do Sul, RS, Brazil.
[4]Departamento de Biologia, Universidade Federal de Santa Maria, RS, Brazil.

* Corresponding Author

**Abstract**

Rice (*Oryza sativa*) is both a model species and an economically relevant crop. The *Oryza* genus comprises 25 species, which constitute a genetic reservoir for cultivated rice breeding and trait introgression. The availability of genomic data for many *Oryza* species, both cultivated and wild, allow for evolutionary studies in this genus, making it a "model-genus" for genetics and evolution. Here we characterized the evolution of a previously described gene named *MIR* (*Mitochondrial Iron-Regulated* gene), which is important for Fe deficiency response in *Oryza sativa* plants. *MIR* was only found in this species until now, and thus was considered an orphan gene. We showed that MIR is also found in other species closely related to *O. sativa* that are part of the AA genome group. Our data indicate that MIR originated from non coding sequences present only in AA genome species, which in turn are derived from an exon fragment of Raffinose Synthase genes, present in several groups of monocots. We also showed that *MIR* genes are regulated by Fe deficiency, a characteristic that is conserved in all species that have a putative functional *MIR*, with the exception of *Oryza barthii*, in which MIR coding sequence was translocated to another position and separated from its regulatory region. This work shows that *MIR* is not an orphan gene, but rather taxonomically restricted, and that it is undergoing genomic changes that can impact its role in Fe deficiency.

## Introduction

Asian rice (*Oryza sativa* L.) is one of the most widely cultivated crops, feeding nearly half of the world population and contributing to 20% of dietary calories consumed by humans (Stein et al. 2018). Rice is also one of the best plant model species, with large resources available for the research community, including reference genome sequences for the two main subspecies, *indica* and *japonica* (Stein et al. 2018), genome resequencing data for more than 3,000 cultivars and landraces worldwide (Wang et al. 2018), and mutant databases (Li et al. 2017). Arguably, rice it is the second most used model plant after *Arabidopsis thaliana*, and based on its synteny with other graminaceous plants, a model for other economically relevant Poaceae species (Gale and Devos 1998).

Rice is a domesticated species, which diverged from its wild progenitors 9,000 to 8,000 years ago (Callaway 2014). The *Oryza* genus includes another domesticated species, *O. glaberrima*, and 25 wild relatives, comprising 11 genome types: six diploid (n = 12: AA, BB, CC, EE, FF and GG) and five polyploid (n = 24: BBCC, CCDD, HHJJ, HHKK and KKLL) (Stein et al. 2018). These species constitute a genetic reservoir which can searched for interesting genes and alleles to improve useful traits in cultivated rice (Menguer et al. 2017). To that end, reference genomes from all AA species, which are closely related to rice, and a few distant wild relatives, were recently sequenced (Stein et al. 2018). These data can also be used to identify orthologous genes of rice in wild species, and thus understand evolution of sequences in the spam of 15 million years comprehending the *Oryza* genus (Zhang et al. 2019).

Genome sequences from closely related species can also shed light in gene novelty generation within particular lineages. Many genes annotated in any given genome are considered orphan, which means they lack detectable similarity to other sequences in related species (Arendsee et al. 2014). These unique genes can be generated by different processes, such as extensive divergence from ancestral orthologous genes (Schlötterer 2015), lateral gene transfer (Husnik and McCutcheon 2018), gene loss in related lineages (Zhao et al. 2015), and *de novo* origination from non-coding sequences (Zhang et al.

2019). Orphan genes are considered important in regulation of species-specific traits, including abiotic stresses (Arendsee et al. 2014). Early work using the rice proteome annotation comprising 59,712 proteins has identified more 19,000 orphan genes, while a recent study showed that at least 175 genes in Asian rice evolved from non-coding sequences found in wild relatives (Guo et al. 2007); (Zhang et al. 2019). These results indicating that newly evolved sequences are quite common in the *Oryza* lineage.

One of the few rice orphan genes to be functionally characterized to date is named *Mitochondrial Iron-Regulated* Gene (*MIR*; ISHIMARU et al., 2009). *MIR* was described as having no homology to other proteins, and was localized to mitochondria. Although its function remains poorly understood, *MIR* was clearly linked to iron (Fe) homeostasis, being transcriptionally upregulated upon Fe deficiency, and repressed under Fe excess (Ishimaru et al. 2009); (Bashir et al. 2014). *mir* loss-of-function mutants were shorter compared to wild type, and were impaired to perceive adequate levels of Fe in tissues, up regulating Fe uptake genes while showing twice as much Fe in roots and shoots (Bashir et al. 2014). Fe uptake in Poaceae species relies on secretion of phytosiderophores that bind $Fe^{3+}$ and are absorbed as $Fe^{3+}$-phytosiderophore complexes, while eudicots and non-Poaceae monocots used $Fe^{3+}$ reduction and $Fe^{2+}$ uptake by specific transporters (Sperotto et al. 2012). Rice plants used a more intricate uptake mechanism, combining features of both mechanisms (Ricachenevsky and Sperotto 2014). Therefore, this specialization in Fe uptake may be aided by newly generated proteins, that can act as modulators of the Fe deficiency response.

Aiming at identifying genes orthologous to *MIR* in closely related wild *Oryza* species, we used the available *Oryza* species genome data (Stein et al. 2018); (Reuscher et al. 2018). Our data indicates that *MIR* was originated before the split of major AA genome lineages, with *O. longistaminata* being the most distant AA species with a *MIR* sequence. Our data show that *MIR* is similar to non-coding sequences found only AA *Oryza* genomes (excluding early diverged *O. meridionalis*), which in turn are distantly related to an exon sequence of *Raffinose Synthase* gene present early in the monocot lineage. We also showed that *MIR* genes are responsive to Fe deficiency in *Oryza*

species, and that in Asian rice *MIR* is directly linked to the Fe deficiency transcriptional network. Our shows that *MIR* not an orphan gene specific to *O. sativa*, but an evolutionary novelty generated within the last one million years in the *Oryza* AA genome lineage.

## Methods

### Data Collection

The omics data (i.e. genomes, proteomes, transcriptomes and genomic annotation files) used in this work were downloaded for the entire set of available species in both Ensembl Plants (Zerbino et al. 2018) and Phytozome (Goodstein et al. 2012) databases. For species occurring in both databases, only the files from Ensembl Plants were kept. This summed up to 89 plant species, from which 10 are within the *Oryza* genus.

### Retrieving *MIR-like* genes by sequence similarity search

A few search strategies were applied using BLAST (Altschul et al. 1990). First, the MIR protein sequence (Os12t0282000-01, LOC Os12g18410.1) was used as query against the available downloaded plant proteomes using BLASTp and PSI-BLAST strategies ($E_{value} < 10^{-6}$) and then against the available downloaded plant genomes using tBLASTn ($E_{value} < 10^{-10}$). Afterwards, multiple sequence alignments of both MIR-like proteins and genes were performed using the MUSCLE software (Edgar 2004).

### Mapping IDE elements

For each of the *MIR-like* genes, the 2kb genomic sequence upstream the first CDS codon was retrieved. The occurrence of the IDE1 core element (ATGCT) and its reverse complement were mapped on these genomic sequences using the R package Biostrings (Pagès et al. 2019). In order to infer statistical significance for the amount of IDE1 elements found in each sequence, a hypergeometric test was applied.

**Synteny Analyses**

Synteny analyses among genomic blocks were performed using the McScanX tool (Wang et al. 2012). A genomic block, in this context, is a chromosomal segment containing 41 genes: *MIR/MIR-like* genes and the 20 first neighboring genes upstream and downstream. First, a BLASTp search is done with the set of proteins encoded by the genes inside the genomic blocks of interest, all against all ($E_{value} < 10^{-6}$). Next, the BLASTp output table and the genomic coordinates of the genes in each genomic block are fed to the McScanX tool for synteny inference ($E_{value} < 10^{-10}$). Graphical displays of syntenic blocks were made through the R package Circlize (Gu et al. 2014).

**Phylogenetic Analysis**

A conserved portion of the *MIR* and *MIR-like* genes multiple sequence alignment was extracted to build a Hidden Markov Model profile (namely, a *MIR*-profile) through the HMMER software (Eddy 1998). The *MIR*-profile was then scanned against the available plant genomes ($E_{value} < 10^{-6}$) using the same software. From those hits, sequences longer than 300 nucleotides were selected for phylogenetic analysis. First, a multiple sequence alignment was generated using the MUSCLE software (Edgar 2004) and non-conserved sites were trimmed from the dataset. The jModelTest (Posada 2008) was applied to infer the evolution model that best fit the data (i.e the one with the lowest A.I.C. value) and then a bayesian analysis was carried out using BEAST (Drummond and Rambaut 2007) ("birth-and-death process" was set as a tree prior), running $300 \times 10^6$ generations through a Markov chain Monte Carlo process. After that, Tracer was used to check for tree convergence. Tree visualization and editing were done using the R packages Phytools (Revell 2012) and ape (Paradis; Schliep, 2018). Finally, domain prediction was performed using the SMART batch perl script provided by the SMART database (Schultz et al., 2000).

**Plant Material and Treatments**

Rice seeds of *O. sativa* L. (cv. Nipponbare), *O. sativa* ssp. *spontanea* (SPO1 11T8

P5), *O. rufipogon* (BRA 00004909-8 accession from Embrapa Rice & Beans), *O. longistaminata* (IRGC 101254) and *O. barthii* (32 P4 Bar1) were submitted to 50°C for seven days to break the dormancy, according to instructions provided by International Rice Research Institute (IRRI). After, seeds were germinated for four days in petri dishes with two layers of filter paper soaked in distilled water at 28°C, two first days in the dark and the two subsequent days in the light (40μmol.m$^2$.s$^{-1}$). After germination, seedlings were transferred to inert soil (vermiculite) for fifteen days. Plants were transferred to plastic pots with 0.5 L of nutrient solution like described in Ricachenevsky et al., 2011 (Ricachenevsky et al. 2011) for seven days for acclimation. After this period, half of plants were transferred to the control condition (CC) (100 M $Fe^{3+}$-EDTA) and half of pants were transferred to Fe deficiency treatment (-Fe) (no iron added in nutrient solution). The nutrient solutions were replaced completely two times a week. The pH of nutrient solutions was adjusted to 5.4. Plants were cultivated in a growth room at 24°C ± 2°C under photoperiod of 16 h day (150μmol.m$^2$.s$^{-1}$) / 8 h dark. Five days after onset of Fe deficiency treatment roots and shoots n = 4 were collected separately, each in composed for three plants, frozen immediately in liquid nitrogen and stored in -80°C until processing, for evaluation of transcriptional profile by RT-qPCR.

**RNA extraction and gene expression analysis by RT-qPCR**

Total RNA of roots and shoots was extracted using the Plant RNA Reagent (Invitrogen®, Carlsbad, CA; USA), following the manufacturer's instructions, quantified by Nanodrop® (Thermo Fischer Scientific, Waltham, USA) and treated with DNAse I, Amplification Grade (Invitrogen®, Carlsbad, CA; USA). cDNA was prepared using OligodT and reverse transcriptase M-MLV (Invitrogen®, Carlsbad, CA; USA) following the manufacturer's instructions, and using 1 g of RNA. For gene expression analysis, the synthesized first strand cDNA was diluted 100 times. RT-qPCRwere carried out in a StepOneTM 7500 real-time cycler (Applied Biosystems, Foster City, California, EUA). One specific primer for *MIR*, for all species was designed (F: GCCCATGCTTGCCTTC, Tm value = 59.9°C; R. GCGATATATAGAGGCCACGA, Tm value = 60.1°C) and to

amplify 121 bp. Reactions were conducted like described in Ricachenevsky et al (2011). Data were analyzed by the comparative CT method (Livak and Schmittgen 2001). The PCR efficiency from the exponential phase was calculated by LingReg software (Ramakers et al. 2003). Ct values were normalized to the Ct value of *OsUBQ5* (F: ACTTCGACCGCCACTACT; Tm value = 61.9° C; R. CTAAGCCTGCTGGTTGTAGAC; Tm value = 61.6° C amplicon size = 62 bp) using the equation as described in (Ricachenevsky et al. 2011). Each data point corresponds to four biological replicates, each replicate composed for roots or shoots of three plants, and three technical replicates.

## Results

### Similarity search identifies *MIR-like* sequences

In order to infer potential *MIR* homologues through sequence similarity, we first run BLASTp and PSI-BLAST using the MIR protein sequence (Os12t0282000-01) as query against the whole plant proteome datasets from both Ensenbl Plants and Phytozome. Both strategies yielded the same hits: ORUFI12G09590, ONIVA12G08390 and OBART10G03520, from *O. rufipogon*, *O. nivara* and *O. barthii*. Next, we used a tBLASTn to find unannotated *MIR-like* genes, which expanded our set of *MIR-like* putative genes for the following species: *Oryza longistaminata*, *Oryza glaberrima* and *Oryza sativa* ssp. *indica*. These results are summarized in Table 1. Thus, we found new *MIR-like* genes only within the AA genome species of the *Oryza* genus (hereafter reference as MIR for the original sequence in *O. sativa* and *MIR-like* for the new homologous sequences). Most sequences were localized in chromosome 12, with the exception of *O. barthii*, in which the *MIR-like* sequence was localized in chromosome 10 (*O. glaberrima* and *O. longistaminata* localization is in unplaced scaffolds).

Table 1: *MIR*-like genes from similarity search

| Species | Chromosome/Contig/Scaffold | Start | End | Strand | Gene ID | CDS Length | Protein Length | Referred as |
|---|---|---|---|---|---|---|---|---|
| *Oryza sativa ssp. japonica* | 12 | 10651961 | 10653342 | – | Os12g0282000 | 492 | 164 | *OsatMIR* |
| *Oryza nivara* | 12 | 8237760 | 8238930 | - | ONIVA12G08390 | 447 | 149 | *OnivMIR* |
| *Oryza rufipogon* | 12 | 8920826 | 8921991 | – | ORUFI12G09590 | 444 | 148 | *OrufMIR* |
| *Oryza barthii* | 10 | 3797773 | 3798174 | – | OBART10G03520 | 399 | 133 | *ObarMIR* |
| *Oryza sativa ssp. indica* | 12 | 8460341 | 8460766 | - | n.a. | 426 | 142 | *OindMIR* |
| *Oryza longistaminata* | KN540552.1 | 951 | 1376 | - | n.a. | 426 | 142 | *OlonMIR* |
| *Oryza glaberrima* | Oglab12 unplaced142 | 60166 | 60564 | - | n.a. | 399 | 133 | *OglaMIR* |

*n.a. = not available

### *Cis*-acting element IDE1 core sequence is enriched upstream *MIR-like* genes

The presence of a high number of IDE1 core sequences (CATGC) in *OsatMIR* promoter region was reported by Ishimaru et. al. (Ishimaru et al. 2009), indicating that IDEF1, a transcription factor that regulated Fe deficiency responses (Kobayashi et al. 2007), binds to *OsatMIR* promoter to up regulate its transcription under low Fe availability. In order to investigate whether *MIR* homologous genes have the same IDE1 enrichment to be regulated by IDEF1, we counted how many times IDE1 elements occur 2kb sequences upstream the start codon of each *MIR* gene. Assuming a hypergeometric distribution, the number of IDE1 elements in these sequences were statistically enriched (pvalue< 0.01), indicating that *MIR-like* genes might also be regulated by IDEF1 and therefore up regulated under Fe deficiency. The exception was *ObarMIR* from *O. barthii*, which did not show such enrichment. Therefore, we searched the *O. barthii* genome using *MIR* and *MIR-like* promoters, and found one hit in chromosome 12 with a cutoff Evalue = $10^{-10}$, which is enriched with IDE1 elements (pvalue< 0.01). These results are summarized in Figure 1, showing a graphical display of the multiple sequence alignment of *MIR* genes plus 2kb long genomic sequences upstream from their start codons.

The IDE1 overall distribution is conserved across *Oryza* species that have *MIR-like* genes. In *OsatMIR*, the first two exons (which are 5'UTRs) and the second exon are specially enriched with IDE1 elements, indicating a possible regulation by IDEF1. The genomic segment comprising these exons is conserved across the studied species,

although the homologous genomic sites are not annotated as exons in the predicted gene structures from Ensembl Plants database for *OrufMIR* (ORUFI12G09590), *OnivMIR* (ONIVA12G08390) and *ObarMIR* (OBART10G03520). Thus, these results indicate that *ObarMIR* (OBART10G03520) is located in chromosome 10, and not in chromosome 12 as the other *MIR-like* genes in *Oryza* species. However, chromosome 12 of *O. barthii* shows a fragment highly similar to the IDE1-enriched region of the *MIR-like* genes found in other species, indicating that while *ObarMIR* coding sequence is located in chromosome 10, while the original promoter region is located in chromosome 12 (Figure 1).

**Synteny in *MIR-like* regions among *Oryza* species chromosomal segments**

Synteny analyses revealed that the *MIR* genes are located in syntenic positions across the *Oryza* species, with the exception of *O. barthii* (Figure 2A). Interestingly, however, the genomic segment in *O. barthii* that holds the an identified IDE1-enriched region on chromosome 12 shows high synteny degree when compared to the genomic segments that hold *MIR* genes in the other *Oryza* species (Figure 2A), indicating that the region where *ObarMIR* promoter is found is syntenic to the regions were *MIR-like* genes are located. Conversely, it is shown by the same method that *O. barthii* genomic segment on chromosome 10 carrying the *ObarMIR* gene is syntenic to the homologous region in chromosomes 10 of other *Oryza* species (Figure 2B). These data strongly suggest that *ObarMIR* (OBART10G03520) moved from chromosome 12 to chromosome 10 with minimal disturbance of the genomic environment in its vicinity, while much of the regulatory region of *ObarMIR* remained in chromosome 12.

**Large-scale genomic alignment reveals chromosomal inversions around *MIR* loci**

In order to further explore presence of collinearity in the genomic context in which *MIR* genes are occurring across *Oryza* species, we performed large-scale DNA alignments of ≈400kb genomic segments carrying a *MIR* gene and/or being syntenic to a

reference chromosome segment (*O. sativa* ssp. *japonica* chromosome 12, from 200kb upstream to 200Kb downstream of *OsatMIR*). As shown in Figure 3A, the genomic segment under analysis is fairly conserved in AA genomes, with a few small segments still preserved in more distantly related species *O. punctata*, *O. brachyantha* and even *Leersia perrieri*, an outgroup of the *Oryza* genus. The only species that appears to have large-scale collinearity with *O. sativa* ssp. *japonica* is *O. rufipogon*, as essentially the whole genomic block is aligned in the same orientation. For the other species with AA genomes, we found two inversions around the *MIR* loci. The first one is observed in *O. sativa* ssp. *indica* and *O. nivara*, where the inverted segment comprises *MIR* plus the first gene downstream (Os12g0282400). The second inversion is observed in *O. barthii*, *O. glaberrima*, *O. longistaminata* and *O. meridionalis* (the last does not have a MIR-Like gene), with inverted segment comprising one gene downstream of *MIR* and three other genes upstream (Os12g0281600, Os12g0281450 and Os12g0281300) (Figure 3A). In Figure 3B, we confirm that the orientation of the genes inside the inverted genomic blocks is in fact opposite in comparison with our reference chromosomal segment (from *O. sativa* ssp. *japonica*). Furthermore, the larger inversion seems to have a connection with the removal of *MIR* from its original position, because it is exclusively observed in species where the *MIR* gene appears at a non-syntenic position compared with *OsatMIR*. As already noticed, the most striking case is in *O. barthii*, where the alignment depicts an almost precise removal of *ObarMIR* from chromosome 12 and its insertion on chromosome 10. This result comes into agreement with our previous synteny analyses. Similar results are observed for both *O. longistaminata* and *O. glaberrima*, but in these species the *MIR* gene is located at a scaffold and an unplaced contig, respectively, so we can not be sure of its position regarding the whole genome. It is worth to notice, however, a collinear block downstream *OlongMIR* on the genomic alignment (Figure 3A), so this scaffold may belong to chromosome 12 in this species. Nevertheless, whole genome assemblies for these species must be improved in order to get higher resolution of such synteny relationships.

**Fe deficiency induce the expression level of *MIR* genes in species of *Oryza* genus**

Gene expression analysis has revealed that transcripts of *OsatMIR* in plants grown under Fe deficiency were more abundant compared to shoot and roots grown under Fe sufficient conditions (Ishimaru et al. 2009). We also observed that IDE1 enrichment is common in *MIR-like* genes in *Oryza* species (Figure 1). To understand the role of *MIR* and *MIR-like* genes we evaluated the expression level of these genes in plants of *O. sativa* ssp. *japonica*, *O. sativa* ssp. *spontanea*, *O. rufipogon*, *O.barthii* and *O. longistaminata*. Plants were cultivated as described in Material and Methods and submitted to Fe deficiency for five days. Expression level in roots for *O. sativa* ssp. *japonica* and *O. sativa* ssp. *spontanea*, *O. rufipogon* and *O. longistaminata* showed an increase in transcripts level of *MIR* genes when plants were grown under Fe deficiency (-Fe) (Figure 4A). The transcripts levels of *OsIRT1*, which is involved in $Fe^{2+}$ uptake in *O. sativa* plants (Eide et al. 1996), *OsNRAMP1*, which is known to be induced in *O. sativa* by low Fe conditions, and YSL15, a Fe(III)- deoxymugineic acid transporter of *O. sativa*(Inoue et al. 2009), were used as positive controls. We observed an increase in transcripts level for all these genes in all species, with exception for *IRT1* in roots of *O. longistaminata*, which was not up regulated (Figure 4A). For *O. barthii*, we did not observed difference in *ObarMIR* expression level in roots grown under Fe sufficiency and Fe deficiency conditions, while *ObarMIR* was up regulated in shoots, but to a lesser degree compared to other *Oryza* species. Taken together, these results indicate that *MIR-like* genes are up regulated under Fe deficiency, with the exception of *ObarMIR*, which lacks a regulatory region enriched in IDE1.

**Phylogenetic analysis reveals possible origins of *MIR* genes**

In order to identify possible distant homologies, we built a DNA hidden markov model from the identified *MIR* genes – referred from now on as *MIR*-profile – and scanned plant genomes with it. Such methodology is more sensitive than local alignment

strategies and hence capable of capturing distant homology relationships among sequences (Pearson 2013). Our approach revealed *MIR*-profile hits in all evaluated grass genomes plus two non-grass monocotyledons: *Musa acuminata* and *Ananascomosus* (Figure 5 and Figure 6). Most of these sequences (≈70%) were found to be part of an exon in genes that belong to the Raffinose Synthase Family, even though there is no detectable sequence similarity between *MIR* and Raffinose Synthase members at the amino acid level. However, by using a frame shift-aware aligner, we were able to align *MIR* genes with the Raffinose Synthase Family members within the *Oryza* genus (Supplementary Figure 1). This suggests that DNA indels may be the main source for the observed divergence between MIR and Raffinose Synthase proteins, which seems to have a distant homology relationship without having a related biological function.

We also found highly similar, non-coding sequences in chromosome 4 and chromosome 6 of *O. sativa* ssp japonica, *O. rufipogon*, *O.barthii*, *O. glaberrima* and *O. nivara*. *O. glumaepatula*, which does not have a *MIR-like* gene, has a similar sequence only in chromosome 6 (Figure 6B). Based on these sequences plus the Raffinose Synthase hits, we reconstructed a phylogenetic tree (Figure 6B). Our tree shows a dichotomy between two major groups. Group I, with three subgroups, one containing the *MIR* genes and the other two containing the internegenic segments on chromosomes 4 and 6. These subgroups are respectively named as "MIR", "Ch4" and "Ch6", as indicated in Figure 6. Sequences in Group I are derived exclusively from AA genomes: *O. sativa* ssp. *japonica*, O. sativa ssp. *indica*, *O. rufipogon*, *O. nivara*, *O. barthii*, *O. glaberrima*, *O. longistaminata* and *O. glumaepatula*. *O. glumaepatula* is the only one with no sequences in branches Ch4 and MIR, which may indicate that the genomic *MIR-like* fragment on chromosome 6 is the one that originated the others by gene duplication events (see Discussion). The second major group, named Group II, clusters *MIR*-profile hits located at an exon of genes that belong to the Raffinose Synthase Family. Among these branches, two of them are exclusively composed by the *Oryza* genus and they are labeled as Ch7 and Ch3, because most *MIR*-profile hits of each are located, respectively, in chromosomes 7 and 3. Proteins encoded by genes bearing *MIR*-profile hits were scanned

against the SMART protein database for domain prediction and all of them were classified as members of the Raffinose Synthase Family (even the ones missing annotation). Annotations for all genes are shown in Figure 6B.

We next confirmed that the genomic blocks in which these *MIR-like* sequences occur are syntenic to each other inside subgroups Ch4 and Ch6, but do not share synteny with each other (i.e., "Ch4" and "Ch6" sequences for the same species are not in syntenic blocks) (Figure 7). We also observed that there is a genomic segment in *O. glumaepatula* chromosome 4 that is syntenic to the genomic blocks in which *MIR-like* sequences of Ch4 subgroup is present, even though *O. glumaepatula* chromosome 4 itself does not show any sequences with similarity to *MIR*.

## Discussion

### *MIR* is not an orphan, but rather a lineage-restricted gene

All sequenced genomes present a set of unique genes. The definition of an orphan gene is sharing no similarity with genes or protein domains in other evolutionary lineages (Tautz and Domazet-Lošo 2011). The identification of orphans depends both on the detection method and the reference set of genomes considered, and thus orphan genes are more likely to be classified as such when only genomes from distantly related species are available. It is hypothesized that, since protein count is fairly constant comparing closely related species, but there is always a certain number of orphan genes, there is equilibrium between gene origin and extinction (Arendsee et al. 2014). It would also be reasonable to assume that most turnover occurs in young genes (Tautz and Domazet-Lošo 2011), which have not yet assumed a central position in metabolism. Thus, genes unique to certain species or lineages can be important for modulating existent pathways, rather than necessary for its function. The *Oryza* genus, which recently had several genomes sequenced (Stein et al. 2018), is especially attractive for identifying gene families for genes considered orphans based on previous analyses of the rice genome (Zhang et al. 2019).

Here we showed that, contrary to previous reports, *MIR* is not a rice-specific, orphan gene, but part of a new gene family that is taxonomically restricted to a subset of AA genome species. Considering the genomes available, we only found *MIR-like* sequences in AA species, but not in *O. punctata* (BB genome), *O. brachyantha* (FF), *L. perrieri* (commonly used as an outgroup for *Oryza* genus studies; STEIN et al., 2018), or any other genome available at public databases. The AA species clade within *Oryza* diverged around 2.5 million years ago (MYA), and *MIR-like* sequences are not present in *O. meridionalis*, a species found only in Australia that diverged early (~2.4 million years ago), as well as in *O. glumaepatula*, which is found only in South America and diverged around to 1 MYA (Stein et al. 2018). *MIR-like* sequences were found in both species of the African lineages, *O. barthii* and *O. glaberrima*, in all species in the Asian lineage, and in *O. longistaminata*, which diverged later than *O. glumaepatula* (REUSCHER et al., 2018; Figure 6A). Since the split between these two lineages occurred right after *O. glumaepatula* speciation, functional *MIR-like* origination can be timed to less than 1 MYA.

**A model for *MIR* origination**

Our data also shows that non-coding sequences similar to *MIR* (which can be detected using BLAST tool) are found in chromosomes 4 and 6 of many *Oryza* genomes, including *O. glumaepatula* (Figure 6). Chromosome 4 and chromosome 6 sequences are more closely related to each other than to *MIR-like* genes, indicating that they are products of a duplication event, which likely took place after *O. glumaepatula* speciation, as only the sequence in chromosome 6 is observed in *O. glumaepatula* genome (Figure 6 and Figure 7). However, we did not find these sequences in *O. meridionalis* or *O. longistaminata*, a result that can be derived from genome gaps in the draft sequence (Stein et al. 2018); (Reuscher et al. 2018). For *O. glumaepatula*, the region where the non-coding sequence similar do *MIR* is found is present in genome draft, implying that duplication occurred after *O. glumaepatula* split. Interestingly, we attempted to detect expression of these sequences in *O. sativa*, and found only a few reads (data not shown).

Although it seems that these non-coding sequences were the source of origin for functional *MIR* genes, it remains to be explored if they have any function in the genome.

Strikingly, we found distant similarities between a MIR-based hidden markov model and sequences derived from an exon inside a *Raffinose Synthase* gene, spread throughout the monocot lineages (Figure 6). These results indicate that *Raffinose Synthase* genes might be the source of genetic material for *MIR* origin, although in an indirect path. We suggest that sequences present in chromosome 6 and chromosome 4 that share similarity with *MIR-like* are directly derived from *Raffinose Synthase* exon, whereas *MIR-like* functional gene is directly derived from chromosome 6 and chromosome 4 sequences.

If correct, this would imply that *MIR* evolutionary origin includes a mixture of two proposed models for orphan (or new, lineage restricted) genes evolution: (1) duplication-divergence model, in which coding sequences are duplicated and undergo mutations that render one adaptive to a new function, but rather different from the original sequence that BLAST tools would not capture similarities; and (2) *de novo* evolution model, in which random sequences would combine to form functional sites and would come under regulatory control to produce a transcribed RNA, which would next acquire a functional open reading frame (Tautz and Domazet-Lošo 2011). *MIR* evolution would be a result of (1) duplication and divergence from a coding sequence of Raffinose Synthase into a non-coding one in chromosome 6 and (2) duplication of the non-coding sequence, and *de novo* evolution into a protein coding *MIR-like* gene.

### *MIR* function may still be evolving in distinct *Oryza* lineages

*MIR* was functionally characterized as important for Fe homeostasis in rice, and independent expression data has corroborated its role in Fe deficiency responses (Ishimaru et al. 2009); (Bashir et al. 2014); (Wairich et al, unpublished). *MIR* loss-of-function mutant seems impaired in correctly perceiving Fe sufficiency status, which leads to exaggerated Fe uptake, increased expression of Fe deficiency response genes, and Fe accumulation (Ishimaru et al. 2009). MIR protein was localized to

mitochondria, and *mir* plants show somewhat similar phenotypic changes to the *mitochondrial iron transporter* (*MIT*) *mit* loss-of-function plants (Bashir et al. 2011). Based on these studies, *MIR* seems to regulate Fe perception in mitochondria, which might be important to Fe status sensing. Since Fe sensing is key for making plants more Fe efficient and accumulate Fe in edible tissues for biofortification (Sperotto et al. 2012), it is important to understand how *MIR* might regulate such function, how this protein became entrenched in the Fe regulon, and when during *Oryza* speciation *MIR* origination occurred.

We found that most functional *MIR* genes have IDE1-enriched promoters, and are up-regulated under Fe deficiency (Figure 1 and Figure 4). The exception is *O. barthii*: *ObarMIR* promoter does not have IDE1 rich regulatory sequences upstream its coding sequence, and Fe response is impaired (Figure 4). *ObarMIR* is localized in chromosome 10, whereas the regulatory region homologous to *MIR-like* gene promoters is in *O. barthii* chromosome 12, in a region syntenic to *MIR-like* gene localization is all other species (Figure 1 and Figure 2). Thus, *ObarMIR* has changed position inside *O. barthii* genome, but only partially, leaving its original promoter. Interestingly, in *O. glaberrima*, the domesticated species to which *O. barthii* is the progenitor (Wang et al. 2014), *OglaMIR* also has changed position within chromosome 12, as evidenced by our large-scale genomic aligment (Figure 3; we were not able to include *O. glaberrima* in MCScan analyses in Figure 2, due to lack of neighboring genes data – *OglaMIR* is located in an unplaced scaffold). Thus, we showed that *ObarMIR* jumped from chromosome 12 to chromosome 10, losing its Fe deficiency responsive capacity. These results indicate that, although *MIR* is might be functional in all *Oryza* species where it is found, its role is not central, and can be changed or lost in specific lineages. Moreover, we have showed that *MIR-like* genes are likely to have a conserved function in *Oryza* within the subset of AA genome specie where it is found, regulating responses to low Fe.

**Conclusion**

We propose the protein coding MIR genes are anciently related to a sequence

within a *Raffinose Synthase* gene, which by duplication and degeneration results in a non-coding sequence with no apparent function. This non-coding sequence was again duplicated and de novo evolution resulted in the protein coding MIR gene around 1 million years ago. *MIR* genes have a function in Fe homeostasis, but are still undergoing disruptive evolutionary changes in at least one species, including changes in its position and loss of Fe-related regulatory sequences, which can drastically change its role within the plant. Our data provide a detailed example of fast evolution of newly evolved, lineage-restricted genes.

## References


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

Arendsee ZW, Li L, Wurtele ES (2014) Coming of age: orphan genes in plants. Trends Plant Sci 19:698–708. doi: 10.1016/j.tplants.2014.07.003

Bashir K, Hanada K, Shimizu M, et al (2014) Transcriptomic analysis of rice in response to iron deficiency and excess. Rice 7:1–15. doi: 10.1186/s12284-014-0018-1

Bashir K, Ishimaru Y, Shimo H, et al (2011) The rice mitochondrial iron transporter is essential for plant growth. Nat Commun 2:322–327. doi: 10.1038/ncomms1326

Bauer P, Ling HQ, Guerinot M Lou (2007) FIT, the FER-LIKE IRON DEFICIENCY INDUCED TRANSCRIPTION FACTOR in Arabidopsis. Plant Physiol Biochem 45:260–261. doi: 10.1016/j.plaphy.2007.03.006

Briat JF, Curie C, Gaymard F (2007) Iron utilization and metabolism in plants. Curr Opin Plant Biol 10:276–282. doi: 10.1016/j.pbi.2007.04.003

Brumbarova T, Bauer P, Ivanov R (2015) Molecular mechanisms governing Arabidopsis iron uptake. Trends Plant Sci 20:124–133. doi: 10.1016/j.tplants.2014.11.004

Callaway E (2014) The birth of rice. Nature 514:S58-59

Castaings L, Caquot A, Loubet S, Curie C (2016) The high-affinity metal transporters NRAMP1 and IRT1 team up to take up iron under sufficient metal provision. Sci Rep 6:1–11. doi: 10.1038/srep37222

Colangelo EP, Guerinot M Lou (2004) The essential Basic Helix-Loop-Helix Protein FIT1 is required for the iron deficiency response. Plant Cell 16:3400–3412. doi: 10.1105/tpc.104.024315

Cui X, Lv Y, Chen M, et al (2015) Young genes out of the male: An insight from evolutionary age analysis of the pollen transcriptome. Mol Plant 8:1–11. doi: 10.1016/j.molp.2014.12.008

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7:1–8. doi: 10.1017/CBO9781139095112.007

Eddy SR (1998) Profile hidden Markov models. Bioinformatics 14:755–763

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. doi: 10.1093/nar/gkh340

Eide D, Broderius M, Fett J, Guerinot M Lou (1996) A novel iron-regulated metal transporter from plants identified by functional expression in yeast. Proc Natl Acad Sci 93:5624–5628. doi: 10.1073/pnas.93.11.5624

Elert E (2014) Rice by the numbers: A good grain. Nature 514:S50–S51. doi: 10.1038/514S50a

FAO (2019) Food and Agriculture Organization of the United Nations

Gale MD, Devos KM (1998) Comparative genetics in the grasses. Proc Natl Acad Sci 95:1971–1974

Goff SA, Ricke D, Lan T-H, et al (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). Science (80- ) 296:92–100. doi: 10.1126/science.1068037

Gómez-Galera S, Rojas E, Sudhakar D, et al (2010) Critical evaluation of strategies for mineral fortification of staple food crops. Transgenic Res 19:165–180. doi: 10.1007/s11248-009-9311-y

Goodstein DM, Shu S, Howson R, et al (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40:D1178-1186. doi: 10.1093/nar/gkr944

Graham RD, Knez M, Welch RM (2012) How much nutritional iron deficiency in humans globally is due to an underlying zinc deficiency?, 1st edn. Elsevier Inc.
Gu H, Zhu P, Jiao Y, et al (2011) PRIN: a predicted rice interactome network. BMC Bioinformatics 12:1–13. doi: 10.1186/1471-2105-12-161
Gu Z, Gu L, Eils R, et al (2014) Circlize implements and enhances circular visualization in R. Bioinformatics 30:2811–2812. doi: 10.1093/bioinformatics/btu393
Guerinot M Lou, Yi Y (1994) Iron: nutritious, noxious, and not readily available. Plant Physiol 104:815–820
Guo W-J, Li P, Ling J, Ye S-P (2007) Significant comparative characteristics between orphan and nonorphan genes in the rice ( Oryza sativa L. ) genome. Comp Funct Genomics 2007:1–7. doi: 10.1155/2007/21676
Helmy M, Sugiyama N, Tomita M, Ishihama Y (2012) The rice proteogenomics database OryzaPG-DB: development, expansion, and new features. Front Plant Sci 3:1–6. doi: 10.3389/fpls.2012.00065
Husnik F, McCutcheon JP (2018) Functional horizontal gene transfer from bacteria to eukaryotes. Nat Rev Microbiol 16:67–79. doi: 10.1038/nrmicro.2017.137
Inoue H, Kobayashi T, Nozoye T, et al (2009) Rice OsYSL15 is an iron-regulated iron(III)-deoxymugineic acid transporter expressed in the roots and is essential for iron uptake in early growth of the seedlings. J Biol Chem 284:3470–3479. doi: 10.1074/jbc.M806042200
IRGA (2017) Instituto Riograndense do Arroz
Ishimaru Y, Bashir K, Fujimoto M, et al (2009) Rice-specific mitochondrial iron-regulated gene (MIR) plays an important role in iron homeostasis. Mol Plant 2:1059–1066. doi: 10.1093/mp/ssp051
Ivanov R, Brumbarova T, Bauer P (2012) Fitting into the harsh reality: regulation of iron-deficiency responses in dicotyledonous plants. Mol Plant 5:27–42. doi: 10.1093/mp/ssr065
Jacquemin J, Bhatia D, Singh K, Wing RA (2013) The International Oryza Map Alignment Project: development of a genus-wide comparative genomics platform to help solve the 9 billion-people question. Curr Opin Plant Biol 16:147–156. doi: 10.1016/j.pbi.2013.02.014
Jakoby M, Wang HY, Reidt W, et al (2004) FRU (BHLH029) is required for induction of iron mobilization genes in Arabidopsis thaliana. FEBS Lett 577:528–534. doi: 10.1016/j.febslet.2004.10.062
Jung KH, Dardick C, Bartley LE, et al (2008) Refinement of light-responsive transcript lists using rice oligonucleotide arrays: evaluation of gene-redundancy. PLoS One 3:1–18. doi: 10.1371/journal.pone.0003337
Kennedy G, Burlingame B (2003) Analysis of food composition data on rice from a plant genetic resources perspective. Food Chem 80:589–596. doi: 10.1016/S0308-8146(02)00507-1
Kobayashi T, Itai RN, Aung MS, et al (2012) The rice transcription factor IDEF1 directly binds to iron and other divalent metals for sensing cellular iron status. Plant J 69:81–91. doi: 10.1111/j.1365-313X.2011.04772.x
Kobayashi T, Itai RN, Ogo Y, et al (2009) The rice transcription factor IDEF1 is essential for the early response to iron deficiency, and induces vegetative expression of late embryogenesis abundant genes. Plant J 60:948–961. doi: 10.1111/j.1365-313X.2009.04015.x
Kobayashi T, Nakayama Y, Itai RN, et al (2003) Identification of novel cis-acting elements, IDE1 and IDE2, of the barley IDS2 gene promoter conferring iron-deficiency-inducible, root-specific expression in heterogeneous tobacco plants. Plant J 36:780–793. doi: 10.1046/j.1365-313X.2003.01920.x

Kobayashi T, Nishizawa NK (2012) Iron uptake, translocation, and regulation in higher plants. Annu Rev Plant Biol 63:131–152. doi: 10.1146/annurev-arplant-042811-105522
Kobayashi T, Ogo Y, Itai RN, et al (2007) The transcription factor IDEF1 regulates the response to and tolerance of iron deficiency in plants. Proc Natl Acad Sci 104:19150–19155. doi: 10.1073/pnas.0707010104
Kovach MJ, Sweeney MT, McCouch SR (2007) New insights into the history of rice domestication. Trends Genet 23:578–587. doi: 10.1016/j.tig.2007.08.012
Lee S, Chiecko JC, Kim SA, et al (2009) Disruption of OsYSL15 leads to iron inefficiency in rice plants. Plant Physiol 150:786–800. doi: 10.1104/pp.109.135418
Li G, Jain R, Chern M, et al (2017) The sequences of 1,504 mutants in the model rice variety Kitaake facilitate rapid functional genomic studies. Plant Cell 29:1218–1231. doi: 10.1105/tpc.17.00154
Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. Methods 25:402–408. doi: 10.1006/meth.2001.1262
Marschner P (2012) Marschner's mineral nutrition of higher plants. Academic Press
Menguer PK, Sperotto RA, Ricachenevsky FK (2017) A walk on the wild side: Oryza species as source for rice abiotic stress tolerance. Genet Mol Biol 40:238–252. doi: 10.1590/1678-4685-gmb-2016-0093
Myers SS, Zanobetti A, Kloog I, et al (2014) Rising CO2 threatens human nutrition. Nature 510:139–142. doi: 10.1038/nature13179
Naranjo-Arcos MA, Maurer F, Meiser J, et al (2017) Dissection of iron signaling and iron accumulation by overexpression of subgroup Ib bHLH039 protein. Sci Rep 7:1–12. doi: 10.1038/s41598-017-11171-7
Nozoye T, Nagasaka S, Kobayashi T, et al (2011) Phytosiderophore efflux transporters are crucial for iron acquisition in graminaceous plants. J Biol Chem 286:5446–5454. doi: 10.1074/jbc.M110.180026
Ogo Y, Kobayashi T, Itai RN, et al (2008) A novel NAC transcription factor, IDEF2, that recognizes the iron deficiency-responsive element 2 regulates the genes involved in iron homeostasis in plants. J Biol Chem 283:13407–13417. doi: 10.1074/jbc
Pagès H, Aboyoun P, Gentleman R, DebRoy S (2019) Biostrings: Efficient manipulation of biological strings. R package version 2.52.0.
Paradis E, Schliep K (2018) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35:526–528
Pearson WR (2013) An introduction to sequence similarity ("homology") searching. In: Current Protocols in Bioinformatics. pp 1–9
Pfeiffer WH, McClafferty B (2008) Biofortification: breeding micronutrient-dense crops
Posada D (2008) jModelTest: Phylogenetic model averaging. Mol Biol Evol 25:1253–1256. doi: 10.1093/molbev/msn083
Ramakers C, Ruijter JM, Lekanne Deprez RH, Moorman AFM (2003) Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neurosci Lett 339:62–66. doi: 10.1016/S0304-3940(02)01423-4
Reuscher S, Furuta T, Bessho-Uehara K, et al (2018) Assembling the genome of the African wild rice Oryza longistaminata by exploiting synteny in closely related Oryza species. Commun Biol 162:1–10. doi: 10.1038/s42003-018-0171-y
Revell LJ (2012) phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol Evol 3:217–223. doi: 10.1111/j.2041-210X.2011.00169.x
Ricachenevsky FK, Sperotto RA (2014) There and back again, or always there? The evolution of

rice combined strategy for Fe uptake. Front Plant Sci 5:1–5. doi: 10.3389/fpls.2014.00189
Ricachenevsky FK, Sperotto RA, Menguer PK, et al (2011) ZINC-INDUCED FACILITATOR-LIKE family in plants: lineage-specific expansion in monocotyledons and conserved genomic and expression features among rice (Oryza sativa) paralogs. BMC Plant Biol 11:11–20. doi: 10.1186/1471-2229-11-20
Römheld V, Nikolic M (2006) Iron. In: Barker A V., Pilbeam DJ (eds) Handbook of Plant Nutrition. Taylor & Francis Group, Boca Raton, p 662
Sang T, Ge S (2013) Understanding rice domestication and implications for cultivar improvement. Curr Opin Plant Biol 16:139–146. doi: 10.1016/j.pbi.2013.03.003
Sato Y, Namiki N, Takehisa H, et al (2013a) RiceFREND: A platform for retrieving coexpressed gene networks in rice. Nucleic Acids Res 41:1214–1221. doi: 10.1093/nar/gks1122
Sato Y, Takehisa H, Kamatsuki K, et al (2013b) RiceXPro Version 3.0: expanding the informatics resource for rice transcriptome. Nucleic Acids Res 41:1206–1213. doi: 10.1093/nar/gks1125
Schlötterer C (2015) Genes from scratch – the evolutionary fate of de novo genes. Trends Genet 31:215–219. doi: 10.1016/j.tig.2015.02.007
Schmid NB, Giehl RFH, Doll S, et al (2014) Feruloyl-CoA 6'-hydroxylase1-dependent coumarins mediate iron acquisition from alkaline substrates in arabidopsis. Plant Physiol 164:160–172. doi: 10.1104/pp.113.228544
Sperotto RA, Ricachenevsky FK, Waldow V de A, Fett JP (2012) Iron biofortification in rice: It's a long way to the top. Plant Sci 190:24–39. doi: 10.1016/j.plantsci.2012.03.004
Stein JC, Yu Y, Copetti D, et al (2018) Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus Oryza. Nat Genet 50:285–296. doi: 10.1038/s41588-018-0040-0
Tautz D, Domazet-Lošo T (2011) The evolutionary origin of orphan genes. Nat Rev Genet 12:692–702. doi: 10.1038/nrg3053
Tsai HH, Schmidt W (2017) Mobilization of iron by plant-borne coumarins. Trends Plant Sci 22:538–548. doi: 10.1016/j.tplants.2017.03.008
Uauy C, Distelfeld A, Fahima T, et al (2006) A NAC gene regulating senescence improves grain protein, Zn, and Fe content in wheat. Science (80- ) 314:1298–1301
Vaughan DA, Morishima H, Kadowaki K (2003) Diversity in the Oryza genus. Curr Opin Plant Biol 6:139–146. doi: 10.1016/S1369-5266(03)00009-8
Wang HY, Klatte M, Jakoby M, et al (2007) Iron deficiency-mediated stress regulation of four subgroup Ib BHLH genes in Arabidopsis thaliana. Planta 226:897–908. doi: 10.1007/s00425-007-0535-x
Wang M, Yu Y, Haberer G, et al (2014) The genome sequence of African rice (Oryza glaberrima) and evidence for independent domestication. Nat Genet 46:982–991. doi: 10.1038/ng.3044
Wang N, Long T, Yao W, et al (2013) Mutant resources for the functional analysis of the rice genome. Mol Plant 6:596–604. doi: 10.1093/mp/sss142
Wang W, Mauleon R, Hu Z, et al (2018) Genomic variation in 3,010 diverse accessions of Asian cultivated rice. Nature 557:43–49. doi: 10.1038/s41586-018-0063-9
Wang Y, Tang H, Debarry JD, et al (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res 40:1–14. doi: 10.1093/nar/gkr1293
Xiao W, Liu H, Li Y, et al (2009) A rice gene of de novo origin negatively regulates pathogen-induced defense response. PLoS One 4:1–12. doi: 10.1371/journal.pone.0004603
Yamaki S, Ohyanagi H, Yamasaki M, et al (2013) Development of INDEL markers to discriminate all genome types rapidly in the genus Oryza. Breed Sci 63:246–254. doi:

10.1270/jsbbs.63.246
Yu J, Hu S, Wang J, et al (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. indica). Science (80- ) 296:79–92. doi: 10.1126/science.1068037
Yuan Y, Wu H, Wang N, et al (2008) FIT interacts with AtbHLH38 and AtbHLH39 in regulating iron uptake gene expression for iron homeostasis in Arabidopsis. Cell Res 18:385–397. doi: 10.1038/cr.2008.26
Zerbino DR, Achuthan P, Akanni W, et al (2018) Ensembl 2018. Nucleic Acids Res 46:D754–D761. doi: 10.1093/nar/gkx1098
Zhang L, Ren Y, Yang T, et al (2019) Rapid evolution of protein diversity by de novo origination in Oryza. Nat Ecol Evol 3:679–690. doi: 10.1038/s41559-019-0822-5
Zhao Y, Tang L, Li Z, et al (2015) Identification and analysis of unitary loss of long-established protein-coding genes in Poaceae shows evidences for biased gene loss and putatively functional transcription of relics. BMC Evol Biol 15:1–10. doi: 10.1186/s12862-015-0345-x

## Figures and legends to figures

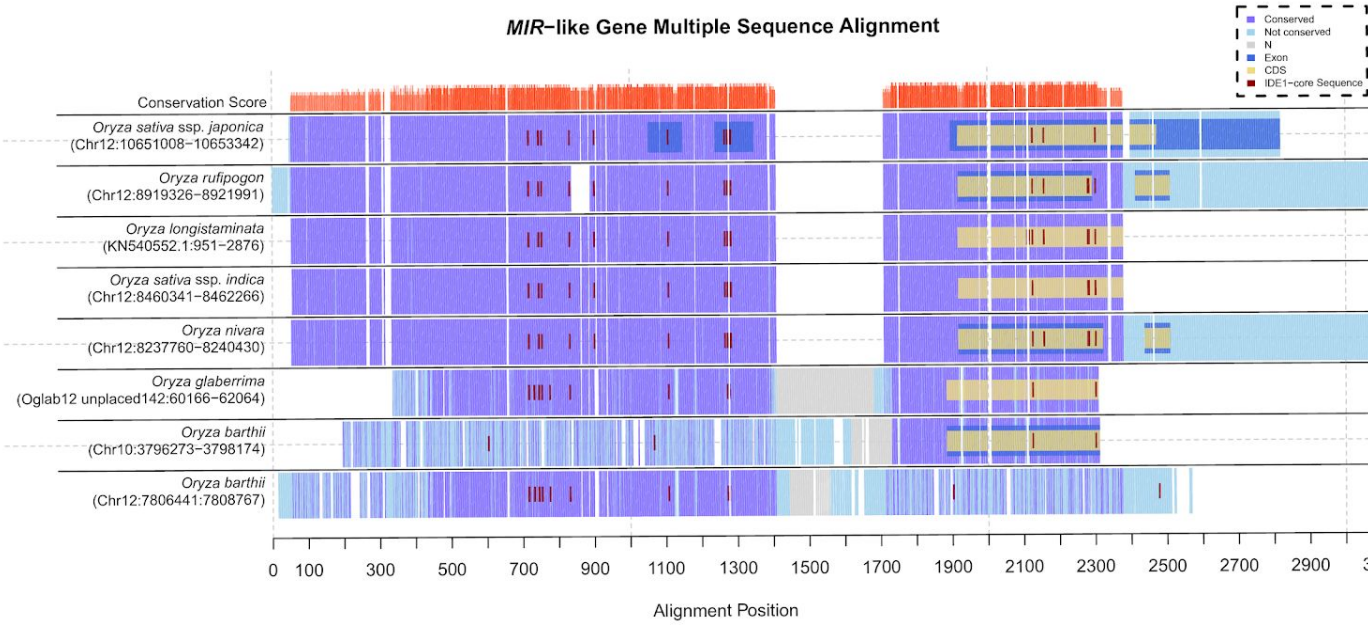


**Figure 1.Graphical representation of the genomic DNA multiple sequence alignment of *MIR-like* genes.** On the top row, red bars indicate conservation score of the alignment, where conservation score the proportion of nucleotides on that particular site that are equal to the consensus nucleotide for that position. Purple represents conserved sites on the sequences (>50% conservation), while light blue represents non-conserved sites. Annotated exons and CDS (according to ENSEMBL plants annotation files) are shown in blue and yellow, respectively. Undetermined nucleotides ("N") are grey and blank spaces represent gaps. Short vertical dark red lines indicate where the IDE1 core sequence occurs across the alignment. Sequences from distinct *Oryza* species are identified, and their chromosome and genomic position is shown.

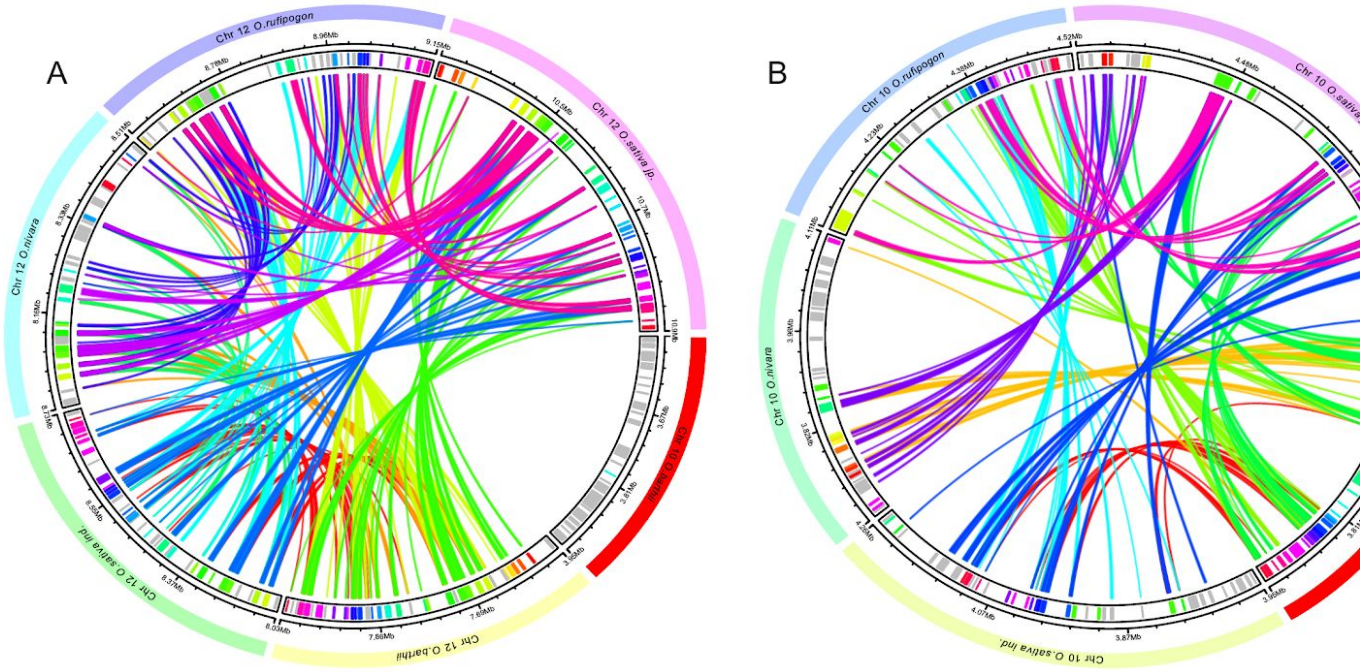


**Figure 2.Circular plots of synteny in the *MIR-like* genomic regions.** The rectangular arcs represent genomic segments. Vertical bars within arcs represent genes in the genomic segment. Ribbons connecting a pair of genomic loci show syntenic pairs of genes (E-value < $10^{-30}$). **(A)** Circle plot containing exclusively genomic segments containing a *MIR-like* gene in *Oryza* species. Genes with the same colors are represented by vertical lines in the same color (*O. sativa* ssp. *japonica* is the reference, and homologous genes are colored accordingly). Grey lines represent genes with no similarity. **(B)** Circle plot showing the region in chromosome 10 where *O. barthii* has a *MIR* genes. The only circle containing a *MIR-like* gene in this scenario is *O.barthii*. Genes with the same colors are represented by vertical lines in the same color (*O. barthii* is the reference, and homologous genes are colored accordingly). Grey lines represent genes with no similarity.

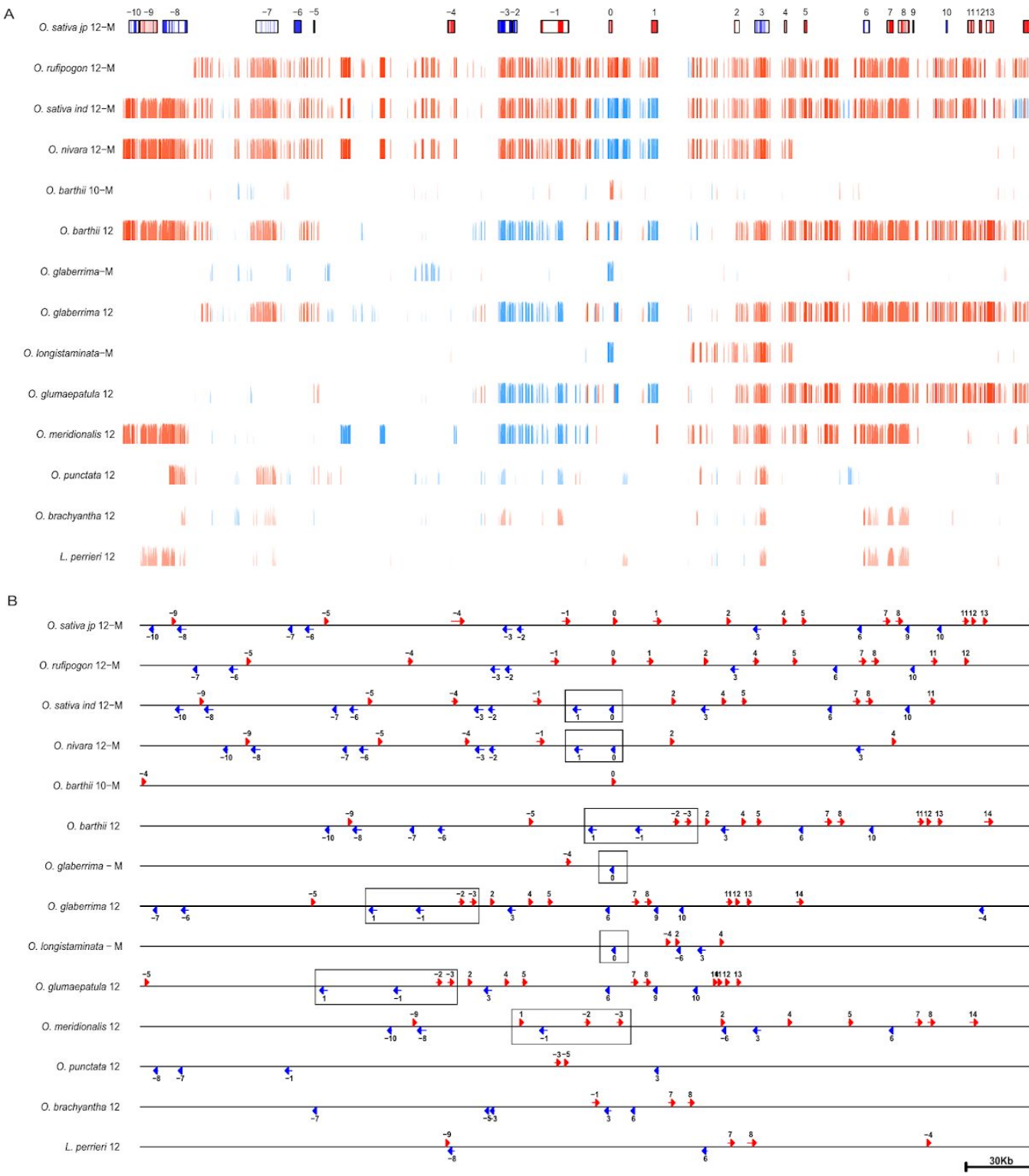


**Figure 3.Large Genomic Segments Multiple Sequence Alignment. (A)** Graphical representation generated using LASTz alignment tool. Boxes on the top row represent genes of the sequence used as

reference to guide the alignment (*O. sativa* ssp. *japonica* as a reference). The red and blue colors indicate gene orientation (red for forward; blue for reverse); blank spaces represent exons. The numbers above each box gives its relative coordinate (negative numbers are upstream the *OsaMIR* while positive numbers are downstream). The red/blue bars below on each genomic site represent an alignment of a block containing at least 300 nucleotides and sequence identity > 50% in a local alignment (bar height is proportional to the alignment percent sequence identity). Genomic segments from different *Oryza* species are identified, and segments labeled with "-M" have a MIR-like sequence. **(B)** Genomic segments in A, not aligned. The arrows indicate BLASTn hits (E value < $10^{-10}$) in the sense (red) or antisense (blue) orientation in reference to *O. sativa jp.* The queries are the genes comprised on the *O. sativa* ssp. *japonica* genomic segment and the best hits on the other ones are labeled accordingly to the query on the top row. Black rectangles indicate loci where there are inverted sequences in reference to *O. sativa*ssp *japonica*.

**Figure 4. RT-qPCR analyses of *MIR-like* gene expression under Fe deficiency in *Oryza* species.**

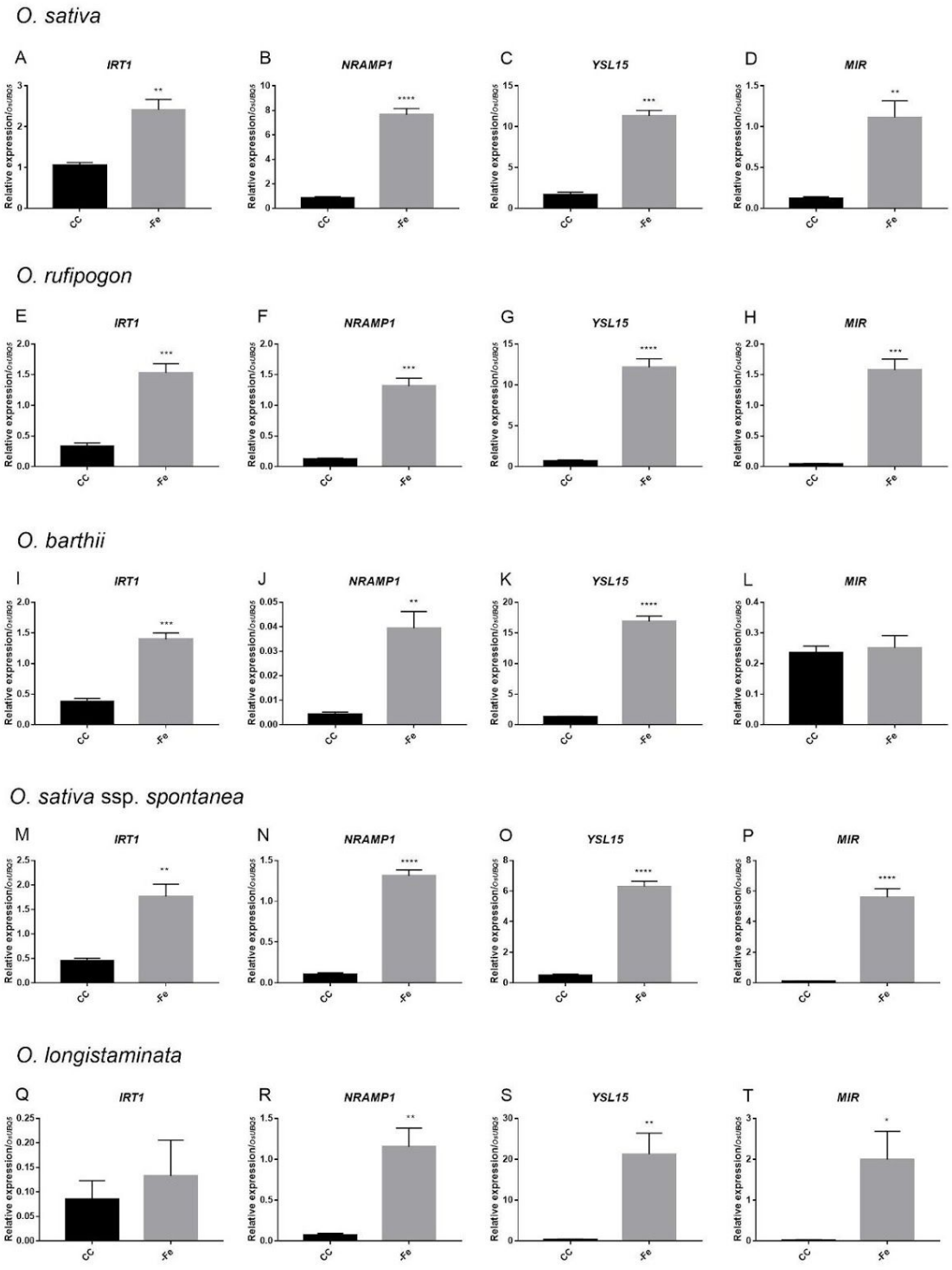

**Figure 4:** Expression analysis of Fe deficiency-related genes of(A-D) *Oryza sativa*; (E-H) *O. rufipogon*; (I-L) *O. barthii*; (M-P) *Oryza sativa* ssp. *spontanea*; (Q-T) *O. longistaminata*. Relative expression levels (RT-qPCR, relative to *OsUBQ5* expression) of selected genes (*IRT1*, *NRAMP1, YSL15* and MIR), in roots of plants submitted to control (CC) or Fe deficiency (-Fe) conditions for five days. Roots were collected from plants grown under non-aerated nutrient solution, at three-leaf stage on both conditions at the time of RNA extraction. Values are the averages of four samples (3 plants each) ± SE. Asterisks indicate statistical difference between plants grown under CC and -Fe conditions (Student T-test, *P-value < 0.05, **P-value < 0.01, ***P-value < 0.001, ****P-value < 0.0001).

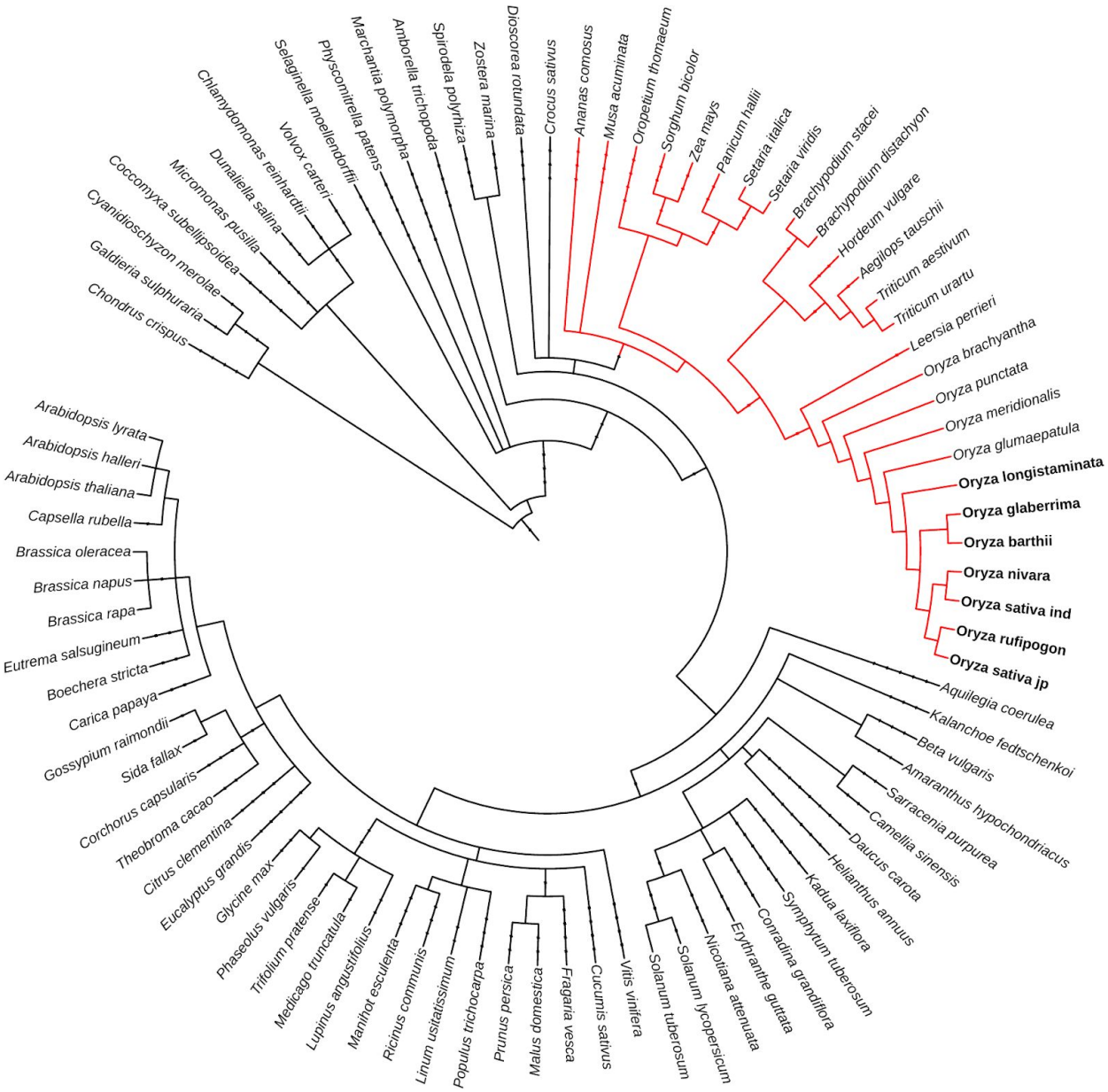


**Figure 5. Phylogeny of the plant species genomes screened using the the MIR hidden markov model sequence profile.** The red branch comprises species in which a hit was found (E-value < $10^{-6}$). The species in bold are the ones with *MIR* genes. Information on the species phylogeny relationships came from https://www.ncbi.nlm.nih.gov/taxonomy.

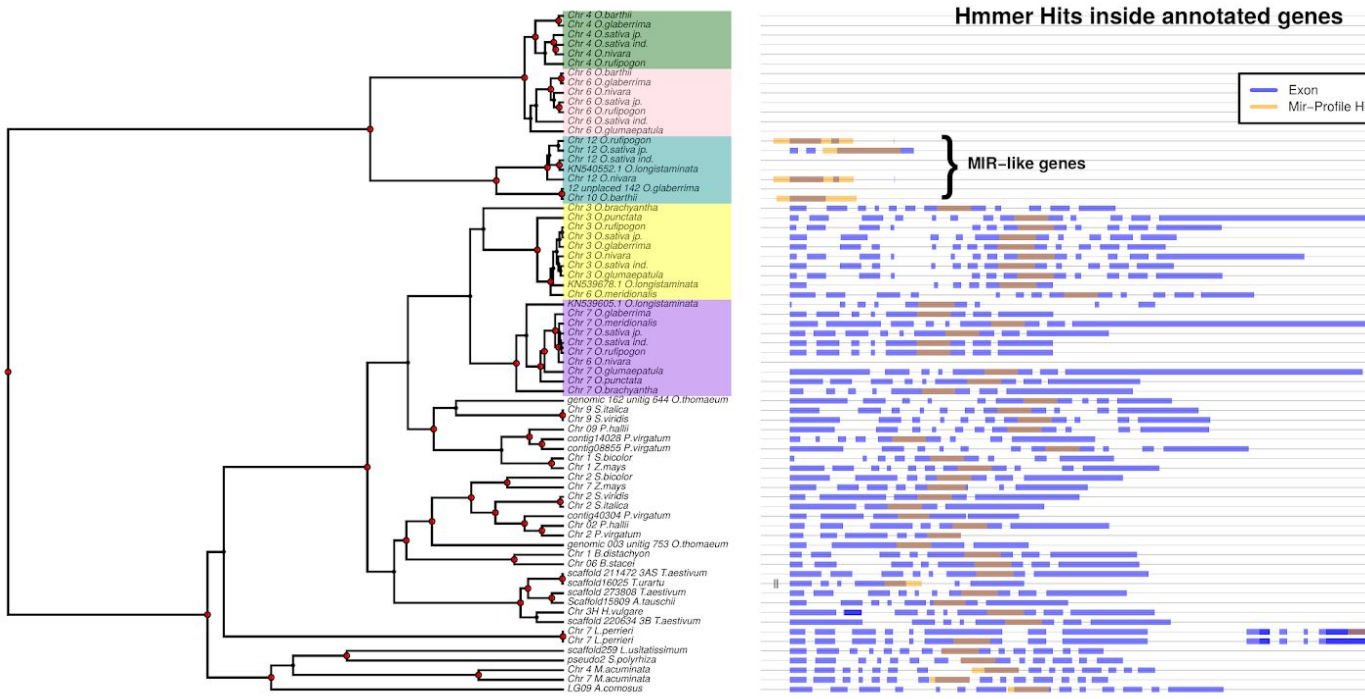


**Figure 6. Phylogenetic tree reconstruction of *MIR* and *MIR-like* genes.** (A) Reconstructed bayesian phylogenetic tree built using MIR-profile hits. Nodes marked with a red dot indicate support > 0.8. Colored branches show groups of exclusively *Oryza* species. (B) Graphical representation of the position of the MIR-profile hits inside genes (when they occur inside a gene). Exons are shown in blue, MIR-profile hits are shown in orange. Annotated *MIR* genes are highlighted.

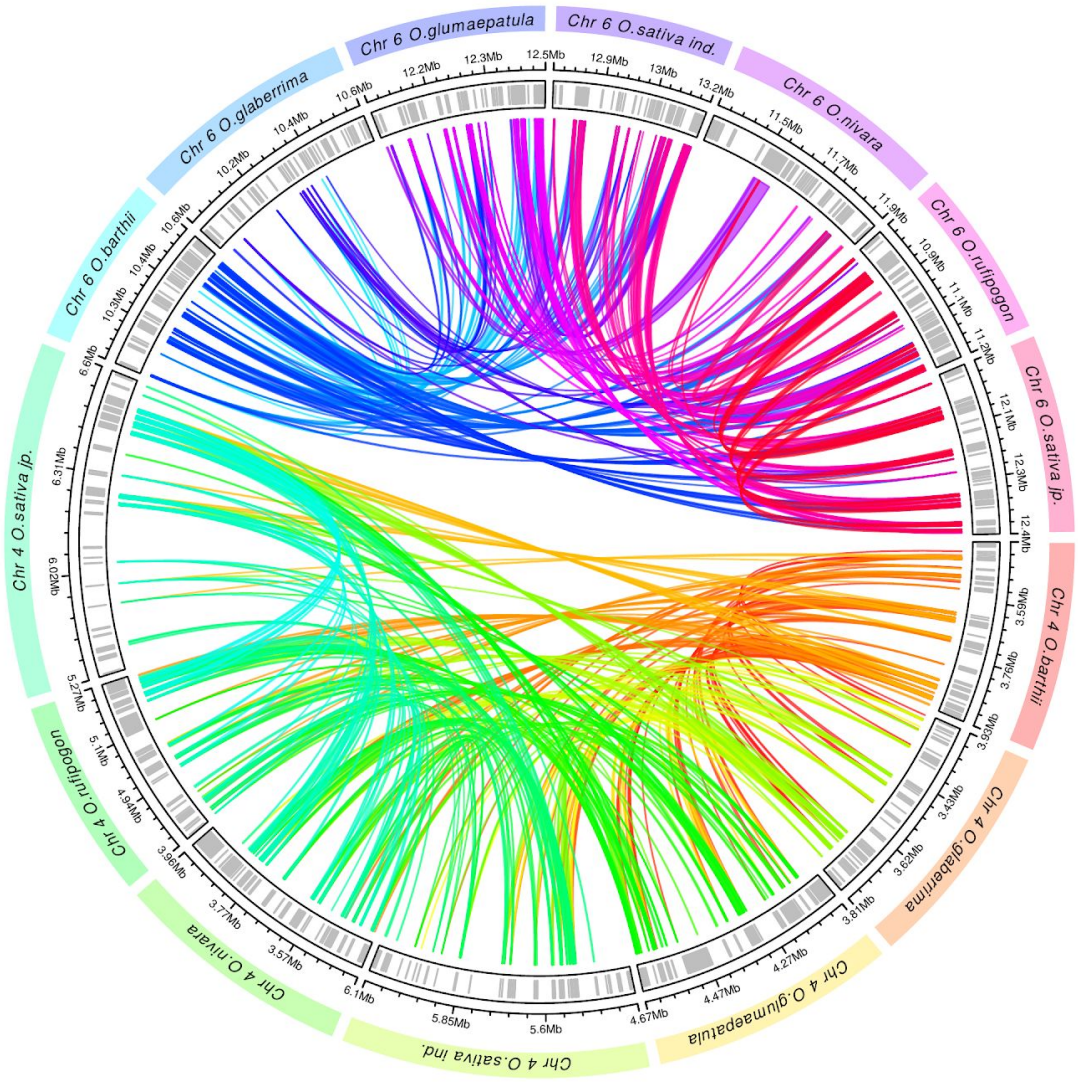


**Figure 7. Circular plot of synteny analyses from *MIR-like* sequences in chromosomes 4 and 6 of *Oryza* species.** Rectangular arcs forming the circles represent genomic segments. Grey vertical bars inside represent genes. Ribbons connect pairs of genomic loci that are syntenic (E-value < 10E-30).

## Referências


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

Arendsee ZW, Li L, Wurtele ES (2014) Coming of age: orphan genes in plants. Trends Plant Sci 19:698–708. doi: 10.1016/j.tplants.2014.07.003

Bashir K, Hanada K, Shimizu M, et al (2014) Transcriptomic analysis of rice in response to iron deficiency and excess. Rice 7:1–15. doi: 10.1186/s12284-014-0018-1

Bashir K, Ishimaru Y, Shimo H, et al (2011) The rice mitochondrial iron transporter is essential for plant growth. Nat Commun 2:322–327. doi: 10.1038/ncomms1326

Bauer P, Ling HQ, Guerinot M Lou (2007) FIT, the FER-LIKE IRON DEFICIENCY INDUCED TRANSCRIPTION FACTOR in Arabidopsis. Plant Physiol Biochem 45:260–261. doi: 10.1016/j.plaphy.2007.03.006

Briat JF, Curie C, Gaymard F (2007) Iron utilization and metabolism in plants. Curr Opin Plant Biol 10:276–282. doi: 10.1016/j.pbi.2007.04.003

Brumbarova T, Bauer P, Ivanov R (2015) Molecular mechanisms governing Arabidopsis iron uptake. Trends Plant Sci 20:124–133. doi: 10.1016/j.tplants.2014.11.004

Callaway E (2014) The birth of rice. Nature 514:S58-59

Castaings L, Caquot A, Loubet S, Curie C (2016) The high-affinity metal transporters NRAMP1 and IRT1 team up to take up iron under sufficient metal provision. Sci Rep 6:1–11. doi: 10.1038/srep37222

Colangelo EP, Guerinot M Lou (2004) The essential Basic Helix-Loop-Helix Protein FIT1 is required for the iron deficiency response. Plant Cell 16:3400–3412. doi: 10.1105/tpc.104.024315

Cui X, Lv Y, Chen M, et al (2015) Young genes out of the male: An insight from evolutionary age analysis of the pollen transcriptome. Mol Plant 8:1–11. doi: 10.1016/j.molp.2014.12.008

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7:1–8. doi: 10.1017/CBO9781139095112.007

Eddy SR (1998) Profile hidden Markov models. Bioinformatics 14:755–763

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. doi: 10.1093/nar/gkh340

Eide D, Broderius M, Fett J, Guerinot M Lou (1996) A novel iron-regulated metal transporter from plants identified by functional expression in yeast. Proc Natl Acad Sci 93:5624–5628. doi: 10.1073/pnas.93.11.5624

Elert E (2014) Rice by the numbers: A good grain. Nature 514:S50–S51. doi: 10.1038/514S50a

FAO (2019) Food and Agriculture Organization of the United Nations

Gale MD, Devos KM (1998) Comparative genetics in the grasses. Proc Natl Acad Sci 95:1971–1974

Goff SA, Ricke D, Lan T-H, et al (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). Science (80- ) 296:92–100. doi: 10.1126/science.1068037

Gómez-Galera S, Rojas E, Sudhakar D, et al (2010) Critical evaluation of strategies for mineral fortification of staple food crops. Transgenic Res 19:165–180. doi: 10.1007/s11248-009-9311-y

Goodstein DM, Shu S, Howson R, et al (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40:D1178-1186. doi: 10.1093/nar/gkr944

Graham RD, Knez M, Welch RM (2012) How much nutritional iron deficiency in humans globally is due to an underlying zinc deficiency?, 1st edn. Elsevier Inc.

Gu H, Zhu P, Jiao Y, et al (2011) PRIN: a predicted rice interactome network. BMC Bioinformatics 12:1–13. doi: 10.1186/1471-2105-12-161

Gu Z, Gu L, Eils R, et al (2014) Circlize implements and enhances circular visualization in R. Bioinformatics 30:2811–2812. doi: 10.1093/bioinformatics/btu393

Guerinot M Lou, Yi Y (1994) Iron: nutritious, noxious, and not readily available. Plant Physiol 104:815–820

Guo W-J, Li P, Ling J, Ye S-P (2007) Significant comparative characteristics between orphan and nonorphan genes in the rice ( Oryza sativa L. ) genome. Comp Funct Genomics 2007:1–7. doi: 10.1155/2007/21676

Helmy M, Sugiyama N, Tomita M, Ishihama Y (2012) The rice proteogenomics database OryzaPG-DB: development, expansion, and new features. Front Plant Sci 3:1–6. doi: 10.3389/fpls.2012.00065

Husnik F, McCutcheon JP (2018) Functional horizontal gene transfer from bacteria to eukaryotes. Nat Rev Microbiol 16:67–79. doi: 10.1038/nrmicro.2017.137

Inoue H, Kobayashi T, Nozoye T, et al (2009) Rice OsYSL15 is an iron-regulated iron(III)-deoxymugineic acid transporter expressed in the roots and is essential for iron uptake in early growth of the seedlings. J Biol Chem 284:3470–3479. doi: 10.1074/jbc.M806042200

IRGA (2017) Instituto Riograndense do Arroz

Ishimaru Y, Bashir K, Fujimoto M, et al (2009) Rice-specific mitochondrial iron-regulated gene (MIR) plays an important role in iron homeostasis. Mol Plant 2:1059–1066. doi: 10.1093/mp/ssp051

Ivanov R, Brumbarova T, Bauer P (2012) Fitting into the harsh reality: regulation of iron-deficiency responses in dicotyledonous plants. Mol Plant 5:27–42. doi: 10.1093/mp/ssr065

Jacquemin J, Bhatia D, Singh K, Wing RA (2013) The International Oryza Map Alignment Project: development of a genus-wide comparative genomics platform to help solve the 9 billion-people question. Curr Opin Plant Biol 16:147–156. doi: 10.1016/j.pbi.2013.02.014

Jakoby M, Wang HY, Reidt W, et al (2004) FRU (BHLH029) is required for induction of iron mobilization genes in Arabidopsis thaliana. FEBS Lett 577:528–534. doi: 10.1016/j.febslet.2004.10.062

Jung KH, Dardick C, Bartley LE, et al (2008) Refinement of light-responsive transcript lists using rice

oligonucleotide arrays: evaluation of gene-redundancy. PLoS One 3:1–18. doi: 10.1371/journal.pone.0003337

Kennedy G, Burlingame B (2003) Analysis of food composition data on rice from a plant genetic resources perspective. Food Chem 80:589–596. doi: 10.1016/S0308-8146(02)00507-1

Kobayashi T, Itai RN, Aung MS, et al (2012) The rice transcription factor IDEF1 directly binds to iron and other divalent metals for sensing cellular iron status. Plant J 69:81–91. doi: 10.1111/j.1365-313X.2011.04772.x

Kobayashi T, Itai RN, Ogo Y, et al (2009) The rice transcription factor IDEF1 is essential for the early response to iron deficiency, and induces vegetative expression of late embryogenesis abundant genes. Plant J 60:948–961. doi: 10.1111/j.1365-313X.2009.04015.x

Kobayashi T, Nakayama Y, Itai RN, et al (2003) Identification of novel cis-acting elements, IDE1 and IDE2, of the barley IDS2 gene promoter conferring iron-deficiency-inducible, root-specific expression in heterogeneous tobacco plants. Plant J 36:780–793. doi: 10.1046/j.1365-313X.2003.01920.x

Kobayashi T, Nishizawa NK (2012) Iron uptake, translocation, and regulation in higher plants. Annu Rev Plant Biol 63:131–152. doi: 10.1146/annurev-arplant-042811-105522

Kobayashi T, Ogo Y, Itai RN, et al (2007) The transcription factor IDEF1 regulates the response to and tolerance of iron deficiency in plants. Proc Natl Acad Sci 104:19150–19155. doi: 10.1073/pnas.0707010104

Kovach MJ, Sweeney MT, McCouch SR (2007) New insights into the history of rice domestication. Trends Genet 23:578–587. doi: 10.1016/j.tig.2007.08.012

Lee S, Chiecko JC, Kim SA, et al (2009) Disruption of OsYSL15 leads to iron inefficiency in rice plants. Plant Physiol 150:786–800. doi: 10.1104/pp.109.135418

Li G, Jain R, Chern M, et al (2017) The sequences of 1,504 mutants in the model rice variety Kitaake facilitate rapid functional genomic studies. Plant Cell 29:1218–1231. doi: 10.1105/tpc.17.00154

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. Methods 25:402–408. doi: 10.1006/meth.2001.1262

Marschner P (2012) Marschner's mineral nutrition of higher plants. Academic Press

Menguer PK, Sperotto RA, Ricachenevsky FK (2017) A walk on the wild side: Oryza species as source for rice abiotic stress tolerance. Genet Mol Biol 40:238–252. doi: 10.1590/1678-4685-gmb-2016-0093

Myers SS, Zanobetti A, Kloog I, et al (2014) Rising CO2 threatens human nutrition. Nature 510:139–142. doi: 10.1038/nature13179

Naranjo-Arcos MA, Maurer F, Meiser J, et al (2017) Dissection of iron signaling and iron accumulation by overexpression of subgroup Ib bHLH039 protein. Sci Rep 7:1–12. doi: 10.1038/s41598-017-11171-7

Nozoye T, Nagasaka S, Kobayashi T, et al (2011) Phytosiderophore efflux transporters are crucial for iron acquisition in graminaceous plants. J Biol Chem 286:5446–5454. doi: 10.1074/jbc.M110.180026

Ogo Y, Kobayashi T, Itai RN, et al (2008) A novel NAC transcription factor, IDEF2, that recognizes the iron deficiency-responsive element 2 regulates the genes involved in iron homeostasis in plants. J Biol Chem 283:13407–13417. doi: 10.1074/jbc

Pagès H, Aboyoun P, Gentleman R, DebRoy S (2019) Biostrings: Efficient manipulation of biological strings. R package version 2.52.0.

Paradis E, Schliep K (2018) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35:526–528

Pearson WR (2013) An introduction to sequence similarity ("homology") searching. In: Current Protocols in Bioinformatics. pp 1–9

Pfeiffer WH, McClafferty B (2008) Biofortification: breeding micronutrient-dense crops

Posada D (2008) jModelTest: Phylogenetic model averaging. Mol Biol Evol 25:1253–1256. doi: 10.1093/molbev/msn083

Ramakers C, Ruijter JM, Lekanne Deprez RH, Moorman AFM (2003) Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neurosci Lett 339:62–66. doi: 10.1016/S0304-3940(02)01423-4

Reuscher S, Furuta T, Bessho-Uehara K, et al (2018) Assembling the genome of the African wild rice Oryza longistaminata by exploiting synteny in closely related Oryza species. Commun Biol 162:1–10. doi: 10.1038/s42003-018-0171-y

Revell LJ (2012) phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol Evol 3:217–223. doi: 10.1111/j.2041-210X.2011.00169.x

Ricachenevsky FK, Sperotto RA (2014) There and back again, or always there? The evolution of rice combined strategy for Fe uptake. Front Plant Sci 5:1–5. doi: 10.3389/fpls.2014.00189

Ricachenevsky FK, Sperotto RA, Menguer PK, et al (2011) ZINC-INDUCED FACILITATOR-LIKE family in plants: lineage-specific expansion in monocotyledons and conserved genomic and expression features among rice (Oryza sativa) paralogs. BMC Plant Biol 11:11–20. doi: 10.1186/1471-2229-11-20

Römheld V, Nikolic M (2006) Iron. In: Barker A V., Pilbeam DJ (eds) Handbook of Plant Nutrition. Taylor & Francis Group, Boca Raton, p 662

Sang T, Ge S (2013) Understanding rice domestication and implications for cultivar improvement. Curr Opin Plant Biol 16:139–146. doi: 10.1016/j.pbi.2013.03.003

Sato Y, Namiki N, Takehisa H, et al (2013a) RiceFREND: A platform for retrieving coexpressed gene networks in rice. Nucleic Acids Res 41:1214–1221. doi: 10.1093/nar/gks1122

Sato Y, Takehisa H, Kamatsuki K, et al (2013b) RiceXPro Version 3.0: expanding the informatics resource for rice transcriptome. Nucleic Acids Res 41:1206–1213. doi: 10.1093/nar/gks1125

Schlötterer C (2015) Genes from scratch – the evolutionary fate of de novo genes. Trends Genet 31:215–219. doi: 10.1016/j.tig.2015.02.007

Schmid NB, Giehl RFH, Doll S, et al (2014) Feruloyl-CoA 6'-hydroxylase1-dependent coumarins mediate iron acquisition from alkaline substrates in arabidopsis. Plant Physiol 164:160–172. doi: 10.1104/pp.113.228544

Sperotto RA, Ricachenevsky FK, Waldow V de A, Fett JP (2012) Iron biofortification in rice: It's a long way to the top. Plant Sci 190:24–39. doi: 10.1016/j.plantsci.2012.03.004

Stein JC, Yu Y, Copetti D, et al (2018) Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus Oryza. Nat Genet 50:285–296. doi: 10.1038/s41588-018-0040-0

Tautz D, Domazet-Lošo T (2011) The evolutionary origin of orphan genes. Nat Rev Genet 12:692–702. doi: 10.1038/nrg3053

Tsai HH, Schmidt W (2017) Mobilization of iron by plant-borne coumarins. Trends Plant Sci 22:538–548. doi: 10.1016/j.tplants.2017.03.008

Uauy C, Distelfeld A, Fahima T, et al (2006) A NAC gene regulating senescence improves grain protein, Zn, and Fe content in wheat. Science (80- ) 314:1298–1301

Vaughan DA, Morishima H, Kadowaki K (2003) Diversity in the Oryza genus. Curr Opin Plant Biol 6:139–146. doi: 10.1016/S1369-5266(03)00009-8

Wang HY, Klatte M, Jakoby M, et al (2007) Iron deficiency-mediated stress regulation of four subgroup Ib BHLH genes in Arabidopsis thaliana. Planta 226:897–908. doi: 10.1007/s00425-007-0535-x

Wang M, Yu Y, Haberer G, et al (2014) The genome sequence of African rice (Oryza glaberrima) and evidence for independent domestication. Nat Genet 46:982–991. doi: 10.1038/ng.3044

Wang N, Long T, Yao W, et al (2013) Mutant resources for the functional analysis of the rice genome. Mol Plant 6:596–604. doi: 10.1093/mp/sss142

Wang W, Mauleon R, Hu Z, et al (2018) Genomic variation in 3,010 diverse accessions of Asian cultivated rice. Nature 557:43–49. doi: 10.1038/s41586-018-0063-9

Wang Y, Tang H, Debarry JD, et al (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res 40:1–14. doi: 10.1093/nar/gkr1293

Xiao W, Liu H, Li Y, et al (2009) A rice gene of de novo origin negatively regulates pathogen-induced defense response. PLoS One 4:1–12. doi: 10.1371/journal.pone.0004603

Yamaki S, Ohyanagi H, Yamasaki M, et al (2013) Development of INDEL markers to discriminate all genome types rapidly in the genus Oryza. Breed Sci 63:246–254. doi: 10.1270/jsbbs.63.246

Yu J, Hu S, Wang J, et al (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. indica). Science (80- ) 296:79–92. doi: 10.1126/science.1068037

Yuan Y, Wu H, Wang N, et al (2008) FIT interacts with AtbHLH38 and AtbHLH39 in regulating iron uptake gene expression for iron homeostasis in Arabidopsis. Cell Res 18:385–397. doi: 10.1038/cr.2008.26

Zerbino DR, Achuthan P, Akanni W, et al (2018) Ensembl 2018. Nucleic Acids Res 46:D754–D761. doi:

10.1093/nar/gkx1098

Zhang L, Ren Y, Yang T, et al (2019) Rapid evolution of protein diversity by de novo origination in Oryza. Nat Ecol Evol 3:679–690. doi: 10.1038/s41559-019-0822-5

Zhao Y, Tang L, Li Z, et al (2015) Identification and analysis of unitary loss of long-established protein-coding genes in Poaceae shows evidences for biased gene loss and putatively functional transcription of relics. BMC Evol Biol 15:1–10. doi: 10.1186/s12862-015-0345-x