UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA INTERUNIDADES DE PÓS-GRADUAÇÃO EM BIOINFORMÁTICA

WENDERSON FELIPE COSTA RODRIGUES

Ferrous iron transport via IRT1 evolved at least twice in green plants

Belo Horizonte – MG

2022

WENDERSON FELIPE COSTA RODRIGUES

Ferrous iron transport via IRT1 evolved at least twice in green plants

Dissertação apresentada ao Programa Interunidades de Pós-graduação em Bioinformática da Universidade Federal de Minas Gerais – UFMG, como requisito parcial para obtenção do título de Mestre em Bioinformática.

Orientador:

Prof. Dr. Luiz Eduardo Vieira Del Bem
Departamento de Botânica – ICB – UFMG

Belo Horizonte – MG

2022

043 Rodrigues, Wenderson Felipe Costa.
Ferrous iron transport via IRT1 evolved at least twice in green plants [manuscrito] / Wenderson Felipe Costa Rodrigues. – 2022.
40 f. : il. ; 29,5 cm.

Orientador: Prof. Dr. Luiz Eduardo Vieira Del Bem.
Dissertação (mestrado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa Interunidades de Pós-Graduação em Bioinformática.

1. Biologia Computacional. 2. Genômica. 3. Ferro. 4. Archaeplastida. I. Del Bem, Luiz Eduardo Vieira. II. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. III. Título.

CDU: 573:004

Ficha catalográfica elaborada pela bibliotecária Fabiane C M Reis – CRB 6 - 2680

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Biológicas
Programa Interunidades de Pós-Graduação em Bioinformática da UFMG

PARECER Nº	**21/2022**
PROCESSO Nº	23072.217014/2020-74

FOLHA DE APROVAÇÃO

"Ferrous iron transport via IRT1 evolved at least twice in green plants"

Wenderson Felipe Costa Rodrigues

Dissertação aprovada pela banca examinadora constituída pelos Professores: Prof. Luiz Eduardo Vieira Del Bem - Orientador (Universidade Federal de Minas Gerais), Prof. Aristóteles Góes Neto (Universidade Federal de Minas Gerais), Prof. Felipe Klein Ricachenevsky (Universidade Federal do Rio Grande do Sul) e Prof. Joni Esrom Lima (Universidade Federal de Minas Gerais)

Belo Horizonte, 22 de fevereiro de 2022.

Documento assinado eletronicamente por **Aristoteles Goes Neto**, **Coordenador(a) de curso de pós-graduação**, em 04/08/2022, às 19:50, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1657614** e o código CRC **7A52976C**.

AGRADECIMENTOS

À Universidade Federal de Minas Gerais e ao Programa Interunidade de Pós-Graduação em Bioinformática da UFMG, pela infraestrutura e apoio no desenvolvimento da dissertação.

À CAPES, pela bolsa concedida e ao Instituto Serrapilheira pelo financiamento do projeto de pesquisa. Agradeço também aos coautores e mentes por trás desse projeto Prof. *Felipe Ricachenevsky* e Prof. *Joni Lima*.

Em especial, gostaria de dedicar esse parágrafo ao Prof. Dr. *Luiz Eduardo Del-Bem*, que não poderia ser mais compreensível e atencioso. A todos os momentos onde conseguimos conversar no além da Academia, como humanos. Eu agradeço principalmente pelo simples ato de ouvir, muito obrigado mesmo.

Sou grato também pela minha família, que apoiou e financiou essa mudança repentina em nossas vidas. Um afastamento geográfico, mas com nenhum dia sem conversar por vídeo chamada. Obrigado por vibrarem quando publiquei meu primeiro artigo, acho que sem vocês esse momento seria um momento qualquer.

Durante esses tempos pandêmicos, perder a sensação de empolgação e vontade de continuar foi algo muito difícil. Passar grande parte do mestrado sem entrar pelos portões da UFMG sorrindo e pulando de alegria por estar em uma Universidade muito boa, ou até mesmo sem pegar o ônibus todo santo dia para ir para o laboratório, foi bastante difícil e tirou muito da minha felicidade em fazer parte da Academia. Mas, graças a companhia (infelizmente a distância) de muitos, eu não teria conseguido seguir nisso e chegado até aqui, escrevendo meus agradecimentos da dissertação. Eu sou muito grato a vocês, cada nome é responsável por uma memória boa. Muito obrigado por terem me feito sorrir: *João Luiz, Marciely Lira, Fenícia Brito, Aline Lima, Larissa Ávila, Thalyta Silva, Ingridy Moura, Emerson Galvão, Suanny Araújo, Thiago Mendonça, Matheus Bonaccorsi, Ângelo Rafael, Filipe Oliveira, Camila Maciel, Zanderluce Gomes, Uriel Nogueira, Jhennifer Gomes, Breno Lisboa, Paulo Jesus, Rafael Dutra, Beatriz Kfoury, Juan David, Thaís Cristina, Rosiane Santos, Edson Afonso, Daniele Costa, Wadson Rodrigues, Alison Menezes e Ashley Rodrigues.*

Por fim, esse último parágrafo eu dedico, de coração, para meu amigo *Ramon Silva*. Obrigado por ter me ajudado e instruído no início disso tudo, eu ainda espero ser seu colega de trabalho novamente. Enfim, a você *Ramon* e a todos nós eu dedico esse trabalho. Obrigado!

“Ano passado eu morri, mas esse ano eu não morro”
Belchior em Sujeito de Sorte (1976).

RESUMO

O Ferro (Fe) é um micronutriente essencial para todos os seres vivos, sendo necessário para diversas atividades metabólicas e praticamente insubstituível pois possui propriedades eletroquímicas únicas, que possibilitam ou facilitam uma série de processos bioquímicos. Possivelmente, o Fe foi um elemento essencial para o surgimento da vida na terra, uma vez que sua utilização em processos bioquímicos essenciais é conservada durante toda a evolução dos organismos vivos. Atualmente na Terra, o Fe é abundante, mas é encontrado em uma forma pouco solúvel (ferro férrico – Fe^{3+}) pelos organismos vivos como as plantas. Esse cenário muda de acordo com o ambiente, onde por exemplo, em ambientes aquáticos o Fe é mais biodisponível (como ferro ferroso – Fe^{2+}). As plantas atuais estabeleceram estratégias de absorção do Fe que são respostas a deficiência desse metal na forma mais solúvel. *A. thaliana*, por exemplo, captura o Fe através de um mecanismo que diminui o pH da região da raiz e aumenta a solubilidade do Fe^{3+}, através do bombeamento de prótons na rizosfera. O Fe férrico mais solúvel é reduzido a Fe ferroso pela FRO2 redutase e transportado para o interior da célula pelo principal transportador de ferro das plantas, o IRT1, pertencente à família proteica ZIP (ZRT/IRT - *Related Protein*), composta por transportadores de membrana de Zn, Fe e outros micronutrientes. Nesse trabalho, investigamos como os mecanismos de absorção do Fe podem ser, possivelmente, um dos fatores que levaram a terrestrialização do planeta pelas estreptófitas iniciais e como essas estratégias de captura de Fe se estabeleceram durante a evolução na linhagem das plantas (Archaeplastida). Através de análises filogenéticas, observamos que o IRT1 de *C. reinhardtii* possui distância suficiente do AtIRT1 para indicar uma possível aparição em paralelo e em diferentes momentos da história evolutiva das plantas desses dois tipos de IRT1. Além disso, nossas análises comparativas demonstram que os aminoácidos de CrIRT1 são completamente diferentes em algumas posições essenciais para que o transporte de Fe aconteça como em AtIRT1. Com os resultados apresentados nesse artigo, nosso trabalho abre caminho para diferentes análises que podem ser empregadas em cima dos dados encontrados aqui, como por exemplo, uma análise estrutural e funcional dos homólogos da família ZIP de espécies de algas que nunca foram estudados antes. A caracterização funcional e estrutural dos membros da família ZIP em Archaeplastida ajudará a entender como se estabeleceram as estratégias de absorção do Fe nas plantas e auxiliará nos estudos biotecnológicos em cultura e saúde oferecidos por essas espécies.

Palavras-chaves: captura de ferro, Archaeplastida, transportadores ZIP, genômica comparativa.

ABSTRACT

Iron (Fe) is an essential micronutrient for all living beings, being necessary for several metabolic activities and practically irreplaceable because it has unique electrochemical properties, which enable or facilitate a series of biochemical processes. Possibly, Fe was an essential element for the emergence of life on earth, since its use in essential biochemical processes is conserved throughout the evolution of living organisms. Currently on Earth, Fe is abundant, but it is found in a poorly soluble form (ferric iron – Fe^{3+}) by living organisms such as plants. This scenario changes according to the environment, where, for example, in aquatic environments Fe is more bioavailable (as ferrous iron – Fe^{2+}). Current plants have established Fe absorption strategies that are responses to the deficiency of this metal in the most soluble form. *A. thaliana*, for example, captures Fe through a mechanism that lowers the pH of the root region and increases the solubility of Fe^{3+}, by pumping protons into the rhizosphere. The most soluble ferric Fe is reduced to ferrous Fe by FRO2 reductase and transported into the cell by the main iron transporter in plants, IRT1, belonging to the ZIP protein family (ZRT/IRT - Related Protein), composed of membrane transporters of Zn, Fe and other micronutrients. In this work, we investigate how Fe uptake mechanisms could possibly be one of the factors that led to terrestrialization of the planet by early streptophytes and how these Fe capture strategies were established during evolution in the plant lineage (Archaeplastida). Through phylogenetic analyses, we observed that the IRT1 of *C. reinhardtii* has sufficient distance from AtIRT1 to indicate a possible appearance in parallel and at different times in the evolutionary history of plants of these two types of IRT1. Furthermore, our comparative analyzes demonstrate that CrIRT1 amino acids are completely different in some essential positions for Fe transport to occur as in AtIRT1. With the results presented in this article, our work opens the way for different analyzes that can be used on top of the data found here, such as a structural and functional analysis of homologs of the ZIP family of algae species that have never been studied before. The functional and structural characterization of members of the ZIP family in Archaeplastida will help to understand how Fe absorption strategies were established in plants and will help in biotechnological studies in culture and health offered by these species.

Keywords: iron uptake, Archaeplastida, ZIP transporters, comparative genomics.

LISTA DE FIGURAS E GRÁFICOS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

DNA	Ácido desoxirribonucleico
pH	Potencial Hidrogeniônico
Fe	Ferro
Fe^{3+}	Ferro férrico
Fe^{2+}	Ferro ferroso
AHA2	H-ATPase 2
FRO2	*Ferric Reduction Oxidase* 2
IRT	*Iron-Regulated Transporter* 1
F6'H1	*Feruloyl* CoA *ortho-hydroxylase* 1
ABCG37	ABC *transporter* G *family member* 37
SAM	S-adenosil-metionina
MA	*Mugineic acid*
DMA	2-*deoxy-mugineic acid*
TOM1	*Transporter of Mugineic acid family phytosiderophores*
YSL	*Yellow Stripe-Like*
FRE1	*Ferric-chelate reductase/oxidoreductase*
FOX1	*Multicopper ferroxidase*
FTR1	*Iron permease*
ZIP	ZRT/IRT *Related Protein*
Zn	Zinco
Mn	Manganês
Cu	Cobre
Cd	Cádmio
ZRT	*Zinc-Regulated Transporter*
HMM	*Hidden Markov Model*
AIC	*Akaike Information Criterion*
NNI	*Nearest Neighbor Interchanges*
SSN	Rede de Similaridade de Sequências
WGD	*Whole Genome Duplication*

SUMÁRIO

1. INTRODUÇÃO GERAL

O Ferro (Fe) é um nutriente mineral fundamental para o crescimento, desenvolvimento e manutenção de quase todos os organismos vivos, exceto por algumas bactérias (CRICHTON, 2016; POSEY; GHERARDINI, 2000). Da síntese de hemoglobina em humanos (AGGETT, 2012) a importante cofator para várias enzimas que estão envolvidas na fotossíntese de plantas (KIM; GUERINOT, 2007; ROUT; SAHOO, 2015), o Fe é essencial pois participa de uma ampla variedade de processos metabólicos, incluindo transporte de oxigênio, síntese de ácido desoxirribonucleico (DNA) e transporte de elétrons (ABBASPOUR; HURRELL; KELISHADI, 2014).

Apesar da grande importância, o excesso Fe é altamente tóxico para diversos organismos. Em plantas, o Fe em grandes quantidades interrompe o desenvolvimento morfológico e fisiológico do indivíduo, causando danos, por exemplo, a organelas celulares como os cloroplastos que retarda o metabolismo de carbono pela fotossíntese (ONYANGO et al., 2019; ZAHRA et al., 2021).

A necessidade do Fe em processos biológicos tão importantes de diferentes grupos de organismos é um mecanismo evolutivo bem caracterizado e conservado. O desempenho desse papel pelo Fe sugere que sua utilização é ancestral, indicando sua ligação com o início da vida na Terra e podendo refletir nas condições ambientais desse período (LILL; FREIBERT, 2020; WADE et al., 2021).

Atualmente, em um cenário onde o oxigênio é presente, o Fe é um dos elementos mais abundante no planeta em sua forma oxidada (Fe^{3+}). Acredita-se que o aumento da fotossíntese e do oxigênio, consequência da evolução das plantas terrestres, resultou em mudanças geológicas que acumulou o Fe^{3+} no ambiente (KROH; PILON, 2020; LENTON et al., 2016). Essa forma do Fe não é prontamente biodisponível para a absorção pelas plantas, tendo que ser reduzido a Fe^{2+} ainda na rizosfera em processos muito bem estabelecidos atualmente por esses organismos (CONNORTON; BALK; RODRÍGUEZ-CELMA, 2017; FREY; REED, 2012). Em ambientes aquáticos de água doce, o Fe^{2+} é mais abundante que em ambientes marinhos, sendo um dos principais fatores limitantes para a produção fotossintética nesses ambientes (BLOMQVIST; GUNNARS; ELMGREN, 2004; FREY; REED, 2012).

Ao longo do tempo evolutivo, os organismos adquiriram mecanismos específicos de absorção de Fe que podem responder à essa deficiência de Fe solúvel no ambiente. Em plantas terrestres, duas estratégias de captura de ferro são utilizadas por diferentes espécies do grupo Embryophyta. Estratégia I, realizada por plantas como *Arabidopsis thaliana* e tomate, e

Estratégia II, usada por plantas como arroz, cevada, milho e gramíneas (CONNORTON; BALK; RODRÍGUEZ-CELMA, 2017; RÖMHELD; MARSCHNER, 1986).

Na Estratégia I, como resposta à deficiência de Fe, as células radiculares bombeiam prótons para o solo através da H-ATPase 2 (AHA2), diminuindo o pH da área circundante e aumentando a solubilidade do Fe^{3+}. Depois disso, o Fe^{3+} é reduzido a Fe^{2+} via FRO2 redutase (Ferric Reduction Oxidase 2) e transportado para dentro das células pelo IRT1 (Iron-Regulated Transporter 1). A etapa de acidificação-redução também pode ser realizada utilizando compostos fenólicos, flavonóides, flavinas e cumarinas, sendo este último sintetizado pelo F6'H1 (Feruloyl CoA ortho-hydroxylase 1) e transportado para fora da célula em direção ao Fe^{3+} pelo ABCG37 (ABC transporter G family member 37). Nas plantas que usam a Estratégia II, as células da raiz secretam fitossideróforos no solo para aumentar a solubilidade do Fe^{3+} por quelação. Os fitosideróforos são sintetizados a partir da S-adenosil-metionina (SAM), que gera compostos da família do ácido mugineico (MA), como o ácido 2'-desoximugiênico (DMA), que é transportado para o solo via TOM1 (Transporter of Mugineic acid family phytosiderophores). O fitossideróforo forma um complexo com Fe^{3+} que é então transportado para dentro da célula com a ajuda de YSL (Yellow Stripe-Like). Espécies do gênero arroz (Oryza) são capazes de combinar ambas as estratégias como adaptação a áreas alagadas, onde a falta de O2 faz com que o Fe seja reduzido a Fe^{2+}, sendo facilmente absorvido através de homólogos de IRT1 (CONNORTON; BALK; RODRÍGUEZ-CELMA, 2017; WAIRICH et al., 2019).

Em algas verdes, como *Chlamydomonas reinhardtii* (Chlorophyta), a absorção de ferro ocorre pela redução do Fe livre através da FRE1 redutase, seguida pela captação e reoxidação do Fe por uma ferroxidase (FOX1). O Fe é então transportado para dentro da célula via permease FTR1. Chlamydomonas também captura Fe^{2+} disponível em seu ambiente de água doce através de duas proteínas ZIP chamadas IRT1/IRT2. O complexo FOX1-FTR1 é um sistema de aquisição de Fe de alta afinidade que adquire Fe em concentrações muito baixas (ALLEN et al., 2007; BLABY-HAAS; MERCHANT, 2012; GLAESENER; MERCHANT; BLABY-HAAS, 2013; KROH; PILON, 2020).

O IRT1, tanto de *C. reinhardtii*, quanto de *O. sativa* e *A. thaliana* é o membro mais conhecido da grande família proteica ZIP (ZRT/IRT - Related Protein) de transportadores de membrana, identificados a partir de um cDNA de Arabidopsis capaz de complementar a função de captação de Fe em mutantes de Saccharomyces cerevisiae (EIDE et al., 1996). A família Zip transporta uma ampla gama de micronutrientes além do Fe, como Zinco (Zn) e Manganês (Mn). De fato, além do IRT1, os membros que deram o nome a esta família são os transportadores de Zn ZRT1 e ZRT2 (EIDE et al., 1996; LEE et al., 2021; MILNER et al., 2013; PITA-BARBOSA et al., 2019). Além desses elementos, estudos recentes demonstram que membros da família ZIP de arroz e *A. thaliana* respondem ao estresse de Cádmio (Cd) e acumulam esse metal, implicando ainda mais seus papéis na absorção de Cd em plantas (ZHENG; CHEN; LI, 2018). Apesar disso, pouco se sabe quanto a função desses membros da família ZIP em diferentes grupos de organismos, como as algas.

Como já mencionado, os ambientes de água doce e terrestre têm distribuição e biodisponibilidade diferentes de Fe. Assim, a evolução da captura de Fe da crosta terrestre e dos solos primitivos pode ter sido um dos passos cruciais na terrestreização das primeiras estreptófitas unicelulares que finalmente deram origem às plantas terrestres. É interessante pensar em como e quando os mecanismos de captura de Fe de plantas verdes existentes evoluíram para lidar com a deficiência de ferro biodisponível. Com base nessa questão, este estudo utilizou uma abordagem de genômica comparativa para entender melhor a origem e diversificação das proteínas IRT1/ZIP na linhagem Archaeplastida, composta pelos grupos de algas Glaucophyta (algas de água doce), pelas algas vermelhas (Rodophyta) e pela divisão Viridiplantae, que incluem as algas verdes Chlorophyta e Charophyta e as plantas terrestres (Embryophyta). As algas verdes Charophyta são mais intimamente relacionadas com as plantas terrestres dentro de Viridiplantae e atualmente os dois grupos compõem um grupo conhecido como Streptophyta.

Nossos resultados sugerem que os mecanismos baseados em IRT1 de captura de Fe^{2+} em plantas verdes evoluíram pelo menos duas vezes de ramos notavelmente distintos da árvore filogenética da família ZIP. Os homólogos de IRT1/ZIP responsáveis pelo transporte de Fe^{2+} na Estratégia I de plantas terrestres e aqueles envolvidos na absorção de Fe^{2+} da água doce em clorófitas provavelmente evoluíram o transporte de Fe independentemente dos transportadores ancestrais de Zn. Apesar disso, faz-se necessário uma série de estudos quanto a captura de Ferro em organismos como as algas, já que estudos nessa área são poucos, sendo restrito a espécies modelo.

2. OBJETIVOS

2.1. OBJETIVO GERAL

Lançar luz sobre a evolução da proteína de membrana IRT1/ZIP, principal proteína transportadora de Ferro, na linhagem Archaeplastida.

2.2. OBJETIVOS ESPECÍFICOS

- Identificar proteínas homologas ZIP em diferentes grupos da linhagem Archaeplastida;
- Construir redes e árvores filogenéticas que evidenciem os caminhos evolutivas das proteínas homologas ZIP na linhagem Archaeplastida;
- Identificar possíveis relações nos mecanismos de captura de Ferro de plantas atuais e a terrestrialização das plantas pelas estreptófitas iniciais.

3. ARTIGO RESULTANTE

O artigo “***Ferrous iron transport via IRT1 evolved at least twice in green plants***”, resultante das análises realizadas nesse trabalho, foi preparado para submissão à revista Genomics (Elsevier), que engloba áreas da genômica comparativa e evolutiva na qual esse trabalho se encontra.

Ferrous iron uptake via IRT1/ZIP evolved at least twice in green plants

Wenderson Felipe Costa Rodrigues[1,2], Joni Lima[3], Felipe Ricachenevsky[4], Luiz-Eduardo Del-Bem[2*]

[1] Graduate Program in Bioinformatics, Institute of Biological Sciences (ICB), Federal University of Minas Gerais (UFMG)
[2] Del-Bem Lab, Department of Botany, Institute of Biological Sciences (ICB), Federal University of Minas Gerais (UFMG)
[3] Department of Botany, Institute of Biological Sciences (ICB), Federal University of Minas Gerais (UFMG)
[4] Department of Botany, Institute of Biosciences (IB), Federal University of Rio Grande do Sul (UFRGS)

[*] *Correspondence: Luiz-Eduardo Del-Bem: delbem@ufmg.com*

Abstract

Iron (Fe) is an essential micronutrient for all living beings, being practically irreplaceable because of its unique electrochemical properties that enable or facilitate a series of biochemical processes, including photosynthesis. Although Fe is abundant on Earth, it is generally found in the poorly soluble form Fe^{3+}. Extant plants have established Fe absorption strategies that are adapted to the deficiency of this metal in the soluble form Fe^{2+}. *Arabidopsis thaliana*, for example, captures Fe through a mechanism that lowers the pH through proton pumping to the rhizosphere that increases Fe^{3+}solubility, which is then reduced by a reductase and transported into the cell by the ZIP (ZRT/IRT - Related Protein) family protein IRT1. ZIP proteins are transmembrane transporters of a variety of metal such as Fe, Zn, Mn and Cd. In this work, we investigate the evolution of functional homologs of the IRT1 in the supergroup of photosynthetic eukaryotes Archaeplastida (Viridiplantae + Rhodophyta + Glaucophyta) using a dataset of 35 high-quality genomes of diverse lineages. Our analyses suggest that Fe is acquired through deeply divergent ZIP proteins in land plants and chlorophyte green algae, indicating that Fe^{2+} uptake evolved at least twice independently during green plant evolution. Sequence analyses indicate that the archetypical IRT proteins from angiosperms likely emerged in streptophyte algae before the origin of land plants and might be an important player in green plant terrestrialization, a process that involved the evolution of Fe acquisition in terrestrial subaerial settings.

Keywords: iron uptake, Archaeplastida, ZIP transporters, comparative genomics.

Introduction

Iron (Fe) is a fundamental element for virtually all cellular organisms. In plants, iron deficiency results in serious physiological problems, such as failures in photosynthetic electron transport and in sulfur and nitrogen metabolism, which impair growth [1]. Fe is not readily bioavailable, besides being the second most abundant metal in Earth's crust (after aluminum). Soil Fe is predominantly found as Fe^{3+} (ferric iron) with oxidants, which is poorly bioavailable to plants. In opposition, Fe^{2+} (ferrous iron) can be absorbed by plant roots without being reduced and without the aid of phytosiderophores. In freshwater environments, Fe can be found dissolved in greater abundance than in marine environments, where its limitation impacts photosynthetic productivity [2,3].

Throughout evolutionary time, organisms acquired specific mechanisms for Fe uptake that can sense and respond to soluble Fe deficiency in the environment. In terrestrial plants, two iron capture strategies are used by different species of land plants (Embryophyta). Strategy I, carried out by plants like *Arabidopsis thaliana* and tomato, and Strategy II, used by plants like rice, barley, maize and grasses (Poaceae) that live in flooded freshwater environments [4,5]. In Strategy I, as a response to Fe deficiency, root cells pump protons to soil through H-ATPase 2 (AHA2), decreasing the pH of the surrounding area and increasing the solubility of Fe^{3+}. After that, Fe^{3+} is reduced by Fe^{2+} via FRO2 reductase (Ferric Reduction Oxidase 2) and transported into cells by IRT1 (Iron-Regulated Transporter 1) (Figure 1A). The acidification-reduction step can also be performed using phenolic compounds, flavonoids, flavins, and coumarins, the latter being synthesized by F6'H1 (Feruloyl CoA ortho-hydroxylase 1) and carried outside the cell towards Fe^{3+} by ABCG37 (ABC transporter G family member 37). In plants that use Strategy II, root cells secrete phytosiderophores into the soil to increase the solubility of Fe^{3+} by chelation (Figure 1B). Phytosiderophores are synthesized from S-adenosyl-methionine (SAM), which generates mugineic acid (MA) family compounds such as 2’-deoxymugineic acid (DMA), which is transported into the soil via TOM1 (Transporter of Mugineic acid family phytosiderophores). The phytosiderophore forms a complex with Fe^{3+} that is then transported into the cell with the help of YSL (Yellow Stripe-Like). Species of the rice genus (Oryza) are able to combine both strategies as an adaptation to flooded areas, where the lack of O_2 causes Fe to be reduced to Fe^{2+}, being easily absorbed through homologues of IRT1 [5,6].

In green algae, such as *Chlamydomonas reinhardtii* (Chlorophyta), iron absorption occurs through the reduction of free Fe through FRE1 reductase, followed by the uptake and reoxidation of Fe by a ferroxidase (FOX1). Fe is then transported into the cell via FTR1 permease. *Chlamydomonas* also capture Fe^{2+} available in its freshwater environment through two ZIP proteins called IRT1 and IRT2 (Figure 1C). The FOX1–FTR1 complex is a high-affinity Fe acquisition system that acquires Fe under very low concentrations [7–10].

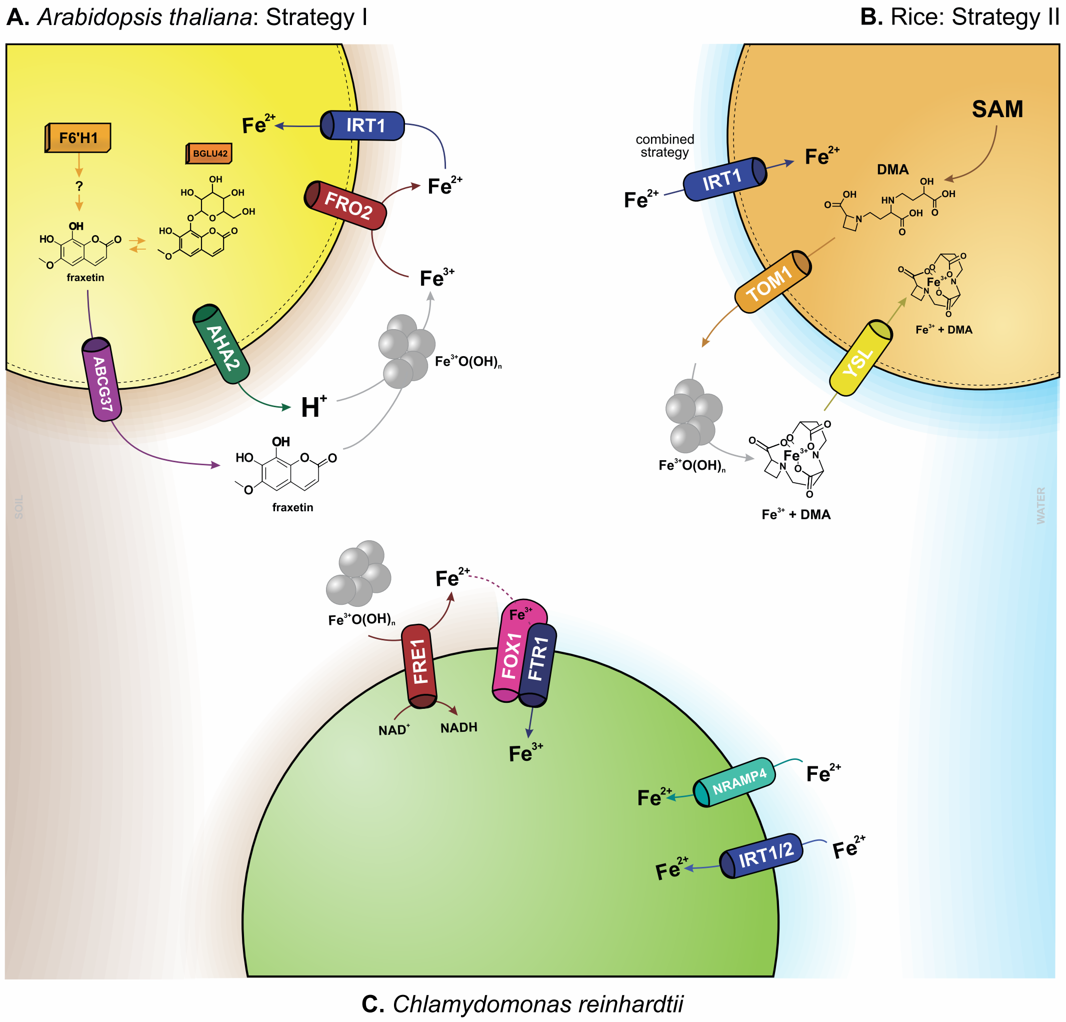


Fig. 1 Iron uptake strategies in (A) *Arabidopsis thaliana* (Strategy I), (B) Rice (Strategy II) and (C*) Chlamydomonas reinhardtii*. Cylindrical shapes indicate membrane proteins involved in different steps of iron capture. Blue shapes represent Fe transporters and red shapes represent reductases. Boxes inside cells represent enzymes involved in the biosynthesis of phytosiderophores. (B) and (C) represent combined strategies, where Fe^{2+} from freshwater environments can be directly transported via IRT1 homologs.

IRT1 is first described and most well-known member of the large ZIP (ZRT/IRT - Related Protein) family of membrane transporters, identified from an *Arabidopsis* cDNA capable of complementing the iron uptake function in *S. cerevisiae* mutants [11]. The Zip family transports a broad range of micronutrients, such as Zinc (Zn), Manganese (Mn) and Copper (Cu). In fact, in addition to IRT1, the members that gave this family its name are the Zn transporters ZRT1 and ZRT2 [11,12]. There are not many studies on the other 12 *Arabidopsis*

ZIP's [12–15], despite their possible importance for different groups of organisms in the absorption of micronutrients such as Fe and Zn.

The great physiological importance that Fe has for plants and all living organisms is possibly explained by the role of iron-based catalysts in early life before the origin of extant domains [16]. Fe bioavailability decrease due to the Great Oxidation Event likely shaped the evolution of life afterwards [17].

Freshwater and terrestrial environments have different distribution and bioavailability of Fe. Thus, the evolution of Fe capture from Earth's crust and early soils may have been one of the crucial steps in the terrestrialization of early unicellular streptophytes that ultimately gave rise to land plants. It is interesting to think about how and when the Fe capture mechanisms of extant green plants evolved to deal with iron bioavailable deficiency. Based on this question, this study used a comparative genomics approach to better understand the origin and diversification of IRT1/ZIP proteins in the Archaeplastida lineage. Our results suggest that IRT1-based mechanisms of Fe^{2+} capture in green plants evolved at least twice from remarkably distinct branches of ZIP family tree. IRT1/ZIP homologs responsible for Fe^{2+} transport in the Strategy I of land plants and those involved in the uptake of Fe^{2+} from water in chlorophytes likely evolved Fe transport independently from ancestral Zn transporters.

Methods

Datasets

The HMM (hidden Markov model) profile for Zip family (PF02535) was obtained from Pfam database (https://pfam.xfam.org/) [18]. The complete predicted proteomes of the analyzed organisms were collected from the databases detailed in Table 1.

Table 1. List of complete predicted proteomes databases used in this study.

Species	Link
Arabidopsis thaliana	https://phytozome-next.jgi.doe.gov/info/Athaliana_TAIR10
Marchantia polymorpha	https://phytozome-next.jgi.doe.gov/info/Mpolymorpha_v3_1
Oryza sativa	https://phytozome-next.jgi.doe.gov/info/Osativa_v7_0
Physcomitrella patens	https://phytozome-next.jgi.doe.gov/info/Ppatens_v3_3
Populus trichocarpa	https://phytozome-next.jgi.doe.gov/info/Ptrichocarpa_v4_1
Sorghum bicolor	https://phytozome-next.jgi.doe.gov/info/Sbicolor_v3_1_1
Selaginella moellendorffii	https://phytozome-next.jgi.doe.gov/info/Smoellendorffii_v1_0
Zea mays	https://phytozome-next.jgi.doe.gov/info/Zmays_RefGen_V4
Anthoceros agrestis [Oxford]	https://www.hornworts.uzh.ch/static/download/a_agr_oxford.zip

Sphagnum fallax	https://phytozome-next.jgi.doe.gov/info/Sfallax_v1_1
Anthoceros angustus	Dryad Data -- The hornwort genome and early land plant evolution (datadryad.org)
Chara braunii	https://bioinformatics.psb.ugent.be/gdb/Chara_braunii/chbra_iso_noTE_23546_pep.fasta
Klebsormidium nitens	http://www.plantmorphogenesis.bio.titech.ac.jp/~algae_genome_project/klebsormidium/kf_download/160614_klebsormidium_v1.1_AA.fasta
Mesotaenium endlicherianum	https://ndownloader.figshare.com/files/17819138
Chlorokybus atmophyticus	ftp://ftp.cngb.org/pub/CNSA/data1/CNP0000228/CNS0021447/CNA0002353/scaffold_Chlorokybus_atmophyticus.pep.gz
Mesostigma viride	ftp://ftp.cngb.org/pub/CNSA/data1/CNP0000228/CNS0021438/CNA0002352/scaffold_Mesostigma_viride.pep.gz
Spirogloea muscicola	https://ndownloader.figshare.com/files/17819147
Auxenochlorella protothecoides	https://genome.jgi.doe.gov/portal/Auxeprot1/download/Auxeprot1_GeneCatalog_proteins_20170909.aa.fasta.gz
Bathycoccus prasinos	https://genome.jgi.doe.gov/portal/Batpra1/download/Batpra1_GeneCatalog_proteins_20180426.aa.fasta.gz
Chlorella variabilis	https://genome.jgi.doe.gov/portal/ChlNC64A_1/download/Chlorella_NC64A.all_proteins.fasta.gz
Coccomyxa subellipsoidea	https://genome.jgi.doe.gov/portal/Coc_C169_1/download/Coccomyxa_C169_v2_all_proteins.fasta.gz
Chlamydomonas reinhardtii	https://genome.jgi.doe.gov/portal/pages/dynamicOrganismDownload.jsf?organism=Phytozome#
Dunaliella salina	https://genome.jgi.doe.gov/portal/Dunsal1/download/Dunsal1_GeneCatalog_proteins_20180508.aa.fasta.gz
Gonium pectorale	https://genome.jgi.doe.gov/portal/Gonpec1/download/Gonpec1_GeneCatalog_proteins_20180501.aa.fasta.gz
Micromonas commoda	https://genome.jgi.doe.gov/portal/MicpuN3v2/download/MicpuN3v2_GeneCatalog_proteins_20160404.aa.fasta.gz
Monoraphidium neglectum	https://genome.jgi.doe.gov/portal/Monneg1/download/Monneg1_GeneCatalog_proteins_20170920.aa.fasta.gz
Micromonas pusilla	https://genome.jgi.doe.gov/portal/MicpuC2/download/MicromonasCCMP1545.allModels.aa.fasta.gz
Ostreococcus lucimarinus	https://phytozome-next.jgi.doe.gov/info/Olucimarinus_v2_0
Ostreococcus tauri	https://genome.jgi.doe.gov/portal/Ostta4221_3/download/Ostta4221_3_GeneCatalog_proteins_20161028.aa.fasta.gz
Picochlorum soloecismus	https://genome.jgi.doe.gov/portal/Picsp_1/download/Picsp_1_GeneCatalog_proteins_20170909.aa.fasta.gz
Volvox carteri	https://phytozome-next.jgi.doe.gov/info/Vcarteri_v2_1
Chondrus crispus	https://greenhouse.lanl.gov/greenhouse/organisms/
Cyanidioschyzon merolae	https://greenhouse.lanl.gov/greenhouse/organisms/
Galdieria sulphuraria	https://greenhouse.lanl.gov/greenhouse/organisms/
Cyanophora paradoxa	https://greenhouse.lanl.gov/greenhouse/organisms/

Identification of ZIP family homologs

We performed a HMMER v3.2.1 [19] search with default e-value to identify Zip family proteins in Archaeplastida. We kept only the most representative protein isoform encoded by each unique gene, defined as the most complete isoform. We tried to discard, through an in-house script written in Perl, partial sequences or pseudogenes by eliminating sequences shorter than 138aa, which represents 40% of the well characterized IRT1 protein of *Arabidopsis thaliana*. Also, we removed 100% redundant sequences with *skipredundant* from EMBOSS v6.6.0 [20] package.

All subsequent analyses were performed with the filtered sequences and results were plotted in different types of graphs with RStudio [21]. The relative genomic frequency of IRT/ZIP genes per genome was calculated by dividing the number of homologs found in each genome by the total number protein-coding genes. P-values were calculated for each contrast using the Wilcoxon statistical test in RStudio and p-values > 0.05 were considered not significant.

Sequence alignment

Protein sequences were aligned using MAFFT (https://mafft.cbrc.jp/alignment/software/) [22] with the following parameters *--thread 10 --reorder --leavegappyregion --maxiterate 1000 --retree 1 --localpair*. The alignment output was exported in Phylip format (*--phylipout*) for subsequent phylogenetic analyses. We also aligned our sequences using MUSCLE in MEGA 11 software [23] to analyze the conservation of amino acids between proteins. This analysis was performed using sequence logos generated by WebLogo 3 (http://weblogo.threeplusone.com/) [24], with standard parameters, except for the y-axis unit which was relative frequency instead of bits.

Phylogenetic analysis

The phylogenetic analyses were performed using the maximum-likelihood method implemented in IQTree v1.6.12 [25]. The amino acid substitution model was selected using the SMS method with Akaike Information Criterion (AIC) as the selection criterion. The Best-fit model chosen according to AIC was WAG+R10. The initial tree was calculated with BIONJ and improved by the tree topology search method NNI (Nearest Neighbor Interchanges). The branch support values were calculated with the likelihood-based method aLRT SH-like.

Sequence Similarity Network (SSN)

To corroborate our phylogenetic results, we performed a network analysis based on sequence similarity. We generated these networks with a threshold equal to 40 in the Enzymatic Similarity Tool (EFI-EST tool) (https://efi.igb.illinois.edu/efi-est/) [26]. After that, we visualize and edit the network with Cytoscape v3.8.2 software [27], keeping only nodes and edges between 50-100% identity as a parameter.

Results & Discussion

Distribution of ZIP family homologs in Archaeplastida

In our analyses looking for members of the ZIP family in Archaeplastida lineage groups, we found 392 homologs proteins (Supplemental Table S1), each corresponding to a unique genomic *locus*, distributed across the 35 species in our dataset (Table 1). We detected ZIP homologs not only in all lineages of green plants but also in the glaucophyte and the rhodophytes, which may indicate that genes encoding IRT1/ZIP proteins in land plants were likely inherited vertically from the last common ancestor of Archaeplastida (Figure 2A). Our data suggest that archaeplastidian algae typically have less than 10 ZIP genes (Figure 2A). We separated the algae by habitat (marine, freshwater or terrestrial) and no difference could be observed between the number of ZIP genes per genome in relation to different habitats (Supplemental Figure 1A). In the aqueous environment, algae adopt specific approaches to iron absorption, such as those described in *C. reinhardtii* (Figure 1C). The solubility of Fe^{2+} is generally temperature-, pH- and oxygen-dependent. In the ocean, when the water is warm and oxygenated, Fe^{2+} is stable for only a few seconds, and can last for hours or days when the temperature and oxygenation are reduced. Much of the ocean is well oxygenated and Fe^{2+} is normally oxidized to Fe^{3+}, which is bound to organic ligands in the environment [28,29]. Understanding the availability of Fe in water is important to understand how the Fe capture mechanism in the first algae may have developed initially.

Among archaeplastidians, land plants generally have a significantly greater number of ZIP genes (Figure 2A). However, we observed a great variability in the ZIP gene repertoire between land plants species, ranging from ~10 to ~40 genes per genome (Figure 2A). Little difference in the distribution of the number of ZIP homologs per genome was detected between bryophytes and tracheophytes (Figure 2A). It indicates that different lineages of bryophytes and tracheophytes might have independently experience increased fitness provided by the fixation of additional copies of ZIP genes. A broader set of ZIP genes might be related to the evolution of more elaborated metal acquisition and internal transport due to compartmentalization of gene expression and metal specialization of transporters.

The copy number of a gene family might be influenced by whole genome duplication (WGD), making it difficult to compare deeply divergent organisms as green plants, red algae and glaucophytes, that have experienced a variety of independent rounds of WGD. Thus, we decided to compare the relative genomic frequency of ZIP genes, calculated as the fraction of the total protein-coding gene set occupied by them. Results indicate that archaeplastidian lineages dedicate a similar fraction of their gene set to encode ZIP proteins (Figure 2B). Also, no significant difference was detected regarding algal habitat (Supplemental Figure 1B). Interestingly, chlorophyte genomes tend to dedicate a greater fraction of genes to encode ZIP proteins when compared to charophytes (Figure 2B), which are more closely related to land plants [30–32]. Chlorophytes from the Class Chlorophyceae apparently tend to have more ZIP genes per genome (Supplemental Figure 1C), although no differences regarding the relative genomic frequency of ZIP genes were found between Chlorophyta Classes Chlorophyceae, Trebouxiophyceae and Mamiellophyceae (Supplemental Figure 1D). In the unicellular freshwater chlorophytes *Chlorella variabilis* and *Coccomyxa subellipsoidea*, both from Class Trebouxiophyceae, > 0.1% of their genes encode for ZIP proteins, a similar fraction is only observed in the top ZIP-enriched genomes of our land plant set like maize and *Anthoceros augustus* (hornwort) (Figure 2B). No significant differences were observed in ZIP genomic frequency related to algal habitats (Supplemental Figure 1B). It is unclear why genomes of simple unicellular chlorophytes have accumulated such a high frequency of ZIP genes, more than double of the median value (~ 0.05%) observed for our set of vascular plants (Tracheophyta) (Figure 2B). *C. subellipsoidea* and *C. variabilis* are promising species for biofuel production and in the accumulation of metals such as cadmium (Cd) [33,34]. In angiosperms it is known that some members of the ZIP family help in the absorption of Cd [13]. So, it is tentative to speculate that the large number of ZIP proteins in these algae might be related to the absorption of a broad range of metals other than iron.

No significant difference in term of genomic frequency of ZIP genes were detected between bryophytes and tracheophytes (Figure 2B), indicating that the variability of ZIP genomic frequency is similar between the two major groups of land plants. Although we accumulated knowledge in angiosperms such as *A. thaliana* and *O. sativa*, the molecular mechanisms of Fe acquisition are less well understood in bryophytes. Our results suggest, however, that ZIP genes might be important players in bryophytes. Experiments with the liverwort *Marchantia polymorpha* show evidence only for the use of Strategy I to capture Fe. *M. polymorpha* uses members of the ZIP family, such as MpZIP3 (Mapoly0169s0010) and MpZIP5 (Mapoly0088s0041) to capture Fe^{2+} [35].

We then wonder if the variation in number of ZIP homologs between species could be correlated with other genomic features. We first compared the number of ZIP genes with the total number of protein-coding genes for each genome. Result indicates that land plants do not exhibit any clear correlation between the two variables, suggesting that ZIP genes might not be preferably retained after WGD events in embryophytes (Figure 2C). The variation in the number of ZIP *loci* observed in land plants might be explained by gene-level duplications rather than large-scale gene duplication events. In opposition, data suggest that green algae species with more protein-coding genes tend to have more ZIP copies, indicating that independent expansions in the gene space of different taxa involved the fixation of duplicated copies of ZIP genes. Thus, the expanded set of ZIP observed in some green algae might be result of gene fixation after lineage specific WGD events (Figure 2C). Unfortunately, the low number of available genomes of non-Viridiplantae algae prevents any further conclusions. Overall, ZIP genes occur in low copy number in the red algae and glaucophyte analyzed. Our data supports the idea that expansions in genome size *per se* might not explain the observed variation in ZIP genes between or within groups of archaeplastidians (Figure 2D).

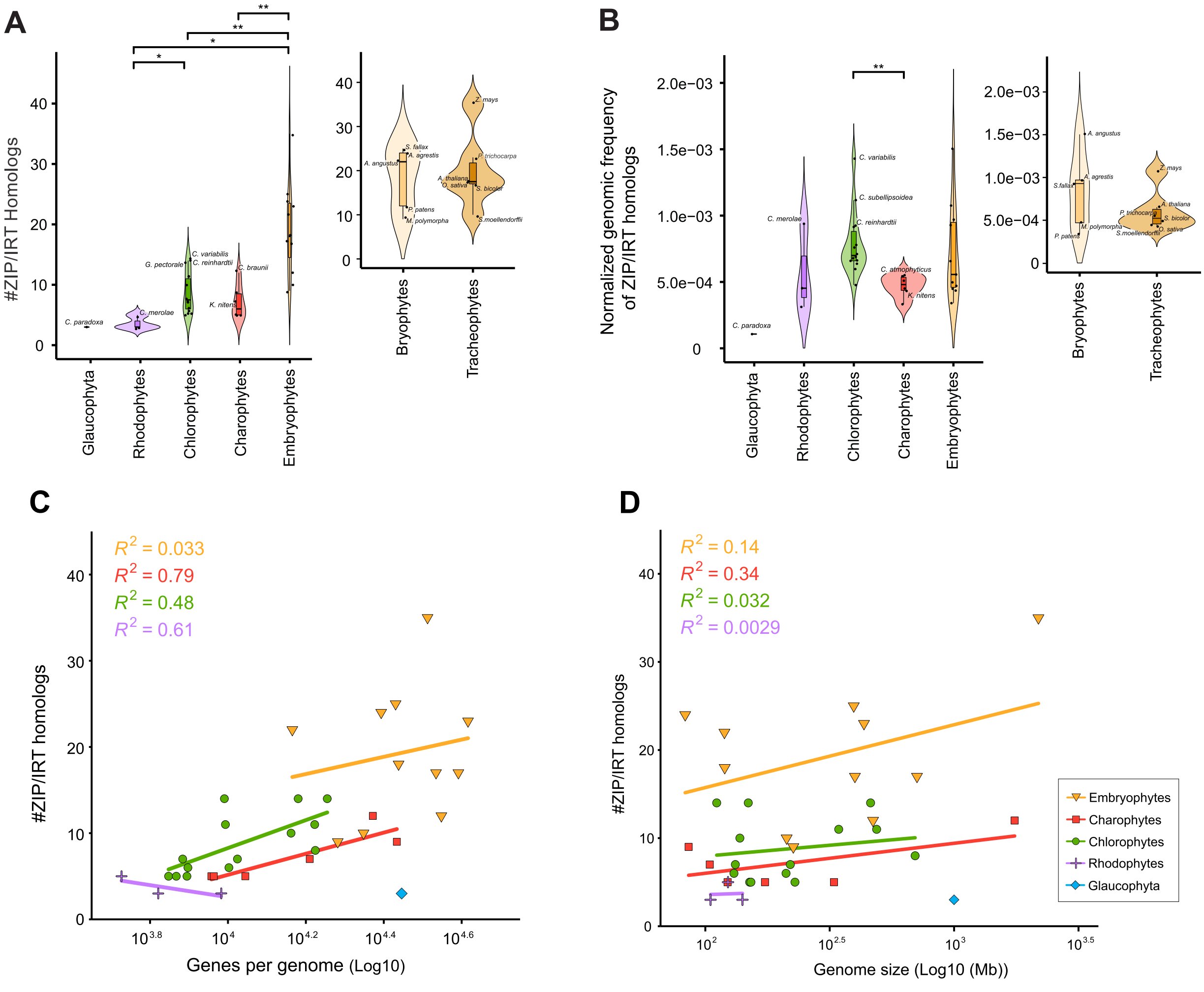


Fig. 2 Distribution of ZIP genes across archaeplastidians. (A) Distribution of the number of ZIP homologs found per genome for diverse taxonomic groups. (B) Distribution of the relative genomic frequency of ZIP genes per genome (normalized by total number of protein-coding genes per genome). ** represents p-value < 0.01 and * for p-value < 0.05. Correlations between the total number of ZIP genes per genome and (C) the number of genes per genome and (D) genome size.

Phylogenetic analysis of ZIP proteins in Archaeplastida

To better understand the phylogenetic relationship between ZIP proteins, we constructed a phylogenetic tree using the most representative protein sequence from the 392 unique ZIP genes identified across our sample of 35 genomes. The resulting phylogenetic tree can be subdivided into two deeply divergent major clades, here called X and Y (Figure 3A). Interestingly, there was no clustering among the IRT1 proteins of land plants and green algae. IRT1/ZIP proteins from the angiosperms *Arabidopsis* and rice, and from the liverwort *Marchantia* all belong to Clade X, while *Chlamydomonas* IRT1 and IRT2 belong to Clade Y (Figure 3A). Interestingly, MpZIP3 is phylogenetically related to angiosperms IRT proteins while MpZIP5 belongs to a subclade more distantly related to *Arabidopsis* and rice IRTs that is shared with charophytes. It indicates that divergent ZIP proteins might be recruited into Strategy I in land plants, possibly due to low specificity to metal transport of the proteins and fast changes in regulatory networks that may including new ZIP genes into the early transcriptional response to environmental Fe depletion. We also performed a sequence similarity network (SSN) analysis with the same protein sequences that yielded similar results, confirmed the distant relationship between Clades X and Y (Figure 3B). This result suggests that the mechanism of ferrous iron acquisition via ZIP proteins evolved at least twice in green plants from deeply divergent ancestral genes of the ZIP family.

Angiosperm proteins AtZIP13 (AT3G08650) and OsZIP13 (LOC_Os02g10230) from *Arabidopsis* and rice, respectively, are likely Zn transporters belonging to Clade Y [36]. It suggests that *Chlamydomonas* IRT1/2 is more closely related to proteins that are dedicated to Zn transport in angiosperms. The *Arabidopsis* proteins from Clade X AtZIP2 (AT5G59520) and AtZIP11 (AT1G55910), are involved in the uptake of Zn or Mn [12,13]. Suggesting that the IRT1 from angiosperms like *Arabidopsis* and rice are more closely related to proteins acting in Zn or Mn transporters than to the functional homologous IRT1 and IRT2 from chlorophytes. These observations suggest that even deeply sequence divergent ZIP proteins can evolve to fulfill the functional role of IRT1 first discovered in *Arabidopsis*. ZIP proteins might be able to have a fast turnover between different metal acquisition mechanisms due to an overall low metal selectivity coupled with transcriptional network rewiring of ZIP genes [37].

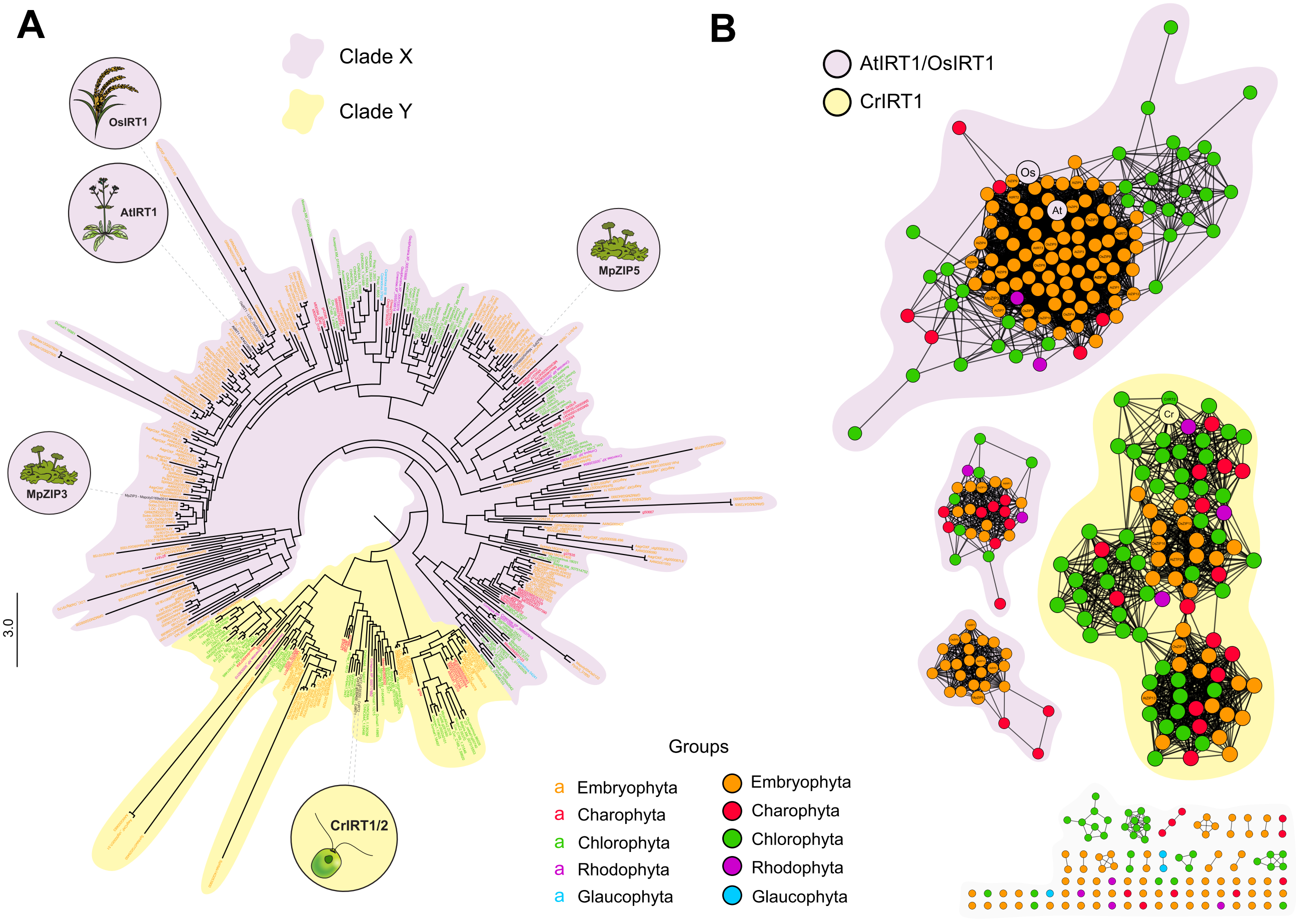


Fig. 3 Maximum-likelihood and Sequence Similarity Network analyses of ZIP family in Archaeplastida. (A) Circular maximum-likelihood phylogenetic tree highlighting Clade X and Clade Y. Proteins that are discussed in the text are highlighted. (B) Sequence Similarity Networks showing clustering of proteins in agreement with the phylogenetic analysis. Color scheme reflects the taxonomic classification of species.

Previous work has demonstrated that Aspartic Acid (D) residues at positions 119 and 155 of the *Arabidopsis* IRT1 are conserved with rice IRT1 (Figure 4A) and critical for Fe transport [38]. If those D residues are substituted with Alanine (A), the *Arabidopsis* IRT1 loses its ability to transport Fe, but not Zn [38]. Zn-transport can be abolished when Glutamic Acid (E) at position 122 is replaced by A, which does not affect Fe transport [38]. These results suggest that small changes in ZIP proteins can affect metal selectivity. However, if we align CrIRT1 amino acid sequence together with IRT1 from *Arabidopsis* and rice (Figure 4A) we observe no conservation in any of those critical positions, suggesting that IRT proteins might exhibit some degree of sequence plasticity. Whether the presence of A in the position 122 in the CrIRT1 means it lacks the ability to transport Zn should be better investigated with functional studies (Figure 4A).

Our results suggest that Strategy I of land plants evolved based on Clade X IRT proteins and could have evolved before the last common ancestor of Embryophyta. To better clarify this possibility, we aligned the *Arabidopsis* IRT1 with its closest related charophyte proteins from Clade X (Figure 4B). We observed D residue conservation in distinct charophyte lineages related to *Arabidopsis* IRT1 position 116. However, most charophytes have an E residue in position 154 (Figure 4B). It is interest to note that E and D residues are both acidic amino acids, with similar structure and chemical properties, indicating that the substitution D154E might not affect Fe transport. These results suggest that charophytes have Clade X ZIP proteins compatible with Fe^{2+} acquisition in Strategy I. It opens the possibility that Strategy I might have evolve in Streptophyta (Embryophyta + Charophyta) during early land colonization by simples charophytes before the origin of land plants. Functional studies in charophytes, especially with terrestrial species, will be needed to confirm this hypothesis.

To check whether Clade Y proteins have evolved diverse metal affinities, we aligned the *Chlamydomonas* IRT1 with its closest related proteins from *Arabidopsis* and rice, together with IRT1 from *Arabidopsis* and rice from Clade X (Figure 4C). No sequence conservation was observed in the critical residues, suggesting that Clade Y and Clade X proteins might have structural differences and still be able to function in the transport the same metals.

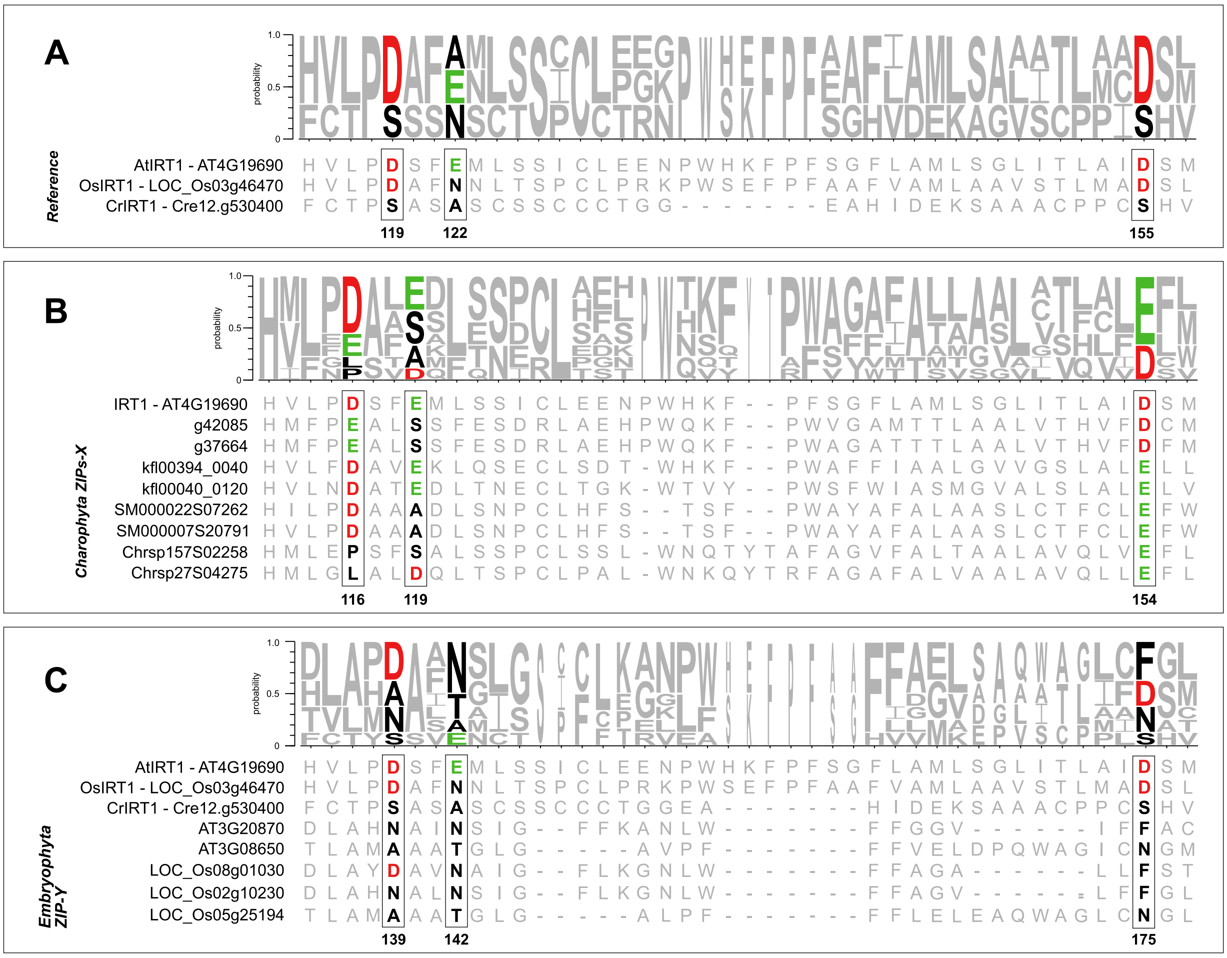


Fig. 4 Multiple sequence alignments and logo sequence comparison of AtIRT1, OsIRT1 and CrIRT1 with their closest related sequences. The size of residues within a stack indicates the relative frequency of each amino acid per position. Aspartic acid residues are marked in red and Glutamic acid residues in green. (A) Alignment between IRT1 sequences from *Arabidopsis*, rice and *Chlamydomonas*. indicating the amino acids essential for Fe transport in *A. thaliana*. (B) Charophyta ZIP proteins from Clade X aligned with the AtIRT1. (C) *Arabidopsis* and rice ZIP proteins from Clade Y aligned with AtIRT1, OsIRT1, and CrIRT1.

Variation in domain structure of ZIP proteins in Archaeplastida

We next wondered about structural variations in ZIP proteins, related to the presence of additional protein domains. Our results indicate that ~93% of the ZIP proteins of archaeplastidians present only the ZIP domain (PF02535). The remaining ~7% present the ZIP domain associated with at least another domain. In rare cases, the ZIP domain was associated with tens of different domains in the same protein (Figure 5A). Overall, land plants and chlorophytes tend to have ZIP proteins associated with other domains more frequently than other taxa (Figure 5B). The domains Presenilin (PF01080) and CDC45-like (PF02724) are found associated with the ZIP domain in 15 proteins (~3.8% of the sample) and Ryanodine Receptor (RR) TM 4-6 (PF06459) appears associated with a ZIP domain in 14 proteins (~3.6%)

(Figure 5C). Some ZIP proteins present combinations between those domains (Figure 5D). Presenilin and RR TM 4-6 are domains associated with transmembrane proteins [18,39], such as ZIP proteins. CDC45 is involved in the initiation of DNA replication in *S. cerevisiae* (Saha et al. 1998), although not much is known about CDC45-like proteins in plants. The role of ZIP domain in such multidomain proteins remains to be further investigated.

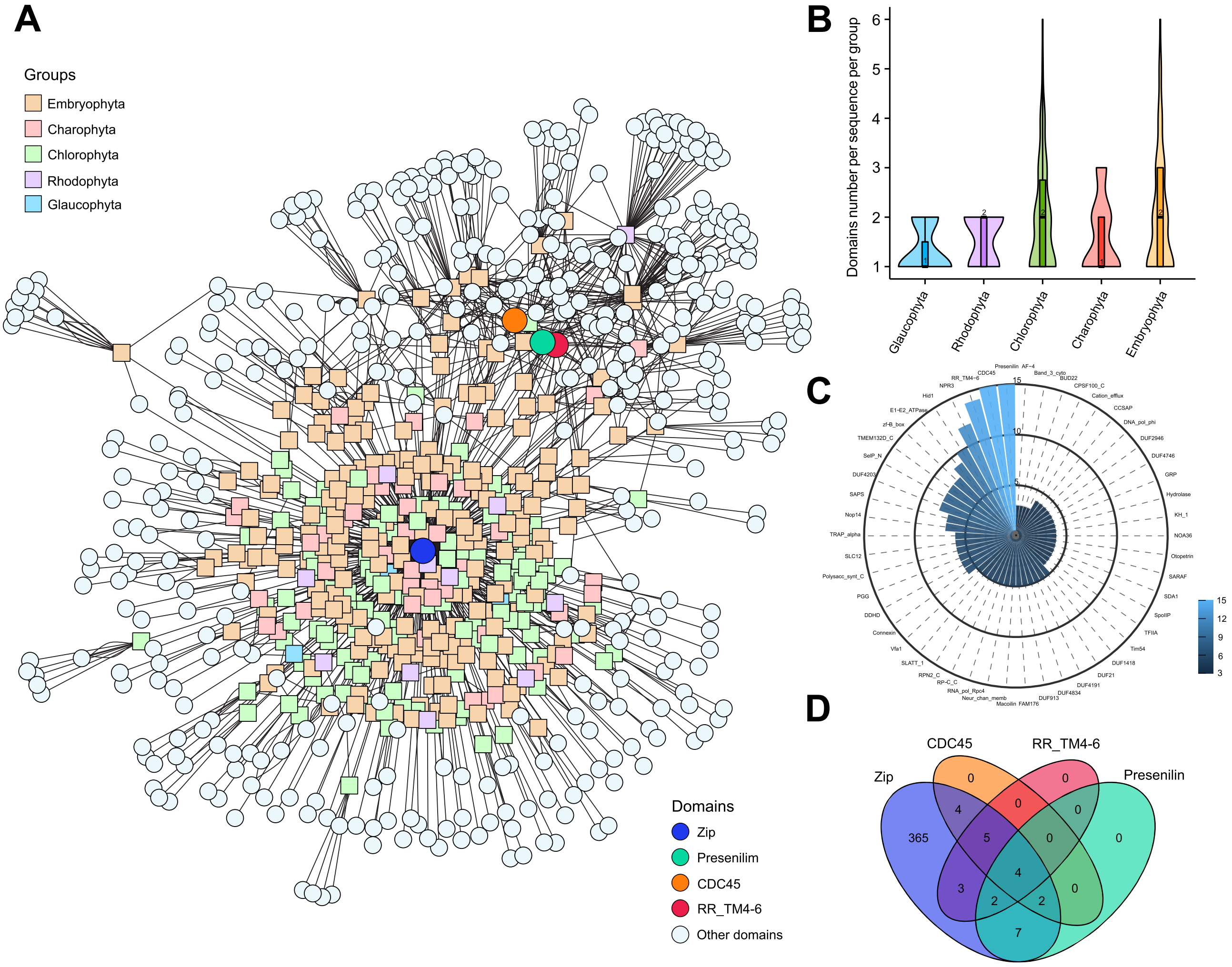


Fig. 5 Domains analysis of ZIP proteins found in Archaeplastida. (A) Network summarizing the sequences (squares) connected to their respective protein domains (circles). In the network, the domains that have the most connections with protein sequences stand out (blue = ZIP, green = Presenilin, orange = CDC45-like and red = RR TM4-6). (B) Distribution of the number of unique domains per sequence in Archaeplastida groups and their medians. Sequences with more than 6 domains were discarded. (C) Bar plot showing the 50 most common domains that are found associated with the ZIP domain by the number of occurrences in different proteins. (D) Venn diagram showing the occurrence of multidomain proteins and the dominance of single ZIP domain in ~93% of ZIP proteins.

Conclusion

Our results support the notion that green plants functional homologs of the archetypical angiosperm IRT1 emerged from distantly related ZIP proteins that are divergent in key features of their amino acid sequences. Furthermore, the typical critical amino acids in the *Arabidopsis* IRT1 protein seems to have evolved in before the split of land plants from streptophytic algae. Functional data on Fe acquisition of soil charophytes (such as *Klebsormidium*) might help understand how land plants ancestors were first able to obtain Fe in terrestrial settings with subaerial exposure and direct contact with the substrate. IRT1/2 from the chlorophyte *Chlamydomonas* seem to have a divergent sequence regarding key residues known in *Arabidopsis*, suggesting that multiple primary amino acid sequences of ZIP proteins might be fitted to transport Fe^{2+} efficiently.

The phylogenetic diversity of ZIP proteins used in Fe acquisition in green plants indicate that networking rewiring of ZIP genes encoding proteins with the ability to transport Fe might be the ultimate cause of IRT1 diversity in the Strategy I of Fe acquisition. Besides its ecological importance for photosynthesis, little is known about Fe acquisition strategies across the superdiverse group Archaeplastida. The results presented here indicate that functional homologs of IRT1 can emerge from highly divergent phylogenetic groups of ZIP transporters, suggesting that much evolutionary diversity is still to be found in Fe uptake in less studied groups such as green and red algae.

Acknowledgments

This work was supported by the Serrapilheira Institute (grant number Serra-1812-26691) and by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) that provided a master’s scholarship to WFCR.

References

[1] R.R. Crichton, Iron metabolism: from molecular mechanisms to clinical consequences, 4th edition, John Wiley and Sons, Inc, Chichester, West Sussex, United Kingdom, 2016.

[2] P.A. Frey, G.H. Reed, The Ubiquity of Iron, ACS Chem. Biol. 7 (2012) 1477–1481. https://doi.org/10.1021/cb300323q.

[3] S. Blomqvist, A. Gunnars, R. Elmgren, Why the limiting nutrient differs between temperate coastal seas and freshwater lakes: A matter of salt, Limnol. Oceanogr. 49 (2004) 2236–2241. https://doi.org/10.4319/lo.2004.49.6.2236.

[4] V. Römheld, H. Marschner, Evidence for a Specific Uptake System for Iron Phytosiderophores in Roots of Grasses, Plant Physiol. 80 (1986) 175–180. https://doi.org/10.1104/pp.80.1.175.

[5] J.M. Connorton, J. Balk, J. Rodríguez-Celma, Iron homeostasis in plants – a brief overview, Metallomics. 9 (2017) 813–823. https://doi.org/10.1039/C7MT00136C.

[6] A. Wairich, B.H.N. de Oliveira, E.B. Arend, G.L. Duarte, L.R. Ponte, R.A. Sperotto, F.K. Ricachenevsky, J.P. Fett, The Combined Strategy for iron uptake is not exclusive to domesticated rice (Oryza sativa), Sci Rep. 9 (2019) 16144. https://doi.org/10.1038/s41598-019-52502-0.

[7] M.D. Allen, J.A. del Campo, J. Kropat, S.S. Merchant, *FEA1* , *FEA2* , and *FRE1* , Encoding Two Homologous Secreted Proteins and a Candidate Ferrireductase, Are Expressed Coordinately with *FOX1* and *FTR1* in Iron-Deficient *Chlamydomonas reinhardtii*, Eukaryot Cell. 6 (2007) 1841–1852. https://doi.org/10.1128/EC.00205-07.

[8] C.E. Blaby-Haas, S.S. Merchant, The ins and outs of algal metal transport, Biochimica et Biophysica Acta (BBA) - Molecular Cell Research. 1823 (2012) 1531–1552. https://doi.org/10.1016/j.bbamcr.2012.04.010.

[9] A.G. Glaesener, S.S. Merchant, C.E. Blaby-Haas, Iron economy in Chlamydomonas reinhardtii, Front. Plant Sci. 4 (2013). https://doi.org/10.3389/fpls.2013.00337.

[10] G.E. Kroh, M. Pilon, Regulation of Iron Homeostasis and Use in Chloroplasts, IJMS. 21 (2020) 3395. https://doi.org/10.3390/ijms21093395.

[11] D. Eide, M. Broderius, J. Fett, M.L. Guerinot, A novel iron-regulated metal transporter from plants identified by functional expression in yeast., Proceedings of the National Academy of Sciences. 93 (1996) 5624–5628. https://doi.org/10.1073/pnas.93.11.5624.

[12] M.J. Milner, J. Seamon, E. Craft, L.V. Kochian, Transport properties of members of the ZIP family in plants and their role in Zn and Mn homeostasis, EXBOTJ. 64 (2013) 369–381. https://doi.org/10.1093/jxb/ers315.

[13] X. Zheng, L. Chen, X. Li, Arabidopsis and rice showed a distinct pattern in ZIPs genes expression profile in response to Cd stress, Bot Stud. 59 (2018) 22. https://doi.org/10.1186/s40529-018-0238-6.

[14] A. Pita-Barbosa, F.K. Ricachenevsky, M. Wilson, T. Dottorini, D.E. Salt, Transcriptional plasticity buffers genetic variation in zinc homeostasis, Sci Rep. 9 (2019) 19482. https://doi.org/10.1038/s41598-019-55736-0.

[15] S. Lee, J. Lee, F.K. Ricachenevsky, T. Punshon, R. Tappero, D.E. Salt, M.L. Guerinot, Redundant roles of four ZIP family members in zinc homeostasis and seed development in *Arabidopsis thaliana*, Plant J. (2021) tpj.15506. https://doi.org/10.1111/tpj.15506.

[16] E. Camprubi, S.F. Jordan, R. Vasiliadou, N. Lane, Iron catalysis at the origin of life, IUBMB Life. 69 (2017) 373–381. https://doi.org/10.1002/iub.1632.

[17] J. Wade, D.J. Byrne, C.J. Ballentine, H. Drakesmith, Temporal variation of planetary iron as a driver of evolution, Proc Natl Acad Sci USA. 118 (2021) e2109865118. https://doi.org/10.1073/pnas.2109865118.

[18] J. Mistry, S. Chuguransky, L. Williams, M. Qureshi, G.A. Salazar, E.L.L. Sonnhammer, S.C.E. Tosatto, L. Paladin, S. Raj, L.J. Richardson, R.D. Finn, A. Bateman, Pfam: The protein families database in 2021, Nucleic Acids Research. 49 (2021) D412–D419. https://doi.org/10.1093/nar/gkaa913.

[19] S.R. Eddy, Accelerated Profile HMM Searches, PLoS Comput Biol. 7 (2011) e1002195. https://doi.org/10.1371/journal.pcbi.1002195.

[20] P. Rice, I. Longden, A. Bleasby, EMBOSS: The European Molecular Biology Open Software Suite, Trends in Genetics. 16 (2000) 276–277. https://doi.org/10.1016/S0168-9525(00)02024-2.

[21] R Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria., 2021. https://www.r-project.org/.

[22] K. Katoh, MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform, Nucleic Acids Research. 30 (2002) 3059–3066. https://doi.org/10.1093/nar/gkf436.

[23] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms, Molecular Biology and Evolution. 35 (2018) 1547–1549. https://doi.org/10.1093/molbev/msy096.

[24] G.E. Crooks, G. Hon, J.-M. Chandonia, S.E. Brenner, WebLogo: A Sequence Logo Generator: Figure 1, Genome Res. 14 (2004) 1188–1190. https://doi.org/10.1101/gr.849004.

[25] L.-T. Nguyen, H.A. Schmidt, A. von Haeseler, B.Q. Minh, IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies, Molecular Biology and Evolution. 32 (2015) 268–274. https://doi.org/10.1093/molbev/msu300.

[26] R. Zallot, N. Oberg, J.A. Gerlt, The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways, Biochemistry. 58 (2019) 4169–4182. https://doi.org/10.1021/acs.biochem.9b00735.

[27] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks, Genome Res. 13 (2003) 2498–2504. https://doi.org/10.1101/gr.1239303.

[28] L. Hoffmann, E. Breitbarth, P. Boyd, K. Hunter, Influence of ocean warming and acidification on trace metal biogeochemistry, Mar. Ecol. Prog. Ser. 470 (2012) 191–205. https://doi.org/10.3354/meps10082.

[29] J. Behnke, J. LaRoche, Iron uptake proteins in algae and the role of Iron Starvation-Induced Proteins (ISIPs), European Journal of Phycology. 55 (2020) 339–360. https://doi.org/10.1080/09670262.2020.1744039.

[30] E. Waters, Molecular adaptation and the origin of land plants, Molecular Phylogenetics and Evolution. 29 (2003) 456–463. https://doi.org/10.1016/j.ympev.2003.07.018.

[31] J. Harholt, Ø. Moestrup, P. Ulvskov, Why Plants Were Terrestrial from the Beginning, Trends in Plant Science. 21 (2016) 96–101. https://doi.org/10.1016/j.tplants.2015.11.010.

[32] S. Wang, L. Li, H. Li, S.K. Sahu, H. Wang, Y. Xu, W. Xian, B. Song, H. Liang, S. Cheng, Y. Chang, Y. Song, Z. Çebi, S. Wittek, T. Reder, M. Peterson, H. Yang, J. Wang, B. Melkonian, Y. Van de Peer, X. Xu, G.K.-S. Wong, M. Melkonian, H. Liu, X. Liu, Genomes of early-diverging streptophyte algae shed light on plant terrestrialization, Nat. Plants. 6 (2020) 95–106. https://doi.org/10.1038/s41477-019-0560-3.

[33] J. Kováčik, G. Rotková, M. Bujdoš, P. Babula, V. Peterková, P. Matúš, Ascorbic acid protects Coccomyxa subellipsoidea against metal toxicity through modulation of ROS/NO balance and metal uptake, Journal of Hazardous Materials. 339 (2017) 200–207. https://doi.org/10.1016/j.jhazmat.2017.06.035.

[34] M.S. Rana, S.K. Prajapati, Resolving the dilemma of iron bioavailability to microalgae for commercial sustenance, Algal Research. 59 (2021) 102458. https://doi.org/10.1016/j.algal.2021.102458.

[35] J. Lo, M. Tsednee, Y. Lo, S. Yang, J. Hu, K. Ishizaki, T. Kohchi, D. Lee, K. Yeh, Evolutionary analysis of iron (Fe) acquisition system in *Marchantia polymorpha*, New Phytol. 211 (2016) 569–583. https://doi.org/10.1111/nph.13922.

[36] R. Ivanov, P. Bauer, Sequence and coexpression analysis of iron-regulated ZIP transporter genes reveals crossing points between iron acquisition strategies in green algae and land plants, Plant Soil. 418 (2017) 61–73. https://doi.org/10.1007/s11104-016-3128-2.

[37] I. Nocedal, A.D. Johnson, How Transcription Networks Evolve and Produce Biological Novelty, Cold Spring Harb Symp Quant Biol. 80 (2015) 265–274. https://doi.org/10.1101/sqb.2015.80.027557.

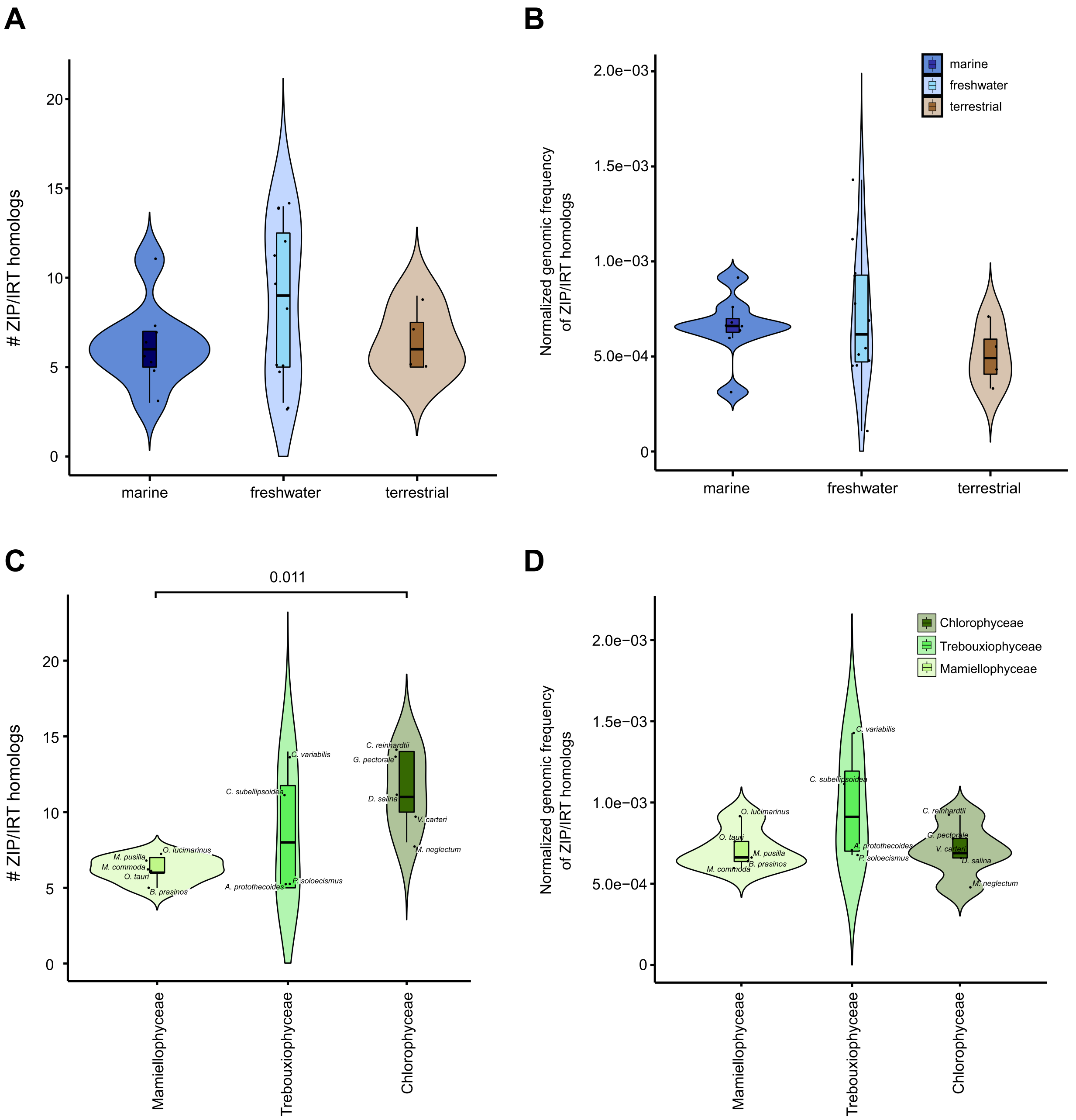
[38] E.E. Rogers, D.J. Eide, M.L. Guerinot, Altered selectivity in an Arabidopsis metal transporter, Proceedings of the National Academy of Sciences. 97 (2000) 12356–12360. https://doi.org/10.1073/pnas.210214197.

[39] M.R. Hass, C. Sato, R. Kopan, G. Zhao, Presenilin: RIP and beyond, Seminars in Cell & Developmental Biology. 20 (2009) 201–210. https://doi.org/10.1016/j.semcdb.2008.11.014.

Supplementary Data

Supplemental Table S1. Table showing the number of ZIP family homologs found by species in our dataset.

Species	Group	ZIP homologs
Arabidopsis thaliana	Embryophyta	18
Marchantia polymorpha	Embryophyta	9
Oryza sativa	Embryophyta	17
Zea mays	Embryophyta	35
Physcomitrella patens	Embryophyta	12
Populus trichocarpa	Embryophyta	23
Sorghum bicolor	Embryophyta	17
Selaginella moellendorffii	Embryophyta	10
Anthoceros agrestis [Oxford]	Embryophyta	24
Sphagnum fallax	Embryophyta	25
Anthoceros angustus	Embryophyta	22
Chlorokybus atmophyticus	Charophyta	5
Chara braunii	Charophyta	12
Klebsormidium nitens	Charophyta	7
Mesotaenium endlicherianum	Charophyta	5
Mesostigma viride	Charophyta	5
Spirogloea muscicola	Charophyta	9
Auxenochlorella protothecoides	Chlorophyta	5
Bathycoccus prasinos	Chlorophyta	5
Chlorella variabilis	Chlorophyta	14
Coccomyxa subellipsoidea	Chlorophyta	11
Chlamydomonas reinhardtii	Chlorophyta	14
Dunaliella salina	Chlorophyta	11
Gonium pectorale	Chlorophyta	14
Micromonas commoda	Chlorophyta	6
Monoraphidium neglectum	Chlorophyta	8
Micromonas pusilla	Chlorophyta	7
Ostreococcus lucimarinus	Chlorophyta	7
Ostreococcus tauri	Chlorophyta	6
Picochlorum soloecismus	Chlorophyta	5
Volvox carteri	Chlorophyta	10
Chondrus crispus	Rhodophyta	3
Cyanidioschyzon merolae	Rhodophyta	5
Galdieria sulphuraria	Rhodophyta	3
Cyanophora paradoxa	Glaucophyta	3
Total		392



Supplemental Figure 1. **Distribution of ZIP genes across algae groups**. Plots showing the global distribution of ZIP/IRT homologs and the frequency of proteins found normalized by the number of genes per genome by algae species (Rhodophytes, Chlorophytes and Charophytes) in differents habitats (A, C) and differents Cholophyta class (B, D).

4. CONCLUSÕES E PERSPECTIVAS

A captura de Fe é baseada em vários mecanismos que diferem de organismo para organismo. Sem dúvida, as proteínas da família ZIP desempenham um papel crucial nesses vários mecanismos. A identificação dessas proteínas e genes candidatos relacionados ao transporte de Fe auxilia na investigação da possível origem das plantas terrestres ou como elas conseguiram dominar o ambiente terrestre através do estabelecimento desses mecanismos. Nossos resultados filogenéticos indicam que o mecanismo de captura de Fe apareceu de forma divergente em algas e plantas. Além disso, os aminoácidos críticos típicos da proteína AtIRT1 parecem ter evoluído antes da separação das plantas terrestres das algas estreptófitas.

Grande parte da biomassa do planeta é constituída por plantas, apresentando grande importância para a sobrevivência de todos os seres vivos. Portanto, este estudo também pode auxiliar no desenvolvimento de culturas de plantas e algas através da biotecnologia, superando problemas como habitats e terrenos com baixa biodisponibilidade de Fe. Para isso, é necessário visualizar essas sequências de proteínas encontradas em nosso trabalho de forma funcional e estrutural. Em geral, o conhecimento dos sistemas de absorção de Fe em algas e muitas plantas ainda é muitas vezes baseado apenas em análises in silico, faltando a caracterização bioquímica dos vários componentes envolvidos neste processo. Nossos achados serão valiosos conhecimentos científicos para estudos futuros.

REFERÊNCIAS

ABBASPOUR, N.; HURRELL, R.; KELISHADI, R. Review on iron and its importance for human health. **Journal of Research in Medical Sciences: The Official Journal of Isfahan University of Medical Sciences**, v. 19, n. 2, p. 164–174, fev. 2014.

AGGETT, P. J. Iron. Em: ERDMAN, J. W.; MACDONALD, I. A.; ZEISEL, S. H. (Eds.). . **Present Knowledge in Nutrition**. Oxford, UK: Wiley-Blackwell, 2012. p. 506–520.

ALLEN, M. D. et al. *FEA1* , *FEA2* , and *FRE1* , Encoding Two Homologous Secreted Proteins and a Candidate Ferrireductase, Are Expressed Coordinately with *FOX1* and *FTR1* in Iron-Deficient *Chlamydomonas reinhardtii*. **Eukaryotic Cell**, v. 6, n. 10, p. 1841–1852, out. 2007.

BLABY-HAAS, C. E.; MERCHANT, S. S. The ins and outs of algal metal transport. **Biochimica et Biophysica Acta (BBA) - Molecular Cell Research**, v. 1823, n. 9, p. 1531–1552, set. 2012.

BLOMQVIST, S.; GUNNARS, A.; ELMGREN, R. Why the limiting nutrient differs between temperate coastal seas and freshwater lakes: A matter of salt. **Limnology and Oceanography**, v. 49, n. 6, p. 2236–2241, nov. 2004.

CONNORTON, J. M.; BALK, J.; RODRÍGUEZ-CELMA, J. Iron homeostasis in plants – a brief overview. **Metallomics**, v. 9, n. 7, p. 813–823, 2017.

CRICHTON, R. R. **Iron metabolism: from molecular mechanisms to clinical consequences**. 4th edition ed. Chichester, West Sussex, United Kingdom: John Wiley and Sons, Inc, 2016.

EIDE, D. et al. A novel iron-regulated metal transporter from plants identified by functional expression in yeast. **Proceedings of the National Academy of Sciences**, v. 93, n. 11, p. 5624–5628, 28 maio 1996.

FREY, P. A.; REED, G. H. The Ubiquity of Iron. **ACS Chemical Biology**, v. 7, n. 9, p. 1477–1481, 21 set. 2012.

GLAESENER, A. G.; MERCHANT, S. S.; BLABY-HAAS, C. E. Iron economy in Chlamydomonas reinhardtii. **Frontiers in Plant Science**, v. 4, 2013.

KIM, S. A.; GUERINOT, M. L. Mining iron: Iron uptake and transport in plants. **FEBS Letters**, v. 581, n. 12, p. 2273–2280, 25 maio 2007.

KROH, G. E.; PILON, M. Regulation of Iron Homeostasis and Use in Chloroplasts. **International Journal of Molecular Sciences**, v. 21, n. 9, p. 3395, 11 maio 2020.

LEE, S. et al. Redundant roles of four ZIP family members in zinc homeostasis and seed development in *Arabidopsis thaliana*. **The Plant Journal**, p. tpj.15506, 8 out. 2021.

LENTON, T. M. et al. Earliest land plants created modern levels of atmospheric oxygen. **Proceedings of the National Academy of Sciences**, v. 113, n. 35, p. 9704–9709, 30 ago. 2016.

LILL, R.; FREIBERT, S.-A. Mechanisms of Mitochondrial Iron-Sulfur Protein Biogenesis. **Annual Review of Biochemistry**, v. 89, n. 1, p. 471–499, 20 jun. 2020.

MILNER, M. J. et al. Transport properties of members of the ZIP family in plants and their role in Zn and Mn homeostasis. **Journal of Experimental Botany**, v. 64, n. 1, p. 369–381, jan. 2013.

ONYANGO, D. A. et al. Mechanistic understanding of iron toxicity tolerance in contrasting rice varieties from Africa: 1. Morpho-physiological and biochemical responses. **Functional Plant Biology**, v. 46, n. 1, p. 93, 2019.

PITA-BARBOSA, A. et al. Transcriptional plasticity buffers genetic variation in zinc homeostasis. **Scientific Reports**, v. 9, n. 1, p. 19482, dez. 2019.

POSEY, J. E.; GHERARDINI, F. C. Lack of a Role for Iron in the Lyme Disease Pathogen. **Science**, v. 288, n. 5471, p. 1651–1653, 2 jun. 2000.

RÖMHELD, V.; MARSCHNER, H. Evidence for a Specific Uptake System for Iron Phytosiderophores in Roots of Grasses. **Plant Physiology**, v. 80, n. 1, p. 175–180, 1 jan. 1986.

ROUT, G. R.; SAHOO, S. ROLE OF IRON IN PLANT GROWTH AND METABOLISM. **Reviews in Agricultural Science**, v. 3, n. 0, p. 1–24, 2015.

WADE, J. et al. Temporal variation of planetary iron as a driver of evolution. **Proceedings of the National Academy of Sciences**, v. 118, n. 51, p. e2109865118, 21 dez. 2021.

WAIRICH, A. et al. The Combined Strategy for iron uptake is not exclusive to domesticated rice (Oryza sativa). **Scientific Reports**, v. 9, n. 1, p. 16144, dez. 2019.

ZAHRA, N. et al. Fe toxicity in plants: Impacts and remediation. **Physiologia Plantarum**, p. ppl.13361, 22 fev. 2021.

ZHENG, X.; CHEN, L.; LI, X. Arabidopsis and rice showed a distinct pattern in ZIPs genes expression profile in response to Cd stress. **Botanical Studies**, v. 59, n. 1, p. 22, 25 set. 2018.