



Serviço Público Federal
Universidade Federal de Minas Gerais
Instituto de Ciências Biológicas – Depto. De Botânica
Programa de Pós-Graduação em Biologia Vegetal

Wenderson Felipe Costa Rodrigues

PARASITISMO VEGETAL E TRANSPORTE DE METAIS:
Narrativas Complementares da Evolução e Inovação Molecular em Plantas

Belo Horizonte
2026

Wenderson Felipe Costa Rodrigues

**PARASITISMO VEGETAL E TRANSPORTE DE METAIS:
Narrativas Complementares da Evolução e Inovação Molecular em Plantas**

Tese apresentada ao Programa de Pós-Graduação em
Biologia Vegetal da Universidade Federal de Minas
Gerais como requisito parcial para obtenção do título
de Doutor em Biologia Vegetal.

Orientador: Prof. Dr. Luiz Eduardo Vieira Del Bem

Coorientadora: Prof.^a Dr.^a Juliane Karine Ishida

Belo Horizonte
2026

043

Rodrigues, Wenderson Felipe Costa.

Parasitismo vegetal e transporte de metais: narrativas complementares da evolução e inovação molecular em plantas [manuscrito] / Wenderson Felipe Costa Rodrigues. – 2026.

181 f. : il. ; 29,5 cm.

Orientador: Prof. Dr. Luiz Eduardo Vieira Del Bem. Coorientadora: Prof.^a Dr.^a Juliane Karine Ishida.

Tese (doutorado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa de Pós-Graduação em Biologia Vegetal.

1. Desenvolvimento Vegetal. 2. Plantas. 3. Evolução Biológica. 4. Parasitismo. I. Del Bem, Luiz Eduardo Vieira. II. Ishida, Juliane Karine. III. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. IV. Título.

CDU: 581



UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
COLEGIADO DO CURSO DE GRADUAÇÃO PÓS-GRADUAÇÃO EM BIOLOGIA VEGETAL

FOLHA DE APROVAÇÃO

WENDERSON FELIPE COSTA RODRIGUES

"Parasitismo Vegetal e Transporte de Metais: Narrativas Complementares da Evolução e Inovação Molecular em Plantas"

Matrícula 2022665354

Data da defesa: 09 de março de 2026.

COMISSÃO EXAMINADORA

Dr. Luiz Eduardo Vieira Del Bem (ESALQ/USP), Dr. Douglas Silva Domingues (ESALQ/USP), Dra. Cristiane Paula Gomes Calixto (IB-USP), Dr. Vinicius Augusto Carvalho de Abreu (Universidade Federal do Pará), Dr. Joni Esrom Lima (UFMG).

Belo Horizonte, 09 de março de 2026

Assinaturas dos membros da comissão



Documento assinado eletronicamente por **Joni Esrom Lima, Membro**, em 10/03/2026, às 12:48, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Cristiane Paula Gomes Calixto, Usuário Externo**, em 11/03/2026, às 07:19, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Luiz Eduardo Vieira Del Bem, Usuário Externo**, em 11/03/2026, às 11:21, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

Documento assinado eletronicamente por **Douglas Silva Domingues, Usuário Externo**, em



12/03/2026, às 10:03, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Vinicius Augusto Carvalho de Abreu, Usuário Externo**, em 17/03/2026, às 15:46, 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 **5008748** e o código CRC **9D3D9EE0**.

Referência: Processo nº 23072.210081/2026-53

SEI nº 5008748

Ao Neném, meu irmão Wadson, que segue em
outro plano, mas continua em meu coração.
Te amo.

AGRADECIMENTOS

Durante todos esses anos recebi apoio de muitos e não conseguiria colocar todos aqui, mas esse primeiro parágrafo eu vou usar como forma de agradecimento a todos que foram luz no meu caminho e se tornaram eternos no meu viver.

Especialmente agradeço aos meus pais, Edson Rodrigues & Daniele Costa, de quem carrego sobrenome e todos os fundamentos que baseio meu caráter. Ao meu irmão Wadson Rodrigues, meu único irmão, que me viu como caçula mesmo sendo mais novo, obrigado pelo tempo que estive aqui. Agradeço a toda minha família, pilares imperfeitos dessa construção, mas que nos mantêm em pé.

Agradeço aos meus amigos mais próximos, na qual conto todos os dias sem nem pestanejar: João Luiz e Filipe Oliveira. Aqueles que me acompanham a mais de uma década: Juan David, Aline Lima, Camila Maciel, Thalyta Silva e Marciely Lira. E ao meu amor, presente em todos os momentos desde o dia em que decidimos ser um casal, obrigado por tudo Vinicius Amaral, principalmente por me escutar.

Agradeço imensamente meu orientador, Prof. Luiz Eduardo, e a minha coorientadora, Prof. Juliane Ishida, por serem meus pais acadêmicos. Se hoje eu me considero um bom pesquisador, isso é graças a vocês. Que possamos continuar com essa parceria incrível.

Agradeço também à UFMG e ao Programa de Pós-Graduação em Biologia Vegetal, que foi minha casa nos últimos 4 anos. À FAPEMIG, CAPES e Instituto Serrapilheira pelo suporte acadêmico e financeiro, me permitindo ter bolsa e auxílios responsáveis pela minha participação ativa no meio científico. À FAPEMIG em especial, por ter me permitido viver a maior experiência pelo qual já passei, meu período de estágio sanduíche na Imperial College London. Agradeço imensamente o aceite do Prof. Pedro Ballester, que me recebeu de braços abertos e me ensinou muito. Agradeço a toda sua equipe, que também foram grandes amigos: Nivya James, Naomi Ratnasingam, Abhavya Raja, Qianrong Guo, Klaudia Caba e Elizabeth Amelia. Meus tempos em Londres foram incríveis e vocês deixaram a cores de lá mais brilhantes.

Por fim, eu agradeço de todo meu coração aos meus amigos cientistas que conheci nesse período de doutorado, aos integrantes do LIVE Lab e Del-Bem Lab, muito obrigado pelo apoio sempre, especialmente: Laura Pires, Thiago Mendonça, Matheus Lewi e Ingridy Moura.

A todas as forças positivas, espirituais ou humanas, o meu mais sincero muito obrigado!

RESUMO

A evolução das plantas terrestres é marcada pela reutilização, modificação e especialização de vias moleculares ancestrais que sustentaram a conquista do ambiente continental e permitiram a diversificação de formas autotróficas e heterotróficas. Esta tese integra dois eixos evolutivos centrais: a emergência e diversificação do parasitismo vegetal e a evolução dos sistemas de aquisição e homeostase de metais. O primeiro eixo examina o haustório, órgão característico das plantas parasitas, destacando a mobilidade de lncRNAs na interface parasita-hospedeiro, a cooptação de vias de defesa e sinalização típicas de plantas não parasitas e o papel de receptores do tipo PRR-RLK na percepção de moléculas que desencadeiam sua formação. Em conjunto, esses processos evidenciam como pressões bióticas moldaram circuitos regulatórios capazes de sustentar estilos de vida altamente especializados, como os observados em plantas parasitas. No segundo eixo, são analisadas as estratégias de captação de ferro e a evolução estrutural de transportadores metálicos, com destaque para a convergência funcional de mecanismos de aquisição de ferro ferroso e para a diversificação da família proteica HMA. A síntese desses temas mostra que mecanismos associados à percepção de sinais, defesa, nutrição mineral e comunicação interespécies emergem de princípios evolutivos comuns baseados em inovação a partir de convergência e cooptação, modularidade e plasticidade estrutural. Em conjunto, os resultados apresentados elucidam como vias ancestrais foram remodeladas ao longo da evolução das plantas terrestres para gerar soluções adaptativas distintas.

Palavras-chave: Cooptação; convergência evolutiva; haustório; parasitismo vegetal; transportadores de metais.

ABSTRACT

The evolution of land plants has been driven by the reuse, modification and specialization of ancestral molecular pathways that enabled the colonization of terrestrial environments and the diversification of both autotrophic and heterotrophic forms. This thesis integrates two major evolutionary axes: the emergence and diversification of plant parasitism, and the evolution of systems for metal acquisition and homeostasis. The first axis examines the haustorium, the key organ of parasitic plants, focusing on the mobility of long non-coding RNAs at the parasite-host interface, the co-option of defense and signaling pathways typical of non-parasitic plants, and the role of PRR-RLK receptors in the perception of molecules that trigger haustorium development. Together, these processes illustrate how biotic pressures shaped regulatory circuits capable of sustaining highly specialized parasitic lifestyles. The second axis investigates the strategies for iron uptake and the structural evolution of metal transporters, highlighting the convergent emergence of mechanisms for ferrous iron acquisition and the functional diversification of the HMA family. The synthesis of these themes reveals that mechanisms underlying signal perception, defense, mineral nutrition, and interspecies communication emerge from shared evolutionary principles shaped by convergence, co-option, modularity, and structural plasticity. Overall, the results presented here elucidate how ancestral molecular pathways were remodeled throughout plant evolution to generate distinct adaptive solutions.

Keywords: Co-option; evolutionary convergence; haustorium; metal transporters; plant parasitism.

LISTA DE FIGURAS

Figura 1 – Representação conceitual integrando processos evolutivos e inovação funcional em Archaeplastida.....	10
Figura 2 – Diversidade e classificação de plantas parasitas.....	14
Figura 3 – Tipos de haustório em plantas parasitas.....	17
Figura 4 – Transporte interespecífico de pequenos RNAs (sRNAs) entre plantas parasitas e seus hospedeiros.....	19
Figura 5 – Homeostase de metais em plantas por proteínas HMA e ZIP.....	118

LISTA DE SIGLAS

CADL	<i>CARD-like protein</i>
CARD1	<i>Cannot Respond to DMBQ 1 (HPCA1)</i>
Co	Cobalto
CO ₂	Dióxido de carbono
Cu	Cobre
DMBQ	<i>2,6-dimethoxy-1,4-benzoquinone</i>
ETI	<i>Effector-Triggered Immunity</i>
Fe	Ferro
Fe ²⁺	Ferro ferroso
Fe ³⁺	Ferro férrico
HGT	Transferência horizontal de genes
HIF	<i>Haustorium-Inducing Factor</i>
HMA	P1B-type Heavy Metal ATPase
HPCA1	<i>Hydrogen Peroxide-Induced Ca²⁺ Increase 1</i>
IRT	<i>Iron-Regulated Transporter</i>
LRR	<i>Leucine-Rich Repeat</i>
lncRNA	<i>Long non-coding RNA</i>
MAPK	<i>Mitogen-Activated Protein Kinase</i>
ML	<i>Machine Learning</i>
Mn	Manganês
mRNA	<i>Messenger RNA</i>
ncRNA	<i>Non-coding RNA</i>
Pb	Chumbo
PRR	<i>Pattern-Recognition Receptor</i>
PRR-RLK	<i>Pattern-Recognition Receptor Receptor-Like Kinase</i>

PTI	<i>Pattern-Triggered Immunity</i>
RLK	<i>Receptor-Like Kinase</i>
ROS	<i>Reactive Oxygen Species</i>
sRNA	<i>Small RNA</i>
YSL	<i>Yellow Stripe-Like transporter</i>
ZIP	<i>Zinc/Iron-regulated Transporter-like Protein</i>
Zn	<i>Zinco</i>

SUMÁRIO

1. INTRODUÇÃO GERAL	8
2. OBJETIVOS	13
2.1. OBJETIVO GERAL.....	13
2.2. OBJETIVOS ESPECÍFICOS	13
3. SEÇÃO I – BASES MOLECULARES DO PARASITISMO VEGETAL: lncRNAs, DEFESA E PERCEPÇÃO DE SINAIS	14
3.1. REFERENCIAL TEÓRICO	14
3.2. CAPÍTULO 1 – INVESTIGATING lncRNA AT DIFFERENT LEVELS OF PARASITISM IN PARASITIC PLANTS.....	22
3.3. CAPÍTULO 2 – ORIGIN AND EVOLUTIONARY HISTORY OF DEFENSE-RELATED PROTEINS IN PLANTS.....	49
3.4. CAPÍTULO 3 – AI-GUIDED TWO-STAGE SCREENING IDENTIFIES PUTATIVE HAUSTORIUM NON-INDUCERS / INHIBITORS MOLECULES IN PARASITIC PLANTS	65
3.5. DISCUSSÃO	112
3.6. CONCLUSÃO.....	115
4. SEÇÃO II – TRANSPORTADORES METÁLICOS: DA CARACTERIZAÇÃO FUNCIONAL À EVOLUÇÃO DE SISTEMAS DE AQUISIÇÃO	116
4.1. REFERENCIAL TEÓRICO	116
4.2. CAPÍTULO 4 – FERROUS IRON UPTAKE VIA IRT1/ZIP EVOLVED AT LEAST TWICE IN GREEN PLANTS.....	120
4.3. CAPÍTULO 5 – COMPREHENSIVE CHARACTERIZATION OF HMA TRANSPORTERS IN COMMON BEAN	149
4.4. DISCUSSÃO	164
4.5. CONCLUSÃO.....	166
REFERÊNCIAS	167
5. ANEXOS – PRODUÇÃO ACADÊMICA	173

1. INTRODUÇÃO GERAL

A história evolutiva das plantas (Archaeplastida) é marcada por uma sucessão de transições profundas que transformaram organismos unicelulares nos principais produtores primários dos ecossistemas terrestres. A colonização do ambiente terrestre por plantas exigiu a superação de desafios fisiológicos e estruturais inéditos, incluindo desidratação, exposição intensa à radiação solar, escassez e heterogeneidade de nutrientes minerais, novas pressões bióticas e a necessidade de desenvolver sistemas eficientes de sustentação e transporte interno (Raven; Edwards, 2001; Morris et al., 2018). Esse processo não apenas permitiu a ocupação definitiva dos ambientes continentais, mas também transformou profundamente os ciclos biogeoquímicos do planeta, uma vez que a chegada das plantas à terra firme alterou de modo significativo o ciclo global do carbono, reduzindo concentrações atmosféricas de CO₂ e estabelecendo as bases ecológicas que sustentam a vasta maioria dos ecossistemas terrestres atuais (Lenton et al., 2012).

A diversificação subsequente das plantas terrestres foi acompanhada por radiações evolutivas que originaram tanto formas fotossintéticas especializadas quanto linhagens que remodelaram profundamente seus genomas, estruturas e fisiologia. Estudos recentes mostram que essas transições foram impulsionadas por pulsos concentrados de inovação genética, incluindo expansões e perdas de famílias gênicas e episódios de duplicação genômica que favoreceram o surgimento de novos modos de vida (Clark; Donoghue, 2018; Landis et al., 2018). Entre essas inovações, algumas envolveram trajetórias evolutivas regressivas, nas quais a perda de características consideradas basais acompanhou a especialização ecológica. Esse padrão é particularmente evidente nas plantas parasitas, um grupo de angiospermas marcado por reduções estruturais extremas e profundas reorganizações funcionais associadas à adaptação ao modo de vida heterotrófico (Westwood et al., 2010).

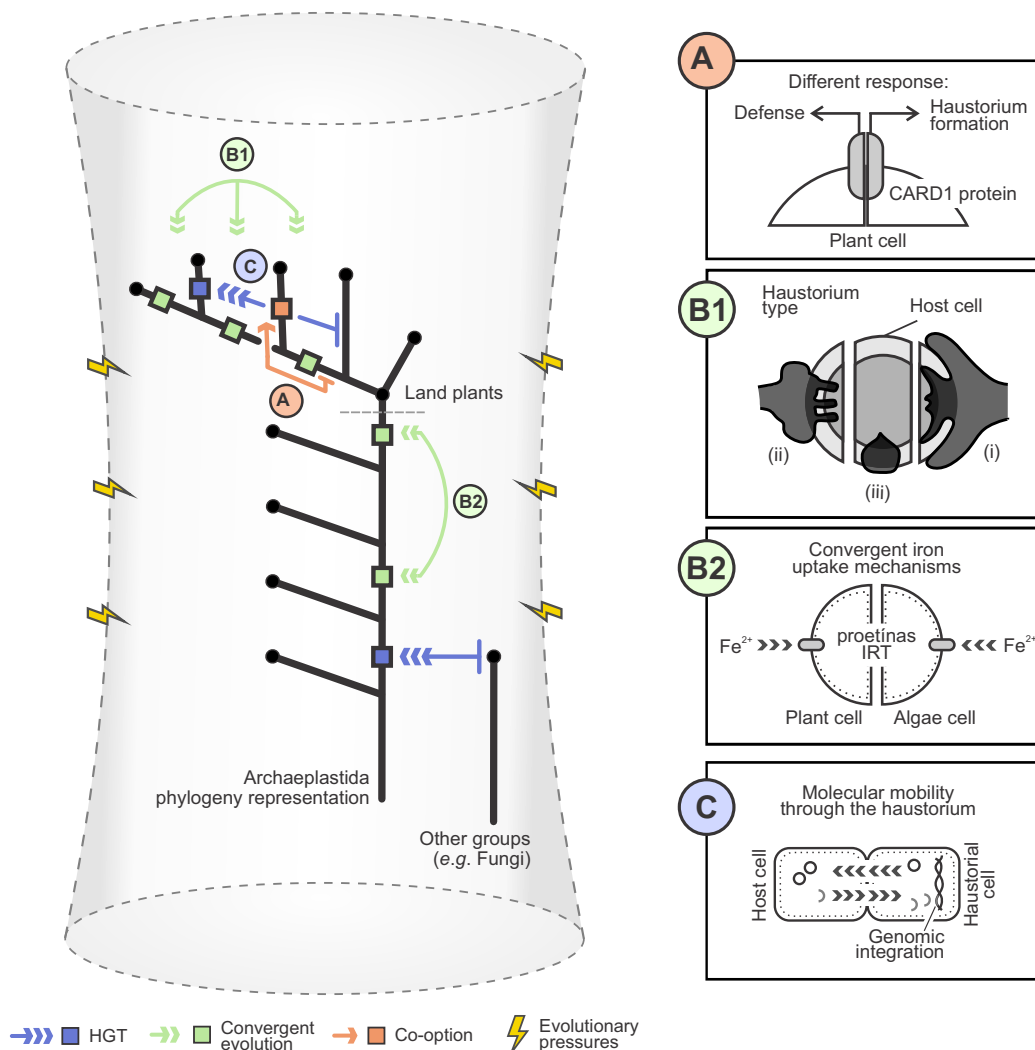
O parasitismo vegetal emergiu nas angiospermas diversas vezes de forma independente, um padrão típico de convergência evolutiva associado a pressões recorrentes como competição por recursos, limitação de nutrientes e intensificação de interações bióticas. Essas origens múltiplas resultaram em linhagens altamente diversas, que incluem desde hemiparasitas ainda fotossintetizantes até holoparasitas totalmente dependentes do hospedeiro, frequentemente acompanhadas por profundas remodelações genômicas (Westwood et al., 2010). Investigações moleculares em linhagens extremas, como nas espécies de plantas parasitas da família Rafflesiaceae, revelam exemplos marcantes desse processo, incluindo perda quase completa do plastoma (genoma plastidial) e múltiplos eventos de transferência horizontal de genes (HGT)

provenientes do hospedeiro. Notavelmente, a frequência desses eventos tende a se correlacionar positivamente com o grau de dependência do parasita. Ou seja, essas transformações indicam que pressões ecológicas intensas levaram não apenas ao surgimento do parasitismo em plantas, mas também à reutilização e reconfiguração de vias moleculares ancestrais por cooptação em múltiplas direções evolutivas (Xi et al., 2012, 2013; Molina et al., 2014a; Kado; Innan, 2018; Cai et al., 2021a).

O desenvolvimento do haustório, órgão que conecta os sistemas vasculares de hospedeiro e parasita, exemplifica essa cooptação funcional, pois depende da percepção de moléculas externas e de gatilhos de sinalização que compartilham paralelos com vias imunes e de crescimento de plantas não parasitas (Yoshida et al., 2016; Yang et al., 2015; Laohavisit et al., 2020). Paralelamente, a interface haustorial também se torna um canal eficiente para a troca bidirecional de moléculas, incluindo RNAs mensageiros (mRNAs), RNAs não codificantes de proteínas (ncRNAs), como RNAs longos não codificantes de proteínas (lncRNAs), e proteínas, revelando que a mobilidade de informação genética é um fenômeno recorrente em interações interespecíes e pode atuar como força adaptativa ao longo da evolução (Shahid et al., 2018; Loreti; Perata, 2022).

Da mesma forma que pressões bióticas moldaram o surgimento e a diversificação do parasitismo, pressões abióticas relacionadas à disponibilidade de nutrientes minerais também exerceram papel central na evolução das plantas terrestres. A adaptação aos ambientes continentais dependeu de inovações nas vias de aquisição de ferro e de outros metais essenciais, cuja baixa solubilidade em solos oxigenados impôs desafios fisiológicos consideráveis (Marschner; Romheld; Kissel, 1986; Marschner; Römheld, 1994; Kobayashi; Nishizawa, 2012; Nanda et al., 2025). Isso favoreceu o surgimento independente de estratégias distintas de captação, como a redução de ferro férrico (Fe^{3+}), altamente biodisponível na rizosfera, porém tóxico, em ferro ferroso (Fe^{2+}) e o uso de fitosideróforos (Kobayashi; Nishizawa, 2012; Connorton; Balk; Rodríguez-Celma, 2017; Vert et al., 2002). Após a absorção, a distribuição interna de metais passou a depender de famílias de transportadores especializados, como a *Zinc/Iron-regulated transporter-like Proteins* (ZIP) e *Heavy Metal ATPases* (HMA), cuja evolução estrutural e funcional refletiu a necessidade simultânea de garantir nutrição adequada e evitar toxicidade, evidenciando como mecanismos de conservação e inovação foram explorados ao longo da história evolutiva das plantas (Smith; Smith; Rosenzweig, 2014; Ma et al., 2021; Ajeesh Krishna et al., 2020; Sánchez-Pérez et al., 2025).

Figura 1 – Representação conceitual integrando processos evolutivos e inovação funcional em Archaeplastida. Representação esquemática da filogenia de Archaeplastida destacando como pressões evolutivas bióticas e abióticas contribuíram para a emergência de novas funções moleculares e interações celulares ao longo da evolução. Eventos como transferência horizontal de genes (HGT), convergência evolutiva e cooptação funcional são indicados ao longo dos ramos, incluindo possíveis contribuições de outros grupos taxonômicos (por exemplo, fungos). (A) Cooptação evolutiva de vias de defesa vegetal, nas quais componentes originalmente associados à resposta imune, como o receptor CARD1, foram recrutados para uma nova função em plantas parasitas: a formação do haustório. (B1) Convergência evolutiva do parasitismo vegetal em angiospermas, ilustrada pela evolução independente de diferentes tipos de haustório em três linhagens: (i) Orobanchaceae, (ii) Convolvulaceae e (iii) Rafflesiaceae. (B2) Convergência funcional na captação de Fe^{2+} mediada por transportadores da família ZIP/IRT em plantas terrestres e algas verdes, evidenciando origens evolutivas independentes para estratégias semelhantes de aquisição de ferro. (C) Mobilidade molecular entre organismos através do haustório, com potencial integração genômica associada a eventos de HGT. Símbolos coloridos indicam processos evolutivos mapeados na filogenia: transferência horizontal de genes (HGT) (azul), convergência evolutiva (verde), cooptação (laranja) e pressões evolutivas (amarelo).



Fonte: Autor da Pesquisa (2026).

Esse panorama torna-se ainda mais diversos quando considerados ambientes aquáticos, particularmente os sistemas marinhos, onde a disponibilidade de ferro atinge concentrações extremamente baixas e a maior parte do metal encontra-se complexada a ligantes orgânicos (Morel; Price, 2003). Nessas condições, algas e fitoplâncton desenvolveram estratégias alternativas de aquisição que diferem das vias clássicas descritas em plantas terrestres. Entre elas destacam-se mecanismos baseados na redução extracelular de ferro férrico complexado, a internalização mediada por proteínas análogas a transferrinas (fitotransferrinas) e, em alguns grupos, a utilização indireta de sideróforos produzidos por bactérias associadas, evidenciando interações ecológicas que modulam a homeostase metálica em escala oceânica (Behnke; LaRoche, 2020). Diferentemente do ambiente terrestre, onde estratégias como a acidificação da rizosfera e a secreção de quelantes são predominantes, muitas algas marinhas dependem de sistemas de alta afinidade capazes de capturar ferro em concentrações subnanomolares, refletindo pressões seletivas impostas pela baixa solubilidade do metal em águas oxigenadas (Tagliabue et al., 2017; Behnke; LaRoche, 2020). Esses contrastes indicam que a evolução das vias de aquisição de ferro ocorreu de forma convergente e modular ao longo da história dos Archaeplastida, com linhagens de algas e plantas terrestres explorando soluções moleculares distintas para um mesmo desafio fisiológico.

A integração desses processos mostra que evolução de defesa, comunicação interespecíes e nutrição mineral não são eixos independentes, mas componentes de uma mesma história adaptativa. A conquista terrestre e a subsequente diversificação das plantas criaram contextos ecológicos nos quais pressões seletivas moldaram, simultaneamente, vias de percepção de sinais, sistemas de transporte, mecanismos de defesa e estratégias de obtenção de nutrientes. Investigar esses temas em conjunto permite compreender como plantas remodelaram vias moleculares ancestrais para sustentar estilos de vida contrastantes, desde organismos autotróficos generalistas até parasitas altamente especializados.

Assim, ao combinar abordagens evolutivas, estruturais e computacionais, esta Tese busca compreender como módulos genéticos ancestrais foram reutilizados e reorganizados ao longo da evolução para manutenção da nutrição mineral, percepção ambiental e desenvolvimento parasítico em plantas. Na primeira seção, são abordados processos essenciais ao parasitismo vegetal, incluindo a formação do haustório, a mobilidade de lncRNAs entre parasita e hospedeiro e a cooptação de vias presentes em plantas não parasitas. A dinâmica de lncRNAs é investigada por meio da identificação comparativa dessas moléculas, através de abordagem de Aprendizado de Máquina (*Machine Learning* – ML), em plantas parasitas de

diferentes famílias e níveis de parasitismo, seguida por análises de expressão e inferências evolutivas. Já a cooptação de vias ancestrais é explorada através da caracterização estrutural de proteínas receptoras de plantas parasitas/não-parasitas e de triagens virtuais (*virtual screening*) com acoplamento (*docking*) molecular, também empregando abordagens baseadas em ML, para posterior identificação de moléculas indutoras de haustório, sempre em análises comparativas (parasitas VS não-parasitas). Na segunda seção, são examinadas a evolução das estratégias de captação de ferro e a organização estrutural de transportadores de ferro, com ênfase na convergência funcional em diferentes linhagens vegetais, e nas especializações fisiológicas da família HMA caracterizadas em feijão (*Phaseolus vulgaris*) sob condições de estresse por metais. Em conjunto, os estudos reunidos nesta tese buscam esclarecer como diferentes vias moleculares foram remodeladas ao longo da evolução para sustentar estilos de vida contrastantes, revelando princípios amplos de adaptação, inovação e reorganização funcional em plantas.

2. OBJETIVOS

2.1. OBJETIVO GERAL

Investigar como componentes do arcabouço genético ancestral das plantas foram remodelados ao longo da evolução para sustentar a adaptação a ambientes limitantes e o estabelecimento do parasitismo vegetal, integrando abordagens evolutivas, estruturais e computacionais.

2.2. OBJETIVOS ESPECÍFICOS

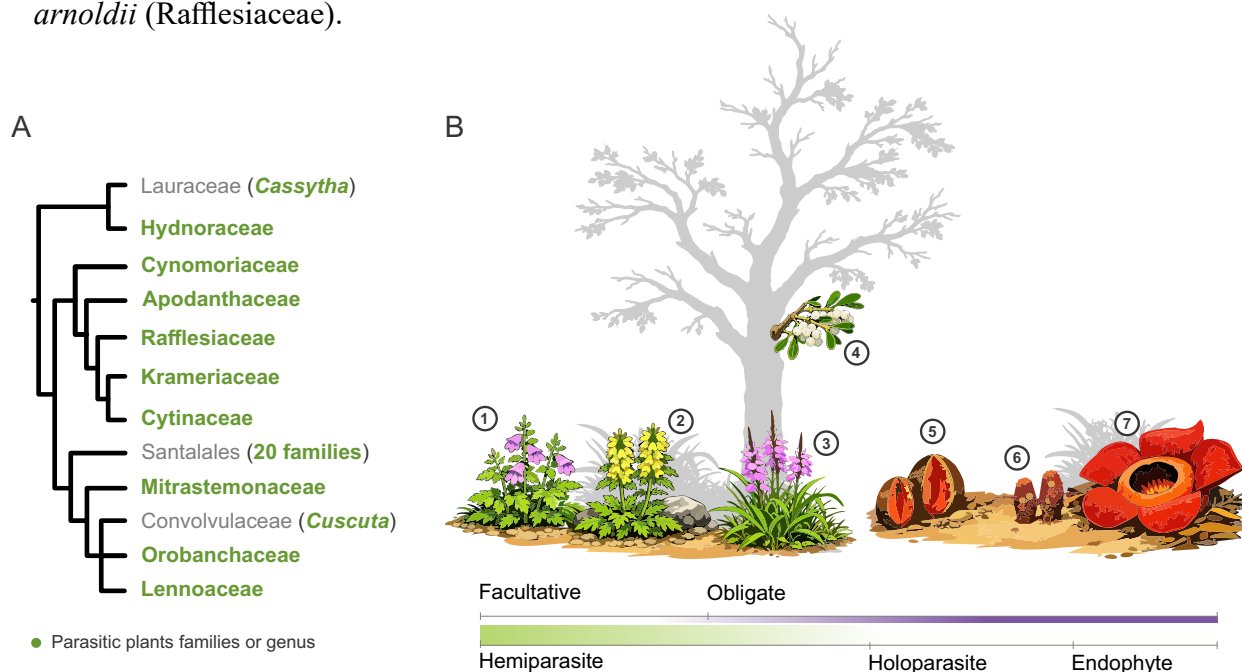
- Analisar a mobilidade e o potencial papel regulatório de lncRNAs na comunicação molecular entre parasita e hospedeiro, através da identificação computacional e de análises filogenéticas;
- Revisar a origem e a história evolutiva de proteínas relacionadas à defesa em plantas, avaliando processos de conservação e diversificação ao longo da evolução de Archaeplastida;
- Investigar os mecanismos moleculares associados à formação do haustório, incluindo a coaptação de vias ancestrais de defesa em plantas parasitas, por meio de análises comparativas de receptores de membrana entre espécies parasitas e não parasitas;
- Identificar e caracterizar computacionalmente regiões potenciais de ligação (*binding sites*) a moléculas indutoras de haustório e a sinais redox associados à defesa em receptores de membrana homólogos de plantas parasitas e não parasitas, por meio de análises estruturais comparativas;
- Identificar e priorizar moléculas putativas não indutoras ou inibidoras da formação do haustório por meio de abordagens integradas de modelagem estrutural, *docking* molecular e *virtual screening* guiada por Inteligência Artificial (IA);
- Investigar a origem profunda e a diversificação funcional de transportadores metálicos da família ZIP/IRT, com foco na aquisição de ferro, em diferentes linhagens vegetais;
- Avaliar especializações estruturais e respostas fisiológicas associadas a transportadores HMA sob condições de estresse por metais;
- Integrar evidências evolutivas e funcionais para compreender como módulos genéticos ancestrais foram reutilizados e reorganizados na transição para estilos de vida contrastantes em plantas.

3. SEÇÃO I – BASES MOLECULARES DO PARASITISMO VEGETAL: lncRNAs, DEFESA E PERCEPÇÃO DE SINAIS

3.1. REFERENCIAL TEÓRICO

Plantas parasitas representam uma das transformações ecológicas e evolutivas mais marcantes dentro das plantas com flores. Embora correspondam a apenas cerca de um por cento das angiospermas, surgiram de maneira independente diversas vezes, com estimativas indicando pelo menos doze origens distintas ao longo da história evolutiva do grupo (Figura 2A) (Westwood et al., 2010; Yoshida et al., 2016). Essas origens múltiplas deram origem a um conjunto extremamente diverso de formas, tamanhos, comportamentos ecológicos e graus de dependência em relação ao hospedeiro, representando um dos casos mais robustos de convergência evolutiva em plantas.

Figura 2 – Diversidade e classificação de plantas parasitas. (A) Cladograma ilustrando a distribuição filogenética de famílias que incluem espécies de plantas parasitas. (B) Exemplos representativos distribuídos segundo diferentes níveis de dependência em relação ao hospedeiro. Hemiparasitas facultativos: 1 - *Phtheirospermum japonicum* (Orobanchaceae); 2 - *Pedicularis foliosa* (Orobanchaceae). Hemiparasitas obrigatórios: 3 - *Striga hermonthica* (Orobanchaceae); 4 - *Viscum album* (Viscaceae, Santalales). Holoparasitas: 5 - *Hydnora africana* (Hydnoraceae); 6 - *Cynomorium coccineum* (Cynomoriaceae). Endófito: 7 - *Rafflesia arnoldii* (Rafflesiaceae).



Fonte: Autor da Pesquisa (2026)

A família Orobanchaceae, por exemplo, abrange desde hemiparasitas verdes capazes de fotossíntese até holoparasitas totalmente desprovidas dessa capacidade, enquanto Convolvulaceae inclui gêneros como *Cuscuta*, caracterizados por um corpo altamente reduzido, perda quase total de fotossíntese e um parasitismo generalista que atinge múltiplas famílias botânicas. Em outros casos extremos, como nas parasitas da família Rafflesiaceae, observa-se a perda quase total de estruturas vegetativas e o surgimento de flores gigantescas, dependentes exclusivamente do hospedeiro para obtenção de recursos (Westwood et al., 2010; Cai et al., 2021)

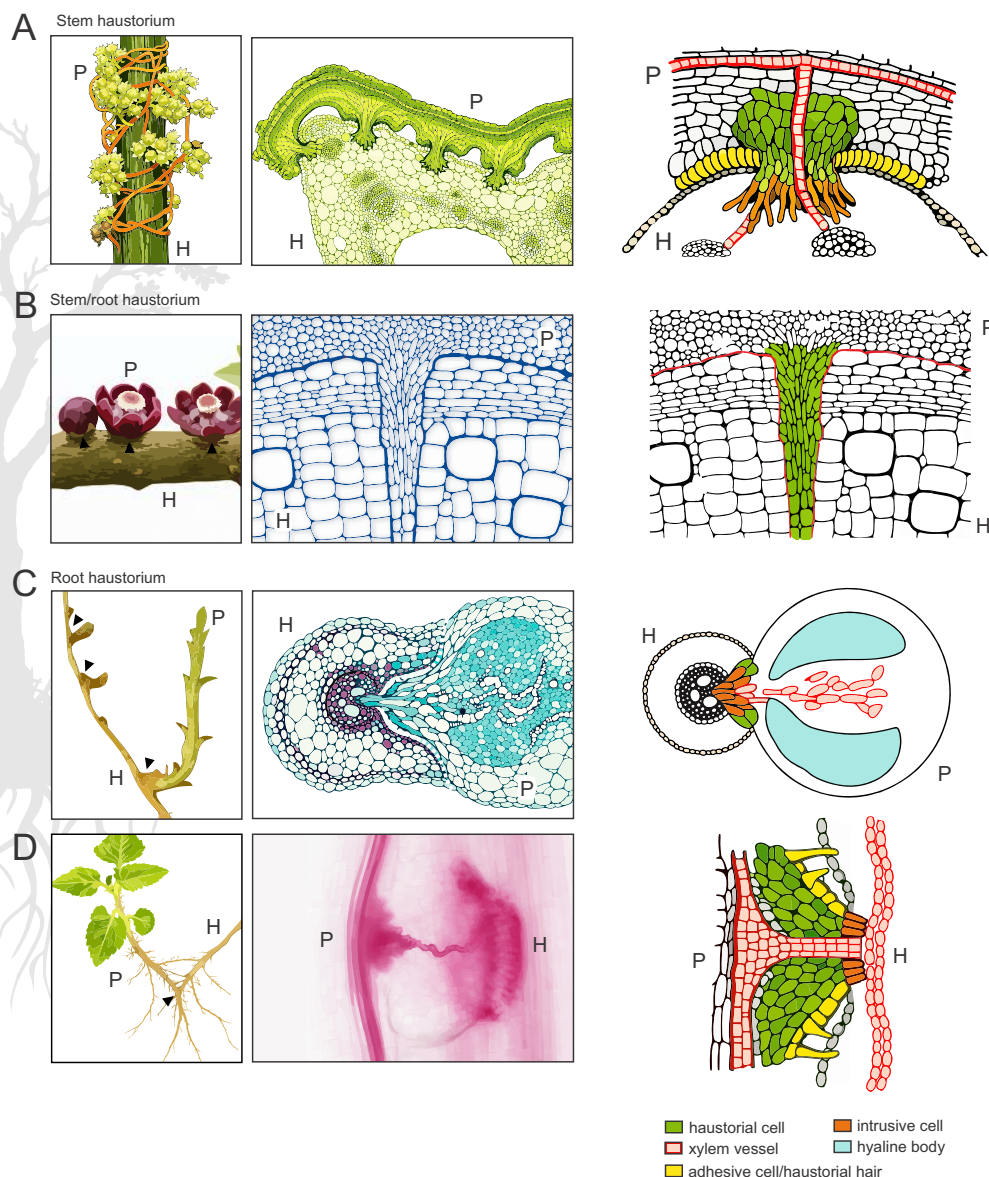
O impacto das plantas parasitas na agricultura é severo e amplamente documentado. Espécies como *Striga hermonthica* (Orobanchaceae) causam prejuízos bilionários anuais em culturas de milho, sorgo e arroz na África Subsaariana, onde são consideradas uma das maiores ameaças à segurança alimentar local (Runo; Kuria, 2018; Runo; Wicke; Thorogood, 2025). O problema tende a se agravar sob condições de mudanças climáticas, aumento populacional e maior pressão por alimentos. Entretanto, apesar de seu impacto negativo, plantas parasitas também são modelos extraordinários para a compreensão da evolução vegetal, da modificação de genomas e da adaptação metabólica. Muitas exibem padrões impressionantes de compactação genômica, perda de vias bioquímicas e aquisição de genes por transferência horizontal (HGT), constituindo sistemas de estudo privilegiados para investigar como genomas se moldam diante de novas pressões ecológicas.

A diversidade de plantas parasitas expressa uma ampla gama de modificações anatômicas, fisiológicas e genômicas que refletem diferentes graus de dependência do hospedeiro (Figura 2B). Em uma extremidade desse gradiente encontram-se os hemiparasitas facultativos, capazes de completar o ciclo de vida de forma autônoma, mas que exploram o parasitismo quando esse modo de obtenção de recursos confere vantagens ecológicas. Em um nível intermediário situam-se os hemiparasitas obrigatórios, que preservam fotossíntese e estruturas aéreas, embora dependam do hospedeiro principalmente para água e nutrientes minerais. Na extremidade oposta do espectro estão as holoparasitas obrigatórias, que perderam completamente a capacidade fotossintética, exibem forte redução de órgãos vegetativos e apresentam plastomas profundamente degradados, com perda extensa de genes (Nickrent, 2020; Yoshida; Kee, 2021). Em casos especialmente derivados, como nas Rafflesiaceae, essa erosão funcional atinge níveis extremos, incluindo relatos de perda quase total do plastoma e degradação ampla de vias metabólicas associadas ao cloroplasto (Cai et al., 2021; Molina et al., 2014). Os parasitas endófitos representam o ápice desse processo dentro do holoparasitismo

(Figura 2B), pois o corpo vegetativo é praticamente eliminado, restando apenas estruturas reprodutivas (flores) emergindo do tecido hospedeiro (Nickrent, 2020; Yoshida; Kee, 2021). Dentro da família Orobanchaceae, o hemiparasitismo é particularmente predominante, abrangendo cerca de 88% das espécies (McNeal et al., 2013). Nesse caso, a combinação entre forte diversificação interna e distribuição quase cosmopolita indica que esse modo de vida representa uma estratégia ecológica evolutivamente estável e bem-sucedida, e não um estágio intermediário inevitável rumo ao holoparasitismo o que levanta questões evolutivas acerca do estabelecimento do estilo de vida “ideal” a um determinado ambiente.

A característica definidora do parasitismo vegetal é a presença do haustório, um órgão especializado que se forma em diferentes regiões do corpo do parasita e estabelece a conexão fisiológica entre o ele e seu hospedeiro (Figura 3). O haustório permite adesão, penetração e contato vascular, constituindo o canal pelo qual água, solutos, metabólitos e moléculas de informação transitam entre os organismos (Kokla; Melnyk, 2018; Yoshida et al., 2016). Diferentes linhagens parasitas desenvolveram tipos distintos de haustório (Figura 3). No gênero *Striga*, por exemplo, o haustório é terminal e se forma na extremidade da raiz após a percepção de moléculas sinalizadoras liberadas pelas raízes do hospedeiro, como estrigolactonas (Toh et al., 2015). Em *Cuscuta* (Convolvulaceae), por outro lado, o haustório é caulinar e se desenvolve diretamente no caule após o contato físico do filamento com o hospedeiro. Nessa linhagem, o parasitismo envolve a perda completa da raiz funcional e uma estratégia baseada em movimento circular até encontrar um hospedeiro adequado, guiada não apenas por estímulos mecânicos, mas também por sinais ambientais. Evidências indicam que plântulas de *Cuscuta* respondem a compostos voláteis liberados por possíveis hospedeiros, além de integrarem sinais luminosos, como altas razões de luz vermelho-distante/vermelho e estímulos de luz azul, que contribuem para a orientação do crescimento, o comportamento de enroscamento do caule e a subsequente indução do haustório (Kelly, 1990, 1992; Clarke et al., 2019; Jhu; Sinha, 2022). *Phtheirospermum japonicum* (Orobanchaceae), um modelo experimental amplamente utilizado no estudo genômico de plantas parasitas, apresenta haustórios laterais que se desenvolvem a partir dos tecidos radiculares em resposta a sinais químicos do hospedeiro, mas mantém capacidade de desenvolvimento autônomo na ausência deste, caracterizando-o como parasita facultativo (Cui et al., 2020).

Figura 3 – Tipos de haustório em plantas parasitas. Caixas à esquerda mostram a interação entre planta parasita e planta hospedeira, com setas indicando a posição do haustório; à direita, apresenta-se o corte anatômico correspondente. O esquema destaca células relevantes para o desenvolvimento haustorial, incluindo células xilemáticas (vermelho), que estabelecem conexão com o sistema vascular do hospedeiro, e células intrusivas (laranja), responsáveis pela penetração nos tecidos hospedeiros. (A) Haustório caulinar de *Cuscuta* (Convolvulaceae); (B) Haustório típico de planta endofítica, exemplificado por *Pilosyles* (Apodanthaceae); (C) Haustório terminal de raiz em *Striga* (Orobanchaceae); (D) Haustório lateral de raiz em *Phtheirospermum japonicum* (Orobanchaceae). P = planta parasita; H = planta hospedeira.



Fonte: Autor da Pesquisa (2026), adaptado de Ichihashi et al., (2015); Alonso, (2016); Ishida et al., (2016); Yoshida et al., (2016); Shimizu; Aoki, (2019); Thorogood et al., (2021).

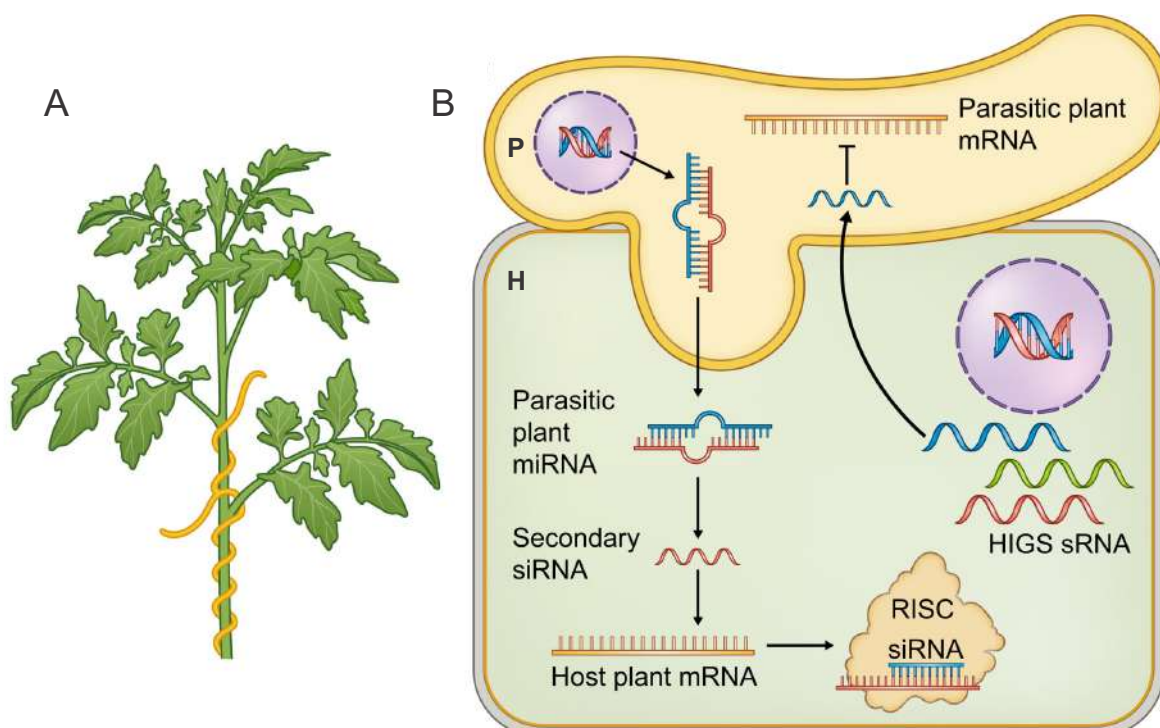
A formação do haustório tem início quando o parasita reconhece compostos denominados Fatores Indutores de Haustório (HIFs), que incluem quinonas, flavonoides e fenóis derivados de lignina e de componentes da parede celular (Bandaranayake; Yoder, 2013;

Clarke et al., 2019; Goyet et al., 2019). Entre esses compostos, 2,6-dimethoxy-1,4-benzoquinone (DMBQ) é um dos mais estudados em parasitas radiculares e atua como um potente sinal indutor em *P. japonicum* e outras hemiparasitas, sendo amplamente utilizado em experimentos que investigam os estágios iniciais do desenvolvimento haustorial (Chang; Lynn, 1986; Ishida et al., 2016; Laohavisit et al., 2020). Embora ainda haja muitas lacunas sobre a natureza exata dos receptores envolvidos na percepção desses sinais, evidências crescentes indicam que a detecção de HIFs está associada a receptores de superfície de membrana pertencentes à ampla família do tipo quinase (*Receptor-like Kinases* - RLKs) (Laohavisit et al., 2020). Esse conjunto de proteínas inclui os receptores de reconhecimento de padrões (*Pattern-Recognition Receptors* - PRRs), moléculas centrais da imunidade inata vegetal, responsáveis por reconhecer padrões moleculares conservados e iniciar respostas de defesa (Ngou; Ding; Jones, 2022). A maioria dos PRRs é formada por RLKs (PRR-RLKs) com domínios extracelulares ricos em repetições de leucina (LRR-RLKs), que funcionam como plataformas estruturais para o reconhecimento de ligantes e a ativação de vias de sinalização dependentes de cálcio, proteínas quinase ativadas por mitógenos (MAP quinases) e hormônios (Jones; Staskawicz; Dangl, 2024). O fato de que receptores estruturalmente semelhantes a esses PRRs (homólogos), tradicionalmente associados ao reconhecimento de patógenos ou moléculas derivadas de dano, tenham sido identificados em parasitas como possíveis sensores de HIFs sugere um exemplo claro de reutilização evolutiva de módulos ancestrais de defesa para fins de desenvolvimento haustorial.

Em *Arabidopsis thaliana*, uma espécie angiosperma modelo e não parasita, o receptor PRR-RLK *cannot respond to* DMBQ 1 (CARD1; também conhecido como HPCA1) desempenha papel importante na percepção de estímulos externos e na ativação de respostas imunes dependentes de cálcio. Estudos comparativos revelaram que proteínas semelhantes a AtCARD1, denominadas proteínas tipo CARD (*CARD-like proteins* - CADLs), estão presentes em plantas parasitas e podem ter sido coaptadas ao longo da evolução para reconhecer HIFs, funcionando como gatilhos moleculares que iniciam a formação do haustório (Laohavisit et al., 2020; Wu et al., 2020). Esse possível redirecionamento de uma via ancestral de defesa para um novo papel no parasitismo representa um caso exemplar de reutilização evolutiva de mecanismos pré-existentes, porém, a identidade completa dos ligantes, os componentes intermediários da via e os detalhes da transdução do sinal, assim como reconhecimento de HIFs, ainda permanecem pouco compreendidos.

O haustório não é apenas uma estrutura de troca de nutrientes, mas também um canal de comunicação molecular bidirecional (Figura 4). Estudos pioneiros demonstraram que, em *Cuscuta*, ocorre transferência maciça de mRNAs, pequenos RNAs (sRNAs) e proteínas entre parasita e hospedeiro, resultando em modulação direta da expressão gênica no organismo receptor (Shahid et al., 2018). Esses fluxos incluem o movimento de sRNAs capazes de silenciar genes no hospedeiro, um fenômeno semelhante ao que ocorre em interações planta-patógeno, como bactéria por exemplo (Figura 4) (González Plaza, 2020). Outros trabalhos revelaram que esse trânsito de moléculas não se restringe a *Cuscuta*, em *Orobanchae aegyptiaca* (Orobanchaceae) também foram detectadas trocas de sRNAs e possíveis eventos regulatórios mediados por moléculas móveis (Dubey et al., 2017).

Figura 4 – Transporte interespecífico de pequenos RNAs (sRNAs) entre plantas parasitas e seus hospedeiros. (A) Representação da planta parasita *Cuscuta campestris* (em amarelo). (B) *Cuscuta* produz microRNAs (miRNAs) específicos de 22 nucleotídeos que são transferidos para as células da planta hospedeira, onde silenciam genes-alvo envolvidos na defesa por meio de interferência por RNA, com produção de pequenos RNAs interferentes secundários (siRNAs). Em resposta, as células da planta hospedeira produzem pequenos RNAs associados ao silenciamento gênico induzido pelo hospedeiro (HIGS), direcionados aos RNAs mensageiros (mRNAs) da planta parasita. P = planta parasita; H = planta hospedeira.



Fonte: Loreti; Perata, (2022).

Outra classe de moléculas recentemente descrita como móvel através do haustório são os lncRNAs. Embora a mobilidade de mRNAs e sRNAs já esteja bem estabelecida em diversas interações parasita-hospedeiro, a participação de lncRNAs nesse processo começou a ser elucidada apenas recentemente. Em *Cuscuta*, espécie conhecida por estabelecer conexões vasculares altamente eficientes com seus hospedeiros, foi demonstrada a transferência bidirecional de lncRNAs entre parasita e planta hospedeira (Wu et al., 2022). Análises funcionais indicaram que os lncRNAs móveis do parasita estão associados a processos metabólicos, sinalização e respostas a estímulos, enquanto lncRNAs provenientes do hospedeiro se relacionam principalmente a funções organelares, sugerindo papéis regulatórios distintos na interação (Wu et al., 2022). Esses achados revelam que lncRNAs não apenas podem atravessar a interface haustorial, mas também podem atuar como reguladores importantes na coordenação fisiológica da relação parasita-hospedeiro. Considerando que lncRNAs exercem funções centrais em plantas não parasitas, como modulação transcricional, controle epigenético e atuação à distância (Chen et al., 2021; Cui, 2023; Yang et al., 2023), torna-se plausível que também desempenhem papéis funcionais no parasitismo vegetal.

A relação entre plantas e plantas parasitas/patógenos é moldada por um constante conflito molecular. As plantas não parasitas dispõem de um sistema imune altamente sofisticado que envolve múltiplas camadas de defesa. A primeira delas, conhecida como imunidade desencadeada por padrões ou PTI (*Pattern-Triggered Immunity*), é ativada quando receptores de superfície reconhecem padrões moleculares conservados associados a microrganismos ou danos, levando à produção de espécies reativas de oxigênio (ROS), ao reforço da parede celular e à ativação de cascatas de sinalização que limitam a invasão. A segunda camada, chamada ETI (*Effector-Triggered Immunity*), é desencadeada em compartimentos intracelulares quando proteínas de resistência reconhecem efetores liberados pelo organismo invasor, resultando em uma resposta ainda mais intensa, frequentemente acompanhada de morte celular localizada como forma de bloquear a progressão da infecção (Yuan et al., 2021; Yu et al., 2024). Parasitas/patógenos, entretanto, precisam suprimir, modular ou contornar esse arsenal defensivo para estabelecer contato funcional com o hospedeiro e garantir o fluxo contínuo de água, nutrientes e sinais regulatórios. Nesse contexto, a troca de RNAs móveis, a secreção de efetores, a interferência em vias hormonais e a manipulação de circuitos de sinalização imune desempenham papéis centrais tanto para o estabelecimento inicial da conexão quanto para a manutenção de relações parasíticas estáveis ao longo da evolução.

Diante desse conjunto de características que envolvem múltiplas origens evolutivas, perda funcional progressiva, reorganização profunda de genomas, remodelação de vias imunes e estabelecimento de comunicação molecular bidirecional com o hospedeiro, plantas parasitas se destacam como modelos excepcionais para investigar processos fundamentais de inovação biológica. A interface haustorial reúne mecanismos de reconhecimento químico, circuitos derivados de vias de defesa, translocação de moléculas regulatórias e reprogramação celular, compondo um sistema altamente integrado no qual diferentes classes de RNAs, receptores, sinais e respostas convergem para sustentar um estilo de vida heterotrófico, mesmo quando os mecanismos de desenvolvimento e tipos de haustório divergem entre linhagens de parasitas. Nesse contexto, compreender como essas plantas percebem seus hospedeiros, como remodelam vias ancestrais e como utilizam moléculas móveis, incluindo lncRNAs, para influenciar a expressão gênica, permite revelar princípios gerais de comunicação interespecíes e evolução de sistemas regulatórios complexos.

Nesta seção, reunimos três linhas de investigação que, em conjunto, ampliam de maneira significativa a compreensão dos mecanismos moleculares que sustentam o parasitismo vegetal. Primeiro, apresentamos evidências que sugerem que lncRNAs podem atravessar a interface haustorial e estabelecer trânsito entre parasita e hospedeiro dentro de um contexto evolutivo, trazendo à tona a possibilidade de que eventos de transferência horizontal envolvendo lncRNAs possam ocorrer entre organismos eucarióticos intimamente conectados por este tipo de interação e não por relações filogenéticas. Em seguida, revisamos a evolução de proteínas relacionadas à defesa nas plantas, destacando como módulos responsáveis pela percepção de padrões moleculares e pela ativação de respostas imunes se diversificaram ao longo da história evolutiva de Archaeplastida. Embora esse conjunto de proteínas seja discutido no contexto de espécies não parasitas, ele fornece uma base conceitual essencial para compreender como vias ancestrais de sinalização podem ser modificadas, redirecionadas ou reutilizadas em diferentes contextos biológicos, incluindo o parasitismo. Por fim, avançamos na caracterização comparativa de receptores do tipo PRR-RLK em plantas parasitas e não parasitas, integrando modelagem estrutural, predição de sítios de ligação e docking molecular para explorar regiões potenciais de interação com HIFs. Essas análises foram associadas a estratégias de *virtual screening*, visando priorizar compostos candidatos a HIFs. Em conjunto, esses resultados revelam um panorama no qual comunicação molecular, evolução de sistemas de defesa e inovação estrutural convergem para permitir o estabelecimento e a diversificação do parasitismo em angiospermas.

3.2. CAPÍTULO 1 – INVESTIGATING lncRNA AT DIFFERENT LEVELS OF PARASITISM IN PARASITIC PLANTS

Este capítulo apresenta o manuscrito ainda inédito “*Investigating lncRNAs at different levels of parasitism in parasitic plants*”, preparado para submissão como VIEWPOINT na revista *Physiologia Plantarum* (fator de impacto 3.6). O artigo discute como diferentes graus de parasitismo moldam a diversidade, a expressão e a origem putativa de lncRNAs em nove espécies parasitas, comparadas a espécies não parasitas. A partir de dados genômicos e transcriptômicos, o trabalho identifica padrões marcantes de abundância, perfis de expressão associados a órgãos como o haustório, diferenças estruturais capturadas por análises de k-mer e PCA, e a forte associação de elementos LTR na origem desses transcritos. O estudo também revela múltiplos indícios de HGT de lncRNAs entre parasitas e seus hospedeiros, incluindo casos robustos em *Sapria himalayana*, configurando evidências que podem representar um dos primeiros registros desse fenômeno entre eucariontes. Esse conjunto de análises sustenta a proposta central do manuscrito: lncRNAs são componentes relevantes da evolução do parasitismo e representam um eixo ainda pouco explorado da comunicação molecular entre parasitas e hospedeiros.

Investigating lncRNAs at different levels of parasitism in parasitic plants

Wenderson Felipe Costa Rodrigues¹; Laura Oliveira Pires¹; Luiz-Eduardo Del-Bem²
Juliane Karine Ishida ^{2*}

¹ Graduate Program in Plant Biology, Department of Botany, Institute of Biological Sciences (ICB), Federal University of Minas Gerais (UFMG), 31270-901 Belo Horizonte, Brazil

² Department of Botany, Institute of Biological Sciences (ICB), Federal University of Minas Gerais (UFMG), 31270-901 Belo Horizonte, Brazil

*Correspondence: Juliane Karine Ishida: julianeishida@ufmg.br

Abstract

Parasitic plants exhibit unique evolutionary adaptations, including the development of haustoria and varying degrees of organ reduction. Here, we analyzed genomic and transcriptomic datasets from nine parasitic species representing hemi-, holo-, and endoparasitic lifestyles to investigate patterns of lncRNA abundance, expression, and evolutionary origin. Holoparasitic species exhibited the highest lncRNA counts, whereas endoparasites showed pronounced reductions. Expression analyses revealed tissue-specific regulatory patterns, with elevated lncRNA activity in haustorial tissues of *Cuscuta australis*, contrasting with predominant repression in hemiparasites. Phylogenetic analyses identified multiple clades consistent with horizontal gene transfer (HGT) involving lncRNA across all levels of parasitism, particularly in the endoparasite *Sapria himalayana* and, unexpectedly, for the first time in the hemiparasite *Phtheirospermum japonicum*, as well as in *Striga asiatica*, *Pedicularis kansuensis*, and the holoparasitic species *Cuscuta australis*, and *Orobanche aegyptiaca*. Notably, our results provide the first evidence of horizontal transfer of lncRNA between plants. Together, these findings indicate that lncRNAs may contribute to parasitic plant adaptation and genome evolution, expanding the scope of HGT beyond protein-coding genes.

Keywords: Parasitic plants; long non-coding RNAs; horizontal gene transfer.

Main

Long non-coding RNAs (lncRNAs) are increasingly recognized as key regulatory molecules in various organisms^{1,2}. These non-coding transcripts, generally exceeding 200 nucleotides, exhibit diverse secondary structures, localization patterns, and regulatory capacities within the cell³. In plants, lncRNAs have been linked to the regulation of diverse developmental processes, including organ development and maturation. They are involved in critical stages such as flower development, root growth, and stress responses, often by modulating gene expression at transcriptional or post-transcriptional

1 levels⁴. Notably, lncRNAs exhibit the ability to function at locations distant from their
2 synthesis site, similar to other types of mobile RNA molecules, such as small interfering
3 RNAs (siRNAs) and messenger RNAs (mRNAs)⁵. This capacity for long-distance
4 movement suggests that lncRNAs may act as signaling molecules, coordinating
5 developmental and physiological processes throughout the plant.

6 The established roles of lncRNAs in non-parasitic plants raise important questions
7 about their potential involvement in parasitic species, particularly considering their
8 capacity for mobility, which may facilitate complex host–parasite interactions. Parasitic
9 plants are angiosperms with specialized adaptations that enable them to extract nutrients
10 from host plants through structures known as haustoria^{6,7}. These plants vary in their
11 degree of dependence on the host, ranging from hemiparasites that can photosynthesize
12 but supplement their nutrition from hosts to holoparasites that rely exclusively on host-
13 derived resources, with endoparasitic lineages representing an extreme within the
14 holoparasitic spectrum, as they live almost entirely embedded within host tissues and
15 typically emerge only transiently for reproduction^{6,8}. Prior studies have demonstrated that
16 lncRNAs play essential roles in plant responses to biotic stress, including interactions with
17 fungi or nematodes, where they contribute to pathogen resistance, gene regulation, and
18 sexual differentiation^{9,10}. These findings suggest that lncRNAs could similarly influence
19 key aspects of parasitism, including haustorial development and function.

20 In addition, beyond their role in parasitic interaction, lncRNAs are likely involved in
21 the broader developmental processes of parasitic plants, potentially shaping the
22 formation and differentiation of organs. Parasitic plants often exhibit modified or reduced
23 structures, such as diminished roots or absent leaves, which are hallmark adaptations to
24 their parasitic lifestyle^{11,12}. *Cuscuta*, a holoparasitic genus, and *Sapria*, an endoparasitic
25 species, exhibit notable alterations in floral growth regulation, reflecting evolutionary
26 adaptations to their parasitic lifestyles. For instance, *Sapria* demonstrates a striking
27 temporal mismatch in flowering relative to its host, which may represent a divergence in
28 developmental timing mechanisms¹³. This mismatch is accompanied by low fertilization
29 rates and the production of exceptionally large, elaborate flowers, likely as compensatory
30 adaptations for its unique reproductive strategy. In contrast, *Cuscuta* exhibits floral growth
31 tightly coordinated with the availability of host-derived resources, often synchronizing its
32 reproductive phase with host availability and environmental cues, potentially mediated by
33 host-derived hormonal signals¹⁴. These traits suggest a fine-tuned regulatory network in
34 which lncRNAs might play a central role, coordinating the expression of genes involved
35 in floral development and compensating for the extensive gene loss documented in the
36 genomes of these parasitic plants^{15,16}. Additionally, in *Cuscuta*, the bidirectional
37 movement of lncRNAs between the parasite and its host has been identified¹⁷, suggesting
38 that these molecules may act as mediators of interspecies communication, facilitating the
39 transfer of regulatory information crucial for parasitism. However, despite their potential

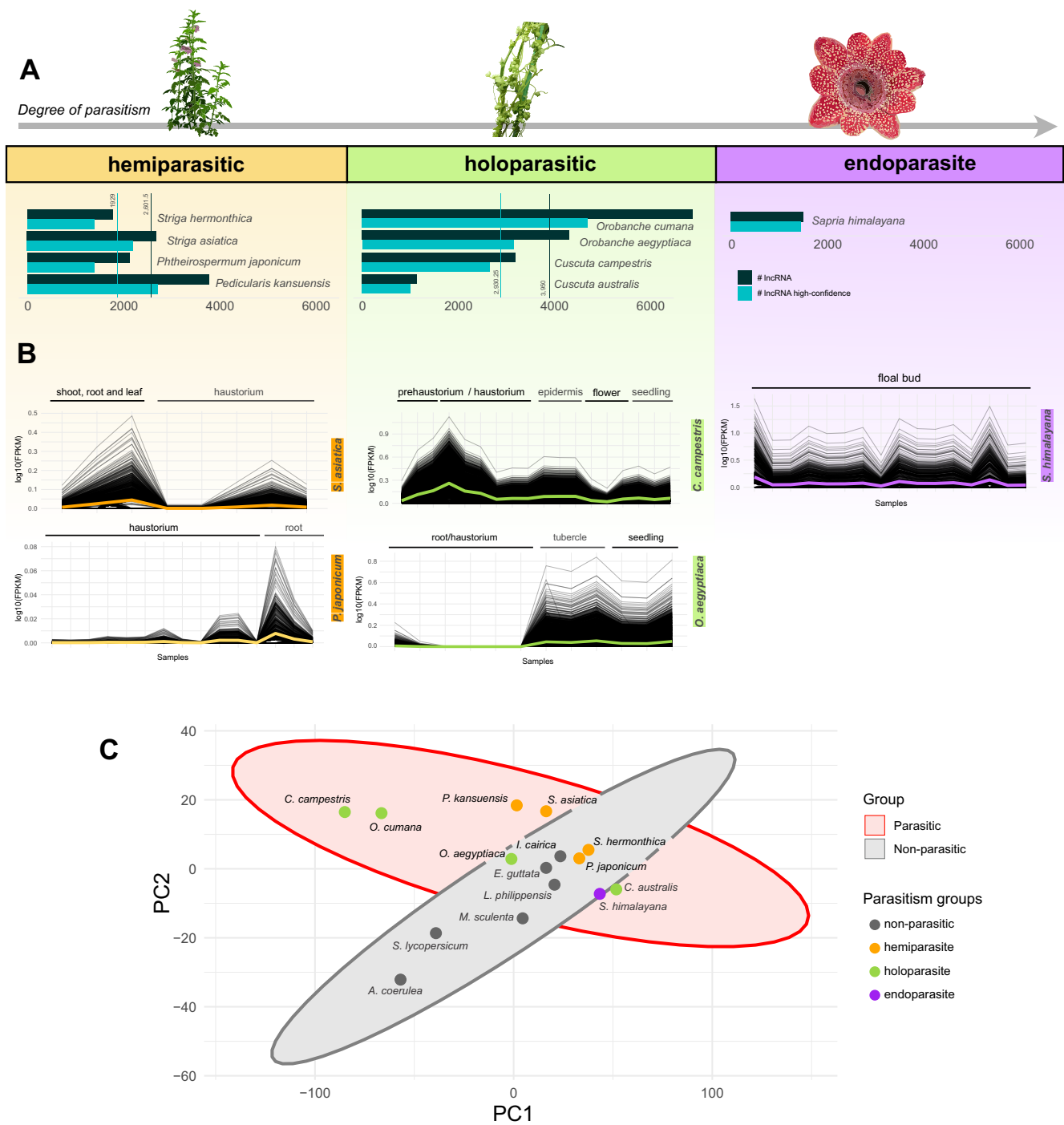
significance, there is a substantial gap in our understanding of the roles of lncRNAs in plant–plant interactions, as no direct evidence currently links these molecules to the specific developmental modifications associated with parasitism, leaving this aspect of their biology largely unexplored.

To begin addressing this knowledge gap, we aimed to identify and characterize the lncRNAs, by leveraging genomic data and bioinformatic tools, in nine parasitic plant species with varying parasitic lifestyles, including hemi-, holo-, and endoparasites (Table S1). Our analysis revealed that there was no consistent trend in lncRNA abundance between parasitic and non-parasitic species. For instance, the basal eudicot *Aquilegia coerulea* contained approximately 4,728 lncRNAs, a number significantly higher than most parasitic species, with the exception of the holoparasitic genus *Orobanche* (Figure S1A). Species within Solanales and Lamiales, a clade that contains the parasitic plant families Orobanchaceae and Convolvulaceae, respectively, exhibited the highest lncRNA counts (Figure S1A). In contrast, the Malpighiales clade, which includes the endoparasitic *Sapria himalayana*, displayed a significantly lower average of around 2,035 lncRNAs (Figure S1A). Phylogenetically related non-parasitic species generally possess higher lncRNA counts compared to parasitic species (Figure S1A), suggesting that parasitism may influence lncRNA abundance.

Examining lncRNA counts across different parasitic lifestyles, holoparasitic species exhibited the highest lncRNA numbers, averaging 51% more than hemiparasites (Figure 1A; Figure S1B; Table S1). Through machine learning–based analyses (see Methods S1), lncRNA identification from *de novo* transcript assemblies aligned to reference genomes followed the same pattern when comparing holo- and hemiparasites (Figure S2A). In addition, when classified, lncRNA genes in both parasitic lifestyles were predominantly antisense, intergenic, and exonic, with no clear differential pattern observed between them (Figure S2B). At the extreme among holoparasitic species, the endoparasite *S. himalayana* showed a marked reduction, with only 1,507 lncRNAs identified (Figure 1A; Figure S1B)¹⁸. This reduction persisted even after normalization by genome size, establishing the trend: holoparasites > hemiparasites > endoparasites (Figure S1C). These findings suggest an association between parasitic lifestyle and lncRNA abundance, with holoparasites displaying notably higher counts, whereas endoparasites such as *Sapria* exhibit a significant reduction despite their large genomes.

To improve accuracy, high-confidence lncRNAs were also identified by removing other classes of non-coding RNAs or small RNA precursors (see Methods S1) from the initial dataset (Table S1), enabling a more robust comparison across parasitic groups. Although this refinement did not alter the overall trends (Figure 1A), restricting the analysis to high-confidence lncRNAs substantially reduced the total number of identified lncRNAs and also markedly increased the divergence relative to those detected using the machine learning-based approach (Figure S2), potentially resulting in the loss of

1 biologically relevant information. Therefore, subsequent analyses were conducted using
 2 the total set of identified in the initial screening approach to ensure a more comprehensive
 3 dataset.



4
 5 **Figure 1. lncRNA landscape across parasitic plant lifestyles.** (A) Bar plot showing the
 6 total number of identified lncRNAs and high-confidence lncRNAs across parasitic plant
 7 species representing different levels of parasitism. (B) Expression profiles of lncRNAs

1 across distinct tissues. (C) Principal component analysis (PCA) highlighting differences
2 in lncRNA profiles between parasitic and non-parasitic species.

3 To further investigate functional dimensions of parasitic lifestyles, we analyzed
4 lncRNA expression profiles using publicly available RNA-seq datasets and examined
5 qualitative features, including k-mer composition, predicted secondary structure, and
6 transcript length distributions across parasitic and non-parasitic species. These
7 complementary analyses allowed us to assess not only expression dynamics but also
8 intrinsic sequence characteristics that may distinguish lncRNA repertoires among
9 lifestyles.

10 In hemiparasitic species such as *Phtheirospermum japonicum*, lncRNA expression
11 was predominantly detected in roots not actively engaged in host interaction (Figure 1B),
12 suggesting that a substantial fraction of these transcripts may be associated with general
13 developmental or physiological processes rather than functions directly related to
14 haustorium formation. A similar trend was observed in *Striga asiatica*, where higher
15 expression levels occurred in leaf, root, and shoot tissues compared to haustorial tissues
16 (Figure 1B). For both species, differential expression analyses revealed a predominance
17 of negatively regulated lncRNAs in haustorial tissues relative to other organs, although
18 this pattern was less pronounced in *Striga* (Figure S3). Together, these results suggest
19 that, in hemiparasites, haustorium formation may involve selective repression of a
20 broader developmental lncRNA program rather than widespread activation of novel
21 lncRNA loci. In contrast, lncRNA expression in holoparasitic species appeared more
22 evenly distributed across organs, although this pattern varied among species (Figure 1B).
23 In *Cuscuta campestris*, lncRNA expression was highest in haustorial tissues, indicating a
24 strong association with haustorial activity and possibly reflecting a greater reliance on
25 transcriptional regulation during host attachment and resource acquisition. Consistently,
26 differential expression analyses showed a predominance of upregulated lncRNAs in
27 *Cuscuta* haustoria, contrasting with the downregulation pattern observed in
28 hemiparasites (Figure S3C). In *Orobancha cumana*, however, lncRNA expression was
29 more prominent in non-haustorial tissues, suggesting broader roles beyond haustorial
30 specialization. Interestingly, floral lncRNA expression was comparatively low in both *S.*
31 *himalayana* and *Cuscuta*, despite the central importance of reproductive structures in
32 their life cycles. This pattern may indicate either a reduced contribution of lncRNA-
33 mediated regulation in floral tissues of obligate parasites or a shift toward more
34 constitutive regulatory programs in these highly specialized lineages.

35 To investigate structural and sequence features of lncRNAs, we performed a
36 Principal Component Analysis (PCA), which revealed that lncRNAs in parasitic species
37 exhibit fundamentally different characteristics compared to those in non-parasitic species,
38 suggesting that parasitism directly influences lncRNA structural and compositional traits
39 (Figure 1C). In addition, to explore the potential origins of lncRNAs in parasitic plants, we

analyzed repetitive motifs within their genomes, given that lncRNA emergence is frequently linked to repetitive sequences across diverse organisms¹⁹. No strong correlation was observed between the overall proportion of repetitive elements and the total number of lncRNAs per genome among parasitic species (Figure S4A). Notably, *S. himalayana*, which has the highest proportion of repetitive elements in its genome¹⁸, exhibits a comparatively low number of lncRNAs (Figure S4A; Table S2), indicating that lncRNA abundance is not simply a function of total repeat content. When specific repeat classes were examined, however, a strong positive correlation emerged between lncRNA abundance and the proportion of LTR (long terminal repeat) elements in parasitic plant genomes (Figure S4B), suggesting that LTR retrotransposons may contribute to lncRNA origination, expansion, or regulatory diversification in these lineages. Interestingly, this repeat-associated pattern may help explain the previously observed predominance of antisense and intergenic lncRNA genes across parasitic lifestyles. Because LTR elements are often inserted in intergenic regions or near gene boundaries²⁰, their expansion could facilitate the emergence of novel antisense and intergenic transcripts without generating clear lifestyle-specific classification biases. Thus, while no differential positional pattern was detected between hemi- and holoparasites, the association with LTR content suggests a shared genomic mechanism shaping lncRNA repertoires across parasitic plants.

Parasitic plant genomes, particularly those of holo- and endoparasites, have incorporated substantial amounts of host-derived sequences through horizontal gene transfer (HGT). For instance, *Sapria himalayana* has approximately 1.2% of its genome attributed to HGT, and species such as *Cuscuta australis*, *Orobanche aegyptiaca*, and *Pedicularis kansuensis* also harbor genes acquired from host lineages^{21–23}. To assess whether lncRNAs in parasitic plants may similarly originate from horizontal transfers, we conducted phylogenetic analyses of the lncRNA sequences identified in this study against species from phylogenetically related groups (Figure 2A; Table S3). These analyses revealed a substantial number of clades consistent with potential HGT events in *Sapria* (Figure 2A; Figure S5A). *Striga* and *Cuscuta* also exhibited putative HGT events involving species from Poaceae and rosid lineages, respectively, groups previously described as gene donors to parasitic plants through HGT^{22,24,25} (Figure 2B-C; Figure S5B-C). Although no direct HGT event involving a confirmed extant host could be established, likely due to limitations in publicly available genomic datasets, lncRNA transcripts that clustered within clades composed predominantly of distantly related species, rather than with phylogenetically close relatives, were considered candidate HGT cases (Figure 2; Figure S5-6). Putative HGT events were also detected in Orobanchaceae species such as *Orobanche aegyptiaca* and *P. kansuensis*, including *P. japonicum*, for which no prior evidence of HGT has been reported (Figure S6). Together, these findings suggest that horizontal gene transfer may have contributed not only to protein-coding gene evolution but also to shaping the lncRNA repertoire of parasitic plants.

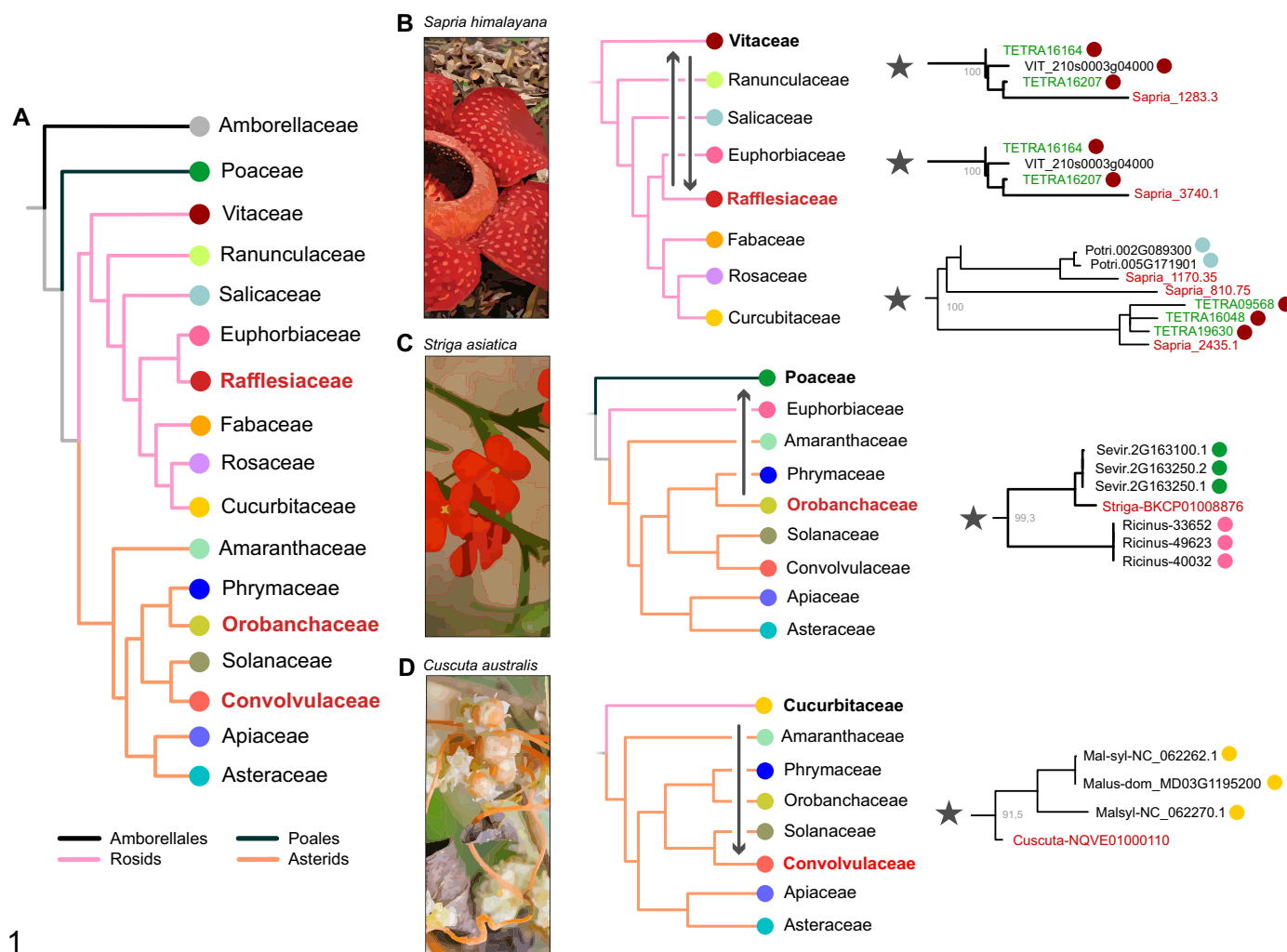


Figure 2. Phylogenetic evidence of potential HGT events involving lncRNAs in parasitic plants. (A) Representative phylogenetic tree of gene families included in the HGT analysis dataset. (B–D) Clades (right) indicating putative HGT events (black stars) and the corresponding inferred transfer models (left) observed in (B) *Sapria himalayana*, (C) *Striga asiatica*, and (D) *Cuscuta australis*. Green labels represent transcripts from confirmed host species of the parasitic plants, whereas red labels denote lncRNAs identified from parasitic plant species.

Our findings provide new insights into the diversity, expression, and potential origins of lncRNAs in parasitic plants, highlighting their functional and evolutionary significance across different parasitic lifestyles. The distinct patterns of lncRNA abundance, structure, and expression between hemi-, holo-, and endoparasites suggest that these molecules are intricately linked to the unique biological and developmental adaptations of parasitism. The strong association between lncRNAs and key parasitic structures, such as haustoria, along with their presence in non-parasitic organs, underscores their multifunctional roles in both parasitism and general plant development. Furthermore, the evidence of horizontal transfer of lncRNA, particularly in *Sapria*, opens new avenues for understanding how host-parasite interactions shape genomic innovation in parasitic plants, representing the first evidence of lncRNA transfer between plants. Together, these results highlight the importance of lncRNAs as central players in the molecular mechanisms underlying parasitism, providing a foundation for further exploration of their regulatory roles and evolutionary origins.

1 **Acknowledgements**

2 This work was supported by the Serrapilheira Institute (grant number Serra-
3 181226691) and by the Fundação de Amparo à Pesquisa do Estado de Minas Gerais
4 (FAPEMIG) – Finance Code 001, that provided a PhD scholarship to WFCR. All data
5 processing was performed in the Striga Server (LIVE Lab) and CENAPAD-SP clusters.

References

1. Kang, C. & Liu, Z. Global identification and analysis of long non-coding RNAs in diploid strawberry *Fragaria vesca* during flower and fruit development. *BMC Genomics* **16**, 815 (2015).
2. Kitagawa, M., Kitagawa, K., Kotake, Y., Niida, H. & Ohhata, T. Cell cycle regulation by long non-coding RNAs. *Cell. Mol. Life Sci.* **70**, 4785–4794 (2013).
3. Rai, M. I., Alam, M., Lightfoot, D. A., Gurha, P. & Afzal, A. J. Classification and experimental identification of plant long non-coding RNAs. *Genomics* **111**, 997–1005 (2019).
4. Chen, Q. *et al.* From “Dark Matter” to “Star”: Insight Into the Regulation Mechanisms of Plant Functional Long Non-Coding RNAs. *Front. Plant Sci.* **12**, 650926 (2021).
5. Song, L., Fang, Y., Chen, L., Wang, J. & Chen, X. Role of non-coding RNAs in plant immunity. *Plant Communications* **2**, 100180 (2021).
6. *Parasitic Plants*. (Chapman & Hall, London Weinheim, 1995).
7. Press, M. C. & Phoenix, G. K. Impacts of parasitic plants on natural communities. *New Phytologist* **166**, 737–751 (2005).
8. Nikolov, L. A. *et al.* Holoparasitic Rafflesiaceae possess the most reduced endophytes and yet give rise to the world’s largest flowers. *Annals of Botany* **114**, 233–242 (2014).
9. Khoei, M. A., Karimi, M., Karamian, R., Amini, S. & Soorni, A. Identification of the Complex Interplay Between Nematode-Related lncRNAs and Their Target Genes in *Glycine max* L. *Front. Plant Sci.* **12**, 779597 (2021).
10. Han, G. *et al.* Identification of Long Non-Coding RNAs and the Regulatory Network Responsive to Arbuscular Mycorrhizal Fungi Colonization in Maize Roots. *IJMS* **20**, 4491 (2019).
11. Yoshida, S., Cui, S., Ichihashi, Y. & Shirasu, K. The Haustorium, a Specialized Invasive Organ in Parasitic Plants. *Annu. Rev. Plant Biol.* **67**, 643–667 (2016).
12. Yoshida, S. & Shirasu, K. Plants that attack plants: molecular elucidation of plant parasitism. *Current Opinion in Plant Biology* **15**, 708–713 (2012).
13. Barkman, T. J. *et al.* Reading between the vines: Hosts as islands for extreme holoparasitic plants. *American J of Botany* **104**, 1382–1389 (2017).

- 1 14. Shen, G. *et al.* *Cuscuta australis* (dodder) parasite eavesdrops on the host plants'
2 FT signals to flower. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 23125–23130 (2020).
- 3 15. Cai, L. *et al.* Deeply Altered Genome Architecture in the Endoparasitic Flowering
4 Plant *Sapria himalayana* Griff. (Rafflesiaceae). *Current Biology* **31**, 1002-1011.e9 (2021).
- 5 16. Sun, G. *et al.* Large-scale gene losses underlie the genome evolution of parasitic
6 plant *Cuscuta australis*. *Nat Commun* **9**, 2683 (2018).
- 7 17. Wu, Y. *et al.* Bidirectional lncRNA Transfer between *Cuscuta* Parasites and Their
8 Host Plant. *IJMS* **23**, 561 (2022).
- 9 18. Guo, X. *et al.* The *Sapria himalayana* genome provides new insights into the
10 lifestyle of endoparasitic plants. *BMC Biol* **21**, 134 (2023).
- 11 19. Lee, H., Zhang, Z. & Krause, H. M. Long Noncoding RNAs and Repetitive
12 Elements: Junk or Intimate Evolutionary Partners? *Trends in Genetics* **35**, 892–902
13 (2019).
- 14 20. Lisch, D. How important are transposons for plant evolution? *Nat Rev Genet* **14**,
15 49–61 (2013).
- 16 21. Lin, Q., Banerjee, A. & Stefanović, S. Mitochondrial Phylogenomics of *Cuscuta*
17 (Convolvulaceae) Reveals a Potentially Functional Horizontal Gene Transfer from the
18 Host. *Genome Biology and Evolution* **14**, evac091 (2022).
- 19 22. Kado, T. & Innan, H. Horizontal Gene Transfer in Five Parasite Plant Species in
20 Orobanchaceae. *Genome Biology and Evolution* **10**, 3196–3210 (2018).
- 21 23. Fang, L. *et al.* Chromosome-level genome assembly of *Pedicularis kansuensis*
22 illuminates genome evolution of facultative parasitic plant. *Molecular Ecology Resources*
23 **24**, e13966 (2024).
- 24 24. Yoshida, S., Maruyama, S., Nozaki, H. & Shirasu, K. Horizontal Gene Transfer by
25 the Parasitic Plant *Striga hermonthica*. *Science* **328**, 1128–1128 (2010).
- 26 25. Yang, Z. *et al.* Convergent horizontal gene transfer and cross-talk of mobile nucleic
27 acids in parasitic plants. *Nat. Plants* **5**, 991–1001 (2019).

Supplementary data

Article Title: **Investigating lncRNAs at different levels of parasitism in parasitic plants**

Authors: Wenderson Felipe Costa Rodrigues; Laura Oliveira Pires; Luiz-Eduardo Del-Bem; Juliane Karine Ishida.

Supplemental Methods

lncRNA Identification

Putative long non-coding RNAs (lncRNAs) were identified using two complementary strategies in order to compare results derived from publicly available transcriptomes and from transcriptomes reconstructed in this study. In the first approach, predicted transcriptomes available in public databases for nine parasitic plant species representing hemi-, holo-, and endoparasitic lifestyles (Table S1) were used as input. Transcripts longer than 200 nucleotides were retained and filtered to exclude those containing open reading frames (ORFs) longer than 120 amino acids. To remove potential protein-coding transcripts, BLASTx¹ searches were performed against the SwissProt database (e-value < 0.001). Additionally, coding potential was further evaluated using Coding Potential Calculator 2 (CPC2, , <https://cpc2.gao-lab.org/>)² and only transcripts classified as non-coding were retained. Additional filtering steps were conducted to eliminate other classes of non-coding RNAs: BLASTn¹ searches (e-value < 0.05) against mature plant miRNAs from miRBase were used to exclude miRNA homologs, CMSCAN tool³ was applied to query the Rfam database to remove structural RNAs, and MIRENA⁴ was employed to identify and discard potential small RNA (sRNAs) precursor transcripts. This workflow resulted in a high-confidence set of lncRNAs for each species based on publicly available transcript annotations.

In parallel, a second and independent strategy was applied to species with available reference genomes and curated gene annotations (*Arabidopsis thaliana*, *Phtheirospermum japonicum*, *Striga asiatica*, *Orobancha cumana*, and *Cuscuta campestris*). For these species, RNA-seq reads were aligned to their respective reference genomes using HISAT2⁵, a splice-aware aligner suitable for genome-guided transcript reconstruction. The resulting alignments were processed with SAMtools⁶ for format conversion, sorting, and indexing. Transcript models were subsequently reconstructed using StringTie⁷ in a reference-guided assembly framework. When multiple assemblies were generated for a given species, transcript annotations were merged using the StringTie merge function to produce a non-redundant consensus transcriptome. Reference gene annotation files obtained from public repositories were standardized and used to guide downstream analyses.

lncRNA identification from these reconstructed transcriptomes was performed using the FEELnc⁸ pipeline, which integrates structural filters and a supervised Machine Learning-based coding potential model. Initially, transcripts were filtered according to length (≥ 200 nt) and genomic context relative to annotated protein-coding genes. FEELnc then estimated coding potential using a model trained on known coding transcripts, incorporating sequence-derived features such as k-mer frequencies, ORF metrics, and compositional properties to discriminate coding from non-coding transcripts. Transcripts classified as non-coding were retained and subsequently categorized according to their genomic position relative to neighboring genes (e.g., intergenic, antisense, or overlapping), generating a final annotated set of high-confidence lncRNAs for each species. This approach resulted in the identification of a substantial number of lncRNAs across the analyzed parasitic species. As a control, we assessed pipeline performance in *Arabidopsis thaliana* and recovered approximately 73% of the lncRNAs annotated in PlncDB⁹ (Figure S7). Moreover, all experimentally validated lncRNAs included in our benchmark were successfully recovered (Figure S7), supporting the robustness and sensitivity of the identification strategy.

Information regarding repetitive elements in the genomes of these species was extracted from the original genome sequencing publications (Table S2) and incorporated to contextualize the genomic landscape of identified lncRNAs. Together, the integration of similarity-based filtering, intrinsic coding potential assessment, genome-guided transcript reconstruction, and machine learning-based classification enabled a robust and comparative characterization of lncRNA repertoires across parasitic plant lineages while minimizing biases associated with a single identification strategy.

Expression and PCA Analysis

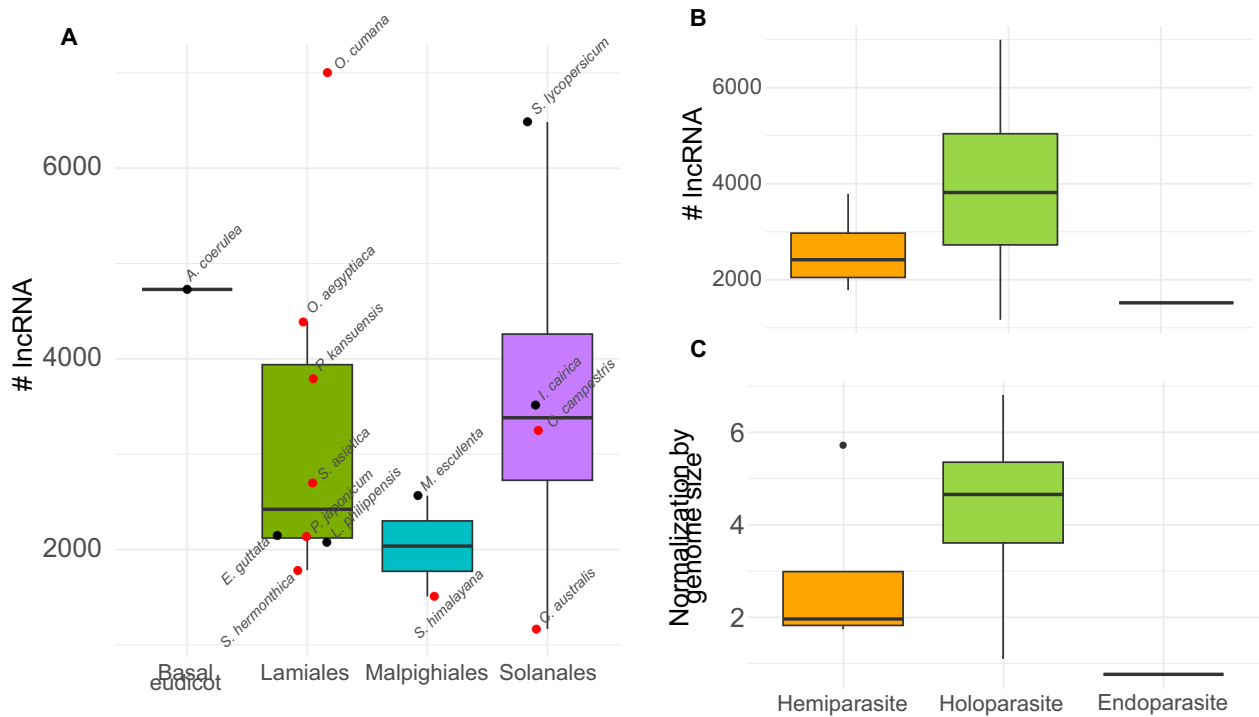
Principal Component Analysis (PCA) was conducted using Jellyfish v2.3¹⁰ to count k-mers and custom scripts to extract nucleotide lengths of the previously identified lncRNAs. Expression analyses were performed using RNA-Seq data available in the Sequence Read Archive (SRA) of NCBI (Table S3). After downloading the data, Salmon¹¹ v1.10.3 was used to quantify the expression levels of lncRNA transcripts. The resulting abundance estimates were imported into R using the tximport¹² package to generate count matrices for downstream analyses. Differential expression analysis between Haustorium and Other samples was performed using DESeq2¹³, applying standard normalization and dispersion estimation procedures. Transcripts with very low counts were filtered prior to statistical testing. Differentially expressed lncRNAs were identified based on adjusted *p*-values and log₂ fold-change thresholds, and results were exported for further interpretation. A volcano plot was generated using ggplot2¹⁴. Visualization of the PCA and all expression data was performed using R 4.4.0 and RStudio¹⁵.

HGT Analysis

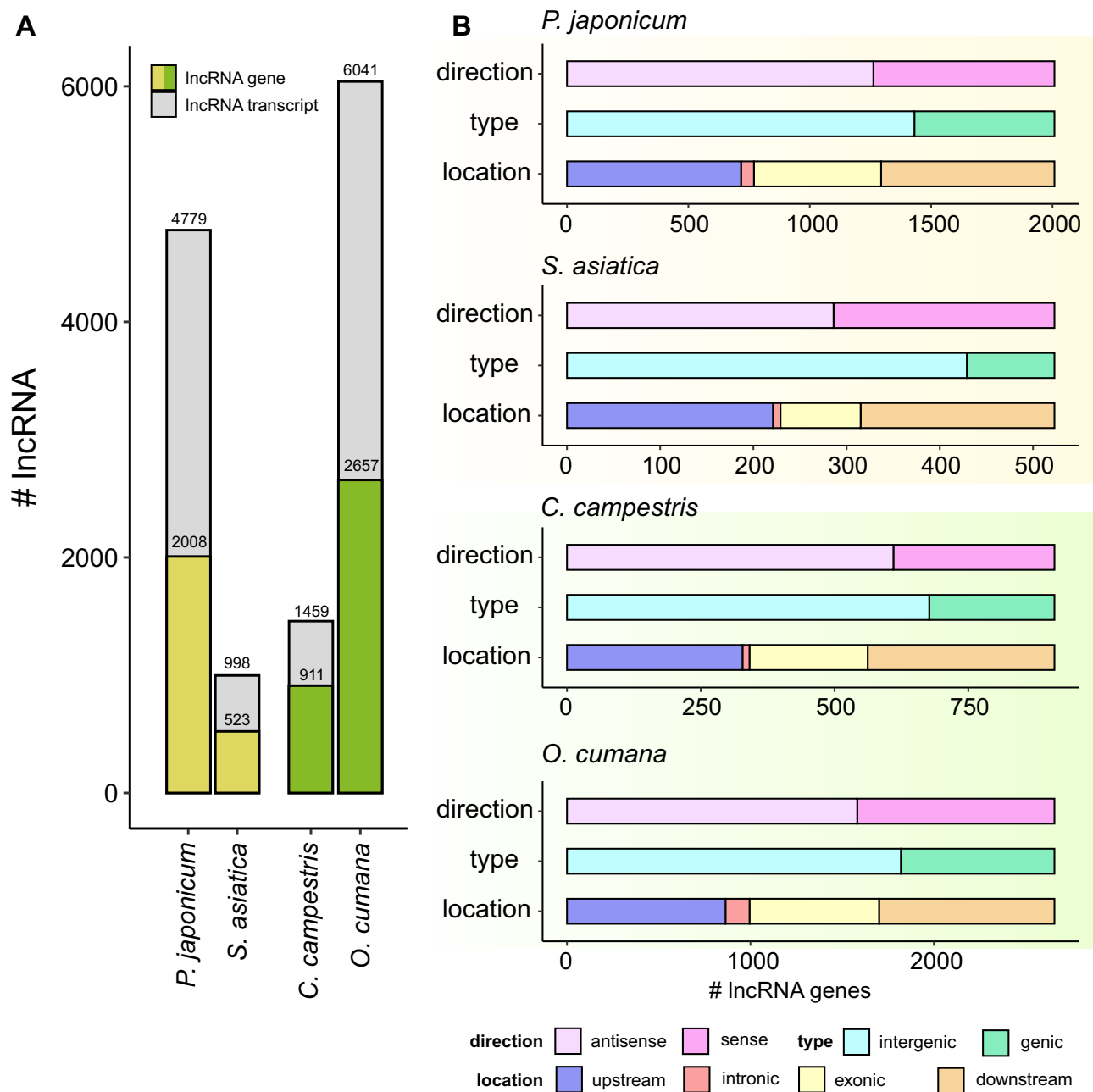
For the horizontal gene transfer (HGT) analysis, parasitic plant species with positive hits were identified using BLASTn (e-value < 1e-15) against a database constructed with transcriptomes of potential host plant species and phylogenetically related non-host species to enable clustering in phylogenetic analyses (Table S3). To ensure data accuracy and reduce redundancy, sequences with 100% identity were merged, and isoforms were removed using the SkipRedundant tool from EMBOSS v6.6.0.0¹⁶.

lncRNA transcripts were aligned using MAFFT¹⁷ v7.490 with the following parameters: --thread 30, --reorder, --leavegappyregion, --maxiterate 1000, --retree 1, and --localpair. Phylogenetic trees were constructed using IQ-TREE¹⁸ v2.0, applying ModelFinder¹⁹ to identify the best substitution model for each family. Branch support was calculated using SH-aLRT with 1000 pseudoreplicates. The resulting phylogenetic trees were visualized using FigTree v1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>) and rooted at the midpoint.

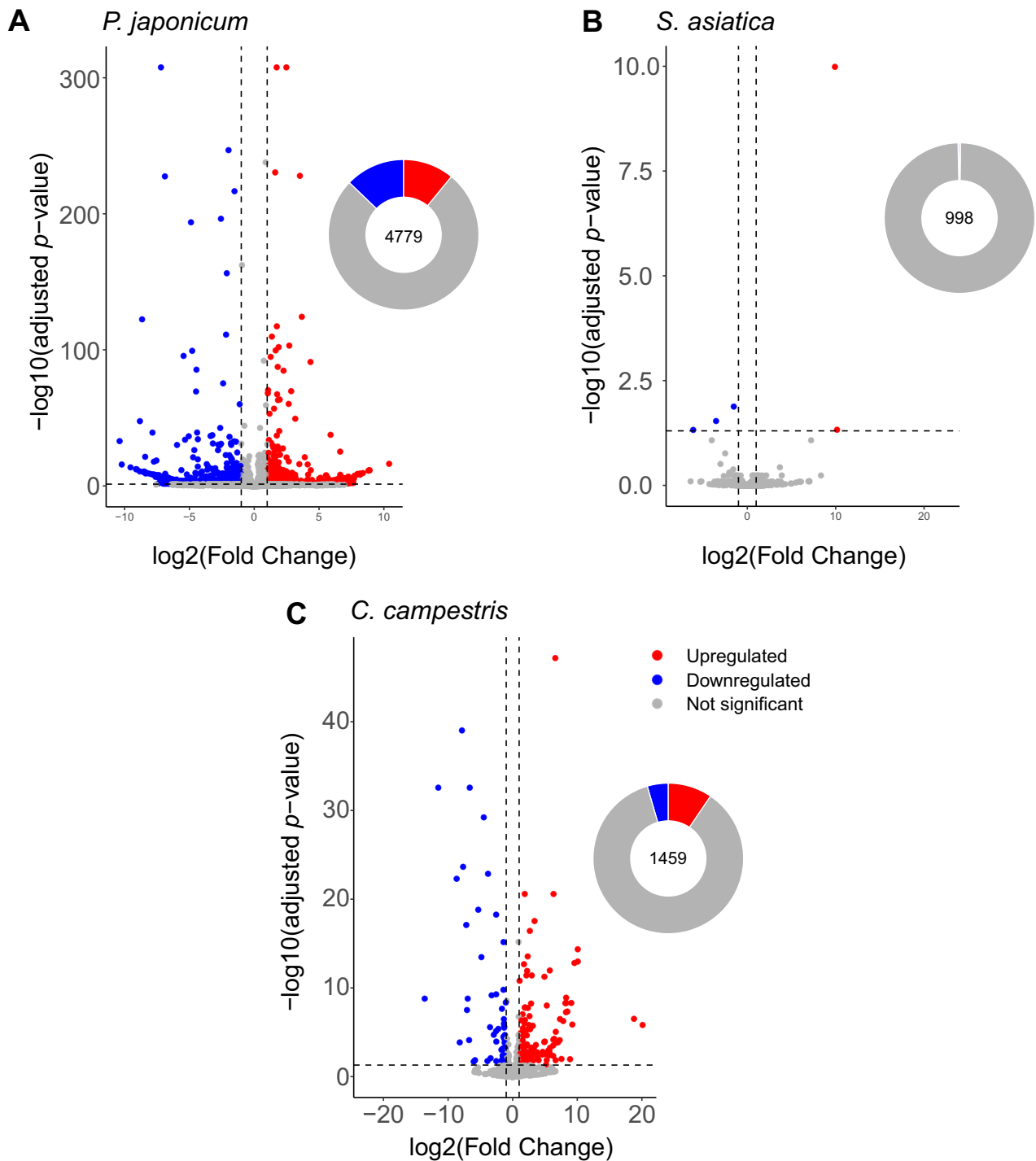
Supplemental Figures 1-7



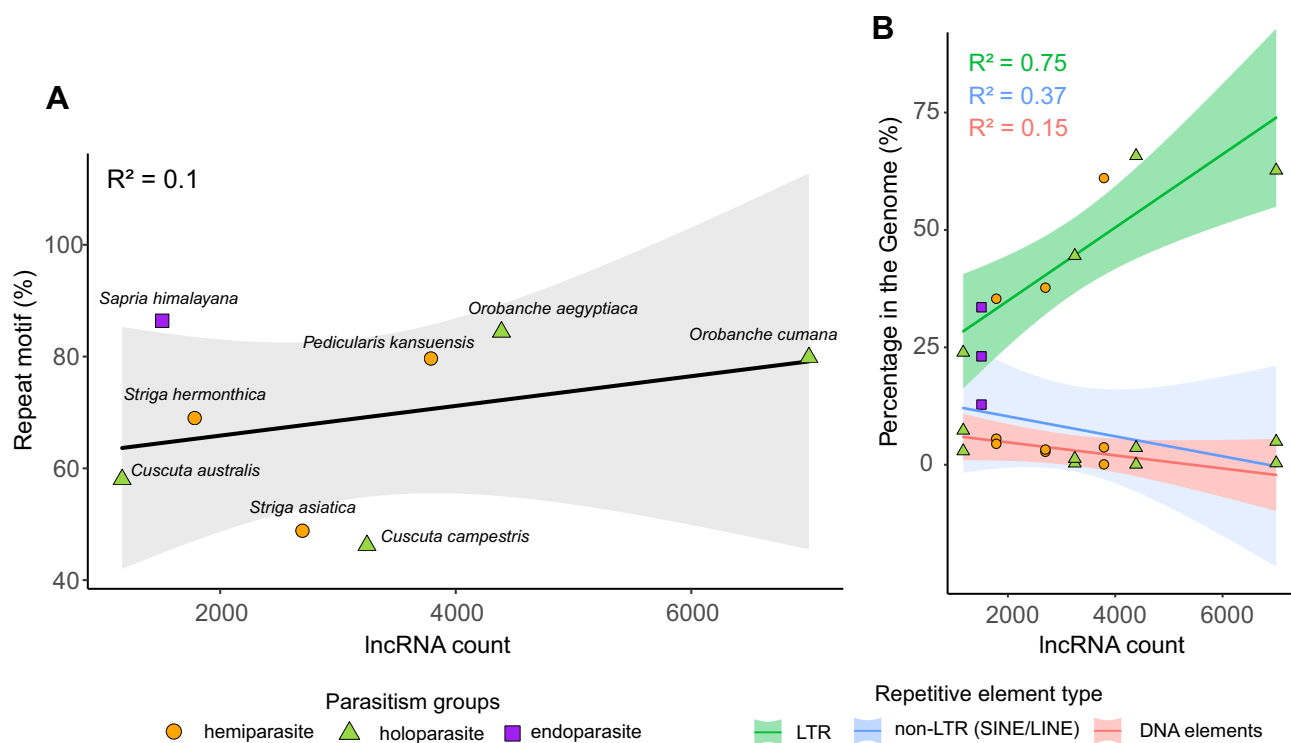
Supplemental Figure 1. Distribution of lncRNA abundance across parasitic and non-parasitic plants. (A) Boxplots showing the distribution of lncRNA counts in phylogenetically related non-parasitic species, with parasitic species indicated by red dots. (B) lncRNA counts grouped according to levels of parasitism. (C) lncRNA abundance normalized by genome size across species.



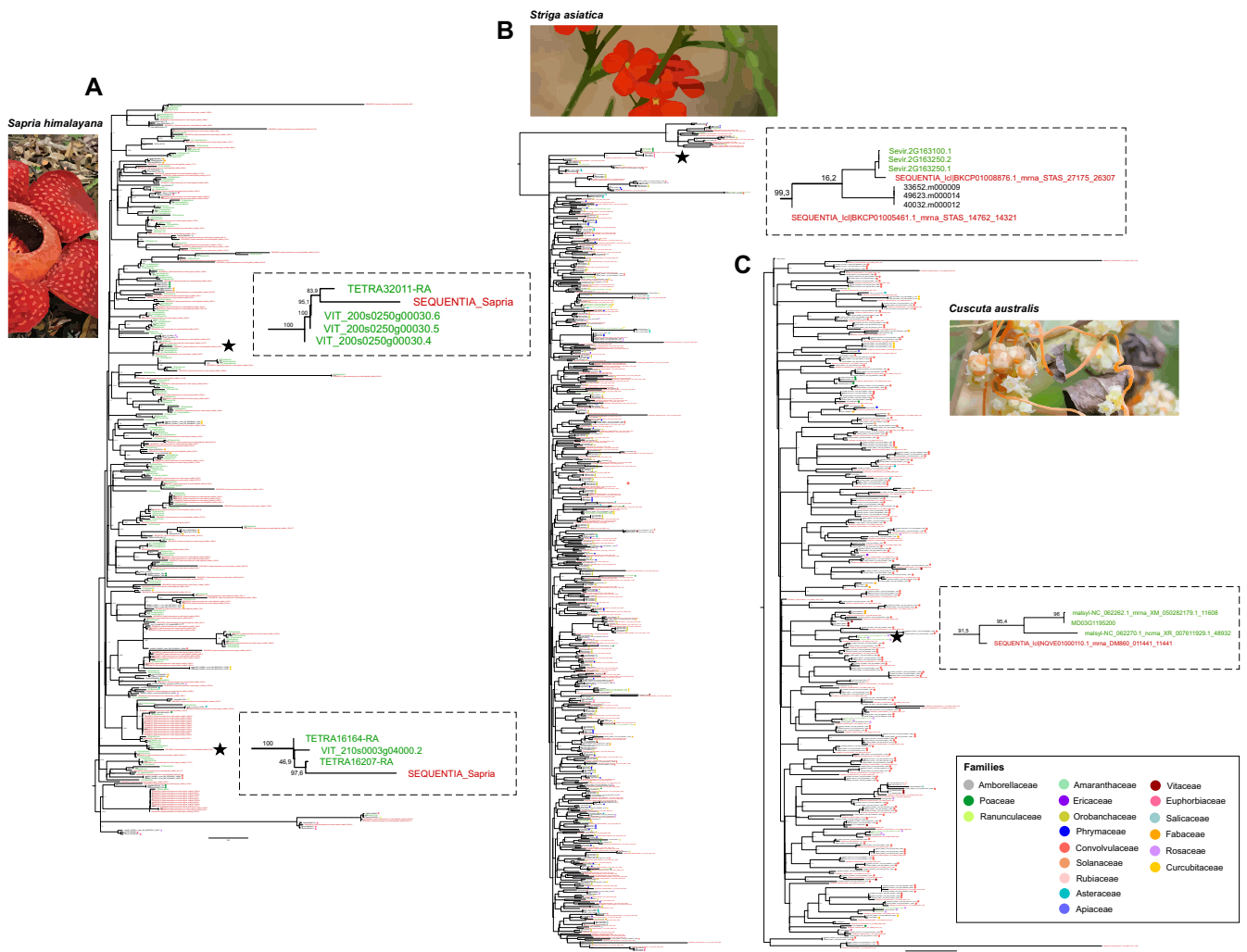
Supplemental Figure 2. FEELnc-based identification and classification of lncRNAs across parasitic lifestyles. (A) Number of lncRNA transcripts and genes identified using FEELnc across species representing different levels of parasitism. (B) Genomic classification of identified lncRNAs by species.



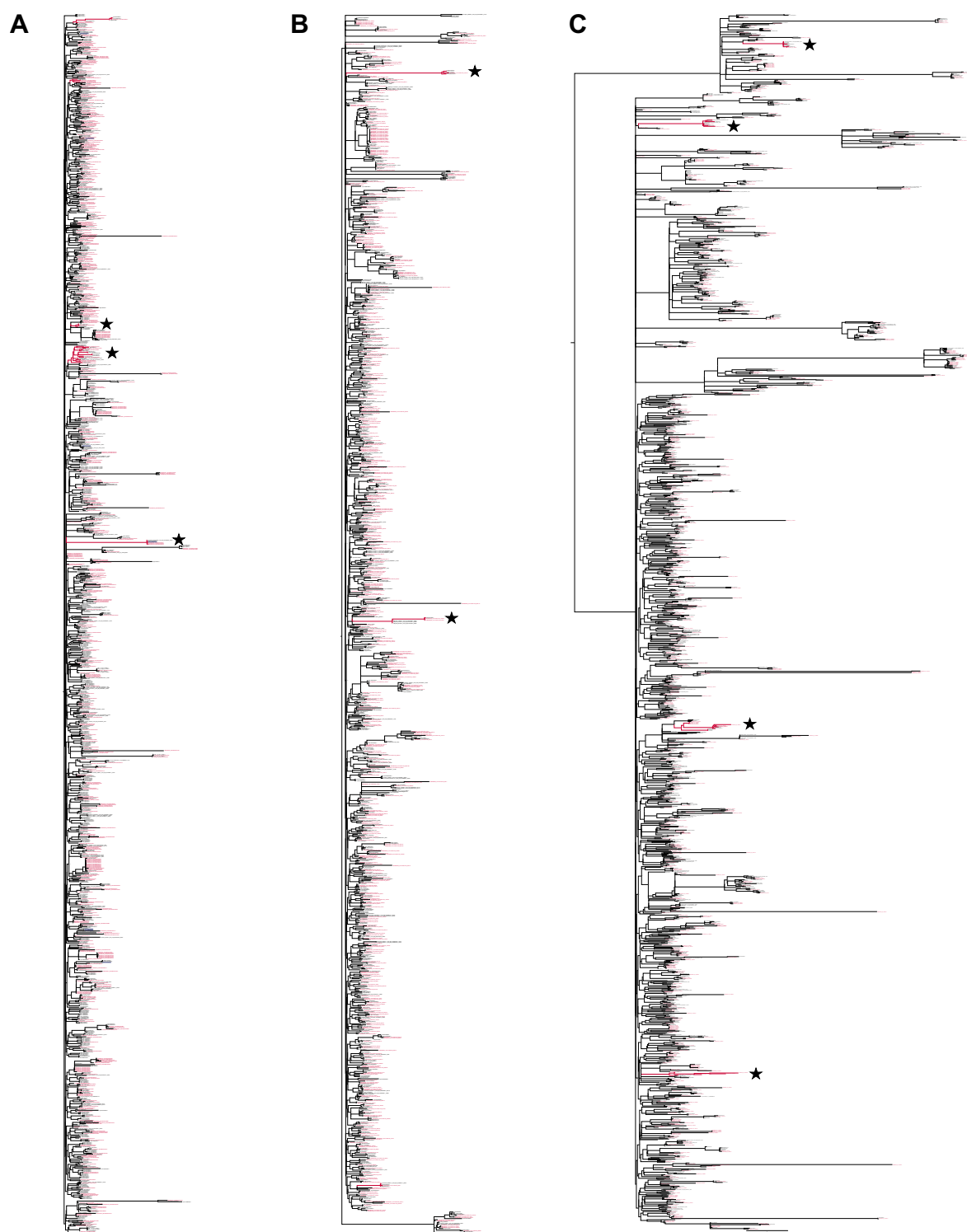
Supplemental Figure 3. Differential expression analysis of lncRNAs in parasitic species. Volcano plots showing differentially expressed lncRNAs between haustorial and non-haustorial tissues in (A) *Phtheirospermum japonicum*, (B) *Striga asiatica*, and (C) *Cuscuta campestris*, highlighting upregulated and downregulated transcripts based on adjusted p -value and $\log_2$ fold-change thresholds.



Supplemental Figure 4. Association between lncRNA abundance and genomic repetitive content in parasitic plants. Scatter plots showing correlations between the total number of identified lncRNAs and (A) the overall percentage of repetitive elements and (B) the proportion of distinct classes of repetitive elements within parasitic plant genomes.

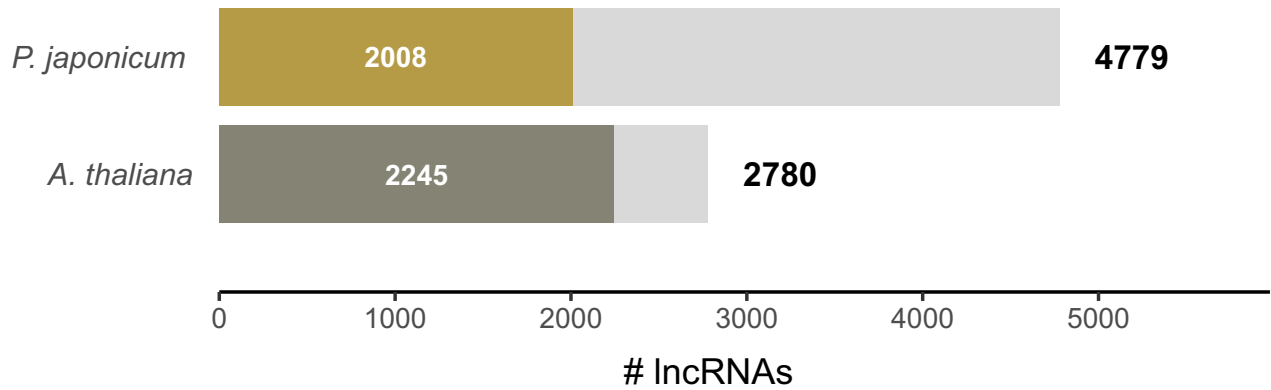


Supplemental Figure 5. Phylogenetic support for putative horizontal gene transfer (HGT) events involving lncRNAs in parasitic plants. Complete phylogenetic trees for (A) *Sapria himalayana*, (B) *Striga asiatica*, and (C) *Cuscuta australis*, highlighting candidate HGT events involving lncRNA transcripts. Putative HGT clades are indicated by black stars and enclosed within dashed boxes. Green labels represent transcripts from host or donor lineages, whereas red labels denote lncRNAs identified from parasitic plant species.

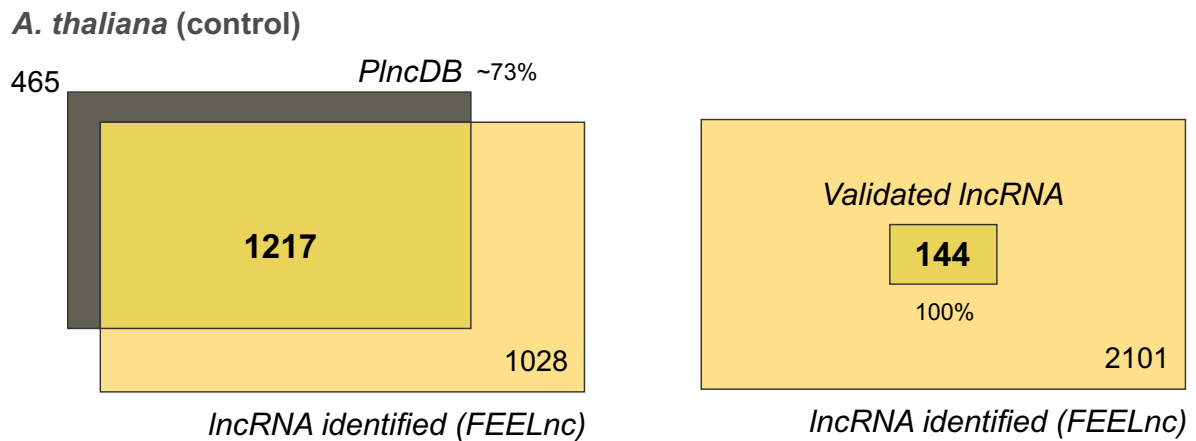


Supplemental Figure 6. Phylogenetic evidence of putative horizontal gene transfer (HGT) events involving lncRNAs in Orobanchaceae species. Phylogenetic trees highlighting candidate HGT events (black stars and red branches) in (A) *Orobanche aegyptiaca*, (B) *Pedicularis kansuensis*, and (C) *Phtheirospermum japonicum*. Red labels denote lncRNAs identified from parasitic plant species.

A



B



Supplemental Figure 7. Validation of FEELnc IncRNA identification performance.

(A) IncRNA identification results in *Phtheirospermum japonicum* and the model plant *Arabidopsis thaliana*. (B) Control benchmarking in *A. thaliana*, comparing identified IncRNAs with those annotated in PIncDB (left) and with experimentally validated IncRNAs (right).

Supplemental Tables 1-4

Supplemental Table 1. Parasitic and non-parasitic plant species analyzed in this study.

Species	lncRNA #1	lncRNA #2 - sRNA precursor	lncRNA [high-confidence]
<i>Aquilegia coerulea</i>	4843	4728	3895
<i>Cuscuta australis</i>	1188	1167	1026
<i>Cuscuta campestris</i>	3328	3246	2706
<i>Erythranthe guttata</i> (<i>Mimulus guttatus</i>)	2236	2145	1680
<i>Ipomoea cairica</i>	3618	3518	3175
<i>Lindenbergia philippensis</i>	2123	2075	1689
<i>Manihot esculenta</i>	2620	2564	2152
<i>Orobancha (Phelipanche) aegyptiaca</i>	4447	4388	3206
<i>Orobancha cumana</i>	7184	6999	4783
<i>Pedicularis kansuensis</i>	3848	3789	2717
<i>Phtheirospermum japonicum</i>	2188	2136	1396
<i>Sapria himalayana</i> (GUO et al., 2023)	1765	1507	1461
<i>Solanum lycopersicum</i>	6617	6485	5534
<i>Striga asiatica</i>	2721	2698	2205
<i>Striga hermonthica</i>	1815	1783	1398

Supplemental Table 2. Quantification of repetitive elements in the genomes of parasitic plants.

Species	IncRNA	Repeat motif (%)	DNA elements	LINE	SINE	LTR	non-LTR
<i>Sapria himalayana</i> (GUO et al., 2023)	1507	86.41	33.53	12.81	0.0	23.08	12.81
<i>Cuscuta australis</i>	1167	58.00	7.3	2.4	0.5	23.90	2.9
<i>Cuscuta campestris</i>	3246	46.20	1.3	0.3		44.50	0.3
<i>Striga asiatica</i>	2698	48.83	3.23	2.70	0.06	37.70	2.76
<i>Orobanche (Phelipanche) aegyptiaca</i>	4388	84.40	3.61	0.06		65.77	0.06
<i>Pedicularis kansuensis</i>	3789	79.65	3.70	0.07	0.01	61.04	0.08
<i>Striga hermonthica</i>	1783	68.99	4.46	5.49	0.05	35.35	5.54
<i>Orobanche cumana</i>	6999	79.82	4.93	0.34		62.68	0.34

Supplemental Table 3. Dataset of non-parasitic species, phylogenetically related or host to parasitic plants, used for HGT analysis.

HGT reference	Clade	Species	Database
All groups	Basal eudicots	<i>Aquilegia coerulea</i>	Pyhtozome
All groups	Basal Angiosperms	<i>Amborella trichopoda</i>	Pyhtozome
All groups	Basal asterids / Ericales	<i>Vaccinium darrowii</i>	Pyhtozome
All groups	Basal rosids / Vitales	<i>Vitis vinifera</i>	Pyhtozome
All groups	Basal monocots / Alismatales	<i>Spirodela polyrhiza</i>	Pyhtozome
All groups	Malvids	<i>Arabidopsis thaliana</i>	Pyhtozome
Orobanchaceae	Rosaceae	<i>Prunus yedoensis</i>	Pyhtozome
Orobanchaceae	Poaceae	<i>Brachypodium distachyon</i>	Pyhtozome
Orobanchaceae	Poaceae	<i>Oryza sativa</i>	Pyhtozome
Orobanchaceae	Poaceae	<i>Setaria viridis</i>	Pyhtozome
Orobanchaceae	Poaceae	<i>Sorghum bicolor</i>	Pyhtozome
Orobanchaceae	Poaceae	<i>Zea mays</i>	Pyhtozome
Orobanchaceae	Poaceae	<i>Miscanthus sinensis</i>	Pyhtozome
Orobanchaceae	Fabaceae	<i>Glycine max</i>	Pyhtozome
Orobanchaceae	Fabaceae	<i>Phaseolus vulgaris</i>	Pyhtozome
Orobanchaceae	Fabaceae	<i>Lotus japonicus</i>	NCBI
Orobanchaceae	Fabaceae	<i>Medicago truncatula</i>	Pyhtozome
Orobanchaceae	Orobanchaceae non-parasitic	<i>Lindenbergia philippensis</i>	Pyhtozome
Orobanchaceae	Orobanchaceae outgroup / Lamiales	<i>Mimulus guttatus</i>	Pyhtozome
Convolvulaceae	Apiaceae	<i>Daucus carota</i>	Pyhtozome
Convolvulaceae	Rosaceae	<i>Prunus persica</i>	Pyhtozome
Convolvulaceae	Cucurbitaceae	<i>Cucumis sativus</i>	Pyhtozome
Convolvulaceae	Convolvulaceae non-parasitic	<i>Ipomoea nil</i>	NCBI
Convolvulaceae	Convolvulaceae non-parasitic	<i>Ipomoea triloba</i>	NCBI
Convolvulaceae	Convolvulaceae outgroup / Solanales	<i>Solanum lycopersicum</i>	Pyhtozome
Convolvulaceae	Basal asterids / Caryophyllales	<i>Beta vulgaris</i>	Pyhtozome
Convolvulaceae	Basal asterids / Caryophyllales	<i>Amaranthus hypochondriacus</i>	Pyhtozome
Convolvulaceae	Rubiaceae	<i>Coffea arabica</i>	Pyhtozome
Convolvulaceae	Asteraceae	<i>Helianthus annuus</i>	Pyhtozome
Convolvulaceae	Asteraceae	<i>Lactuca sativa</i>	Pyhtozome
Convolvulaceae	Rosaceae	<i>Malus domestica</i>	Pyhtozome
Convolvulaceae	Rosaceae	<i>Fragaria vesca</i>	Pyhtozome
Convolvulaceae	Cucurbitaceae	<i>Cucumis melo</i>	Pyhtozome
Convolvulaceae	Cucurbitaceae	<i>Cucurbita pepo</i>	NCBI
Convolvulaceae	Cucurbitaceae	<i>Benincasa hispida</i>	NCBI
Convolvulaceae	Rosaceae	<i>Malus sylvestris</i>	NCBI
Convolvulaceae	Rosaceae	<i>Rosa chinensis</i>	NCBI
Rafflesiaceae	Vitales	<i>Tetrastigma voinierianum</i>	NCBI
Rafflesiaceae	Rafflesiaceae outgroup / Malpighiales	<i>Populus trichocarpa</i>	Pyhtozome
Rafflesiaceae	Rafflesiaceae outgroup / Malpighiales	<i>Manihot esculenta</i>	Pyhtozome
Rafflesiaceae	Malpighiales	<i>Ricinus communis</i>	Pyhtozome
Rafflesiaceae	Malpighiales	<i>Salix purpurea</i>	Pyhtozome

Supplemental Table 4. RNA-Seq projects used for expression analyses.

Species	SRA Project	Instrument	Date	Tissue	Link
<i>Sapria himalayana</i> (GUO et al., 2023)	PRJNA943542	Illumina NovaSeq 6000	20/04/2023	floral bud	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA943542
<i>Cuscuta campestris</i>	PRJDB9165	Illumina Genome Analyzer IIX	07/01/2020	flower; seedling; epidermis and cortex; haustorium	https://www.ncbi.nlm.nih.gov/bioproject/PRJDB9165
<i>Phtheirospermum japonicum</i>	PRJDB10085	Illumina HiSeq 1500	13/08/2020	haustorium (1-7d)	https://www.ncbi.nlm.nih.gov/bioproject/PRJDB10085
<i>Striga asiatica</i>	PRJDB8229	Illumina HiSeq 2000	19/09/2019	shoot; root; leaf; haustorium	https://www.ncbi.nlm.nih.gov/bioproject/PRJDB8229
<i>Cuscuta australis</i>	PRJNA394036	Illumina HiSeq X Ten	06/06/2018	germinating seedling; haustoria; bud; ovary; seeds; stem.	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA394036
<i>Orobancha aegyptiaca</i>	PRJNA929515	Illumina HiSeq 2500	30/01/2023	root	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA929515
	PRJNA1083627	Illumina HiSeq 2500	06/03/2024	tubercle	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1083627
	PRJNA873060	Illumina NovaSeq 6000	26/08/2022	seedling	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA873060

References

1. Altschul, S. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389–3402 (1997).
2. Kang, Y.-J. *et al.* CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Research* **45**, W12–W16 (2017).
3. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013).
4. Mathelier, A. & Carbone, A. MIReNA: finding microRNAs with high accuracy and no learning at genome scale and from deep sequencing data. *Bioinformatics* **26**, 2226–2234 (2010).
5. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**, 907–915 (2019).
6. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
7. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**, 290–295 (2015).
8. Wucher, V. *et al.* FEELnc: a tool for long non-coding RNA annotation and its application to the dog transcriptome. *Nucleic Acids Res* gkw1306 (2017) doi:10.1093/nar/gkw1306.
9. Jin, J., Liu, J., Wang, H., Wong, L. & Chua, N.-H. PLncDB: plant long non-coding RNA database. *Bioinformatics* **29**, 1068–1071 (2013).
10. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k -mers. *Bioinformatics* **27**, 764–770 (2011).
11. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**, 417–419 (2017).

12. Michael Love [Cre, A. tximport. Bioconductor
<https://doi.org/10.18129/B9.BIOC.TXIMPORT> (2017).
13. Michael Love, S. A. DESeq2. Bioconductor
<https://doi.org/10.18129/B9.BIOC.DESEQ2> (2017).
14. Wickham, H. *et al.* ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics. 4.0.2 <https://doi.org/10.32614/CRAN.package.ggplot2> (2007).
15. RStudio Team. *RStudio: Integrated Development Environment for R*. (RStudio, PBC., Boston, MA, 2020).
16. Rice, P., Longden, I. & Bleasby, A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics* **16**, 276–277 (2000).
17. Katoh, K. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**, 3059–3066 (2002).
18. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* **32**, 268–274 (2015).
19. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**, 587–589 (2017).

3.3. CAPÍTULO 2 – ORIGIN AND EVOLUTIONARY HISTORY OF DEFENSE-RELATED PROTEINS IN PLANTS

Este Capítulo 2 apresenta o texto “*Origin and evolutionary history of defense-related proteins in plants*”, publicado como capítulo do livro intitulado como: *Defense-Related Proteins in Plants* (Elsevier, *Academic Press*, 2024; DOI do capítulo: 10.1016/B978-0-443-13236-0.00002-6; DOI do livro: 10.1016/C2022-0-02023-7), no qual divido a primeira autoria com Laura O. Pires, integrante do grupo de pesquisa da Prof. Dr. Juliane Ishida (Laboratório de Interação Vegetal – LIVE). Esse capítulo oferece uma visão geral da origem, diversificação e evolução das principais famílias de proteínas envolvidas na defesa vegetal, assim como da relação planta-patógeno, abordando desde receptores de reconhecimento extracelular até proteínas antimicrobianas, reguladores de sinalização e componentes intracelulares do sistema imune. Embora focado em plantas não parasitas, o conteúdo fornece a base conceitual necessária para compreender como módulos de defesa se expandiram ao longo da evolução das plantas terrestres e como podem ter sido reutilizados ou modificados em diferentes contextos biológicos, incluindo o parasitismo vegetal discutido no Capítulo 3 desta Tese.

Origin and evolutionary history of defense-related proteins in plants

2

Laura O. Pires^a, Wenderson F.C. Rodrigues^a, Luiz-Eduardo Del-Bem, Juliane K. Ishida

Department of Botany, Institute of Biological Sciences (ICB), Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil

Introduction

Plants' association with microorganisms is ancient. The plastid-bearing monophyletic group Archaeplastida is thought to have evolved from an initial endosymbiosis between cyanobacteria and a eukaryote (de Vries & Gould, 2018; Gould et al., 2008). Among Archaeplastida, the Viridiplantae, or green plants are the sister clades alongside Rhodophyta and Glaucophyta. Chlorophyta and Streptophyta, on the other hand, emerged from a significant divergence in the early stages of Viridiplant evolution. The Streptophyta clade is formed by a paraphyletic assemblage consisting of green algae (Charophytes) and terrestrial plants (Embryophyta) (revised by Bowles et al., 2023). Early streptophytes were substantially more exposed to and vulnerable to pathogen attacks with the colonization of the land, and prior genome analyses have demonstrated that gene families relevant to pathogen interactions and stress response were expanded in this group (Wang et al., 2020). The association with fungi was also paramount to land plant evolution, with arbuscular mycorrhizal (AM) being credited with facilitating the colonization of land by plants (Vigneron et al., 2018). Extant and plant species continue to benefit from this interaction, which has enormous ecological and agronomic significance along with nitrogen-fixing rhizobium bacteria. Beneficial and pathogenic microorganisms share the same molecular markers recognized by the plant immune system but with opposite outcomes from the interaction (Wang et al., 2022). Although it is still unclear how plants can distinguish between the two types of interaction, this knowledge would undoubtedly have a positive impact by enabling the management of agricultural pathogens and enhancing the advantages gained from the mutualistic relationship with AM.

Plants employ various strategies to safeguard themselves against potential threats by utilizing a combination of physical, chemical, and cellular mechanisms to ensure defense. These include protective features like the cuticle (Reina-Pinto & Yephremov, 2009), the production of secondary metabolites (Erb & Kliebenstein, 2020),

^a These authors contributed equally to this work.

and the deposition of callose (Wang et al., 2021). At the cellular level, the plant immune system can be categorized into two layers: pattern triggered immunity (PTI) and effector triggered immunity (ETI). The PTI layer recognizes conserved molecules associated with microorganisms or their actions, called microbe-associated molecular patterns (MAMPs), pathogen-associated molecular patterns (PAMPs), damage-associated molecular patterns (DAMPs), or herbivore-associated molecular patterns (HAMPs) (Ngou, Ding, et al., 2022). These molecules are perceived by cell surface receptors, pattern-recognition receptors (PRRs), of the receptor-like kinase (RLK) type and receptor-like protein (RLP) type. The recognition of these pathogen-derived molecular patterns by the receptors triggers a chain of reactions, including mitogen-activated protein kinase (MAPK) cascade, reactive oxygen species (ROS) production, modified ion influx, callose deposition, and transcription reprogramming (DeFalco & Zipfel, 2021; Kaur, 2023; Madhu, 2024; Sharma, 2024; Shekhawat & Upadhyay, 2024; Upadhyay, 2023). To overcome this layer of plant defense response, pathogens insert pathogen-specific effector molecules into the host's cells to suppress PTI. In turn, intracellular receptors will recognize these effectors and activate the ETI response, leading to a set of reactions culminating in the hypersensitive response (HR), causing programmed cell death in the site of infection as a means of decreasing pathogen spread (Meyers et al., 2003). Systemic acquired resistance (SAR) may come later in the ETI response, conferring long-lasting and systemic resistance to the plant, making secondary or future infection in other sites less effective (Durrant & Dong, 2004; Vlot et al., 2009). ETI response is mediated by R proteins (Jones & Dangl, 2006). The R proteins are encoded by R (resistance) genes that form large gene families, numbering around 200 genes in *Arabidopsis thaliana* (Meyers et al., 2003). The most common R proteins are formed by an intracellular nucleotide-binding leucine-rich repeats receptor (NLRs), with central nucleotide binding and C-terminal domains (Martin et al., 2003; Sukarta et al., 2016).

The constant exchange of selective pressure between plants and pathogens drives these organisms' coevolution. The model used to explain plant immune response is called the zig-zag model, consisting of four steps. In the first step, PTI is triggered by the recognition of conserved pathogenic molecules (elicitors) by PRRs; the second step is the pathogen response to PTI, which is the liberation of effector molecules in the host cells to suppress the first immune response; then, the plant recruits intracellular R proteins that recognize pathogens' effectors and triggers the ETI response; to escape ETI, the pathogens will modify its effector molecules, making it unrecognizable by current plant NLRs and, in turn, the plant will adapt its receptors to be able to recognize these modified effectors (Jones & Dangl, 2006).

Although the classical model explains plant immunity, the zig-zag model based on the classification of immune response by the type of pathogen molecule recognized has been challenged. The clear division between PAMPs and effectors is not as clear cut, and in some cases, a molecule classically defined as an effector may also be qualified as a PAMP (Thomma et al., 2011). Plant immunity can also be classified by the location of the plant cell receptor (van der Burgh & Joosten, 2019). In this model, cell

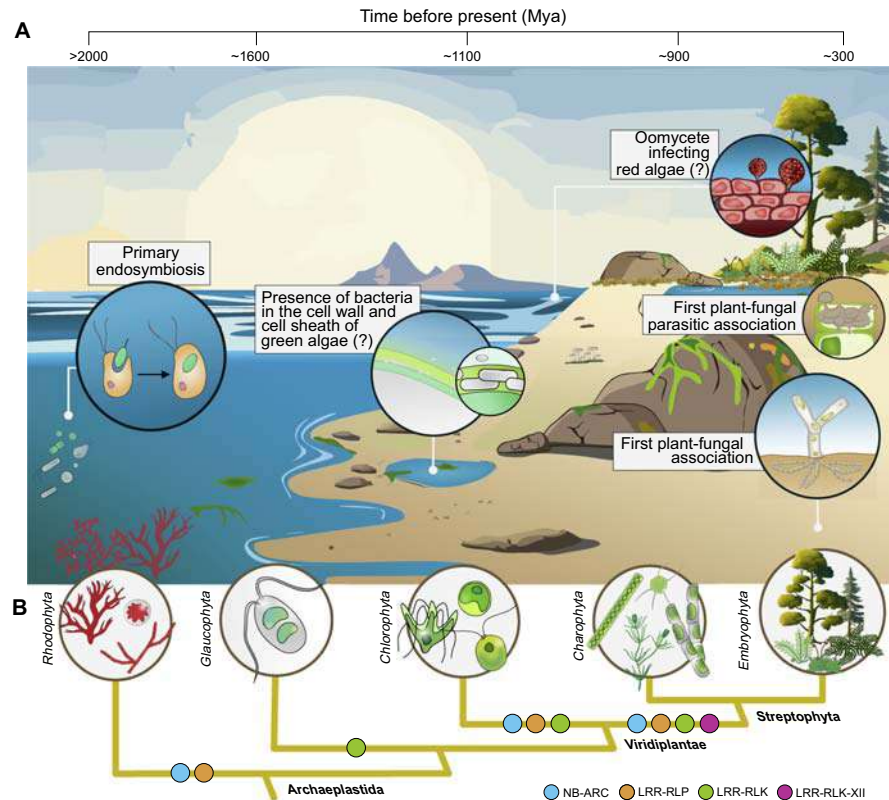
surface receptors would recognize extracellular immunogenic patterns (ExIP), leading to extracellular triggered immunity (ExTI). In contrast, intracellular immunogenic patterns (InIP), identified by cytoplasmic receptors, would trigger intracellular activated immunity (InTI) response.

Regardless of classification, defense proteins serve as receptors and signaling pathway mediators to identify pathogenic and mutualistic microorganisms that may come in contact with the plant and trigger the appropriate immune response (Zhou & Zhang, 2020). The ability to produce and choose resistant cultures makes a thorough grasp of the plant immune system essential. The origins, evolution, and classification of immunity proteins in plants are covered in this chapter.

The complex plant–microbe relationship: Evolutionary adaptations and symbiotic associations

Over the course of evolution, the dynamic relationship between plants and microbes has undergone remarkable development. This intricate interaction is made possible by a diverse range of receptors and protein domains that empower plants to respond to various molecules from multiple species. These interactions have played a vital role in the Earth's ecological balance and the flourishing of plant life. One particularly noteworthy aspect is the pivotal role played by microbes in Earth's colonization by plants, which dates back at least 450 million years ago (Mya) (Delaux & Schornack, 2021; Redecker et al., 2000). However, this intricate interplay between plants and microbes is not without consequences. The appearance of noticeable disease symptoms in plants is, in fact, an extreme result of a continuum of these interactions between plants and microbes (Bulgarelli et al., 2013). This underscores the delicate balance and complexity of their relationship, wherein some associations can lead to mutual benefits, while others may result in detrimental effects, causing diseases that can impact both plants and their surrounding ecosystems. Thus, the evolution of the relationship between plants and microbes has been a critical factor in shaping life on Earth, with mutualistic associations between these two entities being instrumental in the spread and diversification of plant life across the planet.

Plant diseases emerge as an effective survival strategy in response to the selection pressure associated with host coevolution, leading pathogen morphology and physiology in innovative directions capable of establishing its success (Poulin & Morand, 2000; Westwood et al., 2010). Similar to how bacteria can infect terrestrial plants and cause illnesses, other prokaryotes coexist with green algae and may potentially turn into hazardous ones. Prokaryotes and algae are thought to have coexisted since the Earth's formation 1.5 billion years ago, and they have displayed an intricate interaction ever since (Parfrey et al., 2011; Ramanan et al., 2016). Therefore, it is reasonable to hypothesize that symbiotic relationships with bacteria have arisen since the emergence of this group Fig. 2.1A. Additionally, it is speculated that symbiotic interactions between green algae (Chlorophyta) and bacteria were able to provide the basis for the emergence of multicellularity in the Archaeplastida due to

**FIGURE 2.1**

Appearance of Archaeplastida microorganisms' relationships and appearance of the different types of defense proteins to present time in million years ago (mya). (A) Symbiotic relationships developed throughout Archaeplastida's evolutionary history. Associations with the (?) symbol are known to happen in the present day, but the exact time it was first established is not yet precise. The remaining relationship's origins are well defined. (B) Phylogenetic representative tree of Archaeplastida groups, indicating the type of defense protein and their appearance.

Created and based on information from Bowles et al. (2023), de Vries and Archibald (2018), Ngou, Heal, et al. (2022), Yang et al. (2016).

the high presence of bacteria in the cell wall and cell sheath of algae such as those of the genera *Chlamydomonas* (Chlamydomonadaceae) and *Gonium*, *Pandorina*, and *Volvox* (Volvocaceae) (Fig. 2.1A) (de Vries & Archibald, 2018; Gould et al., 2008; Herron et al., 2009; Ramanan et al., 2016). When it comes to pathogenicity, bacteria can also be established in Chlorophyta such as *Chlamydomonas reinhardtii*, whose growth can be inhibited by *Pseudomonas protegens* (Aiyar et al., 2017), or even in red algae (Rhodophyta) of the family Corallinaceae, which can suffer from bacterial infections (Aeby, 2007).

A wide range of *Chlorophyta* and *Rhodophyta* species are susceptible to diseases caused by bacteria, nematodes, and fungi (Li et al., 2010). Among these, fungal and oomycetes diseases stand out as the second most common cause of ailments in algae, following bacterial infections (Nagappa, 2006). For instance, red algae are known to be infected by oomycetes from the genera *Olpidiopsis* and *Pythium* (Fig. 2.1A) (Ding & Ma, 2005; Li et al., 2010). Interestingly, pathogenicity has independently evolved in oomycetes (Stramenopiles) and is considered typical in extant ecosystems, where it can be mutualistic or even cause significant epidemics in crops (Selosse et al., 2015). Notably, the oomycetes that infect land plants, such as *Phytophthora*, and those that target red algae, like *Olpidiopsis*, do not form a monophyletic group. This observation suggests that diseases in oomycetes have originated independently on separate occasions. These findings shed light on the intricate dynamics of disease development and transmission in algae, particularly emphasizing the fascinating evolutionary paths taken by oomycetes to acquire pathogenicity.

Our understanding of plant–microbe associations and coevolution remains limited, making it difficult to conclude. Most research in this area is still focused on model plants or those of agricultural interest, and little is known about pathogens in algae and nonflowering land plants (Upson et al., 2018). Although the origin of ancient plant–microbe associations is challenging, antagonism in these associations is a reality, and the first plants likely developed defense strategies against pathogens (Han, 2019). Oomycetes have evolved genomes with a distinct “two-speed architecture” where genomic regions are rich in repeats and undergoing rapid evolution mediate adaptation and coevolution with host plants over time (Raffaele et al., 2010; Raffaele & Kamoun, 2012; Selosse et al., 2015). Moreover, the question of which came first, the beneficial or harmful interaction is linked to the similarities between these associations. Selosse et al. (2015) suggest that pathogens have explored pathways by which plants establish connections with mutualistic microbes, as mutants of *Medicago trunculata* deficient in colonization by arbuscular mycorrhizae also showed a deficiency in establishing infection by the oomycete *Phytophthora palmivora* (Delaux & Schornack, 2021; Rey et al., 2015).

Regardless of these associations' emergence, land plants have developed two-layered immune systems that recognize and activate defense against pathogens (PTI and ETI) (Jones & Dangl, 2006; Ngou, Ding, et al., 2022). Considered independent for a long time, these systems are now known to be interdependent and share several signaling components that act synergistically in plant defense (Ngou et al., 2021). Data from Ngou, Heal, et al. (2022) suggest an evolutionary relationship between PTI and ETI, as PRRs and NLRs expand or contract under pathogen pressure and ecological adaptations. It indicates that the mutual potentiation of PTI and ETI may be conserved across plant species (Ngou, Ding, et al., 2022).

Defense protein's structure and characteristics

Defense receptors can be classified according to subcellular localization, with cell surface localized PRRs and intracellular NLRs being the two primary categories

(Zhou & Zhang, 2020). PRRs are further divided into RLKs and RLPs. In contrast to RLPs, which are similar but lack the intracellular kinase domain, RLKs have a conserved tripartite structure with a variable ectodomain, a transmembrane domain, and a conserved intracellular kinase domain (Antil & Upadhyay, 2023; Shumayla et al., 2019; Upadhyay & Shumayla, 2022; Han et al., 2014). Both types of PRRs are classified according to their ectodomain, with a few classes described: leucine-rich repeats (LRRs), LysM motifs, lectin-like motifs, and epidermal growth factor-like domain (Fig. 2.2) (Saijo et al., 2018; Shumayla & Upadhyay, 2023). RLPs lack a cytoplasmatic domain, so they form a complex with RLKs to allow intracellular signaling to start the immune reaction after MAMP recognition (Gust & Felix, 2014).

The most common intracellular receptors of the NLR type are organized in a three-domain architecture: a highly conserved NB-ARC domain at the center, surrounded by an N-terminal response domain and a varied LRR domain at the C-terminus (Fig. 2.2) (Collier et al., 2011; Sukarta et al., 2016). The central NB-ARC domain of plants belongs to the same protein superfamily of animal's

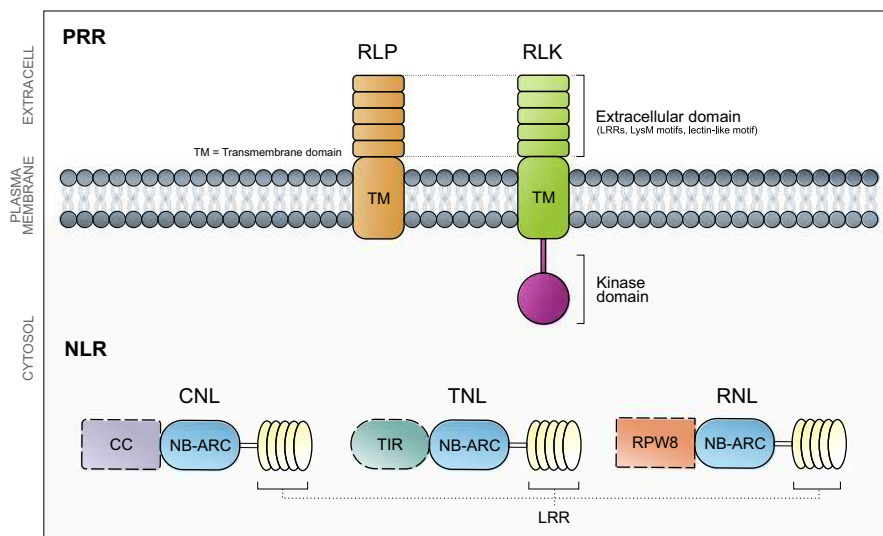


FIGURE 2.2

Overview of NLRs and PRRs' location and structural characteristics within cells. Classification of PRRs as RLKs and RLPs according to their structural differences in the intracellular region, with representation of extracellular domain (LRR, LysM motifs, lectin-like motifs in both RLKs and RLPs), transmembrane domain, and intracellular kinase domain present only in RLKs. Classification of NLRs according to their N-terminal domains: CNLs (either coiled-coil), TNLs (toll/interleukin-1 receptor), and RNLs (RPW8-like coiled-coil). Central nucleotide-binding domain represented in NB-ARC and C-terminal leucine-rich repeats domain.

NLRs, sharing multiple conserved motifs, like the Walker A and Walker B motifs, required for nucleotide binding and ATP hydrolysis, respectively (Steele et al., 2019). The motifs glycine-leucine-proline-leucine (GLPL) and methionine-histidine-aspartate (MHD) have also been seen as conserved in plants (Bendahmane et al., 2002; Sueldo et al., 2015). The N-terminal domain can be either coiled-coil (CC), N-terminal toll/interleukin-1 receptor (TIR), or RPW8-like coiled-coil (RPW8) (Fig. 2.2) (Steele et al., 2019). NRLs are categorized based on their terminal domains. A CC terminal domain is classified as a CNL (CC-NLR) if an NRL possesses a CC terminal domain. Similarly, if it has a TIR terminal domain, it is categorized as a TNL (TIR-NLR), and if it has an RPW8 domain, it falls under the RNL (RPW8-NLR) classification (Collier et al., 2011; Ngou, Heal, et al., 2022). LRR domains are highly diverse at the C-terminal, varying in the number and length of repeats and in noncanonical motifs (Fig. 2.2) (Takken & Goverse, 2012).

Decoding the plant immune system evolution

The first plant *R* gene, *Hm1*, of maize, was cloned in 1992 by Johal and Briggs. Starting in 1994, with the recognition of the first PRRs and NLRs, many immune receptors have been identified, as described by (Ngou, Ding, et al., 2022). Since then, the discovery and description of various lesion mimic mutants (LMM) have proven to be valuable tools in investigating the mechanisms of plant immunity, the physiological response, and the various signaling pathways that prevent pathogen success (Ramírez-Zavaleta et al., 2022).

Evolution of cell-surface immune receptors: PRR

Their origin of RLK receptors predates plants and animals' divergence, but they are notably absent in fungi. These RLKs form a monophyletic gene family (Jamieson et al., 2018; Shiu & Bleecker, 2001, 2003), which implies that they were present in the common ancestors of these organisms but were lost in fungi during evolution. RLKs experienced significant expansion in angiosperms and constitute approximately 60% of the kinases found in Arabidopsis. They can be classified into 44 subfamilies that belong to the RLK/Pelle family and are closely related to *Drosophila* Pelle proteins and animal cytoplasmic kinases (Gong & Han, 2021; Shiu & Bleecker, 2003). The expansion and diversification of RLKs have made their evolutionary history more complex, but it is evident that this occurred since the emergence of Streptophytes and intensified in land plants (Dievart et al., 2020). It is also important to note that both RLKs and RLPs probably expanded and diversified due to coevolution with their cognate MAMPs during the conquer of land plants (Steinbrenner, 2020). Recent phylogenetic studies have shown that the average number of *LRR-RLKs* genes in flowering plants is 250 per species. This specific receptor type can be classified into 20 subgroups, each with a distinct evolutionary history. Angiosperms have a wide range of genes encoding the *LRR-RLK-XII* subgroup (varying from 16 to

1129 genes per species), implying that specific lineages experience contraction or expansion (Dufayard et al., 2017; Ngou, Heal, et al., 2022).

RLPs with LRR ectodomains (LRR-RLPs) are the largest subfamily of RLPs in plants, with a varying number of genes across different plants (ranging from 2 to 585 genes encoding receptors in angiosperms) (Ngou, Ding, et al., 2022; Ngou, Heal, et al., 2022; Shumayla, 2016; Tyagi & Upadhyay, 2023). In contrast, the LysM subfamily, which mediates plant immunity and symbiosis with nitrogen-fixing bacteria, comprises only five LysM-RLK and three LysM-RLP genes in *Arabidopsis* (Ngou, Ding, et al., 2022; Rodríguez Melo et al., 2022; Shiu & Bleeker, 2001; Shumayla, 2021; Shumayla et al., 2022). The origin and evolution of LysM-RLKs in plants remain unclear. However, some studies suggest that they emerged before the divergence of land plants and Charophytes and subsequently underwent multiple duplication events during evolution. Certain LysM-RLKs exhibit lineage specificity, predominantly in legume and *Brassica* species (Abedi et al., 2022; Zhang et al., 2009). Interestingly, it has been observed that PRRs show a relatively lower frequency in Charophytes compared to bryophytes (Fig. 2.1B) (Ngou, Heal, et al., 2022).

Evolution of intracellular immune receptors: NLR

The transition of certain charophytes from aquatic to terrestrial environments represents a pivotal milestone in their evolutionary journey, which played a crucial role in the establishment and evolution of land plants. This shift brought new challenges, including encountering unfamiliar pathogen species and heightened infection levels. In this context, it is hypothesized that the emergence of R proteins facilitates the adaptation of charophytes and early land plants to their newly acquired terrestrial habitats (Hori et al., 2014; Rensing et al., 2008). The R genes in the genomes of Charophytes and all land plants support the hypothesis that they may have played a role in plant-pathogen interaction (Gao et al., 2018). Proteins containing the NB-ARC domain were not found in chlorophytes, rhodophytes, or glaucophytes, except for *Chondrus crispus*, a species of rhodophyte (Ngou, Heal, et al., 2022). In *C. crispus*, specific genes encode proteins that possess the NB-ARC domain. However, these proteins are classified differently from typical R proteins due to the lack of LRRs at the N terminus. Instead, they exhibit WD40 repeats, distinguishing them from the conventional R protein classification (Ngou, Heal, et al., 2022).

Among them, basal angiosperms like *Amborella* possess all types of NLRs. However, it is interesting that most monocots lack these NLRs, indicating a gene loss event after the split between monocots and eudicots (Liu et al., 2021; Tarr & Alexander, 2009). Angiosperms showed high variation in LRR-RLK-XII and LRR-RLP, which resemble NB-ARCs (Fig. 2.1B). Some studies propose that they predate the split of land plants and Charophytes and experienced multiple events of duplication, diversification, loss, and horizontal gene transfer during plant evolution (Adachi et al., 2019; Andolfo et al., 2019; Tamborski & Krasileva, 2020). Andolfo et al. (2019) suggest that the basic unit of the NLR receptor evolved in chlorophytes

(Fig. 2.1B) by horizontal gene transfer from a bacteria and diversified from TNL members. They also hypothesize that NLRs share a common origin with cell surface LRR receptors based on their high LRR homology to extracellular RLKs and RLPs.

The LRR-RLK-XII, LRR-RLP, and NLR gene families have experienced lineage-specific coexpansion or cocontraction since their emergence. The cause of these changes is unclear, but it may be linked to pathogen pressure and ecological specialization. The diversity of immune receptors may reflect the lifestyle and ecological niche of plants (Liu et al., 2021; Ngou, Heal, et al., 2022). For example, carnivorous and parasitic plants have a reduced root system that may limit pathogen entry into the plant. Some aquatic plants live partially or completely submerged, reducing exposure to airborne pathogen spores. These plants have a smaller repertoire of immune receptors than other plants (Liu et al., 2021; Ngou, Heal, et al., 2022). Likewise, life span may influence the evolution of these gene families. Ngou et al. (2022b) found that trees generally have higher percentages of PRRs and NB-ARCs than annual species, which have short periods of contact with pathogens. Therefore, long-term pathogen pressure may drive the expansion of immune receptor gene families in different plants throughout evolution.

Plants have evolved sophisticated defense mechanisms against microbes. However, our knowledge on this topic is still limited, and we need to explore novel aspects of the emergence of defense proteins involved in these mechanisms in Archaeplastida. Besides the evolutionary issue, understanding these molecular mechanisms may enable us to develop pathogen-resistant crops that can improve crop yields and ensure global food security. Some studies have already used this approach and employed synthetic and modified receptors that can trigger an immune response in the plant (Kourelis et al., 2023; Marchal et al., 2022). To move closer to this future, we thus need to conduct these investigations and research on immune receptors and *R* genes in plants.

Conclusions

In conclusion, the study of plant immunity has revealed a complex and dynamic evolutionary history of immune receptors, such as RLKs, RLPs, and NLRs, which have undergone lineage-specific coexpansion or cocontraction since their emergence. The expansion and diversification of these immune receptors have played a crucial role in plant–pathogen interactions, and their evolution has been shaped by coevolution with their cognate MAMPs. Although much progress has been made in understanding the mechanisms of plant immunity, the origin, and evolution of some immune receptors, such as LysM-RLKs, remain unclear. Further research is needed to explore the functional significance of the coevolution between plant immune receptors and pathogens and to gain a deeper understanding of the mechanisms underlying the evolution and diversification of these immune receptors.

References

- Abedi, A., Hajiahmadi, Z., Kordrostami, M., Esmael, Q., & Jacquard, C. (2022). Analyses of lysin-motif receptor-like kinase (Lysm-rlk) gene family in allotetraploid brassica napus L. and its progenitor species: An in silico study. *Cells*, 11(1). <https://doi.org/10.3390/cells11010037>
- Adachi, H., Derevnina, L., & Kamoun, S. (2019). NLR singletons, pairs, and networks: Evolution, assembly, and regulation of the intracellular immunoreceptor circuitry of plants. *Current Opinion in Plant Biology*, 50, 121–131. <https://doi.org/10.1016/j.pbi.2019.04.007>
- Aeby, G. S. (2007). First record of coralline lethal orange disease (CLOD) in the Northwestern Hawaiian Islands. *Coral Reefs*, 26(2). <https://doi.org/10.1007/s00338-007-0223-0>
- Aiyar, P., Schaeme, D., García-Altres, M., Carrasco Flores, D., Dathe, H., Hertweck, C., Sasso, S., & Mittag, M. (2017). Antagonistic bacteria disrupt calcium homeostasis and immobilize algal cells. *Nature Communications*, 8(1). <https://doi.org/10.1038/s41467-017-01547-8>
- Andolfo, G., Di Donato, A., Chiaiese, P., De Natale, A., Pollio, A., Jones, J. D. G., Frusciante, L., & Ercolano, M. R. (2019). Alien domains shaped the modular structure of plant NLR proteins. *Genome biology and evolution*, 11(12), 3466–3477. <https://doi.org/10.1093/gbe/evz248>
- Antil, D., & Upadhyay, S. K. (2023). *Evolutionary history of plant receptor-like kinases*. Elsevier BV. Article. <https://doi.org/10.1016/B978-0-323-90594-7.00010-7>
- Bendahmane, A., Farnham, G., Moffett, P., & Baulcombe, D. C. (2002). Constitutive gain-of-function mutants in a nucleotide binding site-leucine rich repeat protein encoded at the Rx locus of potato. *The Plant Journal*, 32(2), 195–204. <https://doi.org/10.1046/j.1365-313X.2002.01413.x>
- Bowles, A. M. C., Williamson, C. J., Williams, T. A., Lenton, T. M., & Donoghue, P. C. J. (2023). The origin and early evolution of plants. *Trends in Plant Science*, 28(3), 312–329. <https://doi.org/10.1016/j.tplants.2022.09.009>
- Bulgarelli, D., Schlaeppli, K., Spaepen, S., Van Themaat, E. V. L., & Schulze-Lefert, P. (2013). Structure and functions of the bacterial microbiota of plants. *Annual Review of Plant Biology*, 64, 807–838. <https://doi.org/10.1146/annurev-arplant-050312-120106>
- Collier, S. M., Hamel, L. P., & Moffett, P. (2011). Cell death mediated by the N-terminal domains of a unique and highly conserved class of NB-LRR protein. *Molecular Plant-Microbe Interactions*, 24(8), 918–931. <https://doi.org/10.1094/MPMI-03-11-0050>
- DeFalco, T. A., & Zipfel, C. (2021). Molecular mechanisms of early plant pattern-triggered immune signaling. *Molecular Cell*, 81(17), 3449–3467. <https://doi.org/10.1016/j.molcel.2021.07.029>
- Delaux, P. M., & Schornack, S. (2021). Plant evolution driven by interactions with symbiotic and pathogenic microbes. *Science*, 371(6531). <https://doi.org/10.1126/science.aba6605>
- de Vries, J., & Archibald, J. M. (2018). Plant evolution: Landmarks on the path to terrestrial life. *New Phytologist*, 217(4), 1428–1434. <https://doi.org/10.1111/nph.14975>
- de Vries, J., & Gould, S. B. (2018). The monoplastidic bottleneck in algae and plant evolution. *Journal of Cell Science*, 131(2). <https://doi.org/10.1242/jcs.203414>

- Dievart, A., Gottin, C., Périn, C., Ranwez, V., & Chantret, N. (2020). Origin and diversity of plant receptor-like kinases. *Annual Review of Plant Biology*, 71(1), 131–156. <https://doi.org/10.1146/annurev-arplant-073019-025927>
- Ding, H., & Ma, J. (2005). Simultaneous infection by red rot and chytrid diseases in *Porphyra yezoensis* Ueda. *Journal of Applied Phycology*, 17(1), 51–56. <https://doi.org/10.1007/s10811-005-5523-6>
- Dufayard, J. F., Bettembourg, M., Fischer, I., Droc, G., Guiderdoni, E., Périn, C., Chantret, N., & Diévar, A. (2017). New insights on Leucine-Rich repeats receptor-like kinase orthologous relationships in angiosperms. *Frontiers in Plant Science*, 8. <https://doi.org/10.3389/fpls.2017.00381>
- Durrant, W. E., & Dong, X. (2004). Systemic acquired resistance. *Annual Review of Phytopathology*, 42(1), 185–209. <https://doi.org/10.1146/annurev.phyto.42.040803.140421>
- Erb, M., & Kliebenstein, D. J. (2020). Plant secondary metabolites as defenses, regulators, and primary metabolites: The blurred functional Trichotomy1. *Plant Physiology*, 184(1), 39–52. <https://doi.org/10.1104/PP.20.00433>
- Gao, Y., Wang, W., Zhang, T., Gong, Z., Zhao, H., & Han, G. Z. (2018). Out of water: The ORI in and earl diversification of plant R-genes. *Plant Physiology*, 177(1), 82–89. <https://doi.org/10.1104/pp.18.00185>
- Gong, Z., & Han, G. Z. (2021). Flourishing in water: The early evolution and diversification of plant receptor-like kinases. *The Plant Journal*, 106(1), 174–184. <https://doi.org/10.1111/tbj.15157>
- Gould, S. B., Waller, R. F., & McFadden, G. I. (2008). Plastid evolution. *Annual Review of Plant Biology*, 59, 491–517. <https://doi.org/10.1146/annurev.arplant.59.032607.092915>
- Gust, A. A., & Felix, G. (2014). Receptor like proteins associate with SOBIR1-type of adaptors to form bimolecular receptor kinases. *Current Opinion in Plant Biology*, 21, 104–111. <https://doi.org/10.1016/j.pbi.2014.07.007>
- Han, G. Z. (2019). Origin and evolution of the plant immune system. *New Phytologist*, 222(1), 70–83. <https://doi.org/10.1111/nph.15596>
- Han, Z., Sun, Y., & Chai, J. (2014). Structural insight into the activation of plant receptor kinases. *Current Opinion in Plant Biology*, 20, 55–63. <https://doi.org/10.1016/j.pbi.2014.04.008>
- Herron, M. D., Hackett, J. D., Aylward, F. O., & Michod, R. E. (2009). Triassic origin and early radiation of multicellular volvocine algae. *Proceedings of the National Academy of Sciences of the United States of America*, 106(9), 3254–3258. <https://doi.org/10.1073/pnas.0811205106>
- Hori, K., Maruyama, F., Fujisawa, T., Togashi, T., Yamamoto, N., Seo, M., Sato, S., Yamada, T., Mori, H., Tajima, N., Moriyama, T., Ikeuchi, M., Watanabe, M., Wada, H., Kobayashi, K., Saito, M., Masuda, T., Sasaki-Sekimoto, Y., Mashiguchi, K., ... Ohta, H. (2014). Klebsormidium flaccidum genome reveals primary factors for plant terrestrial adaptation. *Nature Communications*, 5(1), 3978. <https://doi.org/10.1038/ncomms4978>
- Jamieson, P. A., Shan, L., & He, P. (2018). Plant cell surface molecular cypher: Receptor-like proteins and their roles in immunity and development. *Plant Science*, 274, 242–251. <https://doi.org/10.1016/j.plantsci.2018.05.030>
- Jones, J. D. G., & Dangl, J. L. (2006). The plant immune system. *Nature*, 444(7117), 323–329. <https://doi.org/10.1038/nature05286>

- Kaur, A., et al. (2023). Ascorbate oxidases in bread wheat: Gene regulatory network, transcripts profiling, and interaction analyses provide insight into their role in plant development and stress response. *Plant Growth Regulation*. <https://doi.org/10.1007/s10725-023-01103-z>
- Kourelis, J., Marchal, C., Posbeykian, A., Harant, A., & Kamoun, S. (2023). NLR immune receptor—nanobody fusions confer plant disease resistance. *Science*, 379(6635), 934–939. <https://doi.org/10.1126/science.abn4116>
- Li, W., Zhang, T., Tang, X., & Wang, B. (2010). Oomycetes and fungi: Important parasites on marine algae. *Acta Oceanologica Sinica*, 29(5), 74–81. <https://doi.org/10.1007/s13131-010-0065-4>
- Liu, Y., Zeng, Z., Zhang, Y. M., Li, Q., Jiang, X. M., Jiang, Z., Tang, J. H., Chen, D., Wang, Q., Chen, J. Q., & Shao, Z. Q. (2021). An angiosperm NLR Atlas reveals that NLR gene reduction is associated with ecological specialization and signal transduction component deletion. *Molecular Plant*, 14(12), 2015–2031. <https://doi.org/10.1016/j.molp.2021.08.001>
- Madhu, et al. (2024). Modulation in gene expression and enzyme activity suggested the roles of monodehydroascorbate reductase in development and stress response in bread wheat. *Plant Science*, 338, 111902. <https://doi.org/10.1016/j.plantsci.2023.111902>
- Marchal, C., Pai, H., Kamoun, S., & Kourelis, J. (2022). Emerging principles in the design of bioengineered made-to-order plant immune receptors. *Current Opinion in Plant Biology*, 70, 102311. <https://doi.org/10.1016/j.pbi.2022.102311>
- Martin, G. B., Bogdanove, A. J., & Sessa, G. (2003). Understanding the functions of plant disease resistance proteins. *Annual Review of Plant Biology*, 54, 23–61. <https://doi.org/10.1146/annurev.arplant.54.031902.135035>
- Meyers, B. C., Kozik, A., Griego, A., Kuang, H., & Michelmore, R. W. (2003). Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. *The Plant Cell*, 15(4), 809–834. <https://doi.org/10.1105/tpc.009308>
- Nagappa, R. (2006). A review on fungal diseases of algae, marine fishes, shrimps and corals. *Indian Journal of Marine Sciences*, 35.
- Ngou, B. P. M., Ahn, H. K., Ding, P., & Jones, J. D. G. (2021). Mutual potentiation of plant immunity by cell-surface and intracellular receptors. *Nature*, 592(7852), 110–115. <https://doi.org/10.1038/s41586-021-03315-7>
- Ngou, B. P. M., Ding, P., & Jones, J. D. G. (2022). Thirty years of resistance: Zig-zag through the plant immune system. *The Plant Cell*, 34(5), 1447–1478. <https://doi.org/10.1093/plcell/koac041>
- Ngou, B. P. M., Heal, R., Wyler, M., Schmid, M. W., & Jones, J. D. G. (2022). Concerted expansion and contraction of immune receptor gene repertoires in plant genomes. *Nature Plants*, 8(10), 1146–1152. <https://doi.org/10.1038/s41477-022-01260-5>
- Parfrey, L. W., Lahr, D. J. G., Knoll, A. H., & Katz, L. A. (2011). Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proceedings of the National Academy of Sciences of the United States of America*, 108(33), 13624–13629. <https://doi.org/10.1073/pnas.1110633108>
- Poulin, R., & Morand, S. (2000). The diversity of parasites. *The Quarterly Review of Biology*, 75(3), 277–293. <https://doi.org/10.1086/393500>
- Raffaele, S., Farrer, R. A., Cano, L. M., Studholme, D. J., MacLean, D., Thines, M., Jiang, R. H. Y., Zody, M. C., Kunjeti, S. G., Donofrio, N. M., Meyers, B. C., Nusbaum, C., & Kamoun, S. (2010). Genome evolution following host jumps in the Irish

- potato famine pathogen lineage. *Science*, 330(6010), 1540–1543. <https://doi.org/10.1126/science.1193070>
- Raffaele, S., & Kamoun, S. (2012). Genome evolution in filamentous plant pathogens: Why bigger can be better. *Nature Reviews Microbiology*, 10(6), 417–430. <https://doi.org/10.1038/nrmicro2790>
- Ramanan, R., Kim, B. H., Cho, D. H., Oh, H. M., & Kim, H. S. (2016). Algae-bacteria interactions: Evolution, ecology and emerging applications. *Biotechnology Advances*, 34(1), 14–29. <https://doi.org/10.1016/j.biotechadv.2015.12.003>
- Ramírez-Zavaleta, C. Y., García-Barrera, L. J., Rodríguez-Verástegui, L. L., Arrieta-Flores, D., & Gregorio-Jorge, J. (2022). An Overview of PRR- and NLR-mediated immunities: Conserved signaling components across the plant kingdom that communicate both pathways. *International Journal of Molecular Sciences*, 23(21). <https://doi.org/10.3390/ijms232112974>
- Redecker, D., Kodner, R., & Graham, L. E. (2000). Glomalean fungi from the ordovician. *Science*, 289(5486), 1920–1921. <https://doi.org/10.1126/science.289.5486.1920>
- Reina-Pinto, J. J., & Yephremov, A. (2009). Surface lipids and plant defenses. *Plant Physiology and Biochemistry*, 47(6), 540–549. <https://doi.org/10.1016/j.plaphy.2009.01.004>
- Rensing, S. A., Lang, D., Zimmer, A. D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., Perroud, P. F., Lindquist, E. A., Kamisugi, Y., Tanahashi, T., Sakakibara, K., Fujita, T., Oishi, K., Shin-I, T., Kuroki, Y., Toyoda, A., Suzuki, Y., Hashimoto, S. I., ... Boore, J. L. (2008). The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. *Science*, 319(5859), 64–69. <https://doi.org/10.1126/science.1150646>
- Rey, T., Chatterjee, A., Buttay, M., Toulotte, J., & Schornack, S. (2015). Medicago truncatula symbiosis mutants affected in the interaction with a biotrophic root pathogen. *New Phytologist*, 206(2), 497–500. <https://doi.org/10.1111/nph.13233>
- Rodríguez Melo, J., Tonelli, M. L., Barbosa, M. C., Ariel, F., Zhao, Z., Wang, J., Fabra, A., & Ibañez, F. (2022). Evolution of LysM-RLK gene family in wild and cultivated peanut species. *Horticulturae*, 8(11). <https://doi.org/10.3390/horticulturae8111000>
- Saijo, Y., Loo, E. P. I., & Yasuda, S. (2018). Pattern recognition receptors and signaling in plant–microbe interactions. *The Plant Journal*, 93(4), 592–613. <https://doi.org/10.1111/tpj.13808>
- Selosse, M. A., Strullu-Derrien, C., Martin, F. M., Kamoun, S., & Kenrick, P. (2015). Plants, fungi and oomycetes: A 400-million year affair that shapes the biosphere. *New Phytologist*, 206(2), 501–506. <https://doi.org/10.1111/nph.13371>
- Sharma, Y., et al. (2024). Decoding the features and potential roles of respiratory burst oxidase homologs in bread wheat. *Current Plant Biology*, 37, 100315. <https://doi.org/10.1016/j.cpb.2023.100315>
- Shekhawat, J., & Upadhyay, S. K. (2024). DPY1 as an osmosensor for drought signaling. *Trends in Plant Science*. <https://doi.org/10.1016/j.tplants.2023.12.011>
- Shiu, S. H., & Bleecker, A. B. (2003). Expansion of the receptor-like kinase/Pelle gene family and receptor-like proteins in Arabidopsis. *Plant Physiology*, 132(2), 530–543. <https://doi.org/10.1104/pp.103.021964>
- Shiu, S. H., & Bleecker, A. B. (2001). Receptor-like kinases from Arabidopsis form a monophyletic gene family related to animal receptor kinases. *Proceedings of the National Academy of Sciences of the United States of America*, 98(19), 10763–10768. <https://doi.org/10.1073/pnas.181141598>

- Shumayla, et al. (2016). Genomic dissection and expression profiling revealed functional divergence in *Triticum aestivum* leucine rich repeat receptor like kinases (TaLRRKs). *Frontiers in Plant Science*, 7, 1374. <https://doi.org/10.3389/fpls.2016.01374>
- Shumayla, et al. (2021). LysM domain-containing proteins modulate stress response and signalling in *Triticum aestivum* L. *Environmental and Experimental Botany*, 183, 104558. <https://doi.org/10.1016/j.envexpbot.2021.104558>
- Shumayla, Tyagi, S., & Upadhyay, S. K. (2019). Receptor-like kinases and environmental stress in plants. *Energy, Environment, and Sustainability*, 79–102. https://doi.org/10.1007/978-981-15-0690-1_4
- Shumayla, & Upadhyay, S. K. (2023). *An overview of receptor-like kinases in plants*. Elsevier BV. Article. <https://doi.org/10.1016/B978-0-323-90594-7.00005-3>
- Shumayla, S. T., Tyagi, S., & Upadhyay, S. K. (2022). *Structural and functional analysis of lysin motif receptor-like kinases in plants*. India: Elsevier. <https://doi.org/10.1016/B978-0-323-90594-7.00002-8>
- Steele, J. F. C., Hughes, R. K., & Banfield, M. J. (2019). Structural and biochemical studies of an NB-ARC domain from a plant NLR immune receptor. *PLoS One*, 14(8). <https://doi.org/10.1371/journal.pone.0221226>
- Steinbreinner, A. D. (2020). The evolving landscape of cell surface pattern recognition across plant immune networks. *Current Opinion in Plant Biology*, 56, 135–146. <https://doi.org/10.1016/j.pbi.2020.05.001>
- Sueldo, D. J., Shimels, M., Spiridon, L. N., Caldararu, O., Petrescu, A. J., Joosten, M. H. A. J., & Tameling, W. I. L. (2015). Random mutagenesis of the nucleotide-binding domain of NRC1 (NB-LRR Required for Hypersensitive Response-Associated Cell Death-1), a downstream signalling nucleotide-binding, leucine-rich repeat (NB-LRR) protein, identifies gain-of-function mutations in the nucleotide-binding pocket. *New Phytologist*, 208(1), 210–223. <https://doi.org/10.1111/nph.13459>
- Sukarta, O. C. A., Slootweg, E. J., & Goverse, A. (2016). Structure-informed insights for NLR functioning in plant immunity. *Seminars in Cell & Developmental Biology*, 56, 134–149. <https://doi.org/10.1016/j.semcdb.2016.05.012>
- Takken, F. L. W., & Goverse, A. (2012). How to build a pathogen detector: Structural basis of NB-LRR function. *Current Opinion in Plant Biology*, 15(4), 375–384. <https://doi.org/10.1016/j.pbi.2012.05.001>
- Tamborski, J., & Krasileva, K. V. (2020). Evolution of plant NLRs: From natural history to precise modifications. *Annual Review of Plant Biology*, 71, 355–378. <https://doi.org/10.1146/annurev-arplant-081519-035901>
- Tarr, D. E. K., & Alexander, H. M. (2009). TIR-NBS-LRR genes are rare in monocots: Evidence from diverse monocot orders. *BMC Research Notes*, 2. <https://doi.org/10.1186/1756-0500-2-197>
- Thomma, B. P. H. J., Nürnberger, T., & Joosten, M. H. A. J. (2011). Of PAMPs and effectors: The blurred PTI-ETI dichotomy. *The Plant Cell*, 23(1), 4–15. <https://doi.org/10.1105/tpc.110.082602>
- Tyagi, S., & Upadhyay, S. K. (2023). *Role of leucine-rich repeat receptor-like kinases in abiotic and biotic stress responses in plants*. Elsevier BV. <https://doi.org/10.1016/b978-0-323-90594-7.00017-x>
- Upadhyay, S. K. (2023). Calcium channels, OST1 and stomatal defence: Current status and beyond. *Cells*, 12(1), 127. <https://doi.org/10.3390/cells12010127>
- Upadhyay, S. K., & Shumayla. (2022). *Plant receptor-like kinases: Role in development and stress*. Elsevier. <https://doi.org/10.1016/C2020-0-03988-5>

- Upson, J. L., Zess, E. K., Białas, A., Wu, C.h., & Kamoun, S. (2018). The coming of age of EvoMPMI: Evolutionary molecular plant–microbe interactions across multiple timescales. *Current Opinion in Plant Biology*, 44, 108–116. <https://doi.org/10.1016/j.pbi.2018.03.003>
- van der Burgh, A. M., & Joosten, M. H. A. J. (2019). Plant immunity: Thinking outside and inside the box. *Trends in Plant Science*, 24(7), 587–601. <https://doi.org/10.1016/j.tplants.2019.04.009>
- Vigneron, N., Radhakrishnan, G. V., & Delaux, P. M. (2018). What have we learnt from studying the evolution of the arbuscular mycorrhizal symbiosis? *Current Opinion in Plant Biology*, 44, 49–56. <https://doi.org/10.1016/j.pbi.2018.02.004>
- Vlot, A. C., Dempsey, D. A., & Klessig, D. F. (2009). Salicylic acid, a multifaceted hormone to combat disease. *Annual Review of Phytopathology*, 47(1), 177–206. <https://doi.org/10.1146/annurev.phyto.050908.135202>
- Wang, D., Dong, W., Murray, J., & Wang, E. (2022). Innovation and appropriation in mycorrhizal and rhizobial Symbioses. *The Plant Cell*, 34(5), 1573–1599. <https://doi.org/10.1093/plcell/koac039>
- Wang, S., Li, L., Li, H., Sahu, S. K., Wang, H., Xu, Y., Xian, W., Song, B., Liang, H., Cheng, S., Chang, Y., Song, Y., Çebi, Z., Wittek, S., Reder, T., Peterson, M., Yang, H., Wang, J., Melkonian, B., ... Liu, X. (2020). Genomes of early-diverging streptophyte algae shed light on plant terrestrialization. *Nature Plants*, 6(2), 95–106. <https://doi.org/10.1038/s41477-019-0560-3>
- Wang, Y., Li, X., Fan, B., Zhu, C., & Chen, Z. (2021). Regulation and function of defense-related callose deposition in plants. *International Journal of Molecular Sciences*, 22(5), 2393. <https://doi.org/10.3390/ijms22052393>
- Westwood, J. H., Yoder, J. I., Timko, M. P., & dePamphilis, C. W. (2010). The evolution of parasitism in plants. *Trends in Plant Science*, 15(4), 227–235. <https://doi.org/10.1016/j.tplants.2010.01.004>
- Yang, E. C., Boo, S. M., Bhattacharya, D., Saunders, G. W., Knoll, A. H., Fredericq, S., Graf, L., & Yoon, H. S. (2016). Divergence time estimates and the evolution of major lineages in the florideophyte red algae. *Scientific Reports*, 6. <https://doi.org/10.1038/srep21361>
- Zhang, X. C., Cannon, S. B., & Stacey, G. (2009). Evolutionary genomics of LysM genes in land plants. *BMC Evolutionary Biology*, 9(1). <https://doi.org/10.1186/1471-2148-9-183>
- Zhou, J. M., & Zhang, Y. (2020). Plant immunity: Danger perception and signaling. *Cell*, 181(5), 978–989. <https://doi.org/10.1016/j.cell.2020.04.028>

3.4. CAPÍTULO 3 – AI-GUIDED TWO-STAGE SCREENING IDENTIFIES PUTATIVE HAUSTORIUM NON-INDUCERS / INHIBITORS MOLECULES IN PARASITIC PLANTS

Este capítulo apresenta o manuscrito **em preparação** “*AI-guided two-stage screening identifies putative haustorium non-inducers / inhibitors molecules in parasitic plants*”, desenvolvido durante período sanduíche (junho a dezembro de 2025) na universidade *Imperial College London* (Londres, Reino Unido), sob supervisão do Prof. Dr. Pedro J. Ballester, cujo grupo trabalha com o desenvolvimento e aplicação de métodos computacionais, IA e big data para acelerar descobertas em biologia e saúde. Esse estágio foi apoiado pela Chamada Interna PRPG 04/2024 – Doutorado Sanduíche no Exterior, Programa FAPEMIG de Fomento à Internacionalização das ICTMGs, financiada pela Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG). O manuscrito integra modelagem estrutural, triagem virtual e ML para investigar interações conservadas entre a PRR AtCARD1 e seus homólogos em plantas parasitas, assim como identificação de potenciais novas HIFs.

Project Title: Harnessing Artificial Intelligence to Combat Parasitic Weeds

Proposed paper title: **AI-guided two-stage screening identifies putative haustorium non-inducers / inhibitors molecules in parasitic plants**

Wenderson Felipe Costa Rodrigues¹, Pedro J. Ballester², Juliane K. Ishida^{1,3}

¹ Graduate Program in Plant Biology, Department of Botany, Institute of Biological Sciences (ICB), Federal University of Minas Gerais (UFMG), Brazil.

² Faculty of Engineering, Department of Bioengineering, White City Campus, Imperial College London, UK.

³ Department of Genetics, Luiz de Queiroz College of Agriculture (ESALQ), University of São Paulo (USP), Brazil.

Abstract

Parasitic plants rely on the perception of host-derived chemical signals, such as quinones, to initiate haustorium formation, an organ essential for host invasion and resource acquisition. The receptor CARD1, originally described in *Arabidopsis thaliana* as a sensor of H₂O₂ and quinones in plant immune signaling, has been co-opted by parasitic plants to regulate this process through the recognition of haustorium-inducing factors (HIFs). Here, we investigate the structural and computational basis of the interaction between CARD1 and HIFs, as well as non-haustorium-inducing or inhibitory molecules. Using AlphaFold-predicted structural models of AtCARD1 and its parasitic homologs from *Phtheirospermum japonicum* and *Striga asiatica*, we identified a conserved ligand-binding site associated with a histidine triad previously described as coordinating Cu⁺, as well as an adjacent secondary site stabilized by a SEA-like domain. Molecular docking revealed that Cu⁺ presence modulates quinone binding geometry and stability, with clear differences between AtCARD1 and its parasitic homologs. Molecular dynamics simulations further indicated distinct ligand stabilization modes across these sites. To prioritize potential haustorium inhibitors, we evaluated virtual screening strategies based on molecular docking and machine learning approaches using QSAR descriptors (Morgan fingerprints) and protein-ligand (PLEC) features. Docking alone showed limited early enrichment. In contrast, machine learning models, particularly Support Vector Machines (SVM) and neural networks, exhibited superior and robust performance even under highly imbalanced conditions with decoys, largely independent of the receptor or species analyzed. Based on these results, we propose a two-stage screening strategy combining QSAR-based large-scale prioritization with PLEC-based structural refinement, resulting in the ranking of more than 50,000 compounds as putative non-haustorium-inducing or inhibitory candidates for subsequent *in vitro* validation.

Introduction

CARD1, from *Arabidopsis thaliana* (also known as AtHPCA1; gene ID AT5G49760), encodes a leucine-rich repeat receptor-like kinase (LRR-RLK) that plays a central role in both H₂O₂ sensing and immune signaling in plants¹ (Fig. 1A). These functions are essential for detecting and responding to environmental and pathogenic stress. One of CARD1's most remarkable features is the perception of quinones such as 2,6-dimethoxy-p-benzoquinone (DMBQ), a small redox-active molecule known to (i) induce haustorium formation, specialized invasive organs used by parasitic plants to attach to and extract resources from their hosts, and (ii) act as pathogen-associated molecular signals in host immune responses². This dual role of DMBQ highlights the evolutionary significance of CARD1: parasitic plants appear to have co-opted CARD1-mediated immune signaling pathways, repurposing them to detect haustorium-inducing factors (HIFs) such as DMBQ, and trigger the developmental programs required for haustorium organogenesis.

Structurally, AtCARD1 consists of an extracellular LRR domain, a transmembrane helix, and an intracellular kinase domain (Fig. 1A). The extracellular region is essential for quinone perception, and several conserved cysteine residues (C395, C405, C421, C424, C434 and C436) (Fig. 1A; Sup. Fig. 1) within this domain are considered critical for signal transduction. Previous mutagenesis studies showed that alteration of these cysteines abolishes DMBQ-induced calcium signaling, consistent with their proposed involvement in thiol–quinone interactions that are typical of redox-based perception mechanisms^{1,2}. This is further supported by the electrophilic nature of quinones, which enables them to react with nucleophilic thiol groups via covalent or redox-mediated modifications^{3–5}. Interestingly, *in vitro* assays have detected no direct binding between DMBQ and the AtCARD1 ectodomain (ECD), suggesting that DMBQ might not act as a true ligand, or alternatively, that its recognition may require additional *in vivo* factors such as co-receptors or membrane context. Moreover, some tested quinones triggered stronger physiological responses and exhibited lower apparent K_m values than DMBQ^{2,6}, implying that multiple quinone-like molecules could serve as effective signals within the CARD1-mediated signaling pathway.

Very recently, a preprint described the cryo-EM structure of the AtCARD1 ECD, revealing that the previously assumed cysteine residues play a predominantly structural role rather than functioning as redox sensors⁷. Mutations in these cysteines primarily reduced protein stability rather than ligand perception, as they form intramolecular disulfide bridges that stabilize the LRR fold⁷. Instead, the study identified a histidine triad (H197, H199 and H222) (Fig. 1A; Sup. Fig. 1) coordinating a copper ion (Cu⁺) in a T-shaped geometry, the first evidence of metal-dependent signaling in a plant LRR-RLK. This His-triad region, located at the interface between the LRR and SEA-like domains, forms an exposed cavity capable of accommodating small redox-active molecules, including quinones⁷. The SEA-like domain stabilizes this interface through hydrogen bonds and glycan contacts, suggesting that this domain may modulate the accessibility or reactivity

of the metal site. Although the authors associated this copper-binding site mainly with H_2O_2 perception, its structural and electrostatic properties are also consistent with a potential quinone-interacting pocket.

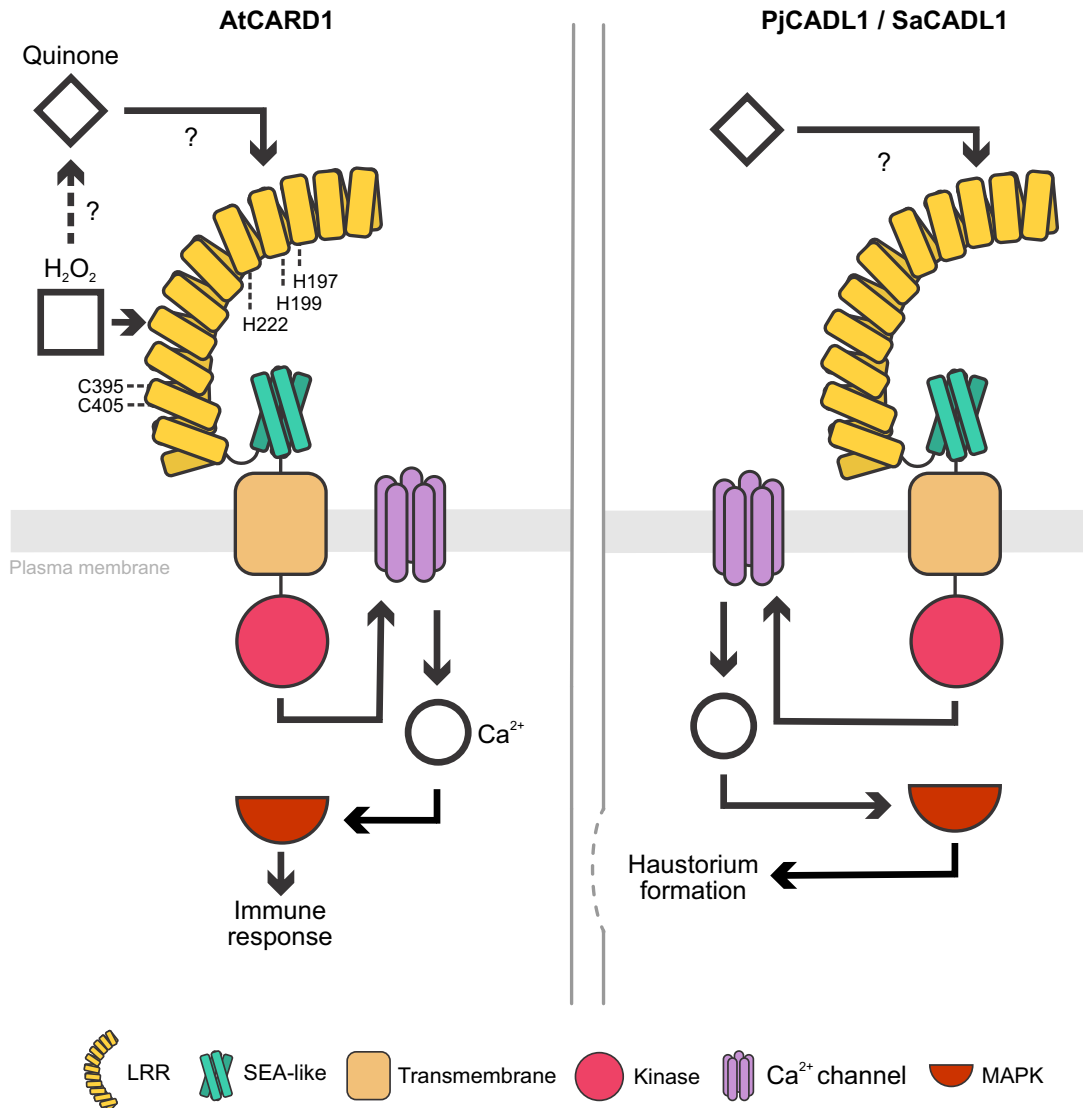


Figure 1. Schematic representation of quinone and H_2O_2 recognition by AtCARD1 and its homologs. Diagram summarizing the models proposed by Körner, Slaby & Albert (2020) and Mutuku et al. (2020). In *Arabidopsis*, CARD1 perceives quinones or H_2O_2 , triggering Ca^{2+} influx and MAPK-dependent immune responses. In parasitic plants, the CARD1 homologs PjCADL1 (from *Phtheirospermum japonicum*) and SaCADL1 (from *Striga asiatica*) are thought to recognize quinones, leading to Ca^{2+} signaling and haustorium formation. Question marks indicate unknown recognition mechanisms, and dashed lines represent hypothetical interactions.

Despite these recent advances, no crystallographic structure of AtCARD1 or any of its homologs is currently available. The only full-length structural model comes from AlphaFold, which predicts the three-dimensional conformation of these proteins with high confidence, particularly in its C-terminal extracellular region with a pLDDT (predicted Local Distance Difference Test) > 90, indicating high

confidence in the atomic positions. Although this prediction provides a reliable framework for hypothesis testing, no experimentally validated ligand-binding site has been described to date. Consequently, AtCARD1 remains structurally unresolved regarding the precise molecular determinants of quinone perception.

Across land plants, AtCARD1 and its homologs (CADLs) are conserved and represent a potential evolutionary link between ancestral immune perception and developmental signaling in parasitic lineages. The ability of parasitic species to detect host-derived quinones through CARD1-like receptors suggests that haustorium initiation may have evolved by recruiting components of pre-existing redox and immune signaling pathways. To investigate the molecular basis of this transition, we will apply molecular docking and structure-based virtual screening to explore putative ligand-binding regions in AtCARD1. By integrating these analyses with machine learning approaches trained on datasets of known molecules, mostly HIFs, we further aimed to predict novel candidate molecules predicted to bind AtCARD1 and its homologs in parasitic plants. Together, these analyses provide a structural and evolutionary framework to understand how redox signaling, metal coordination, and ligand perception may have been repurposed to regulate parasitic plant development.

Materials and Methods

Homologs search and alignment

To identify homologs of *A. thaliana* CARD1 (AtCARD1), a BLASTp⁸ search (E-value $\leq 1e-100$) was performed using AtCARD1 as query against a custom database comprising representative species from the major Archaeplastida (plant) lineages, including Glaucophyta (glaucophyte algae), Rhodophyta (red algae), Chlorophyta (green algae), Charophyta (green algae; the sister clade of land plants), and Embryophyta (land plants) (species list in Supplementary Table S1). Sequences with $\geq 80\%$ identity were retained using custom scripts. Multiple sequence alignment was performed with MAFFT⁹ (--reorder --leavegappyregion --maxiterate 1000 --retree 1 --localpair --anysymbol), visualized in ESPript 3¹⁰ (<https://esprict.ibcp.fr/ESPript/ESPript/>) (Supplementary Fig. S1), and sequence logos were generated using WebLogo 3¹¹ (<https://weblogo.threeplusone.com/>).

Given the lack of experimentally determined structures for CARD1 or its homologs, we additionally conducted a structural homology search using FoldSeek and BLASTp against the RCSB Protein Data Bank (PDB) to identify possible three-dimensional analogs. Structural similarity was evaluated using the FoldSeek web server¹² (<https://search.foldseek.com/search>) in 3Di/AA mode against the PDB100 (20240101) database. This approach, which integrates geometric and sequence-based features, revealed partial structural coverage, with higher similarity in the intracellular kinase region (Supplementary Fig. S2A). Complementary BLASTp^{8,13} searches (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) (PDB release 2025/07/06) confirmed that most hits corresponded to the intracellular domain (Supplementary Fig. S2B). To focus on the extracellular

portion, where quinone binding is hypothesized, results were filtered by sequence identity (>30%) and structural coverage of extracellular regions. Non-native proteins or structures containing engineered mutations or artificial complexes were excluded. Only native entries, either apo or bound to small molecules (e.g., ATP, ADP), were retained for downstream comparison (listed in Supplementary Table S2). Multiple sequence alignments were performed using Clustal Omega¹⁴ (v1.2.4) (<https://www.ebi.ac.uk/jdispatcher/msa/clustalo>) and visualized with ESPript 3. The topology of AtCARD1 was delineated using DeepTMHMM¹⁵ (v1.0.44) (<https://dtu.biolib.com/DeepTMHMM>), a deep learning model for transmembrane topology prediction and classification. To ensure consistency with the ECD-only nature of the selected templates, the intracellular kinase and transmembrane domains were excluded in PyMOL¹⁶ (v3.1.6.1) and aligned with the *align* plugin. Despite local conservation, particularly the alignment of a histidine residue in 8WEG (Supplementary Fig. S3), the structural comparison revealed limited homology within the extracellular region. Consequently, all subsequent analyses were carried out using AlphaFold-predicted structures of AtCARD1 and its parasitic plant homologs, PjCADL1 (from *Phtheirospermum japonicum*) and SaCADL1 (from *Striga asiatica*).

Binding site prediction

To predict potential ligand-binding pockets in CARD1 and its homologs, we employed PrankWeb¹⁷ (<https://prankweb.cz/>), which builds upon P2Rank, a machine learning based method for the prediction of ligand-binding sites directly from protein structure. This algorithm is tolerant to AlphaFold-predicted models and performs well on elongated or shallow grooves such as those found in LRR-RLKs. To evaluate the reliability of this approach, we first used FLAGELLIN-SENSITIVE 2 (FLS2) as a positive control. FLS2 is a well-characterized LRR-RLK that recognizes the bacterial peptide flg22 through its extracellular LRR domain and forms a complex with its co-receptor BRASSINOSTEROID INSENSITIVE 1–ASSOCIATED KINASE 1 (BAK1)¹⁸. The ECDs of both receptors adopt a solenoidal LRR fold, comparable to AtCARD1. We used the 4MN8 crystal structure as input, including chain A (FLS2) and chain C (BAK1), and the AlphaFold (AF) model of FLS2. Two independent PrankWeb runs were performed for each model, one with and one without evolutionary conservation. In the conservation-based mode, PrankWeb integrates per-residue conservation scores obtained through phmmer¹⁹ searches against UniRef50 (Release 2021-03), weighting sequences via the Gerstein/Sonnhammer/Chothia algorithm and computing per-column information content (0 - 4 bits). This mode enhances pocket scoring when functionally conserved residues cluster spatially.

Despite PrankWeb being primarily optimized for small-molecule ligands rather than peptides, predictions using the crystal structure (PDB: 4MN8) correctly identified pockets that overlapped with the experimentally verified flg22-binding site (residues Y148, R152, Y272, Y296) described by Sun et al. (2013) (Supplementary Table S3). In contrast, when using the AF-FLS2 model, which

lacks the BAK1 co-receptor, the corresponding pocket was recovered only in the non-conservation mode and with substantially lower confidence scores, likely reflecting the absence of the stabilizing FLS2–BAK1 interface.

Given this positive control outcome, we applied the same pipeline to AtCARD1, PjCADL1, and SaCADL1 AF models. For each receptor, the extracellular domain (ECD) was isolated in PyMOL, retaining residues with pLDDT > 70, and exported in .pdb format for input into PrankWeb.

Based on the findings of Ishihama et al. (2025), which identified a Cu⁺-binding histidine triad essential for H₂O₂ perception, an additional AtCARD1-specific model was prepared containing a Cu⁺ ion manually positioned (PyMOL) according to the reported coordination geometry (Supplementary Fig. S4). This “+Cu” model was treated as a structural proxy and analyzed independently from the Cu⁺-free (“-Cu”) model. No explicit metal-ligand coordination or electronic effects of Cu⁺ were modeled.

To also assess the contribution of the SEA-like domain (which has been shown to interact with the LRR region and potentially modulate receptor activity⁷), further predictions were performed using AtCARD1ΔSEA constructs, in which the SEA-like domain was removed in PyMOL based on the structural boundaries reported by Ishihama et al. (2025). All corresponding coordinates were saved in .pdb format and used as inputs for PrankWeb analyses.

Molecular docking calibration

Following pocket prediction, a docking calibration step was performed to evaluate whether the chosen parameters could discriminate biologically active quinones that induce haustorium formation from inactive or toxic analogs that do not. The ligands and their corresponding biological classifications are summarized in Supplementary Table S4.

Docking simulations were carried out using Smina²⁰ (based on AutoDock Vina²¹ v1.1.2), focusing on three regions of AtCARD1 previously defined by pocket prediction and structural analysis: the predicted Pocket 1 (PrankWeb top-ranked site), the tri-histidine (triHIS) pocket (Cu⁺-binding site), and the C-C pocket (centered on the C395-C405 region) (Supplementary Fig. S5). Additionally, a “whole-protein” (WholeP) docking box encompassing the entire ECD was included as a blind-docking control to explore potential alternative or noncanonical binding sites not captured by the predefined pockets. Although blind docking is intrinsically more challenging due to the enlarged search space and reduced discrimination power of approximate scoring functions, this strategy was considered informative given the absence of a well-defined binding pocket for quinones in AtCARD1.

For target preparation, both Cu⁺-bound (“+Cu”) and metal-free (“-Cu”) versions of AtCARD1, PjCADL1, and SaCADL1 were generated. The Cu⁺ ion was manually positioned at the histidine triad site according to the coordination

geometry described by Ishihama et al. (2025). Each receptor was processed in UCSF Chimera²² using the DockPrep tool (Tools\Structure Editing\Dock Prep) with all options enabled except “Delete non-complexed ions” and “Add charges”. Hydrogen atoms were added considering hydrogen-bonding networks, and protonation states of histidines were manually adjusted in Cu-bound models to maintain correct metal coordination (His197 = ϵ -nitrogen, His199 = δ -nitrogen, His222 = ϵ -nitrogen, other histidines unprotonated). For the metal-free receptors, the “residue-name-based” protonation option was used. All prepared receptors were saved in .mol2 format. Ligands (Supplementary Table S4) were prepared under identical conditions using Dock Prep, with hydrogen addition and charge assignment optimized for each molecule, and exported as .mol2 files.

Geometric centers (X, Y, Z coordinates) were defined in PyMOL based on key structural residues or predictor output. For the predicted pocket, the center coordinates provided directly by PrankWeb were used. For the remaining sites, centers were defined as follows: the midpoint between Cys395 and Cys405 for the C-C pocket, the centroid of His197, His199, and His222 for the triHIS pocket, and the center of mass of the entire ECD for the WholeP configuration. The docking box corresponding to the WholeP setup, encompassing the full ECD volume, was generated using a custom Python script prior to running Smina. Multiple parameter combinations were tested to calibrate the docking setup, including exhaustiveness values of 1, 8, 16, and 32, random seeds of 0 and 42, and box sizes of 15, 20, 25, 30, and 35 Å (except for WholeP, which was defined as a single large box encompassing the entire ECD).

These configurations collectively established the docking calibration framework, allowing comparison of binding affinities (docking score) and poses among compounds with contrasting biological activities, and ensuring that the selected docking parameters were consistent and reliable enough to reproduce the expected activity trends for subsequent virtual screening analyses.

Following docking, protein-ligand interaction profiling was performed using the Protein-Ligand Interaction Profiler (PLIP)²³ to systematically identify and compare noncovalent interactions across the final docking poses. This analysis was applied to the WholeP and other selected box configurations after the docking parameters had been established, in order to evaluate ligand positioning and contact networks within the AtCARD1 ECD and its homologs. Custom Python scripts were implemented to automate PLIP execution for each receptor-ligand complex. Prior to analysis, Cu atoms were standardized within PDB files to ensure proper recognition of metal coordination by PLIP. The pipeline generated detailed XML and text reports summarizing hydrogen bonds, hydrophobic contacts, π - π stacking, salt bridges, and metal interactions, which were later parsed to quantify contact frequencies and compare binding profiles across ligands and receptors.

Molecular Dynamics simulations

Molecular dynamics (MD) simulations were performed to evaluate the stability of representative docking poses of DMBQ bound to the AtCARD1 ectodomain (ECD). Two ligand-protein complexes corresponding to distinct ligand positions were selected for MD analysis: one derived from docking using the parameters selected during calibration, and a second obtained from the WholeP docking configuration. This comparative analysis was included because the WholeP docking pose overlapped Pocket 2, the second-ranked site predicted by PrankWeb (Supplementary Table S5).

Docking poses were converted from .pdbqt to .pdb format using Open Babel²⁴, and ligand parameters were generated with ACPYPE²⁵, which interfaces Antechamber and the General AMBER Force Field (GAFF)^{26,27}, treating the ligand as neutral. The protein was prepared in GROMACS 2023²⁸ using the AMBER99SB-ILDN^{29,30} force field and the TIP3P³¹ water model. The Cu⁺ ion present in the docking models was removed prior to MD simulations due to the lack of reliable classical force-field parameters for monovalent copper, and no metal-specific interactions were modeled.

Protein-ligand complexes were assembled by structural alignment in PyMOL, solvated in a cubic box with a 1.0 nm (nanometer) buffer, and neutralized with Na⁺ and Cl⁻ ions. Energy minimization was followed by equilibration under NVT conditions (constant number of particles N, volume V, and temperature T) for 200 ps (picoseconds) at 300 K (kelvin) using the V-rescale thermostat³² and under NPT conditions (constant number of particles N, pressure P, and temperature T) for 100 ps using the Berendsen barostat³³. Production MD simulations were conducted under NPT conditions at 300 K and 1 bar using the Parrinello-Rahman barostat³⁴, with a 2 fs (femtoseconds) time step. Initial 1 ns (nanosecond) runs were followed by extended simulations of 10 ns.

Trajectory analyses included backbone and ligand RMSD, hydrogen bond analysis, and intermolecular contact monitoring within a 0.45 nm cutoff. All analyses were performed using standard GROMACS³⁵ tools, and figures were generated with custom Python scripts.

Molecule dataset collection and preparation

The ligand dataset was curated based on previously published studies reporting molecules with experimentally characterized haustorium-inducing or non-inducing activity. Compounds reported to induce haustorium formation were classified as active, whereas non-inducing compounds were classified as inactive, reflecting the goal of identifying potential inhibitors of the process. Importantly, non-inducing compounds were not assumed to be inhibitory, but rather to lack haustorium-inducing activity under the reported conditions. A comprehensive table (Supplementary Table S6) was constructed containing ligand names, SMILES strings, PubChem compound identifiers (CIDs), chemical classes, and literature references. Molecular information, including canonical

SMILES and identifiers, was retrieved from the PubChem database (<https://pubchem.ncbi.nlm.nih.gov/>). In total, 53 molecules were selected, encompassing quinones, lignin-derived phenolics, flavonoids, and other related chemical classes.

Ligand preparation was performed using RDKit³⁶. SMILES strings were standardized by removing salts and counterions, followed by the generation of three-dimensional conformations using the ETKDGv3 algorithm with a fixed random seed to ensure reproducibility. Explicit hydrogen atoms were added prior to 3D embedding. Duplicate molecules were identified and removed to avoid redundancy. Prepared ligands were exported in .sdf format with embedded metadata and subsequently converted to .mol2 format using OpenBabel for downstream docking and molecular dynamics analyses.

An additional “unseen” ligand library was generated to evaluate model generalization beyond compounds with known haustorium-inducing activity. This library was constructed by retrieving structurally related molecules from PubChem based on three-dimensional shape similarity to compounds labeled as active in the primary dataset. For each active reference, similarity searches were performed in PubChem using the 3D similarity tool³⁷ to identify molecules with comparable shape. The retrieved similar compounds were downloaded from PubChem in tabular (.csv) format and subsequently processed to generate three-dimensional conformers. The unseen library was curated to avoid overlap with the original dataset. All compounds overlapping with previously defined active or inactive ligands were removed using PubChem compound identifiers (CIDs) and exact SMILES matching, followed by duplicate elimination to ensure a non-redundant library. Ligand preparation followed the same protocol described for the primary dataset, including salt removal, duplicate elimination, and standardization. After filtering, the final unseen library comprised 50,182 unique molecules. Importantly, compounds included in this library had no prior experimental annotation regarding haustorium induction or inhibition.

Preparation of training and test sets

Given the limited size of the primary ligand dataset, stratified train-test splits were applied to evaluate model robustness. Two split ratios were used (80:20 and 60:40, training:test) (Supplementary Table S7), generated with the *train_test_split* function from scikit-learn³⁸ using stratification on the binary class label (*y_class*) and a fixed random seed to ensure reproducibility. Splitting was performed exclusively on the primary dataset, preserving the class distribution between active and inactive molecules across subsets.

Generation of property-matched decoys

Property-matched decoy molecules were generated using DeepCoy³⁹, a graph-based deep learning framework. Starting from the active compounds in the primary dataset, DeepCoy generated *de novo* decoys through an iterative graph-construction process, optimizing similarity to the corresponding actives across multiple physicochemical descriptors while minimizing trivial structural similarity. Decoy generation was performed using the pre-trained DeepCoy DUDE model⁴⁰ (<https://opig.stats.ox.ac.uk/resources>), producing 50 decoys per active compound (Supplementary Table S6). Decoys were generated exclusively for active molecules present in the test sets of both train-test splits, ensuring that no decoys were included during model training. Active molecules were provided as SMILES files and converted into the internal DeepCoy .json format prior to generation, and the resulting decoys were exported as SMILES and subsequently converted to .csv format for downstream analyses. The generated decoys were subsequently subjected to the same ligand preparation workflow applied to the primary and unseen datasets.

Docking of ligand datasets

All ligand datasets (primary, unseen, and DeepCoy-generated decoys) were docked using Smina against the Cu⁺-containing ECD models of AtCARD1, PjCADL1, and SaCADL1. Docking was performed using the parameters selected during calibration, targeting the PrankWeb-predicted pocket with a 35 × 35 × 35 Å search box, exhaustiveness of 16, a fixed random seed, and one output pose per ligand. Docking outputs were saved in .mol2 and .pdbqt for downstream analysis.

QSAR and PLEC feature generation

Quantitative Structure-Activity Relationship (QSAR) and structure-based Protein-Ligand Extended Connectivity (PLEC) features were computed for both train-test splits (80:20 and 60:40) and used in three evaluation scenarios: (i) standard test set (actives + inactives), (ii) test positives + DeepCoy decoys, and (iii) test positives + unseen library (treated as unlabeled/background for screening). Feature generation and dataset assembly were performed using custom Python pipelines that (i) computed molecular descriptors, (ii) standardized features using training-set statistics, and (iii) assembled training, test, and screening datasets for downstream analysis.

For QSAR, molecules were encoded as 512-bit Morgan fingerprints⁴¹ (RDKit; radius = 2) computed from SMILES strings. Failed SMILES parsing was tracked, and descriptors were still exported for traceability. Fingerprints were then standardized using StandardScaler (scikit-learn library), fitted on the training set and applied to all evaluation sets to avoid information leakage. For each split, the following descriptor tables were exported: *_train_descriptors.csv and

*_test_descriptors.csv. When decoys or the unseen library were provided, an additional standardized dataset and descriptor table were generated for these molecules (e.g., *_decoys_dataset.csv and *_decoys_descriptors.csv). In this case, when a decoy/unseen CSV was provided, the pipeline assembled the screening evaluation set as (test actives only) + (background set), assigning $y = 0$ to all background molecules. This ensured enrichment-style metrics reflected recovery of the held-out positives in a large inactive background.

For PLEC, interaction fingerprints were computed using Open Drug Discovery Toolkit (ODDT) PLEC⁴² from .mol2 structures. The receptor was loaded from the receptor .mol2 file, and ligands were mapped by .mol2 title matching the ligand ID. For each docked ligand-receptor complex, PLEC interaction fingerprints were computed and represented as numerical feature vectors. Feature scaling was performed using StandardScaler fitted on the training set and subsequently applied to the test and decoy datasets to prevent information leakage. For each split and each receptor (AtCARD1, PjCADL1, SaCADL1), features were generated independently using the corresponding receptor .mol2 and the ligand .mol2 produced by docking. When --decoys_csv and --decoys_mol2 were provided, decoys were assigned $y = 0$, featurized against the same receptor, and evaluated in a combined set consisting of test positives + decoys, analogous to the QSAR decoy evaluation.

Machine Learning model evaluation and performance metrics

All machine learning analyses were performed in classification mode, with the objective of distinguishing active molecules from inactive or background compounds. Regression models were not used due to the absence of continuous quantitative activity data and the limited size of the curated dataset. The following classification algorithms were evaluated for both QSAR- and PLEC-based feature sets: Random Forest (RF)⁴³, Support Vector Machine (SVM)⁴⁴, Artificial Neural Network (ANN; multilayer perceptron)⁴⁵, Deep Neural Network (DNN)^{46,47} and Extreme Gradient Boosting (XGBoost)⁴⁸. For each model, training was performed exclusively on the training set, while all performance metrics were computed on the corresponding held-out test set. Predicted class probabilities obtained for test compounds were used to rank molecules for enrichment-based evaluation. All models were trained and evaluated independently for each data split, receptor (when applicable), and feature representation (QSAR or PLEC), using fixed random seeds to ensure reproducibility.

Evaluation strategy and top-k-based metrics

Model performance was assessed using a ranking-based evaluation framework designed for low-data and imbalanced virtual screening scenarios. Evaluations were conducted under two conditions: (i) standard test sets containing active and inactive molecules, and (ii) test positives combined with decoys. Given the small number of active compounds, performance was primarily evaluated using top-k-based metrics, which better reflect early recognition ability than global accuracy metrics. Compounds were ranked according to predicted class probabilities, and performance was evaluated at multiple ranking thresholds (top-k = 3, 4, 5, 10, 15, 20, 25, and 30). The following metrics were computed at each top-k level: Enrichment Factor (EF), Normalized Enrichment Factor (NEF), and Hit Rate. In addition, Precision-Recall Area Under the Curve (PR-AUC) was calculated as a complementary, threshold-independent metric. Precision-recall and recall curves were also generated to visualize model performance across ranking thresholds. For comparison, the same ranking-based metrics were computed using Smina docking scores, allowing direct assessment of machine-learning models relative to docking-based virtual screening. For each model, split, and evaluation scenario, detailed output files were generated, including predicted probabilities, ranked ligand lists, metric tables, and graphical summaries.

Results

Pocket prediction identifies a high-probability binding site at the histidine triad

Due to the high conservation of residues among AtCARD1 and its homologs, particularly histidines (H197, H199, and H222 in AtCARD1 and SaCADL1; H195, H197, and H220 in PjCADL1) and cysteines (C395 and C405 in AtCARD1; C393 and C402 in PjCADL1; and C397 and C406 in SaCADL1) (Fig. 1A; Supplementary Fig. S1), it was hypothesized^{2,7} that these conserved regions may delineate the potential ligand-binding site for quinone-type HIFs. To test this hypothesis, we predicted the ligand-binding pockets of AtCARD1, PjCADL1 and SaCADL1 extracellular domain using PrankWeb, a machine learning-based pocket predictor.

For all proteins analyzed, PrankWeb identified at least one predicted pocket overlapping with two histidines of the triad, whereas no pocket was positioned directly over the cysteine pair (Supplementary Table S5). In all three receptors, the top-ranked (rank 1) pockets were consistently located near the His-triad region and exhibited comparable probability and score values across species. Both Cu⁺-bound and Cu⁺-free models of the receptors were evaluated, given the central role of Cu⁺ coordination by a conserved histidine triad in AtCARD1⁷. Notably, the predicted pocket in AtCARD1 + Cu⁺ model coincided spatially with the one identified in the Cu⁺-free structure (Supplementary Table S5), indicating that Cu⁺ coordination does not alter the overall pocket geometry or position. The prediction scores improved when the evolutionary conservation filter was disabled, suggesting that the local structural geometry, rather than residue

conservation, primarily drives pocket identification (Fig. 2B). Despite minor variation in the number and composition of residues forming the predicted pockets in SaCADL1 and PjCADL1, their biochemical properties remained consistent with AtCARD1 (Fig. 2C).

To evaluate the contribution of the SEA-like domain (Fig. 1A), pocket prediction was also performed using a truncated AtCARD1 model lacking the SEA-like region (AtCARD1 Δ SEA). In this case, no pocket was detected when conservation was applied, and only one low-confidence pocket (still involving H199) appeared without conservation (Supplementary Table S5). The SEA-like domain of AtCARD1, previously described by Ishihama et al. (2025), forms a compact, stabilizing interface with the terminal LRR region, mediated by N-glycan (Asn298) and hydrogen-bond interactions. This interface occupies part of the external surface where, in other LRR-RLKs, ligands typically bind. Consequently, removing the SEA-like domain likely disrupts the ECD geometry and surface accessibility, eliminating structural cavities that could otherwise be recognized as pockets. This observation supports the notion that the SEA-like domain contributes to shaping or stabilizing the binding site, potentially acting as a regulatory cap or conformational modulator, similar to SEA-domain-mediated mechanisms in animal receptors.

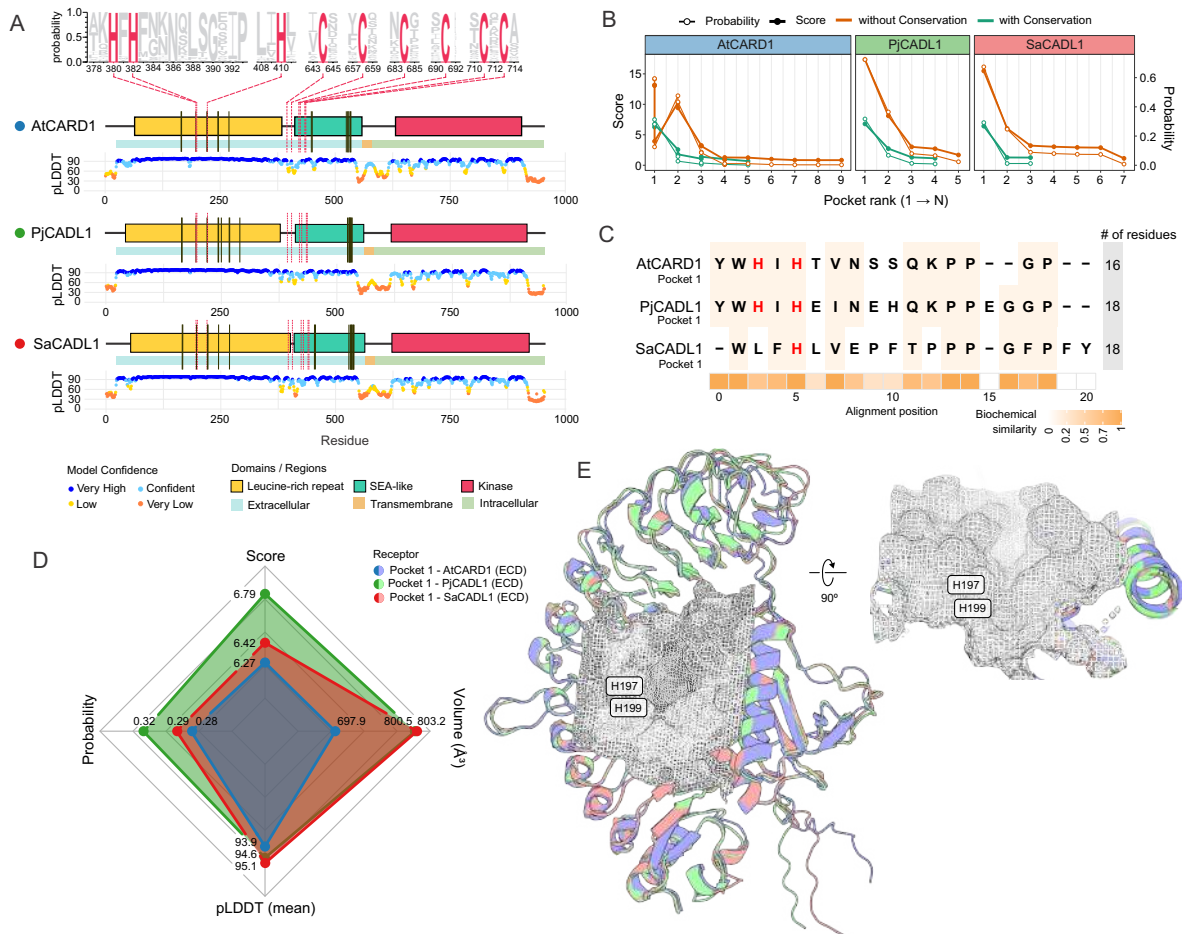


Figure 2. Predicted ligand-binding pockets in AtCARD1 and its homologs. (A) Domain architecture of AtCARD1, PjCADL1, and SaCADL1 showing conserved structural features, including extracellular, transmembrane, and

intracellular regions. Conserved histidines and cysteines are highlighted (red dashed lines; corresponding logo graphs shown above). Black lines within the domain bars indicate the top-ranked (rank 1) predicted pocket for each protein. pLDDT values along the protein length are shown below each bar. (B) Ranking of predicted pockets for each protein obtained from PrankWeb analysis. (C) Biochemical similarity of residues forming the top-ranked pocket (rank 1) and total number of residues involved for each protein. (D) Comparative radar plot showing score, probability, pLDDT, and volume values and (E) Surface representation after structural alignment for the rank 1 pocket across the ectodomains (ECDs) of AtCARD1 and its homologs.

When considering only the top-ranked pockets (rank 1), all predicted sites corresponded to regions with pLDDT > 93.9 and PrankWeb score > 6.2, with AtCARD1 showing slightly lower values compared to its homologs (Fig. 2D). The pocket volumes were consistently larger in PjCADL1 and SaCADL1 than in AtCARD1, possibly reflecting structural adaptations or increased surface accessibility in the parasitic homologs. In all cases, 3D alignment of ECD and surface visualization of the predicted pockets revealed a distinct cavity surrounding the histidine triad, spatially consistent with the proposed Cu⁺-binding site⁷ (Fig. 2E).

Docking calibration and identification of alternative binding regions

Following pocket prediction, we focused docking simulations primarily on the top-ranked predicted pocket but did not exclude the possibility of alternative ligand-binding regions. To comprehensively evaluate potential interaction sites, four docking boxes were defined: C–C (centered between Cys395–Cys405), triHIS (centroid of His197–His199–His222), Predicted (coordinates provided by PrankWeb), and WholeP (covering the entire ECD) (Supplementary Fig. S4). Seven ligands with documented activity in the literature related to AtCARD1 were selected, including compounds reported as haustorium inducers (positives) and non-inducers (negatives), and were used for docking calibration (Supplementary Table S4).

To identify the optimal docking configuration, we systematically tested multiple exhaustiveness values (1, 8, 16, and 32), random seeds (0 and 42), and box sizes (15–35 Å). The docking scores (kcal/mol) values plateaued around an exhaustiveness of 16, indicating convergence of sampling without significant gain from higher values (Fig. 3A). Thus, an exhaustiveness value of 16 was selected as the optimal setting, as it provided reproducible docking results while significantly reducing computational cost. Higher exhaustiveness values did not improve binding affinity convergence and may even increase the occurrence of redundant or low-probability poses due to over-sampling. Likewise, varying the random seed (0 or 42) produced negligible changes in docking scores, confirming the stochastic stability of the selected parameters (Fig. 3A). Independent of box size, docking scores appeared to stabilize, with the exception of some ligands (e.g., benzoquinone) and for docking boxes centered on the cysteine pair (C–C

pocket), which exhibited improved scores with larger boxes (35 Å) (Fig. 3A). Based on these results, we selected seed = 0, exhaustiveness = 16, and box size = 35 Å as the final parameters for all subsequent docking and virtual screening analyses.

Comparative docking across configurations revealed that, in AtCARD1, the presence of Cu⁺ consistently improved predicted docking scores within the Predicted and triHIS pockets, largely independent of ligand identity (Fig. 3B). A similar trend was also observed in PjCADL1 and SaCADL1 (Supplementary Fig. S6A). When analyzed on a per-ligand basis, this effect persisted with the exception of the positive ligand TBHQ, reinforcing a stabilizing role of Cu⁺ in the His-triad region (Fig. 3C; Supplementary Fig. S6B). Interestingly, Cu⁺ increased docking scores in the C-C pocket for the majority of ligands in parasitic CADL homologs, whereas this effect was not observed in AtCARD1 (Fig. 3C; Supplementary Fig. S6B). In addition, PjCADL1 exhibited a higher median docking score in the C-C pocket compared to the other evaluated pockets, independent of Cu⁺ presence (Fig. 3D; Supplementary Fig. S6A). This difference may reflect subtle structural differences in the local environment surrounding the His-triad and the adjacent C-C surface region. In AtCARD1, Cu⁺ coordination may influence the accessibility of the nearby C-C pocket, whereas in PjCADL1 and SaCADL1 this region appears to remain more permissive to ligand binding, consistent with the larger predicted pocket volumes observed in parasitic CADLs compared to AtCARD1 (Fig. 2D). Among all ligands, TFBQ consistently exhibited stronger affinity than DMBQ, despite being a functional antagonist² (Fig. 3C; Supplementary Fig. S6B). This suggests that inhibitory quinones might occupy the same pocket with higher stability, reflecting a potential competitive antagonism mechanism, where stronger but non-productive binding could prevent receptor activation.

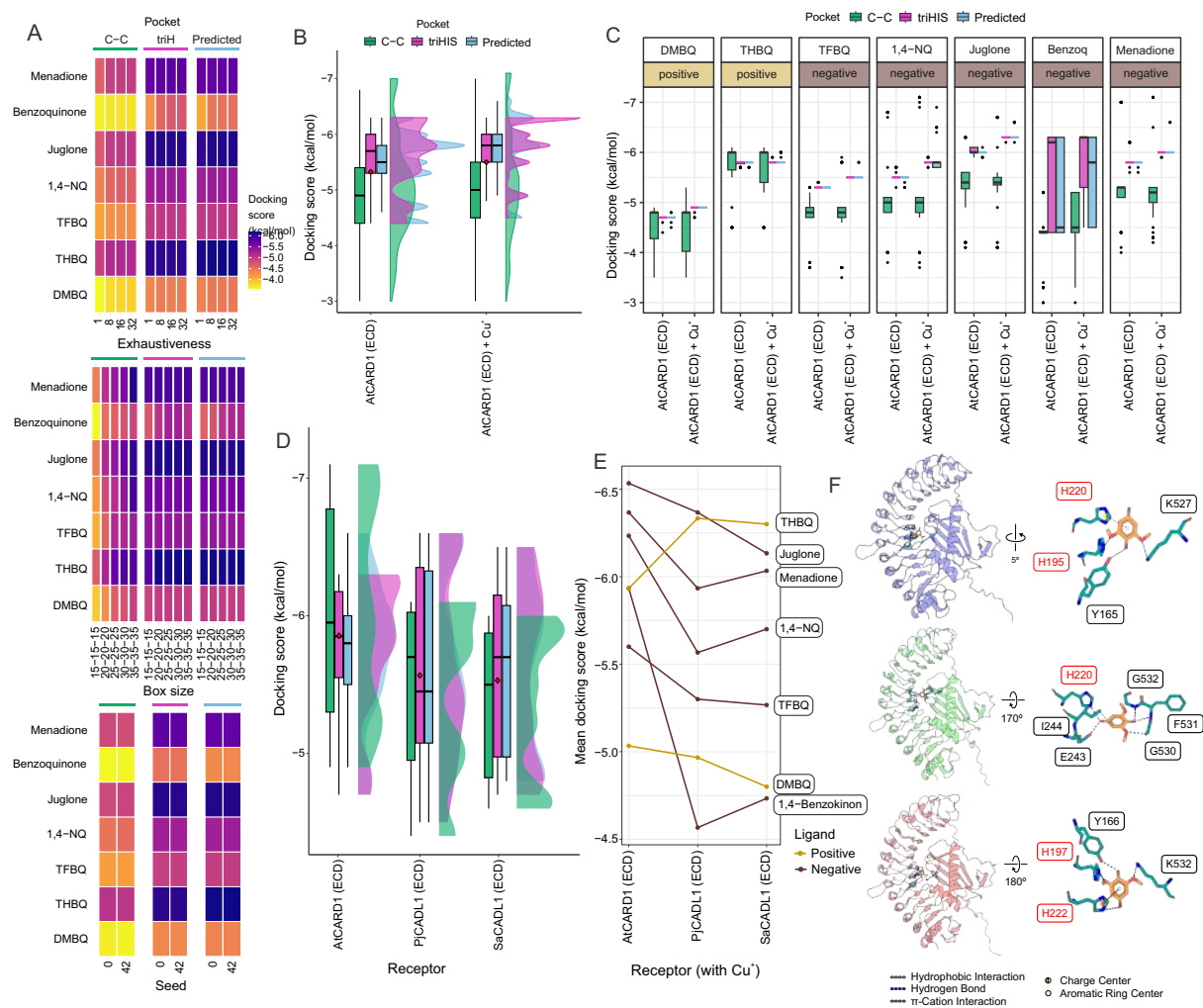


Figure 3. Calibration and comparative analysis of docking performance across AtCARD1 and homologs. (A) Summary of parameter optimization for docking calibration showing the effect of exhaustiveness, box size, and seed values on docking scores (kcal/mol) across all ligands and pockets. (B) Distribution of docking scores for AtCARD1 (ECD) with and without Cu^+ , across the three pocket configurations (C-C, triHIS, and Predicted). (C) Ligand-specific affinity distributions grouped by functional classification (positive and negative). (D) Cross-species comparison of docking scores (AtCARD1, PjCADL1, and SaCADL1; with and without Cu^+ combined) and the three pocket configurations. (E) Mean docking scores across pockets for each ligand type in the Cu^+ -bound configuration, classified according to their reported biological activity. (F) PLIP interaction profiles for DMBQ bound to AtCARD1 (light-blue), PjCADL1 (light-green), and SaCADL1 (salmon), highlighting the main interacting residues and interaction types.

As expected, the Predicted and triHIS pockets exhibited highly similar median docking affinity values across ligands in AtCARD1 and its homologs, indicating that both regions likely correspond to the same site (Fig. 3B-D; Supplementary Fig. S6A-B). Although negative ligands displayed slightly stronger (more negative) median docking scores, particularly in AtCARD1, no clear selectivity pattern between positive and negative ligands emerged (Fig. 3E; Supplementary Fig. S6C).

After parameter calibration and confirmation of the predicted pocket as the most plausible binding region, ligand-receptor interactions were analyzed using PLIP. In AtCARD1 and its homologs, DMBQ was consistently positioned in proximity to two histidines of the His-triad (Fig. 3F). In PjCADL1, DMBQ interactions were primarily mediated by hydrogen bonds and one hydrophobic contact. In contrast, AtCARD1 and SaCADL1 displayed a more similar interaction pattern, not only for DMBQ but also across other quinone ligands (see Supplementary Tables S8). Notably, PjCADL1 appeared to engage a larger number of surrounding residues to stabilize DMBQ binding, whereas AtCARD1 and SaCADL1 additionally formed π -cation interactions, suggesting differences in the local electrostatic environment and aromatic residue positioning that may contribute to ligand stabilization (Fig. 3F). When examining all ligands collectively, most docked near the His-triad region across homologs (Supplementary Fig. S7). However, in AtCARD1, two ligands were localized to an alternative region, hereafter referred to as Region Z.

The WholeP docking configuration consistently yielded improved docking scores for ligands that were previously positioned near the His-triad, revealing this “new” alternative binding region (Region Z). In AtCARD1, DMBQ preferentially bound to this region (Supplementary Fig. S8; Supplementary Table S9). In PjCADL1, Cu^+ presence redirected DMBQ toward Region Z, whereas in the absence of Cu^+ the ligand instead interacted with His220 within the His-triad (Supplementary Fig. S9; Supplementary Table S9). In SaCADL1, Cu^+ also enhanced DMBQ docking scores but shifted its binding site toward the N-terminal LRR region (Supplementary Table S9). Across all ligands, DMBQ, juglone, and menadione clustered in Region Z in PjCADL1, while in AtCARD1 all ligands localized to Region Z in Cu^+ -bound models (Supplementary Fig. S10). In contrast, in SaCADL1 only DMBQ and TFBQ showed a Cu^+ -dependent displacement from the His-triad toward alternative regions, whereas the remaining ligands predominantly remained near the Predicted site (Supplementary Fig. S10; Supplementary Table S9). To further investigate differences in DMBQ binding between these regions, molecular dynamics analyses were performed. Although ligand stability in Region Z became more evident after 6-10 ns, hydrogen bonds were rarely maintained during this interval. In contrast, the number of heavy-atom contacts (<0.45 nm) was slightly more stable for DMBQ bound in the predicted pocket, indicating distinct stabilization modes between the two regions (Supplementary Fig. S11).

Taken together, these results indicate that Cu^+ coordination modulates quinone-binding geometry, stabilizing interactions near the His-triad in parasitic CADLs while favoring an alternative surface-exposed pocket in AtCARD1. Although Region Z was not described as a ligand-binding site by Ishihama et al. (2025), its topological correspondence to the SEA-LRR interface, enriched in hydrophobic and aromatic residues such as phenylalanine, histidine, and glutamine, supports its role as a plausible secondary pocket stabilized by the SEA-like domain. This dual-site behavior is consistent with structural divergence between AtCARD1 and

its homologs and may have implications for ligand recognition and Cu⁺-dependent modulation.

Therefore, the PrankWeb-predicted pocket, using a 35 Å docking box and Cu⁺-bound protein models, was selected for subsequent virtual screening analyses. This configuration encompasses both the functional His-triad site and the putative secondary Region Z, maximizing the detection of active ligands and potential modulators. Notably, Region Z was already detected in AtCARD1 prior to WholeP docking (Supplementary Fig. S7) and corresponds to the rank-2 pocket predicted by PrankWeb (Supplementary Table S5), further supporting its relevance as an alternative binding surface.

Machine learning performance is largely independent of feature representation and receptor species

A total of 53 molecules previously described in the literature as haustorium inducers or non-inducers/inhibitors were collected from published studies (Supplementary Table S6). Compounds were grouped according to their chemical class, including quinones, lignin-derived phenolics, flavonoids, and others. For the purposes of this study, non-inducing or inhibitory compounds were labeled as active, reflecting our focus on identifying potential haustorium inhibitors, whereas haustorium-inducing compounds were labeled as inactive (Fig. 4A).

Docking results obtained with Smina indicated that inactive ligands generally displayed more favorable (more negative) docking scores than active ligands, largely independent of receptor or chemical class (Fig. 4A). This trend likely reflects the fact that compounds labeled as active in this study are defined by the absence of haustorium induction, rather than by experimentally validated inhibitory activity. Despite this apparent separation at the score distribution level, docking-based ranking showed limited early enrichment of active compounds, as reflected by Enrichment Factor (EF) values close to zero at low-ranking thresholds. Specifically, up to top-k = 7, no active ligands were recovered for AtCARD1 and PjCADL1, with SaCADL1 being the only receptor for which at least one active compound was identified at very early ranks (Supplementary Fig. S12A). From top-k = 10 onward, active ligands began to appear more consistently across receptors, although enrichment values remained far below the theoretical EF_{max}.

Precision-recall analysis further characterized the performance of docking-based ranking, yielding Precision-Recall Area Under the Curve (PR-AUC) values ranging from ~0.31 to 0.33 across receptors (Supplementary Fig. S12B). Given the strong class imbalance of the dataset and the focus on early recovery of active compounds, PR-AUC provides a more informative assessment than score distributions alone, as it directly reflects the balance between precision and recall for the active compounds. In agreement with the modest enrichment observed in EF and Normalized Enrichment Factor (NEF) analyses at low top-k values, the

PR-AUC results indicate limited discrimination between active and inactive ligands when ranking is based solely on Smina docking scores.

Given the limited early enrichment observed for docking-based ranking, subsequent analyses focused on evaluating whether machine learning (ML) models could improve the prioritization of active compounds. QSAR models based on Morgan fingerprints were first assessed using an 80:20 train-test split of the primary dataset, chosen to maximize the number of active compounds in the test set. Under this configuration, SVM and DNN showed superior early-recognition performance, with EF and NEF values at top-k = 5 reaching up to 1.65 (EFmax = 2.2) (Supplementary Fig. 13A; Supplementary Table S11). This trend was supported by PR-AUC values of 0.543 (DNN) and 0.542 (SVM) (Supplementary Fig. 13B; Supplementary Table S11), indicating improved separation between active and inactive compounds compared to docking-based ranking. When decoys were incorporated into the evaluation, all models exhibited a marked reduction in performance, reflected by substantially lower PR-AUC values (0.262 for ANN/SVM down to 0.017 for XGBoost; Supplementary Fig. 14; Supplementary Table S11). This decline is consistent with the extreme class imbalance imposed by decoys and highlights the increased difficulty of recovering rare active compounds in a large inactive background. To further assess robustness, a 60:40 split was evaluated. Although enrichment metrics (EF and Hit Rate) did not improve substantially under this configuration, PR-AUC increased for several models, most notably DNN (0.770) and Random Forest (0.683), while SVM showed a relative decrease (0.626) (Supplementary Fig. 15; Supplementary Table S11). Despite these differences, the inclusion of decoys again revealed ANN and SVM as the most reliable models for ranking, albeit with increased variability (Fig. 4B-C). Overall, QSAR-Morgan models showed limited sensitivity to the choice of split strategy and maintained stable ranking performance in the presence of decoys, supporting their use for initial screening.

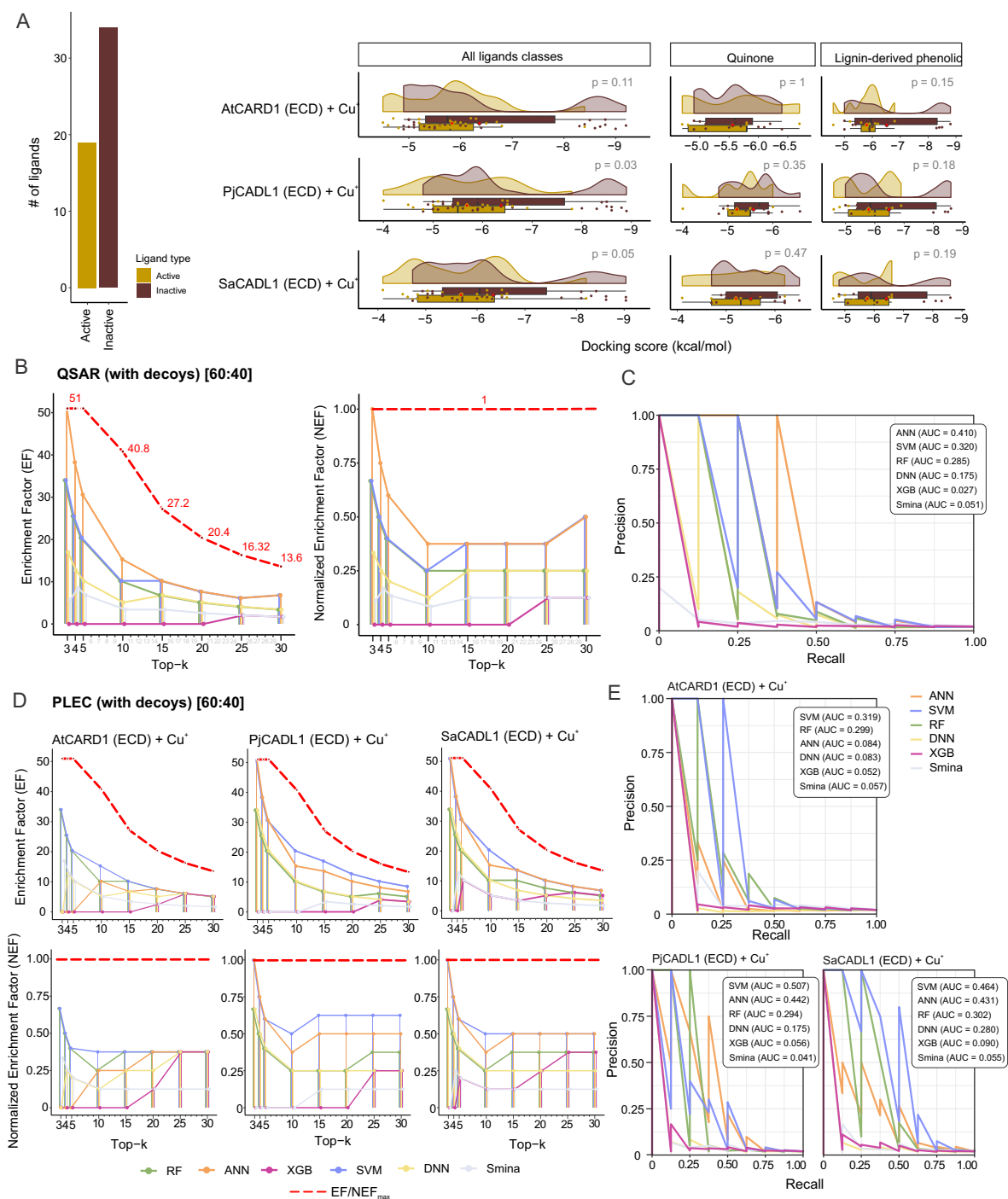


Figure 4. Docking- and machine learning-based prioritization of putative haustorium inhibitors across AtCARD1 and parasitic homologs. (A)

Composition of the curated ligand dataset used in this study, showing the number of compounds classified as active (non-haustorium inducer / putative inhibitors) and inactive (haustorium inducers) (left panel). Boxplots (right panel) show the distribution of Smina docking scores (kcal/mol) for active and inactive ligands docked against the Cu⁺-bound ectodomains of AtCARD1, PjCADL1, and SaCADL1, stratified by the complete ligand set and by individual chemical classes. *p*-values correspond to two-sided statistical comparisons between ligand classes. (B) Enrichment Factor (EF) and Normalized Enrichment Factor (NEF) as

a function of top-k, and (C) Precision-recall curves showing model-specific PR-AUC values for QSAR models based on Morgan fingerprints evaluated with decoys using the 60:40 split. (D) EF and NEF as a function of top-k, and (E) Precision-recall curves showing model-specific PR-AUC values for PLEC-based machine learning models across AtCARD1, PjCADL1, and SaCADL1, evaluated with decoys using the 60:40 split. Dashed lines indicate the theoretical maximum enrichment (EFmax and NEFmax). Across panels, results are shown for multiple classification models (ANN, SVM, Random Forest, DNN, and XGBoost), with Smina docking-based ranking included for comparison.

Docking poses were subsequently used to derive PLEC interaction features, enabling evaluation of structure-informed ML models. In the 80:20 split, SVM achieved the strongest performance for AtCARD1, while DNN performed better for the parasitic homologs, with NEF values stabilizing performance at top-k = 5 for PjCADL1 and SaCADL1, despite large variability in EF (Supplementary Figs. 16-17; Supplementary Table S12). Across all three receptors, SVM consistently yielded PR-AUC values above 0.64, indicating improved generalization across related targets. This pattern persisted after the inclusion of decoys, with SVM remaining the best-performing model for ranking (Supplementary Figs. 17-18; Supplementary Table S12). Notably, PLEC-based models tended to show strong enrichment when successful, with actives concentrated at the very top of the ranked lists. In the 60:40 split, XGBoost exhibited unstable behavior, performing well in the absence of decoys but collapsing when decoys were included. In contrast, SVM maintained consistent ranking performance, reinforcing its robustness under noisy and highly imbalanced conditions (Supplementary Figs. 17, 19; Supplementary Table S12; Fig. 4D-E).

Taken together, because the ligands span a wide range of chemical classes, structural scaffolds, and binding modes, likely including both agonists and antagonists, these results support a two-stage screening strategy in which QSAR models based on Morgan fingerprints (ANN/SVM) are applied for initial large-scale ranking, followed by PLEC-based models, particularly SVM, for refinement of top-ranked candidates. Importantly, model performance did not require receptor-specific tuning for AtCARD1, PjCADL1, or SaCADL1, as comparable enrichment and ranking behavior was observed across all three receptors, even under increased noise in some evaluation scenarios. This consistency indicates that the selected models generalize well across homologous targets and remain robust despite the presence of decoys and the limited size of the training dataset. Based on these observations, Morgan fingerprint-based ranking was used to screen an expanded compound library constructed from three-dimensional similarity to the active compounds considered in this study, comprising approximately 50,182 molecules. The resulting ranking is shown in Supplementary Table S13. Docking was subsequently performed for this library to enable PLEC feature generation against AtCARD1 and its parasitic homologs, followed by ranking using the selected PLEC-based model (**in progress**).

Supplementary Information

Supplementary Figure S1. Alignment of AtCARD1 homologs across Archaeplastida.

Supplementary Figure S2. Summary of structural and sequence-based homology searches for AtCARD1.

Supplementary Figure S3. Structural alignment of AtCARD1 with selected ectodomain homologs.

Supplementary Figure S4. Positioning of the Cu^+ ion within the predicted histidine triad of AtCARD1.

Supplementary Figure S5. Docking box configurations tested for molecular docking calibration.

Supplementary Figure S6. Comparison of docking scores across receptors, pockets, and ligand types.

Supplementary Figure S7. Predicted binding poses of all ligands derivatives across AtCARD1 and homologs.

Supplementary Figure S8. AtCARD1 + Cu^+ (WholeP docking configuration) highlighting the secondary ligand-binding region (Region Z).

Supplementary Figure S9. Predicted interactions of DMBQ with PjCADL1 and SaCADL1 ectodomains, with and without Cu^+ (WholeP docking configuration).

Supplementary Figure S10. Predicted docking poses of quinone derivatives across AtCARD1 and homologs in the presence of Cu^+ (WholeP docking configuration).

Supplementary Figure S11. Molecular dynamics analysis of DMBQ bound to AtCARD1 in two distinct binding configurations.

Supplementary Figure S12. Performance of Smina docking-based ranking across AtCARD1 and parasitic homologs in the presence of Cu^+ .

Supplementary Figure S13. Performance of QSAR-based machine-learning models using Morgan fingerprints in the absence of decoys (80:20 split).

Supplementary Figure S14. Performance of QSAR-based machine-learning models using Morgan fingerprints in the presence of decoys (80:20 split).

Supplementary Figure S15. Performance of QSAR-based machine-learning models using Morgan fingerprints in the absence of decoys (60:40 split).

Supplementary Figure S16. Performance of PLEC-based machine-learning models without decoys (80:20 split).

Supplementary Figure S17. Precision-recall performance of PLEC-based machine-learning models across receptors and dataset configurations.

Supplementary Figure S18. Performance of PLEC-based machine-learning models with decoys (80:20 split).

Supplementary Figure S19. Performance of PLEC-based machine-learning models without decoys (60:40 split).

Supplementary Tables S1-13

Supplementary Table S1. List of Archaeplastida species used for CARD1 homolog search.

Supplementary Table S2. List of selected protein structures from BLASTp and FoldSeek analyses with their identity, resolution, and predominantly covered region.

Supplementary Table S3. Results of PrankWeb pocket predictions for the FLS2 control.

Supplementary Table S4. List of quinone derivatives used for molecular docking calibration and their reported biological classification.

Supplementary Table S5. Predicted ligand-binding pockets in CARD1 and its homologs obtained from PrankWeb analysis.

Supplementary Table S6. Literature-curated ligands reported as inducers, inhibitors, or non-inducers of haustorium formation*.

Supplementary Table S7. Dataset composition, train–test splits, and decoy setup for virtual screening evaluation.

Supplementary Table S8. PLIP-derived protein–ligand interaction analysis of docking complexes.

Supplementary Table S9. PLIP-derived protein–ligand interaction analysis of full-ectodomain docking complexes (WholeP box configuration).

Supplementary Table S10. Virtual screening performance metrics based on Smina docking scores.

Supplementary Table S11. Virtual screening performance metrics of QSAR models based on 512-bit Morgan fingerprints.

Supplementary Table S12. Virtual screening performance metrics using PLEC descriptors.

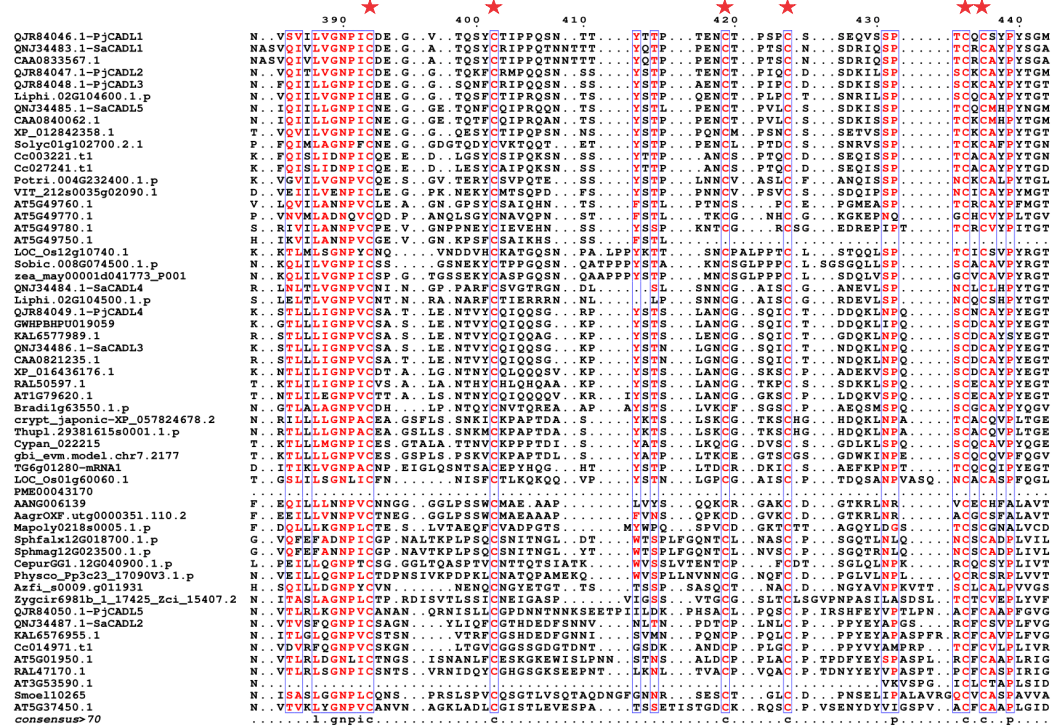
Supplementary Table S13. Top-ranked compounds predicted by the SVM QSAR model using Morgan fingerprints.

[All tables can be accessed by clicking here.](#)

Supplementary Figures S1-19

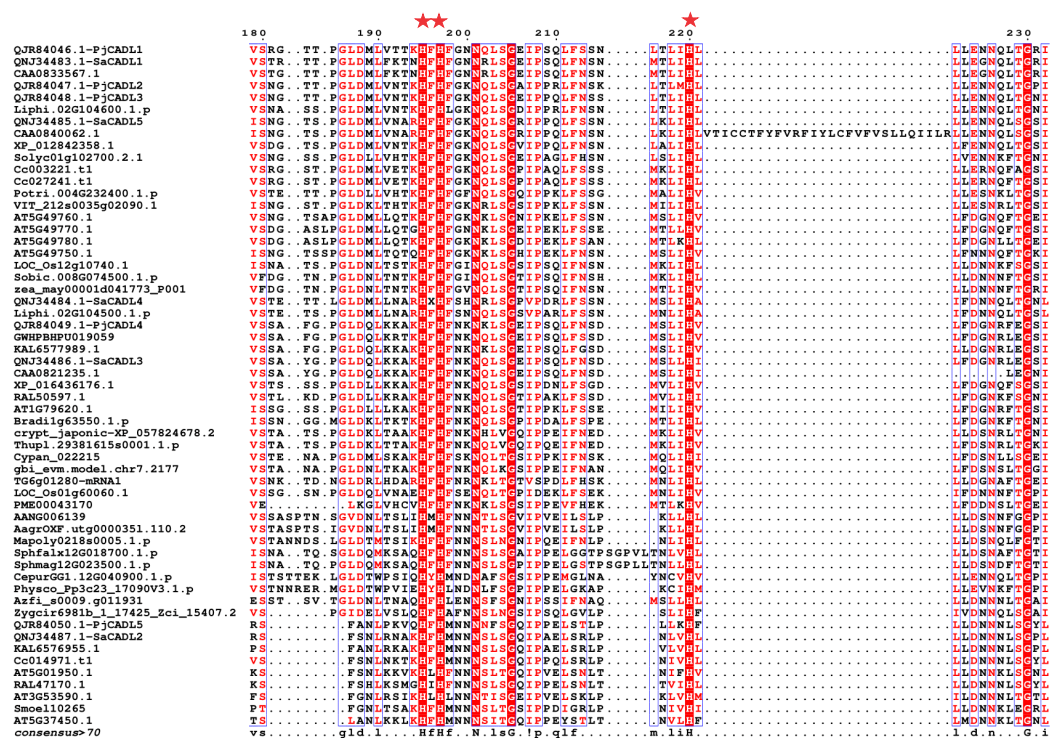
A

Cysteines region

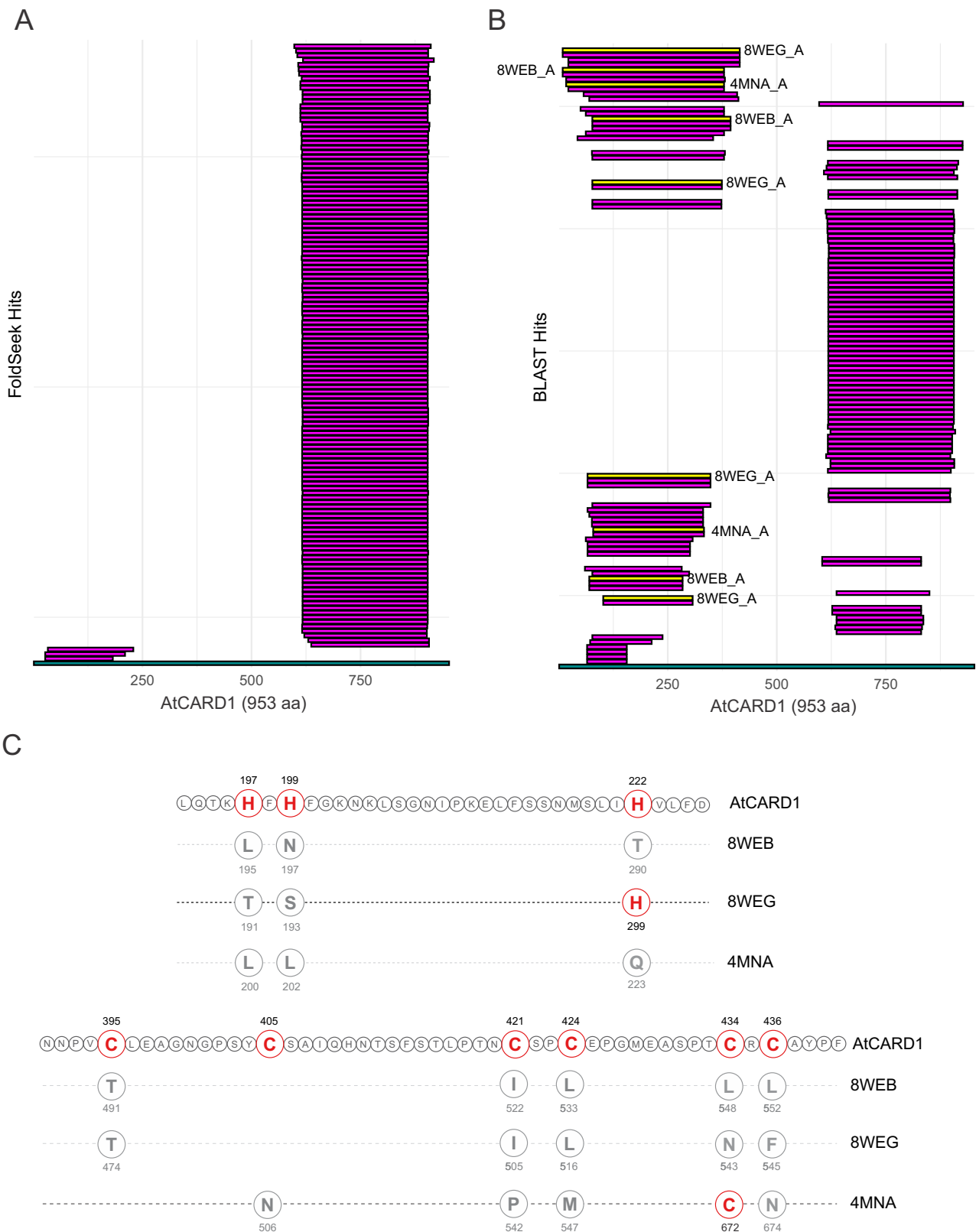


B

Histidine triad region

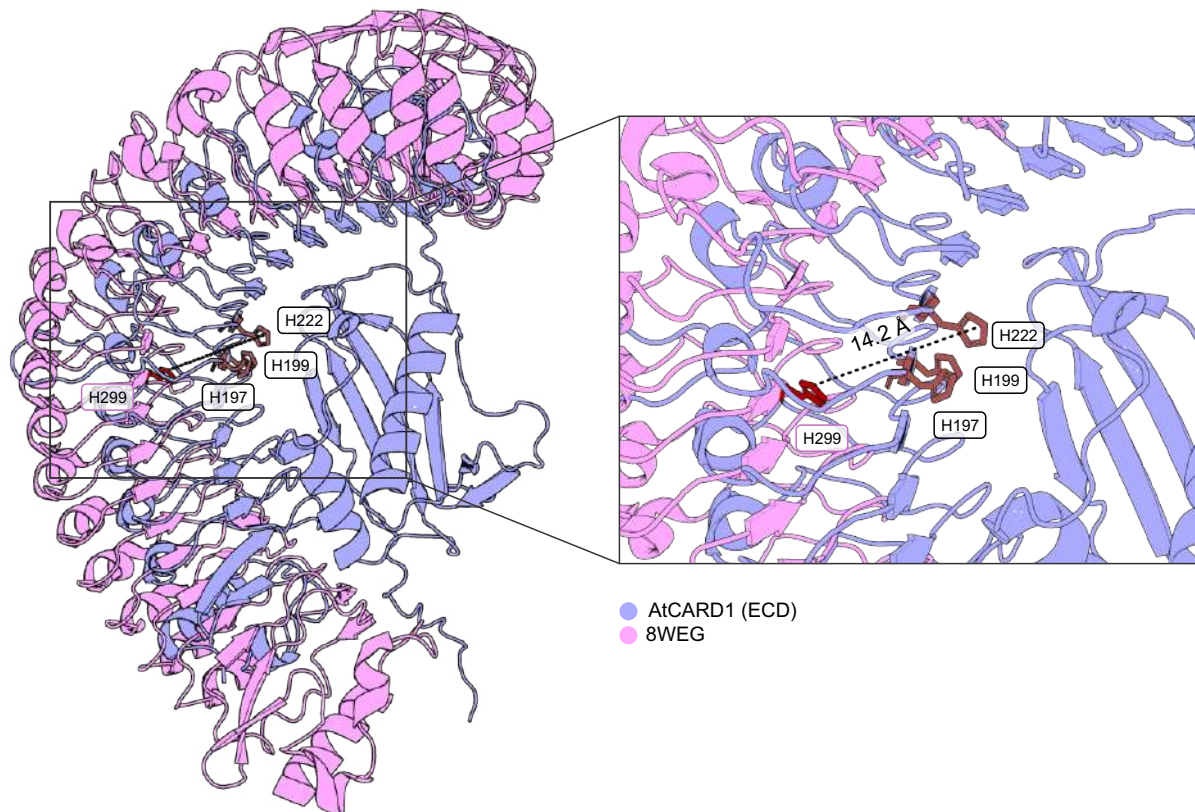


Supplementary Figure S1. Alignment of AtCARD1 homologs across Archaeplastida. Conserved residues are highlighted to illustrate sequence conservation within (A) the cysteine-rich region and (B) the histidine triad. Red stars indicate the positions of conserved cysteine and histidine residues, respectively.

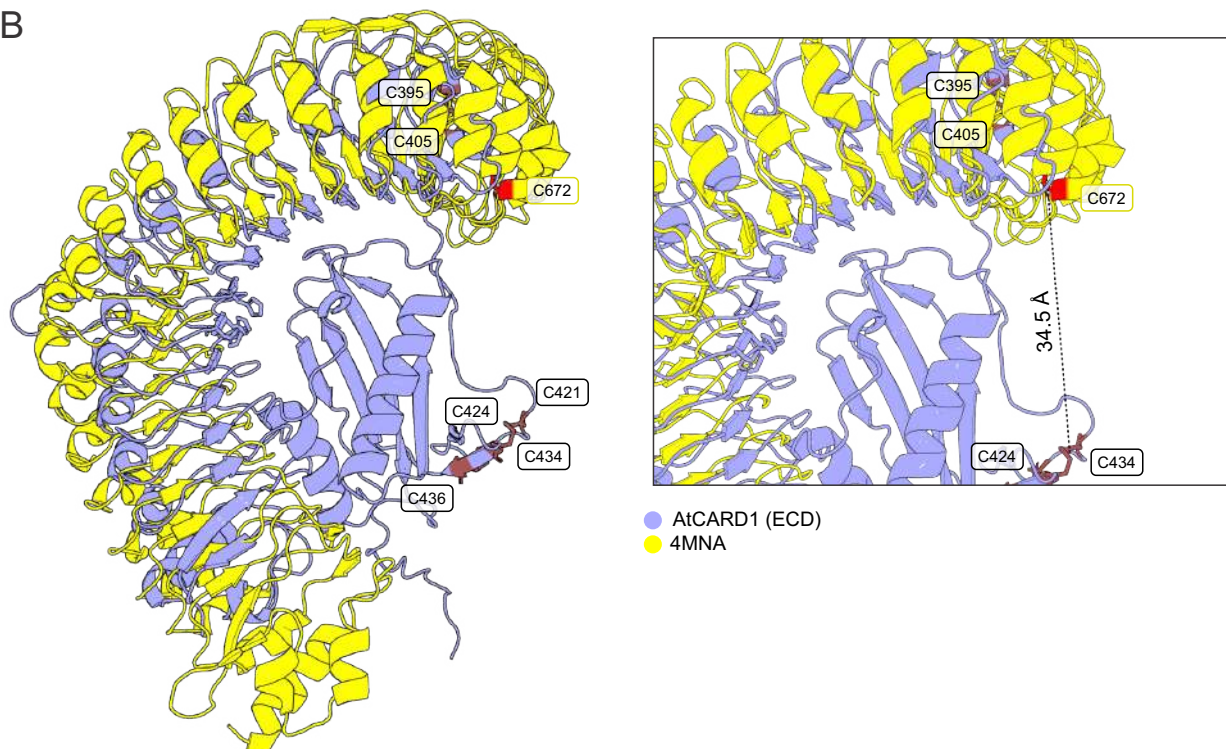


Supplementary Figure S2. Summary of structural and sequence-based homology searches for AtCARD1. (A) FoldSeek and (B) BLASTp (sequence identity >30%) search results against the RCSB PDB. The green bar represents the AtCARD1 query; pink bars correspond to identified structural hits; yellow bars indicate proteins selected for further analysis (Supplementary Table S2). The X-axis represents the full-length AtCARD1 sequence (953 amino acids), showing the relative positions of aligned regions. (C) Multiple sequence alignment of AtCARD1 with selected homologs, highlighting conserved histidine (top) and cysteine (bottom) residues potentially involved in DMBQ recognition.

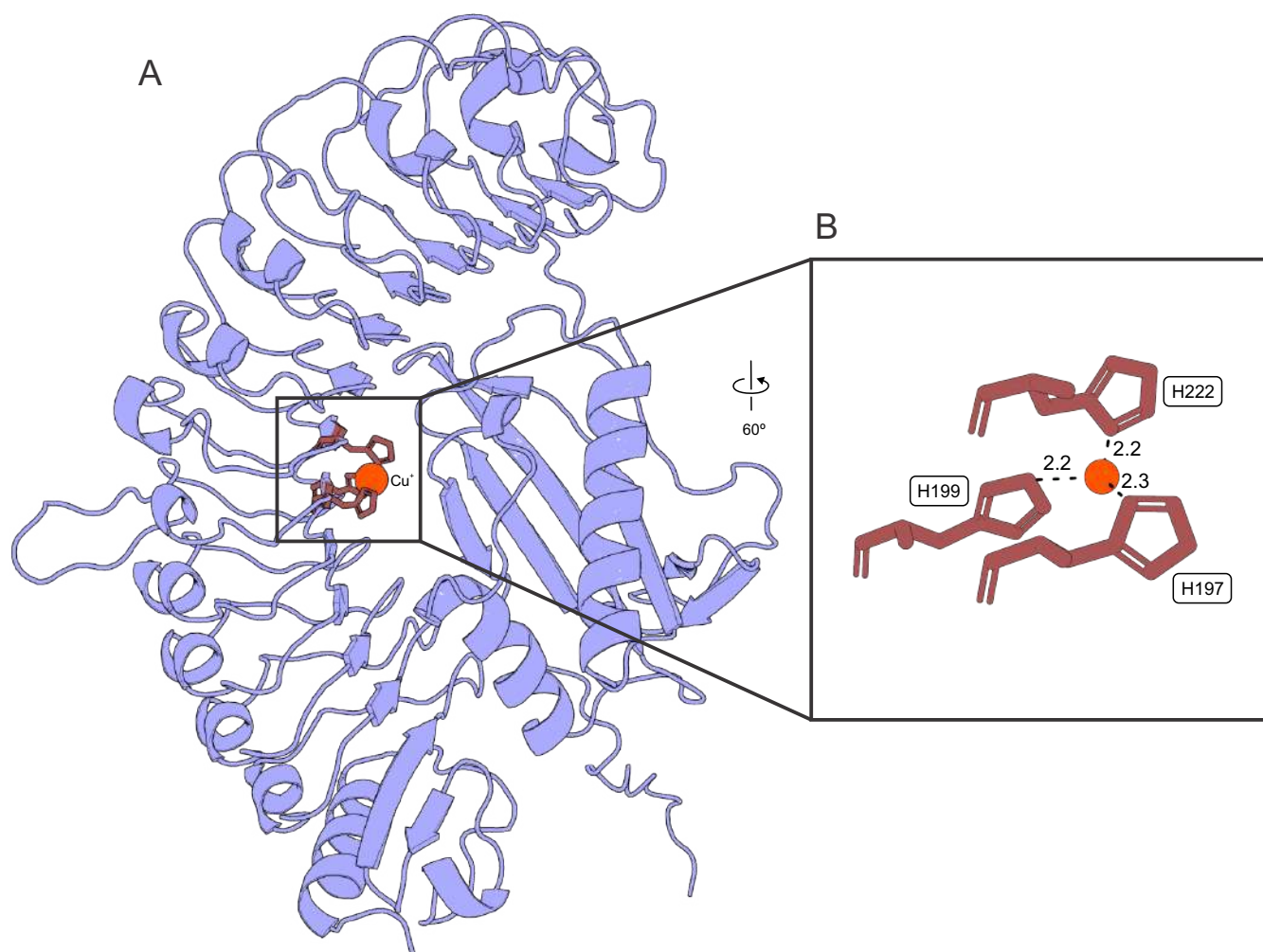
A



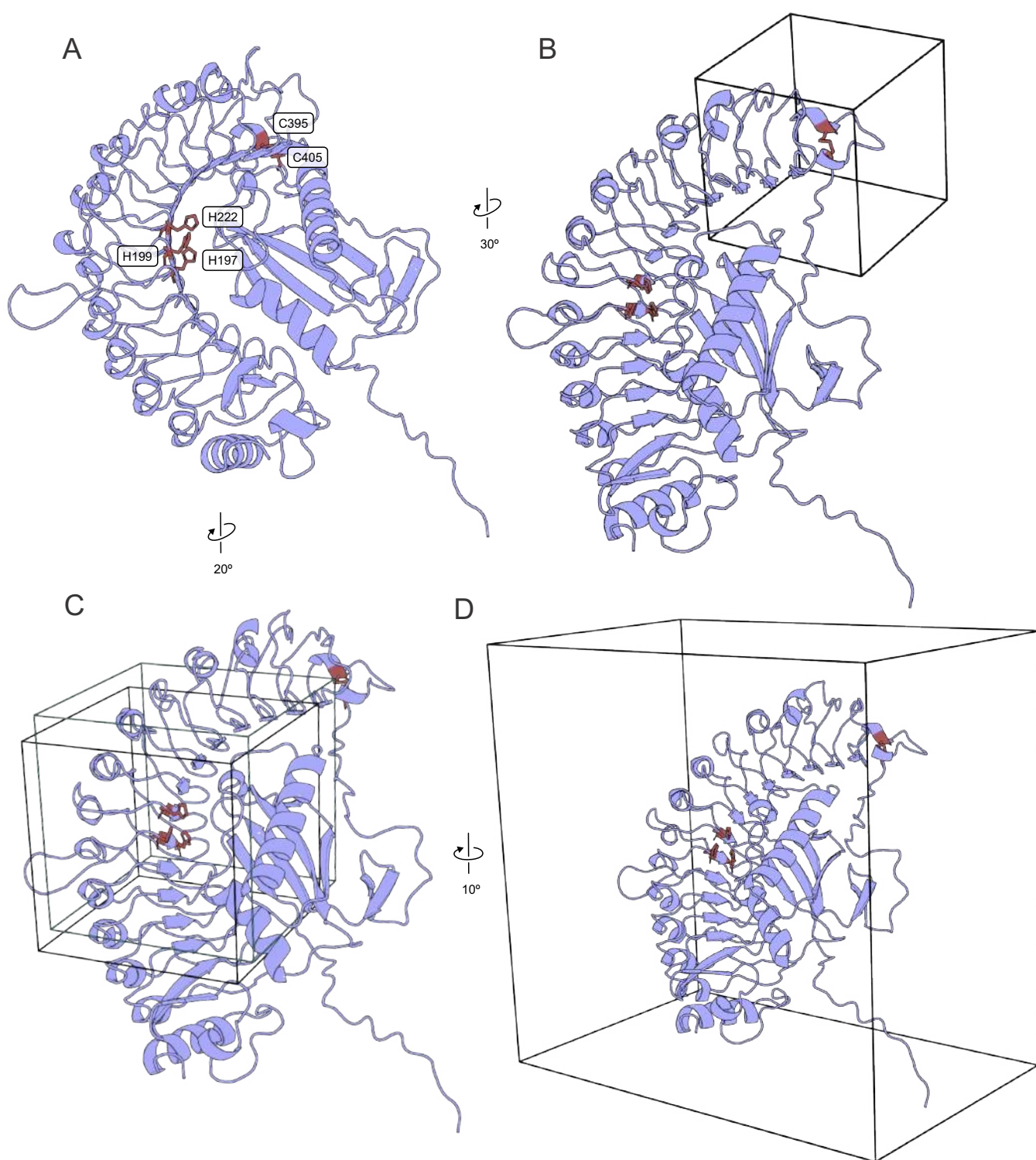
B



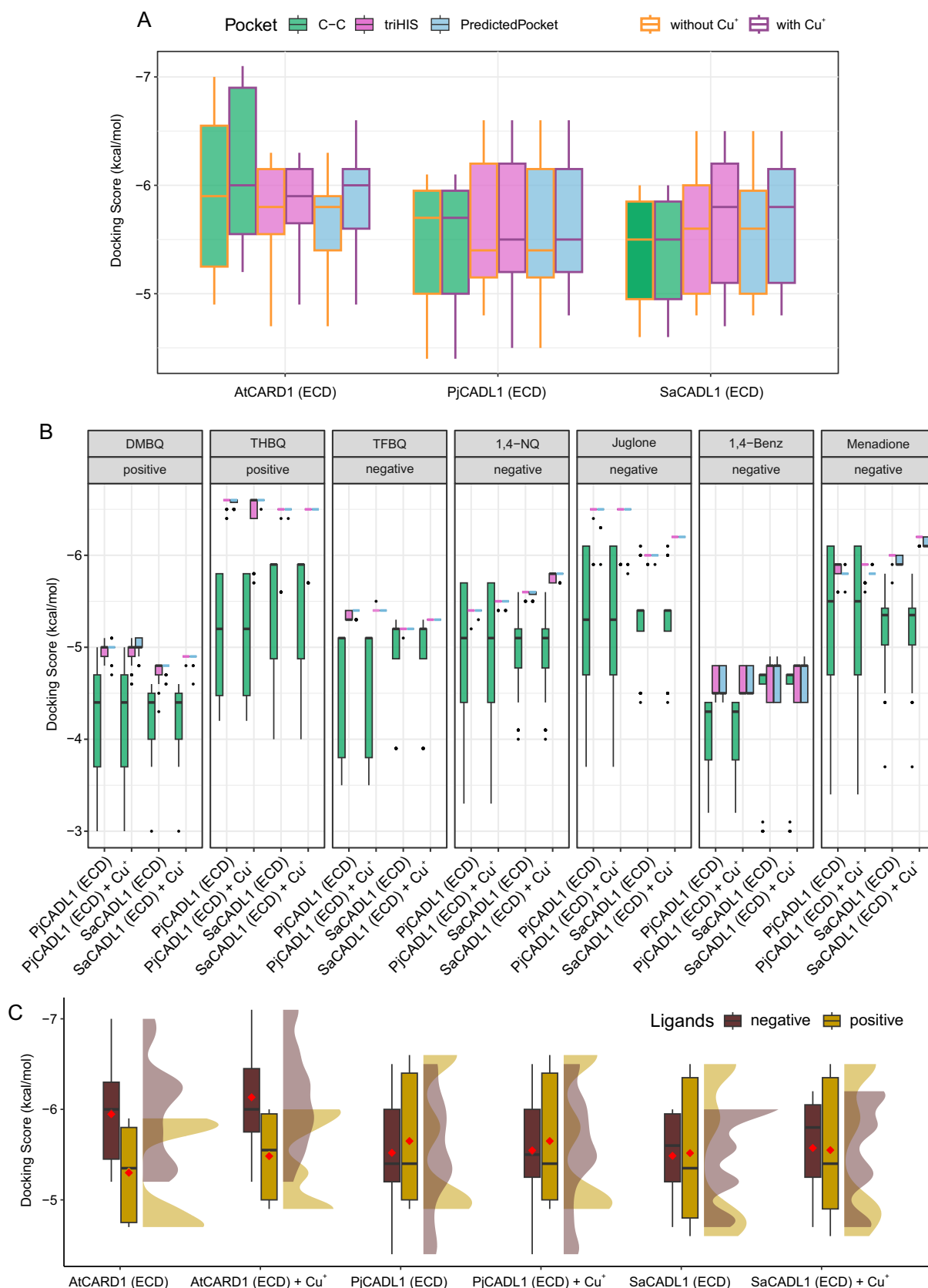
Supplementary Figure S3. Structural alignment of AtCARD1 with selected ectodomain homologs. (A) Superposition of AtCARD1 and 8WEG (MIK2 ectodomain). (B) Superposition of AtCARD1 and 4MNA (AtFLS2 ectodomain). Highlighted regions indicate the spatial distances between conserved residues identified in the previous sequence alignment (Supplementary Fig. S2).



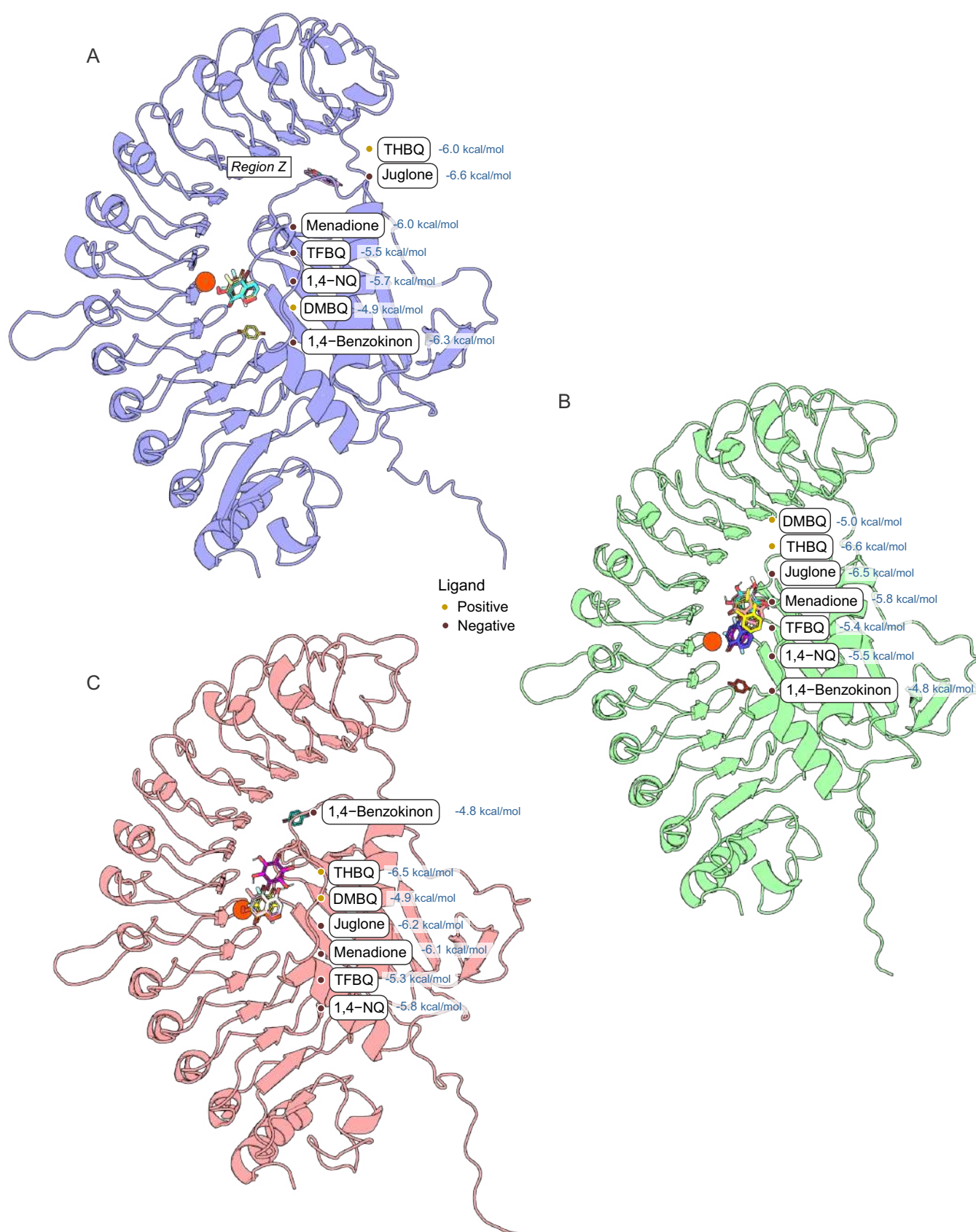
Supplementary Figure S4. Positioning of the Cu^+ ion within the predicted histidine triad of AtCARD1. (A) Ribbon representation of the AtCARD1 extracellular domain showing the modeled Cu^+ ion (orange sphere) coordinated by three histidine residues. (B) Close-up view highlighting the Cu^+ coordination site formed by H197, H199, and H222, with distances (in Å) indicated between the metal center and the imidazole nitrogens of each histidine residue.



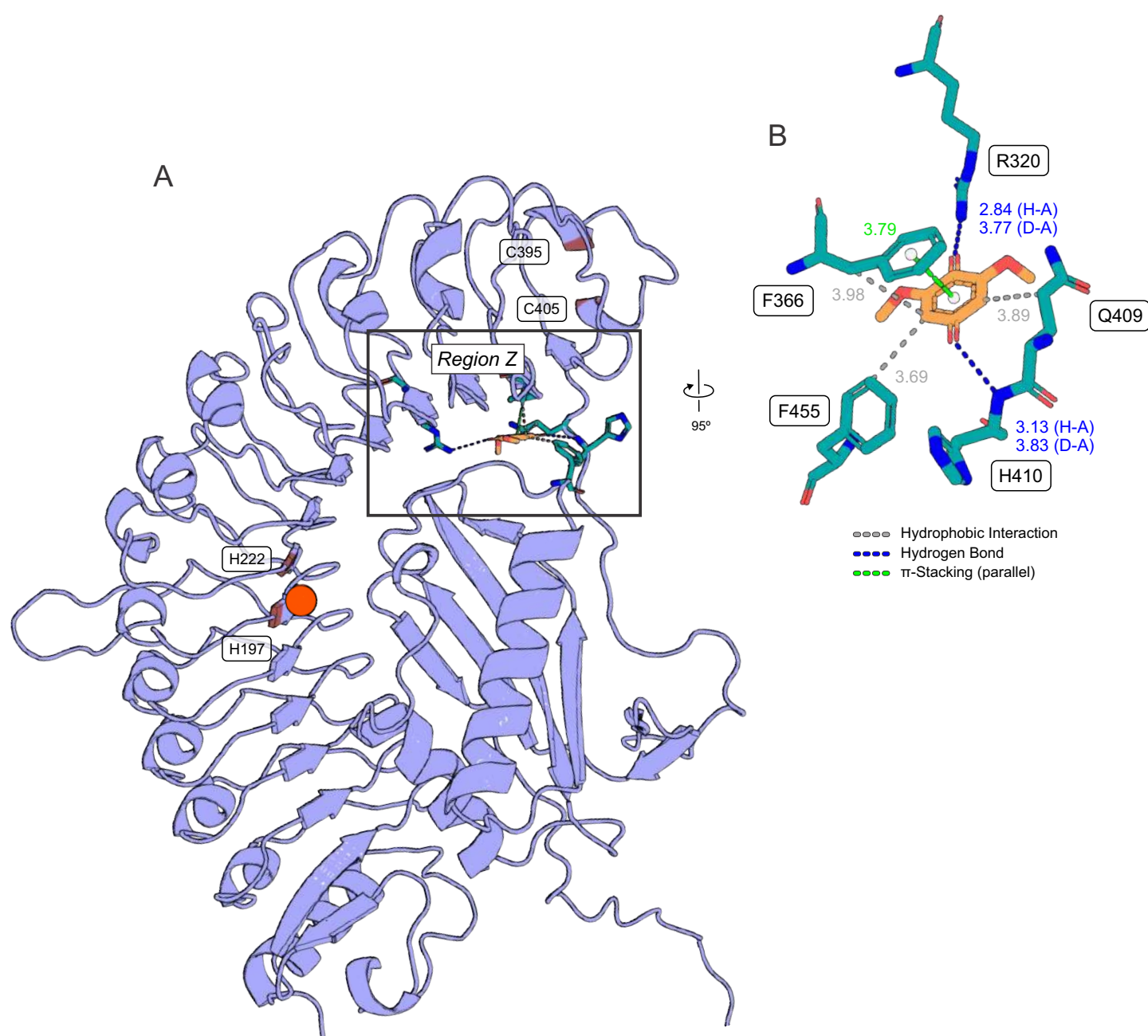
Supplementary Figure S5. Docking box configurations tested for molecular docking calibration. (A) AtCARD1 ectodomain (ECD) showing the positions of the C-C (Cys395 and Cys405) and the histidine triade (His197, His199, His222) region. (B) Docking box centered on the C-C region. (C) Docking box centered on the triHIS site (black) and the predicted pocket from PrankWeb (aquamarine). (D) Whole-protein (WholeP) docking box encompassing the entire ECD, used to explore potential alternative binding regions. The same box configurations were generated for PjCADL1 and SaCADL1. Boxes shown in panels (B) and (C) were set to 35 Å.



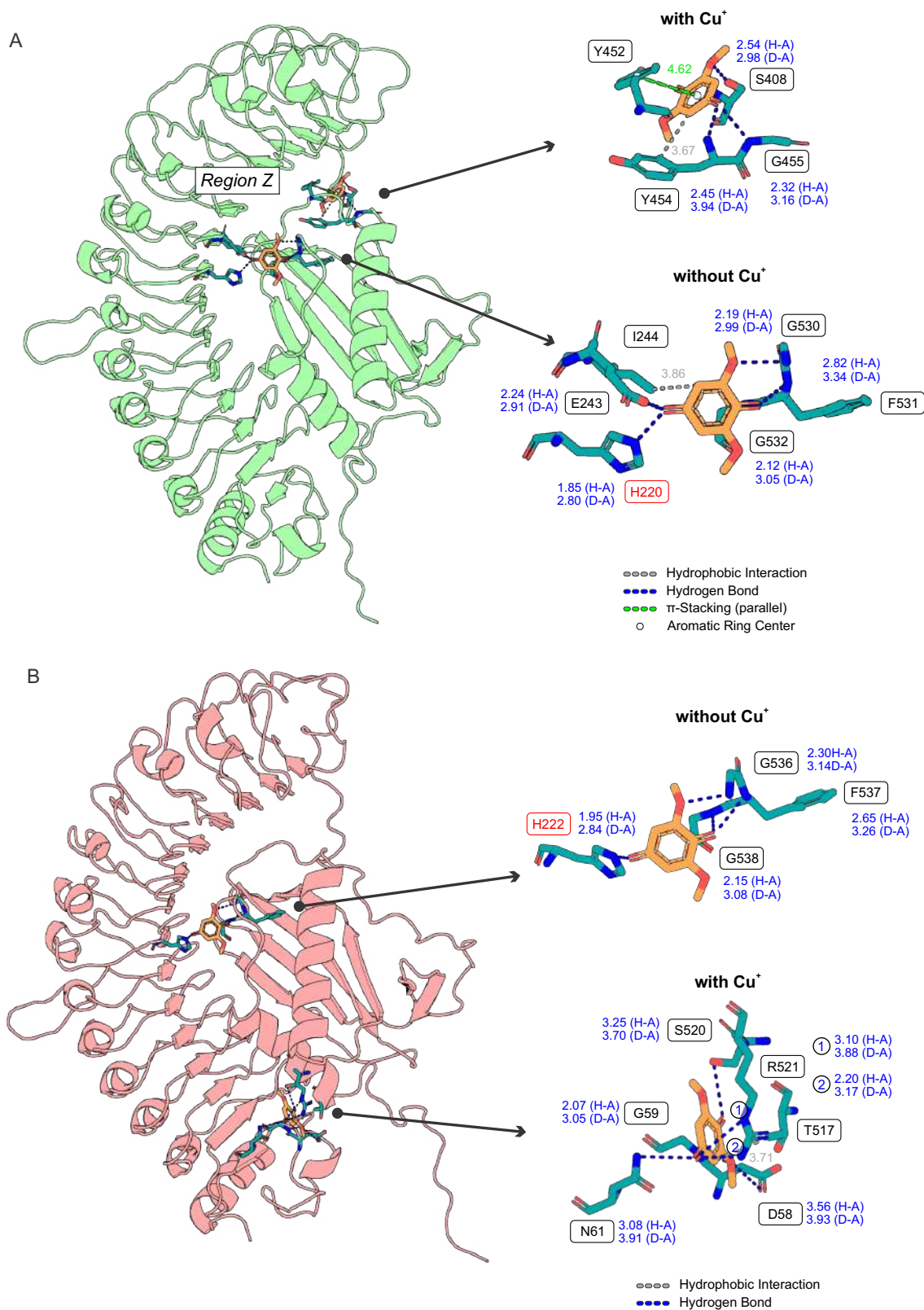
Supplementary Figure S6. Comparison of docking scores across receptors, pockets, and ligand types. (A) Docking scores grouped by species and pocket, comparing models with and without Cu⁺. (B) Docking scores by ligand and pocket for each receptor from *Striga* and *P. japonicum*, shown for both Cu⁺-bound and Cu⁺-free forms. (C) Comparison of docking scores across receptors (with and without Cu⁺) grouped by ligand type. Red diamonds indicate the mean.



Supplementary Figure S7. Predicted binding poses of all ligands derivatives across AtCARD1 and homologs. Docking results are shown for (A) AtCARD1 + Cu⁺, (B) PjCADL1 + Cu, and (C) SaCADL1 + Cu⁺, highlighting the docking scores (kcal/mol) for each ligand showed in blue ligand label.

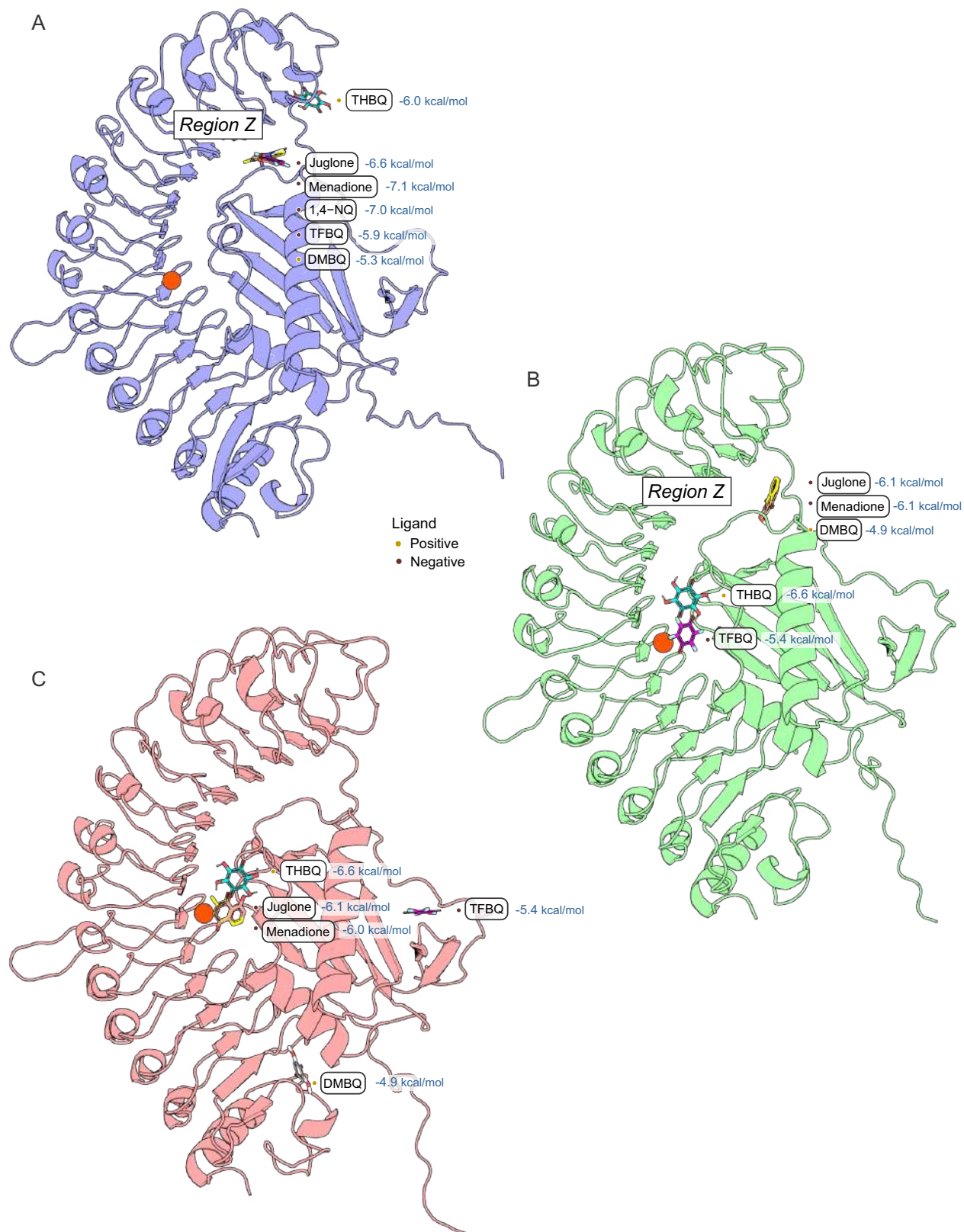


Supplementary Figure S8. AtCARD1 + Cu⁺ (WholeP docking configuration) highlighting the secondary ligand-binding region (Region Z). (A) Overall structure of the AtCARD1 ectodomain showing the Cu⁺-coordinated His-triad (H197, H222) and the C-C region (C395, C405), with the ligand bound in the secondary pocket identified from WholeP docking analysis. (B) Close-up of DMBQ interactions in Region Z as detected by PLIP. Hydrogen bonds (blue dashed lines), hydrophobic interactions (gray dashed lines), and π -stacking (green dashed lines) are shown with their respective distances (Å).

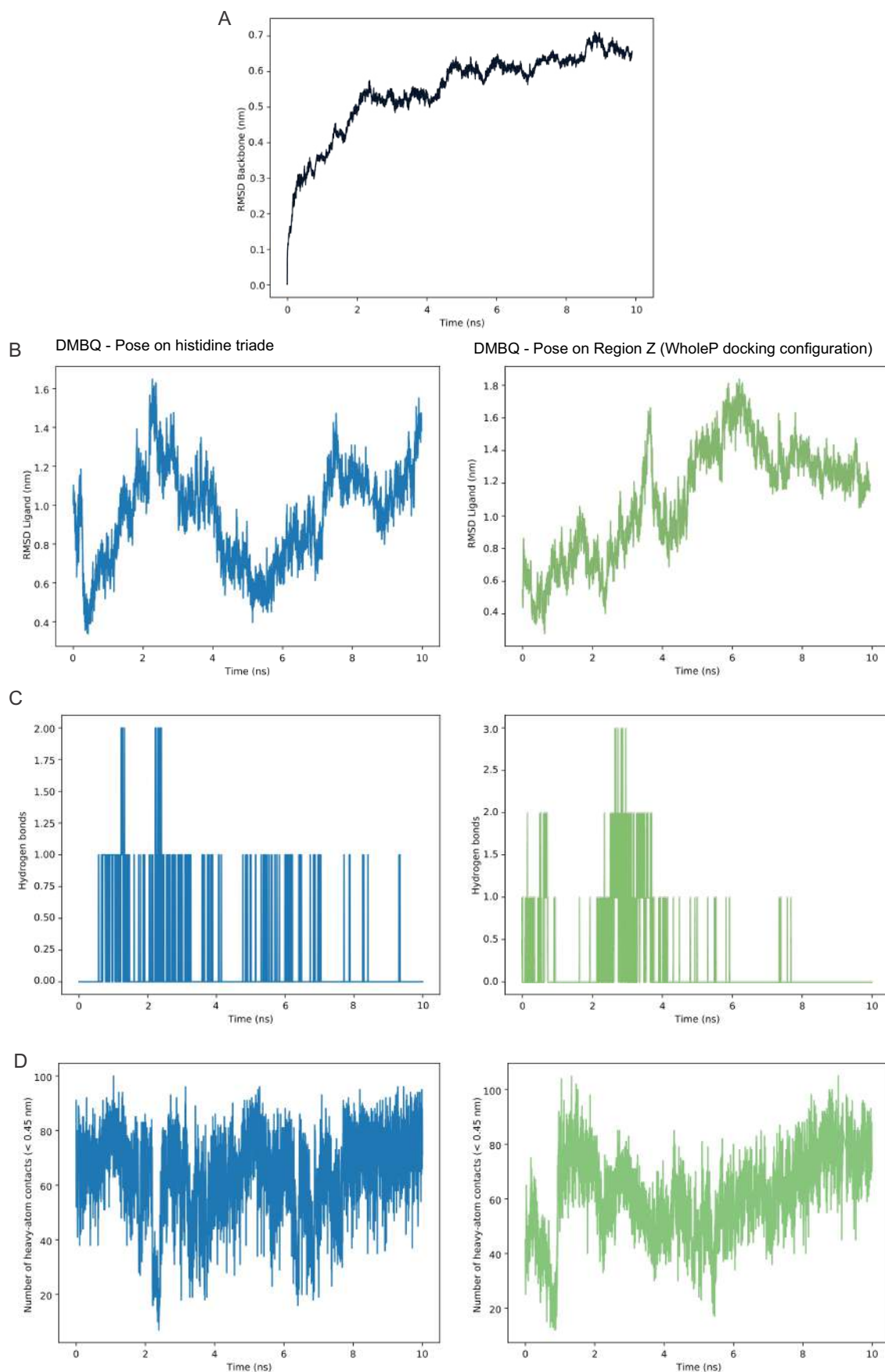


Supplementary Figure S9. Predicted interactions of DMBQ with PjCADL1 and SaCADL1 ectodomains, with and without Cu⁺ (WholeP docking configuration).

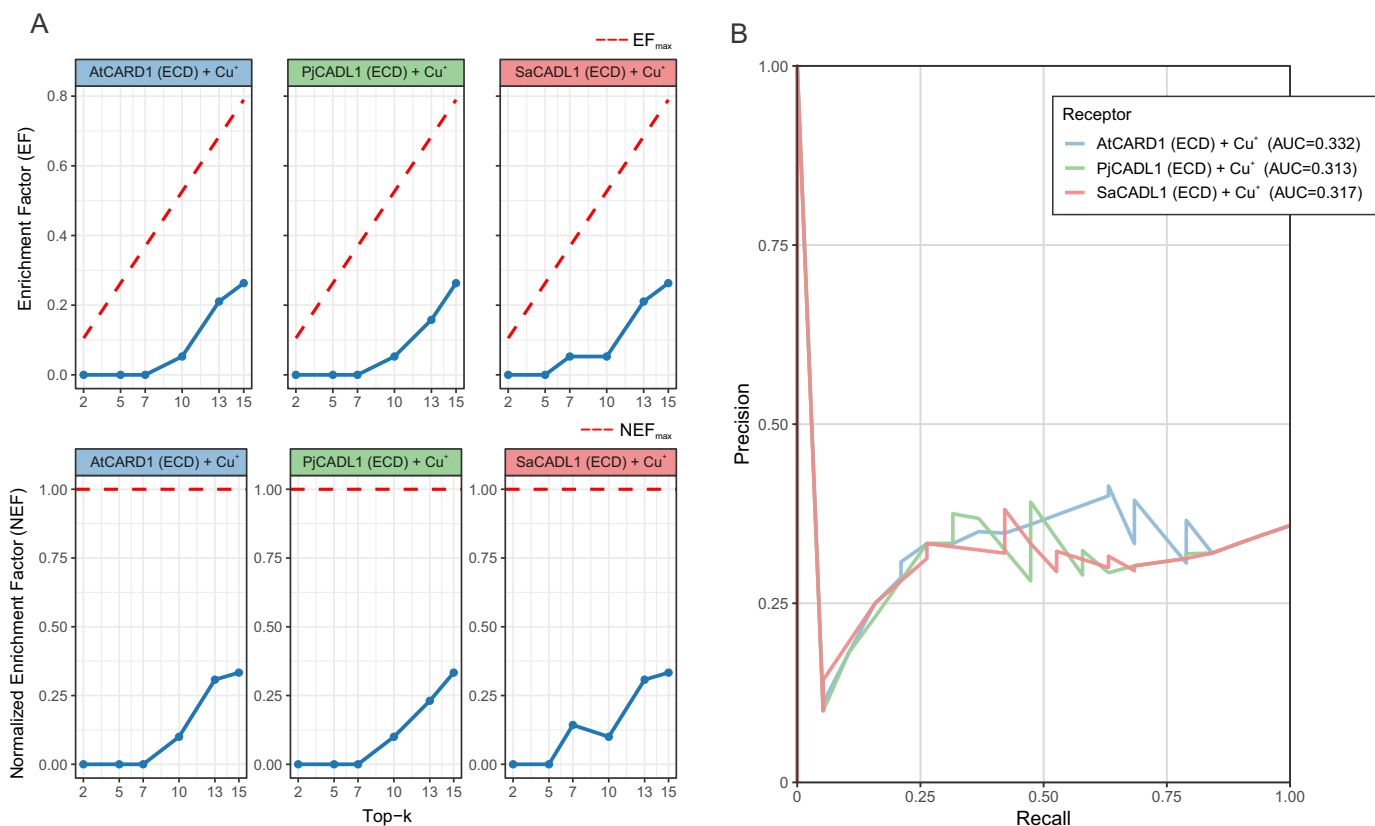
(A) PjCADL1 showing the predicted DMBQ binding site in the presence (top) and absence (bottom) of Cu⁺. (B) SaCADL1 showing the predicted DMBQ binding site in the absence (top) and presence (bottom) of Cu⁺. Hydrogen bonds (blue dashed lines), hydrophobic interactions (gray dashed lines), and π -stacking (green dashed lines) are shown with corresponding distances in Å.



Supplementary Figure S10. Predicted docking poses of quinone derivatives across AtCARD1 and homologs in the presence of Cu^+ (WholeP docking configuration). (A) AtCARD1, (B) PjCADL1, and (C) SaCADL1 showing the relative positioning of all docked ligands within the predicted binding pocket. Ligands are colored according to functional classification (positive or negative), with corresponding docking scores (kcal/mol) shown next to each compound label.

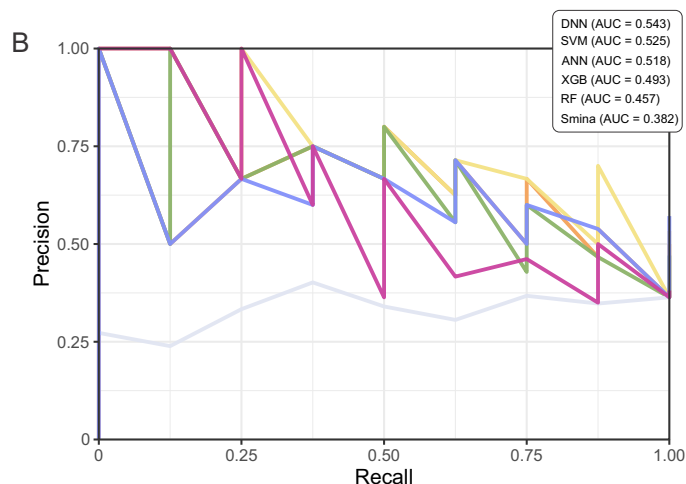
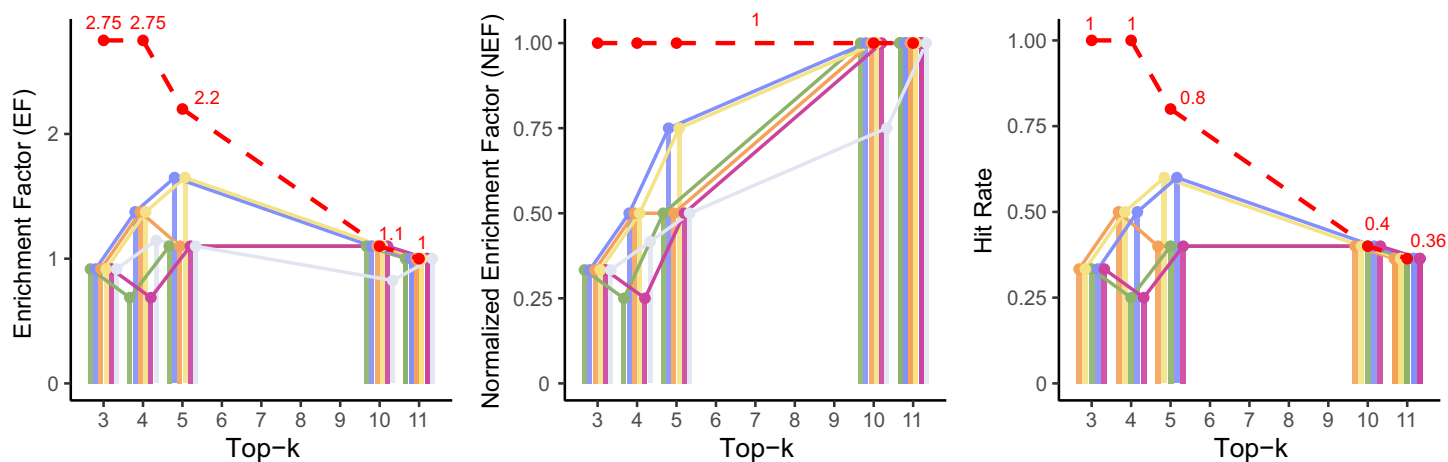


Supplementary Figure S11. Molecular dynamics analysis of DMBQ bound to AtCARD1 in two distinct binding configurations. (A) Backbone RMSD of AtCARD1 over 10 ns of MD simulation, indicating overall protein stability. (B) Ligand RMSD relative to the initial docking pose for DMBQ bound to the histidine triad (blue) and to Region Z (green). (C) Number of protein-ligand hydrogen bonds over time for each binding configuration. (D) Number of heavy-atom contacts within a 0.45 nm cutoff between DMBQ and AtCARD1 during the simulation.



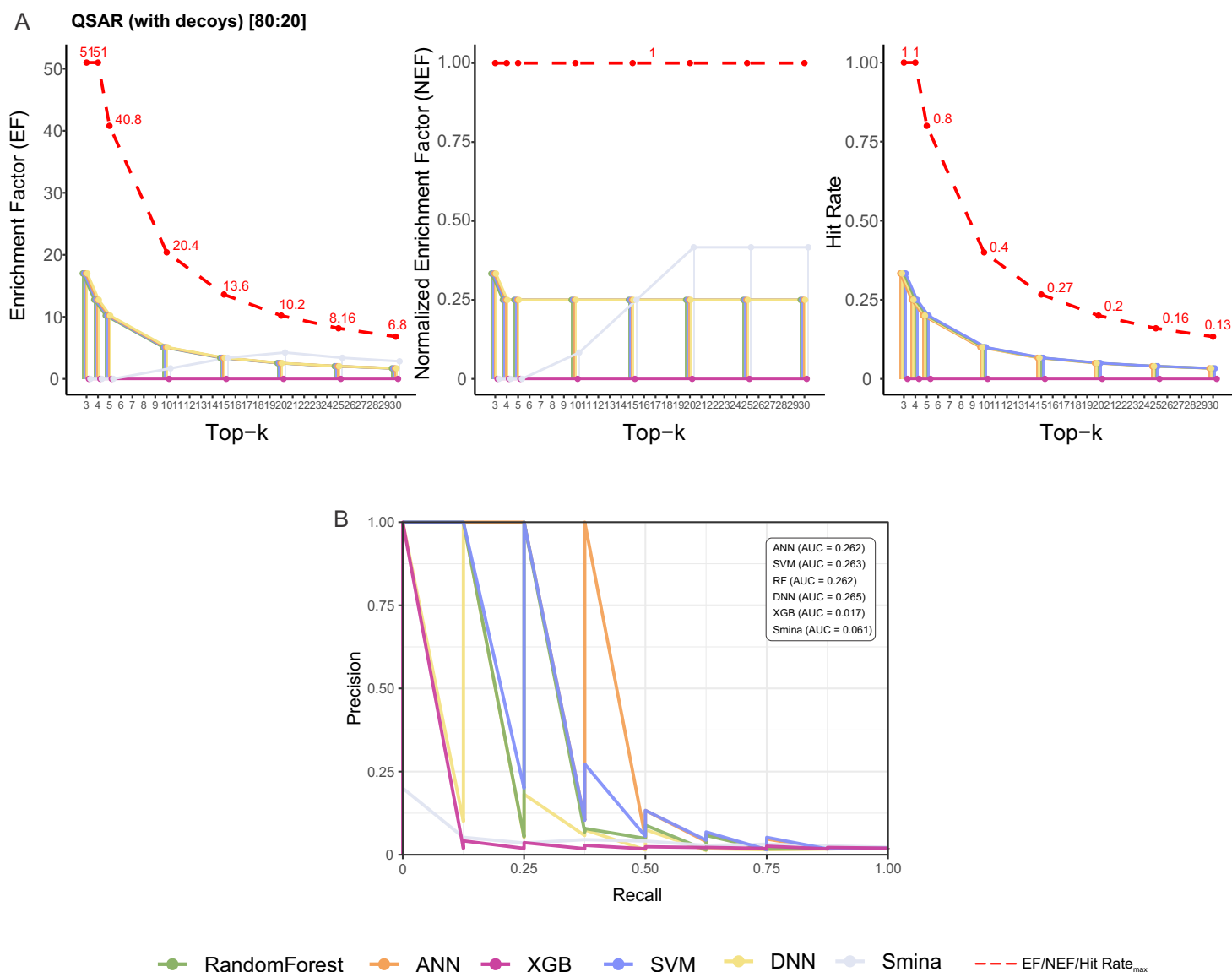
Supplementary Figure S12. Performance of Smina docking-based ranking across AtCARD1 and parasitic homologs in the presence of Cu⁺. (A) Enrichment Factor (EF, top panels) and Normalized Enrichment Factor (NEF, bottom panels) as a function of top-k for AtCARD1 (left), PjCADL1 (center), and SaCADL1 (right) ectodomains docked in the Cu⁺-bound configuration. Blue lines indicate observed enrichment, while red dashed lines denote the theoretical maximum values (EF_{max} and NEF_{max}) at each top-k threshold. (B) Precision-recall curves derived from Smina docking scores for AtCARD1 (blue), PjCADL1 (green), and SaCADL1 (red), with corresponding PR-AUC values indicated in the legend.

A QSAR (without decoys) [80:20]

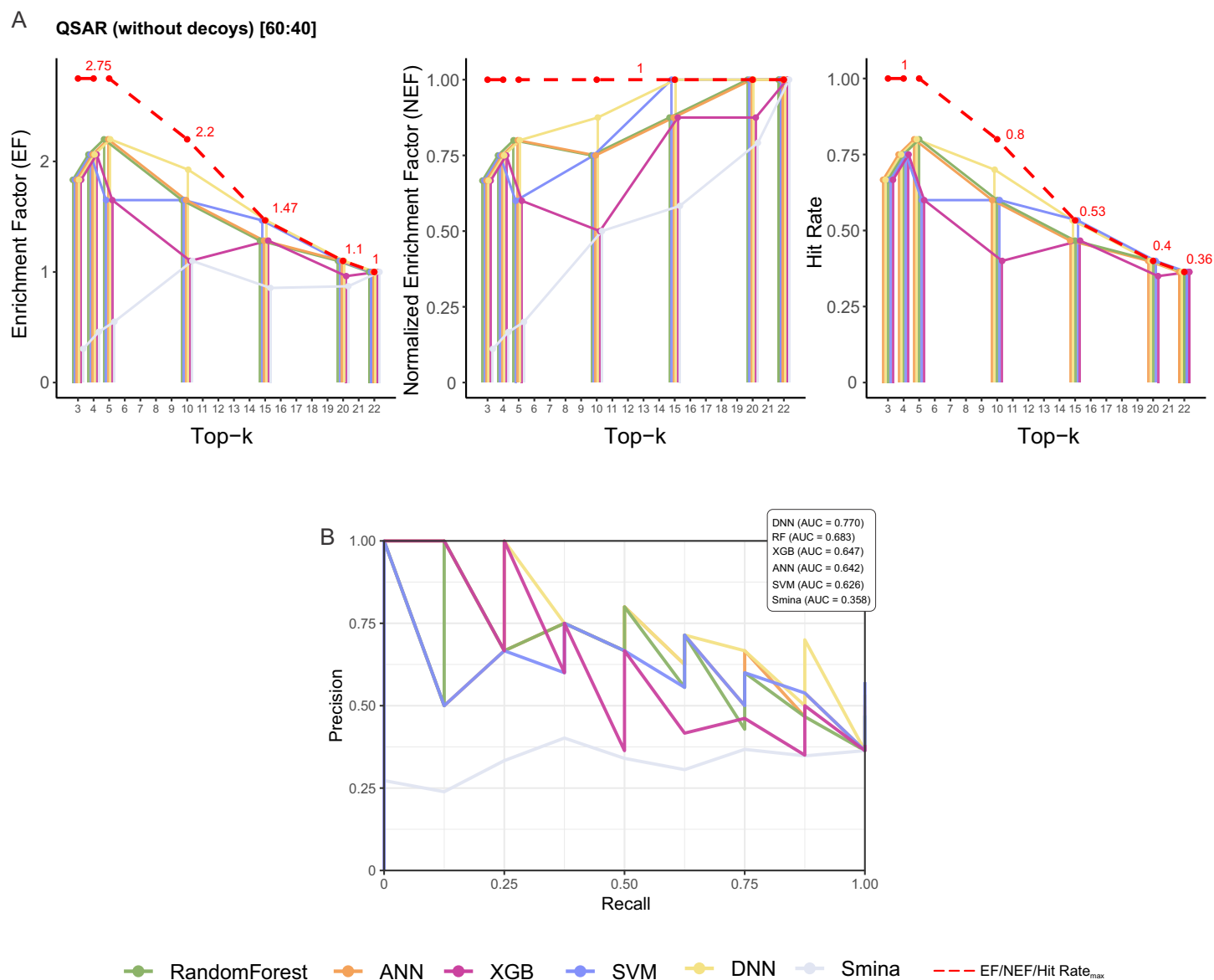


— RandomForest — ANN — XGB — SVM — DNN — Smina — EF/NEF/Hit Rate_{max}

Supplementary Figure S13. Performance of QSAR-based machine-learning models using Morgan fingerprints in the absence of decoys (80:20 split). (A) Enrichment Factor (EF), Normalized Enrichment Factor (NEF), and Hit Rate as a function of top-k for different classification models (Random Forest, ANN, XGBoost, SVM, and DNN) and Smina. Dashed red lines indicate the theoretical maximum values (EFmax, NEFmax, and HitRatemax) achievable at each top-k threshold. (B) Precision-recall curves for all evaluated models, including Smina docking scores shown for comparison, with corresponding PR-AUC values reported in the legend.

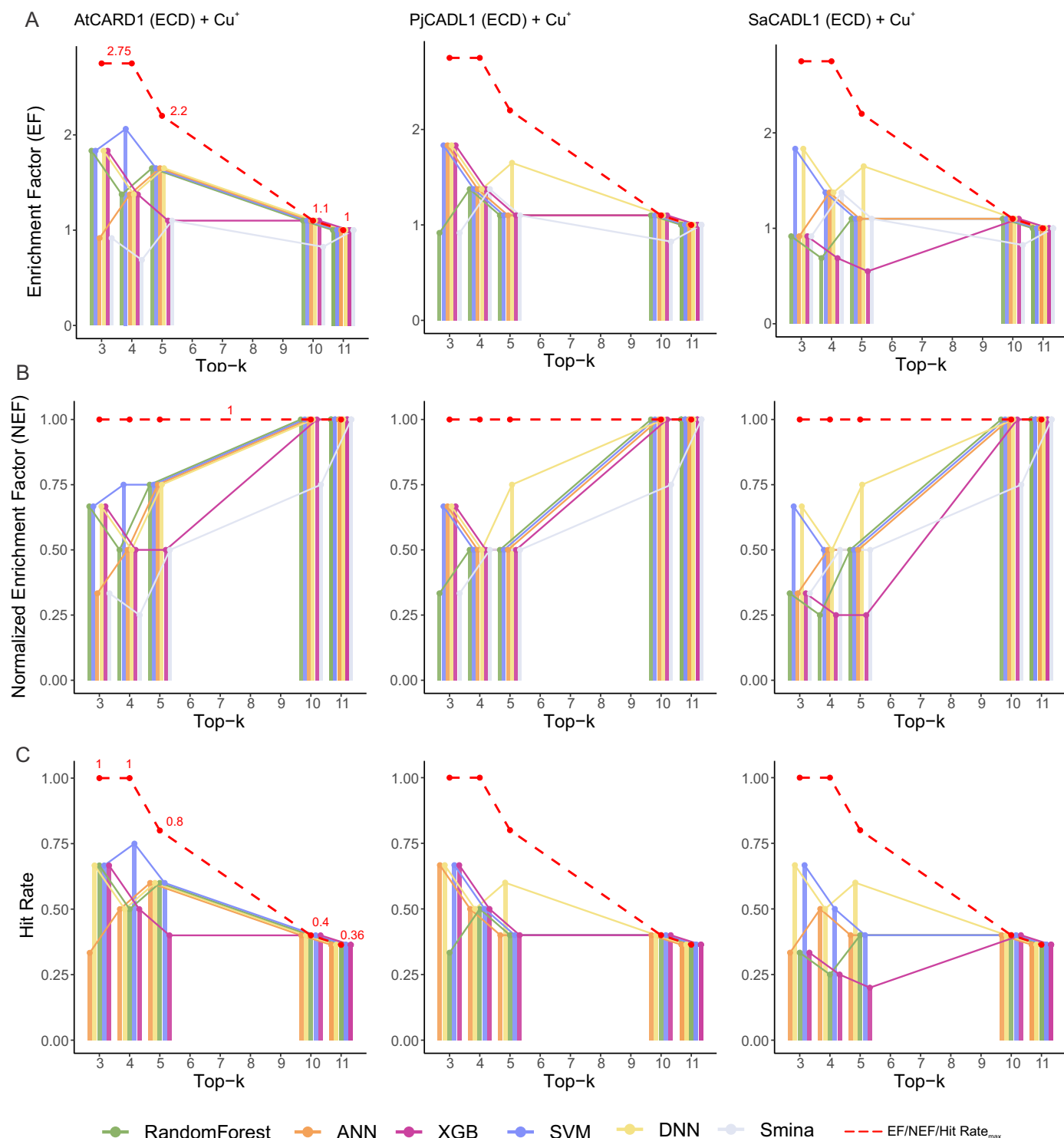


Supplementary Figure S14. Performance of QSAR-based machine-learning models using Morgan fingerprints in the presence of decoys (80:20 split). (A) Enrichment Factor (EF), Normalized Enrichment Factor (NEF), and Hit Rate as a function of top-k for different classification models (Random Forest, ANN, XGBoost, SVM, and DNN) and Smina. Dashed red lines indicate the theoretical maximum values (EF_{max}, NEF_{max}, and HitRate_{max}) achievable at each top-k threshold. (B) Precision-recall curves for all evaluated models, including Smina docking scores shown for comparison, with corresponding PR-AUC values reported in the legend.

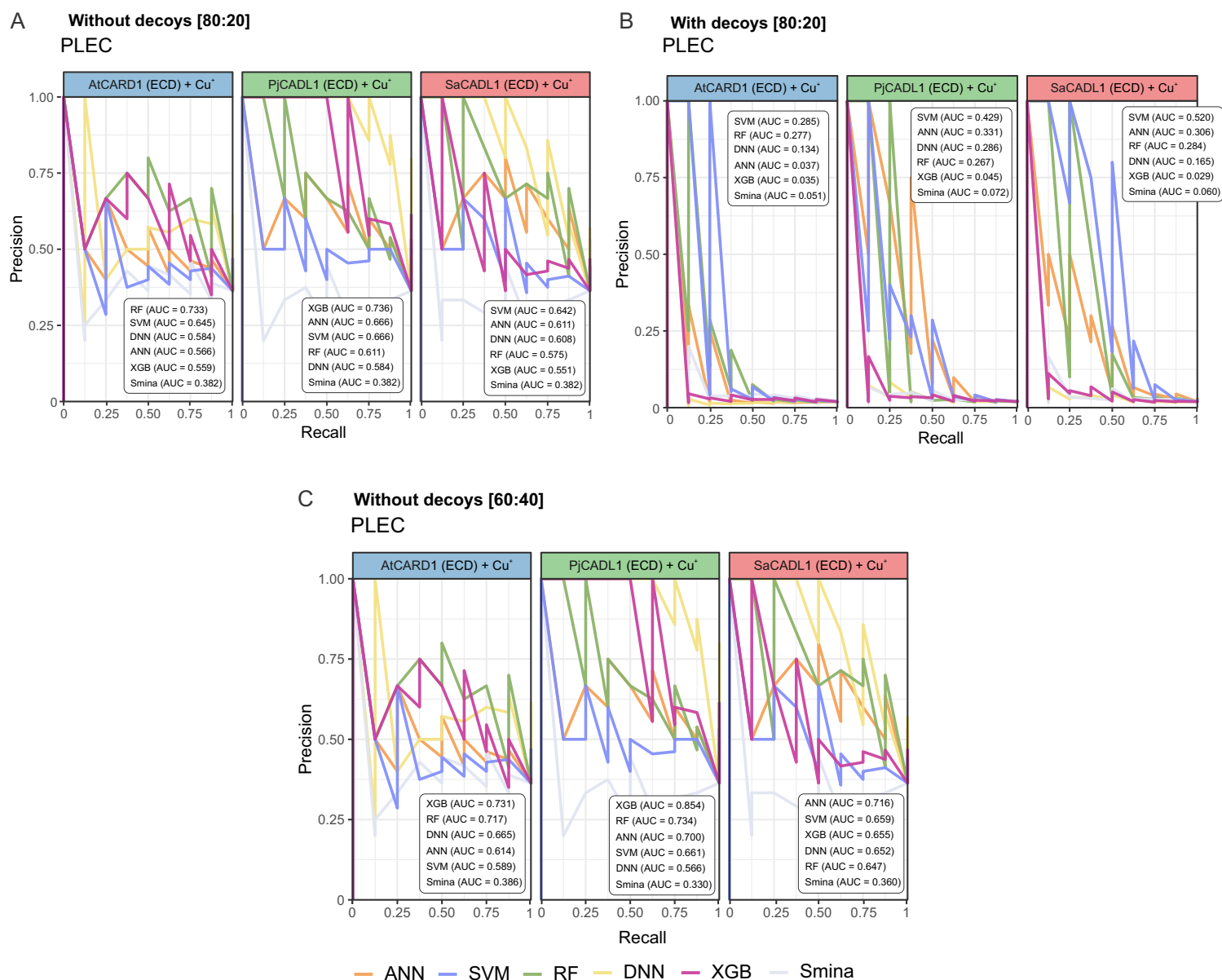


Supplementary Figure S15. Performance of QSAR-based machine-learning models using Morgan fingerprints in the absence of decoys (60:40 split). (A) Enrichment Factor (EF), Normalized Enrichment Factor (NEF), and Hit Rate as a function of top-k for different classification models (Random Forest, ANN, XGBoost, SVM, and DNN) and Smyna. Dashed red lines indicate the theoretical maximum values (EF_{max}, NEF_{max}, and HitRate_{max}) achievable at each top-k threshold. (B) Precision-recall curves for all evaluated models, including Smyna docking scores shown for comparison, with corresponding PR-AUC values reported in the legend.

PLEC (without decoys) [80:20]

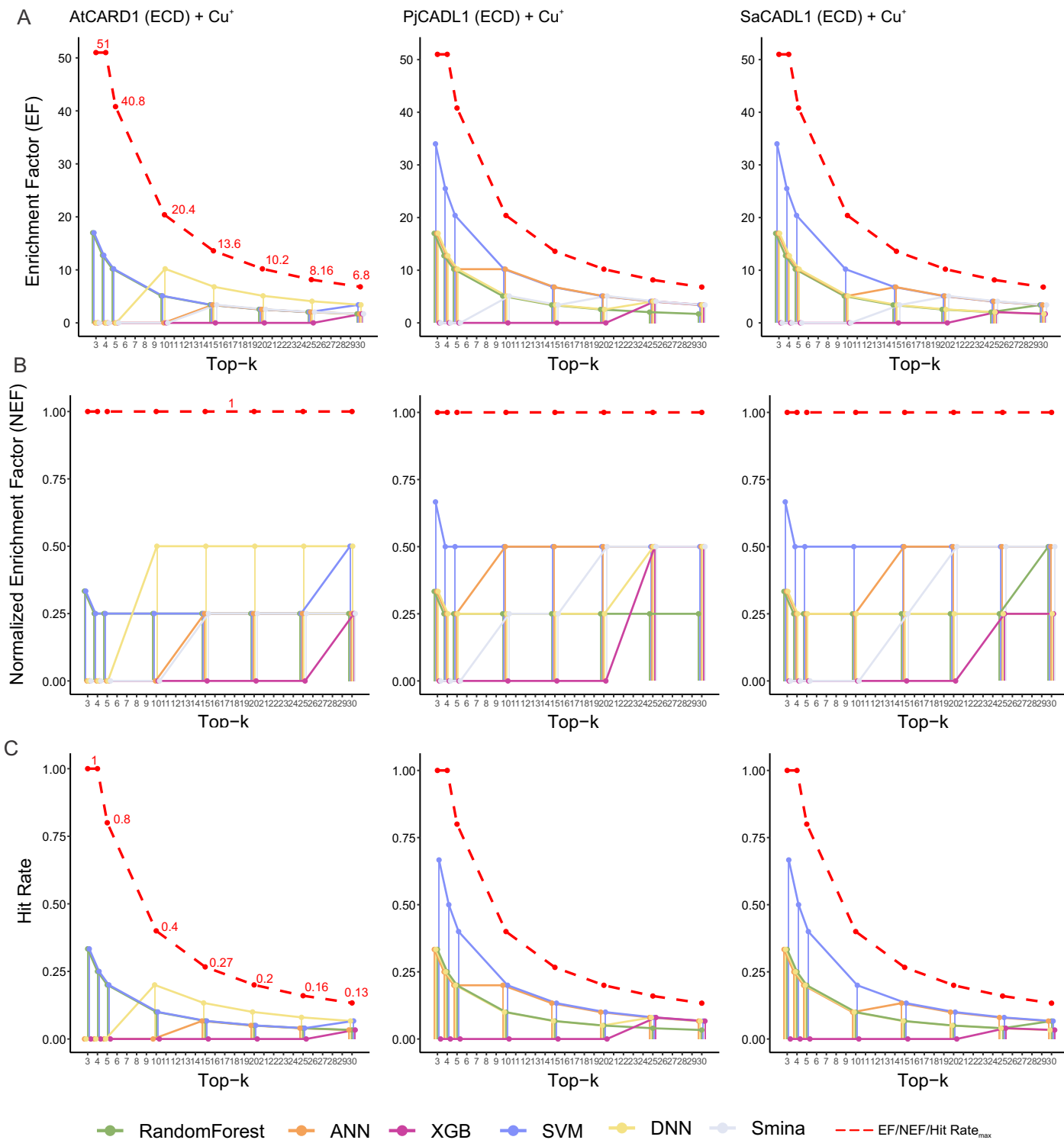


Supplementary Figure S16. Performance of PLEC-based machine-learning models without decoys (80:20 split). Evaluation of classification models using PLEC interaction descriptors and (A) Enrichment Factor (EF), (B) Normalized Enrichment Factor (NEF), and (C) Hit Rate as a function of top-k for AtCARD1 (left), PjCADL1 (center), and SaCADL1 (right) ectodomains in the Cu⁺-bound configuration. Colored lines represent individual machine-learning models (Random Forest, ANN, XGBoost, SVM, and DNN), while Smina docking-based ranking is shown for comparison. Red dashed lines denote the theoretical maximum values (EF_{max}, NEF_{max}, and HitRate_{max}) achievable at each top-k threshold.



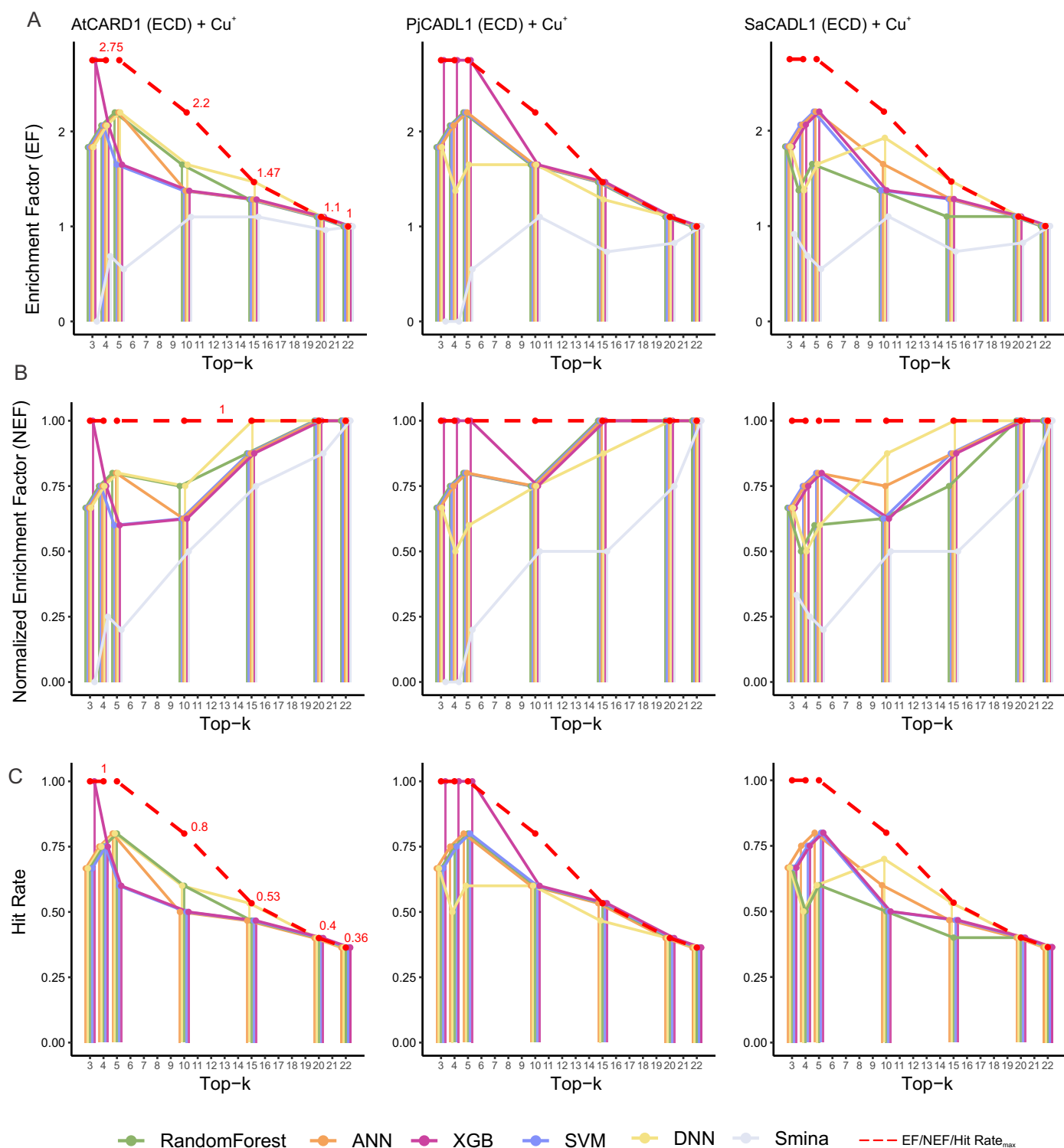
Supplementary Figure S17. Precision-recall performance of PLEC-based machine-learning models across receptors and dataset configurations. Precision-recall curves obtained from PLEC interaction descriptors for AtCARD1 (left), PjCADL1 (center), and SaCADL1 (right) ectodomains in the Cu⁺-bound configuration. (A) Evaluation without decoys using the 80:20 train-test split. (B) Evaluation with decoys using the 80:20 train-test split. (C) Evaluation without decoys using the 60:40 train-test split. Curves correspond to different classification models (ANN, SVM, Random Forest, DNN, and XGBoost), with Smina docking-based ranking shown for comparison. PR-AUC values for each model are reported in the legends.

PLEC (with decoys) [80:20]



Supplementary Figure S18. Performance of PLEC-based machine-learning models with decoys (80:20 split). Evaluation of classification models using PLEC interaction descriptors and (A) Enrichment Factor (EF), (B) Normalized Enrichment Factor (NEF), and (C) Hit Rate as a function of top-k for AtCARD1 (left), PjCADL1 (center), and SaCADL1 (right) ectodomains in the Cu⁺-bound configuration. Colored lines represent individual machine-learning models (Random Forest, ANN, XGBoost, SVM, and DNN), while Smina docking-based ranking is shown for comparison. Red dashed lines denote the theoretical maximum values (EF_{max}, NEF_{max}, and HitRate_{max}) achievable at each top-k threshold.

PLEC (without decoys) [60:40]



Supplementary Figure S19. Performance of PLEC-based machine-learning models without decoys (60:40 split). Evaluation of classification models using PLEC interaction descriptors and (A) Enrichment Factor (EF), (B) Normalized Enrichment Factor (NEF), and (C) Hit Rate as a function of top-k for AtCARD1 (left), PjCADL1 (center), and SaCADL1 (right) ectodomains in the Cu⁺-bound configuration. Colored lines represent individual machine-learning models (Random Forest, ANN, XGBoost, SVM, and DNN), while Smina docking-based ranking is shown for comparison. Red dashed lines denote the theoretical maximum values (EF_{max}, NEF_{max}, and HitRate_{max}) achievable at each top-k threshold.

References

1. Wu, F. *et al.* Hydrogen peroxide sensor HPCA1 is an LRR receptor kinase in Arabidopsis. *Nature* **578**, 577–581 (2020).
2. Laohavisit, A. *et al.* Quinone perception in plants via leucine-rich-repeat receptor-like kinases. *Nature* **587**, 92–97 (2020).
3. Hillion, M. & Antelmann, H. Thiol-based redox switches in prokaryotes. *Biological Chemistry* **396**, 415–444 (2015).
4. Yamamoto, M., Kensler, T. W. & Motohashi, H. The KEAP1-NRF2 System: a Thiol-Based Sensor-Effector Apparatus for Maintaining Redox Homeostasis. *Physiological Reviews* **98**, 1169–1203 (2018).
5. Körner, M., Slaby, P. & Albert, M. Quinones shuffling the CARDS. *Nat. Plants* **6**, 1074–1075 (2020).
6. Mutuku, J. M., Cui, S., Yoshida, S. & Shirasu, K. Orobanchaceae parasite–host interactions. *New Phytologist* **230**, 46–59 (2021).
7. Ishihama, N. *et al.* A copper-dependent, redox-based hydrogen peroxide perception in plants. Preprint at <https://doi.org/10.1101/2025.07.22.666036> (2025).
8. Altschul, S. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389–3402 (1997).
9. Katoh, K. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**, 3059–3066 (2002).
10. Robert, X., Guillon, C. & Gouet, P. FoldScript: a web server for the efficient analysis of AI-generated 3D protein models. *Nucleic Acids Research* **53**, W277–W282 (2025).
11. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: A Sequence Logo Generator. *Genome Res.* **14**, 1188–1190 (2004).
12. Van Kempen, M. *et al.* Fast and accurate protein structure search with Foldseek. *Nat Biotechnol* **42**, 243–246 (2024).
13. Altschul, S. F. *et al.* Protein database searches using compositionally adjusted substitution matrices. *The FEBS Journal* **272**, 5101–5109 (2005).
14. Madeira, F. *et al.* The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. *Nucleic Acids Research* **52**, W521–W525 (2024).
15. Hallgren, J. *et al.* DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. Preprint at <https://doi.org/10.1101/2022.04.08.487609> (2022).
16. Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 3.1.6.1. (2015).
17. Polák, L. *et al.* PrankWeb 4: a modular web server for protein–ligand binding site prediction and downstream analysis. *Nucleic Acids Research* **53**, W466–W471 (2025).

18. Sun, Y. *et al.* Structural Basis for flg22-Induced Activation of the *Arabidopsis* FLS2-BAK1 Immune Complex. *Science* **342**, 624–628 (2013).
19. Potter, S. C. *et al.* HMMER web server: 2018 update. *Nucleic Acids Research* **46**, W200–W204 (2018).
20. Koes, D. R., Baumgartner, M. P. & Camacho, C. J. Lessons Learned in Empirical Scoring with Smina from the CSAR 2011 Benchmarking Exercise. *J. Chem. Inf. Model.* **53**, 1893–1904 (2013).
21. Eberhardt, J., Santos-Martins, D., Tillack, A. F. & Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf. Model.* **61**, 3891–3898 (2021).
22. Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605–1612 (2004).
23. Schake, P., Bolz, S. N., Linnemann, K. & Schroeder, M. PLIP 2025: introducing protein–protein interactions to the protein–ligand interaction profiler. *Nucleic Acids Research* **53**, W463–W465 (2025).
24. O'Boyle, N. M. *et al.* Open Babel: An open chemical toolbox. *J Cheminform* **3**, 33 (2011).
25. Sousa Da Silva, A. W. & Vranken, W. F. ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Res Notes* **5**, 367 (2012).
26. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J Comput Chem* **25**, 1157–1174 (2004).
27. Wang, J., Wang, W., Kollman, P. A. & Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling* **25**, 247–260 (2006).
28. Abraham, M. *et al.* GROMACS 2023 Source code. Zenodo <https://doi.org/10.5281/ZENODO.7588619> (2023).
29. Hornak, V. *et al.* Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **65**, 712–725 (2006).
30. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950–1958 (2010).
31. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**, 926–935 (1983).
32. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **126**, 014101 (2007).
33. Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **81**, 3684–3690 (1984).

34. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **52**, 7182–7190 (1981).
35. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
36. Landrum, G. *et al.* rdkit/rdkit: 2017_03_1 (Q1 2017) Release. Zenodo <https://doi.org/10.5281/ZENODO.556226> (2017).
37. Bolton, E. E. *et al.* PubChem3D: a new resource for scientists. *J Cheminform* **3**, 32 (2011).
38. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
39. Imrie, F., Bradley, A. R. & Deane, C. M. Generating property-matched decoy molecules using deep learning. *Bioinformatics* **37**, 2134–2141 (2021).
40. Mysinger, M. M., Carchia, M., Irwin, John. J. & Shoichet, B. K. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* **55**, 6582–6594 (2012).
41. Morgan, H. L. The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service. *J. Chem. Doc.* **5**, 107–113 (1965).
42. Wójcikowski, M., Zielenkiewicz, P. & Siedlecki, P. Open Drug Discovery Toolkit (ODDT): a new open-source player in the drug discovery field. *J Cheminform* **7**, 26 (2015).
43. Breiman, L. Random Forests. *Machine Learning* **45**, 5–32 (2001).
44. Cortes, C. & Vapnik, V. Support-vector networks. *Mach Learn* **20**, 273–297 (1995).
45. Wang, S.-C. Artificial Neural Network. in *Interdisciplinary Computing in Java Programming* 81–100 (Springer US, Boston, MA, 2003). doi:10.1007/978-1-4615-0377-4_5.
46. Awad, M. & Khanna, R. Deep Neural Networks. in *Efficient Learning Machines* 127–147 (Apress, Berkeley, CA, 2015). doi:10.1007/978-1-4302-5990-9_7.
47. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
48. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (ACM, San Francisco California USA, 2016). doi:10.1145/2939672.2939785.

3.5. DISCUSSÃO

A integração dos resultados obtidos nesta seção revela que o parasitismo vegetal é sustentado por um conjunto de mecanismos profundamente interconectados, nos quais evolução, defesa e regulação por moléculas móveis se entrelaçam para moldar diferentes graus de dependência em relação ao hospedeiro. O panorama traçado pelos três capítulos demonstra que compreender o parasitismo requer uma abordagem multiescalar, indo desde padrões genômicos amplos até interações moleculares altamente específicas, passando por transformações evolutivas que redefiniram a função de vias ancestrais.

O primeiro eixo dessa integração está na caracterização de lncRNAs em espécies com distintos graus de parasitismo. A análise conduzida no Capítulo 1 revela que lncRNAs constituem um componente expressivo e estruturalmente diverso dos transcriptomas de plantas parasitas, com padrões que se correlacionam não apenas ao estilo de vida, mas também a estruturas fundamentais como o haustório. A possível contribuição substancial de elementos LTR para sua origem reforça o papel da inovação transcricional e da plasticidade repetitiva na evolução desses organismos. Expandir de forma rápida e localizada, como demonstrado em diferentes plantas com transições ecológicas marcantes, esses retrotransposons podem remodelar regiões regulatórias, criar promotores e gerar transcritos inéditos, incluindo lncRNAs que emergem diretamente de cópias LTR degeneradas ou de fronteiras entre elementos móveis e genes ancestrais (Sabot; Schulman, 2006; Yang et al., 2019; Cai et al., 2021). Além disso, os indícios consistentes de transferência horizontal de lncRNAs em parasitas extremos, como *S. himalayana*, ampliam de forma significativa o espectro de moléculas capazes de circular entre parasita e hospedeiro, posicionando lncRNAs como potenciais agentes regulatórios na interface haustorial. A demonstração experimental de mobilidade bidirecional de lncRNAs em *Cuscuta* reforça a plausibilidade funcional dessa interação, sugerindo que o movimento de transcritos não codificadores pode contribuir tanto para modulação de vias metabólicas quanto para adaptação a diferentes hospedeiros (Wu et al., 2022).

O segundo eixo, apresentado no Capítulo 2, aprofunda a defesa vegetal como um arcabouço evolutivo amplo. A análise da origem e diversificação de proteínas de defesa em plantas não parasitas evidencia que a habilidade de reconhecer padrões moleculares e acionar respostas imunes já estava bem estabelecida e havia passado por múltiplas expansões antes mesmo do surgimento das linhagens parasitas atuais. No entanto, receptores do tipo PRR pertencentes à subfamília XII de RLKs, grupo no qual se insere AtCARD1 e seus homólogos CADLs, representam uma inovação evolutiva muito mais recente. Esses receptores surgem

apenas na base das estreptófitas, indicando que componentes essenciais de percepção extracelular associados à imunidade moderna foram adquiridos tardiamente e podem ter fornecido substrato molecular propício para posterior cooptação por plantas parasitas (Ngou et al., 2024). Módulos ancestrais como PRRs, vias de PTI e ETI, proteínas antimicrobianas e efetores endógenos constituíram um alicerce molecular sobre o qual diferentes plantas puderam modificar ou redirecionar funções, especialmente em contextos ecológicos novos e complexos. Embora o capítulo não trate diretamente de parasitas, o eixo discutido no Capítulo 3 indica que a maquinaria usada por parasitas para reconhecer quinonas e iniciar a formação do haustório provavelmente deriva de vias imunológicas e de desenvolvimento já presentes em plantas não parasitas. Esse cenário é consistente com modelos que propõem a cooptação de programas endógenos, como os envolvidos na formação de raízes laterais dependentes de auxina, que teriam sido gradualmente redirecionados para sustentar a organogênese haustorial (Tomilov et al., 2005; Yang et al., 2015; Ishida et al., 2016).

O Capítulo 3 por si só aprofunda o entendimento estrutural e funcional dessa cooptação. Ao empregar modelagem estrutural, *virtual screening* e ML para investigar AtCARD1 e seus homólogos, o estudo revela que as características estruturais que permitem o reconhecimento de quinonas são amplamente conservadas, especialmente elementos do domínio extracelular e do sítio coordenador de cobre, mesmo que a posição de ligação de HIFs pareça ser diferente entre AtCARD1 e CADLs. De qualquer forma, essa conservação sugere que a capacidade de detectar HIFs por receptores do tipo PRR tem uma base evolutiva comum, consistente com o cenário proposto no Capítulo 2. Além disso, a identificação de regiões alternativas de ligação em CADLs parasitas e um conjunto diferente de HIFs putativos quando comparados a AtCARD1 indica que esses receptores podem ter passado por adaptações específicas, ajustando-se a diferentes assinaturas químicas derivadas dos hospedeiros e ampliando a sensibilidade a moléculas associadas ao parasitismo.

Quando observamos o contexto genômico das plantas, a reutilização de repertórios gênicos pré-existentes para a aquisição de novas funções não representa um fenômeno isolado, mas sim um padrão recorrente ao longo da evolução vegetal (Bowman; Briginshaw; Florent, 2019). No caso das plantas parasitas, diversos estudos sugerem que genes associados ao desenvolvimento do haustório apresentam perfis de expressão semelhantes aos observados em raízes de espécies não parasitas, além de ortólogos que também são regulados em tecidos reprodutivos, como pólen e flores, sugerindo a reutilização de programas moleculares originalmente envolvidos em processos reprodutivos (Yang et al., 2015; Yoshida et al., 2016).

A presença desses homólogos em diferentes contextos biológicos indica que funções antes restritas a outros tecidos foram progressivamente redirecionadas e refinadas, permitindo que cópias gênicas assumissem novos papéis sem necessariamente perder suas funções originais, por meio de regulação diferencial ao longo do desenvolvimento. Nesse cenário, a emergência de genes relacionados ao haustório tem sido frequentemente associada a eventos de duplicação genômica completa (*Whole-Genome Duplications* - WGDs), que ampliaram o repertório gênico disponível para reutilização evolutiva, um processo recorrente em linhagens parasitas e discutido no Capítulo 1 (Xu et al., 2022).

Uma questão central que emerge dessas observações é se genes associados ao desenvolvimento do haustório mantêm suas funções ancestrais ou se passaram a atuar exclusivamente nesse novo contexto. Evidências indicam que muitos desses genes não perderam seus papéis originais, mas passaram a atuar em múltiplos cenários biológicos por meio de mudanças nos padrões de expressão e na regulação espacial e temporal, permitindo que processos relacionados ao remodelamento da parede celular, percepção de sinais e crescimento polarizado continuem ativos em raízes, tecidos reprodutivos e outros contextos de desenvolvimento, ao mesmo tempo em que são recrutados para a formação do haustório em espécies parasitas (Cosgrove, 2000; Honaas et al., 2013; Ranjan et al., 2014; Yang et al., 2015b; Yoshida et al., 2019). Esse redirecionamento funcional sugere que a evolução do parasitismo não depende necessariamente do surgimento de genes inteiramente novos, mas da reorganização de redes regulatórias capazes de integrar programas moleculares preexistentes a novos contextos fisiológicos e ecológicos.

Quando considerados em conjunto, os três capítulos reforçam que o parasitismo vegetal não pode ser compreendido apenas como resultado da perda da fotossíntese ou da dependência nutricional, mas como um fenômeno molecular complexo, no qual vias de defesa são reconfiguradas, novos reguladores emergem a partir da plasticidade genômica e moléculas móveis passam a atuar como mediadoras de comunicação entre organismos distintos podendo ser aderidos ao seu genoma definitivamente. Assim, a Seção I evidencia que a evolução das plantas parasitas envolveu simultaneamente simplificação metabólica e inovação funcional, sustentadas pela reutilização e reorganização de sistemas regulatórios ancestrais.

3.6. CONCLUSÃO

A partir da integração dos resultados apresentados, esta seção demonstra que o parasitismo vegetal emerge como um processo evolutivo multifacetado, sustentado pela reutilização e reconfiguração de mecanismos moleculares ancestrais em diferentes níveis biológicos. A caracterização de lncRNAs em espécies com distintos graus de parasitismo evidenciou a importância de elementos móveis e da plasticidade transcricional na geração de novas camadas regulatórias, enquanto a análise evolutiva de proteínas de defesa revelou que sistemas imunes pré-existentes forneceram o arcabouço funcional necessário para a percepção de sinais químicos associados ao hospedeiro. Complementarmente, a investigação estrutural de receptores do tipo PRR e a identificação computacional de potenciais moléculas indutoras reforçam a hipótese de que vias originalmente relacionadas à defesa e ao desenvolvimento foram cooptadas para sustentar a organogênese haustorial. Em conjunto, os capítulos indicam que a evolução das plantas parasitas não se limita à perda de funções metabólicas, mas envolve uma profunda reorganização de redes regulatórias, na qual duplicações gênicas, modulação espaço-temporal da expressão e a circulação de moléculas móveis contribuem para a emergência de novas estratégias ecológicas. Dessa forma, a Seção I consolida um modelo no qual o parasitismo vegetal resulta da integração entre inovação e conservação molecular, hipotetizando que a adaptação parasítica foi construída sobre a base de sistemas imunes e de desenvolvimento já presentes nas plantas ancestrais.

4. SEÇÃO II – TRANSPORTADORES METÁLICOS: DA CARACTERIZAÇÃO FUNCIONAL À EVOLUÇÃO DE SISTEMAS DE AQUISIÇÃO

4.1. REFERENCIAL TEÓRICO

A compreensão da evolução das vias de defesa e da comunicação molecular apresentada na Seção I revelou como pressões seletivas distintas podem remodelar profundamente redes proteicas ancestrais. O mesmo princípio evolutivo, baseado em modificação funcional, reutilização de módulos moleculares e inovação paralela em diferentes linhagens, também moldou as proteínas envolvidas na aquisição e no transporte interno de íons metálicos. O estudo desses sistemas torna possível compreender como plantas ajustaram sua fisiologia mineral diante de ambientes variáveis, revelando paralelos conceituais com os padrões evolutivos observados em parasitismo, defesa e regulação por moléculas móveis.

Metais essenciais como ferro (Fe), zinco (Zn), cobre (Cu), manganês (Mn) e cobalto (Co) são indispensáveis para o metabolismo vegetal, pois atuam como cofatores em enzimas, componentes estruturais em complexos fotossintéticos e respiratórios e elementos chave em vias redox, caminho metabólico baseado em reações de oxidação e redução (Mehra; Farago, 1994). Entretanto, a disponibilidade desses metais no solo é frequentemente limitada, sobretudo a do ferro, cuja forma oxidada predominante (Fe^{3+}) é pouco solúvel. Essa limitação impôs pressões seletivas intensas, levando plantas com flores a desenvolver sistemas distintos de aquisição de ferro, conhecidos como: (i) Estratégia I, típica de dicotiledôneas e gimnospermas, baseia-se na redução de Fe^{3+} para Fe^{2+} antes da absorção. Esse processo envolve a acidificação da rizosfera por bombas de prótons, a redução enzimática do ferro por redutases de membrana e a subsequente importação de Fe^{2+} por proteínas transportadoras da família ZIP, especialmente IRT1; (ii) Estratégia II, encontrada em gramíneas, que opera de modo distinto: plantas sintetizam moléculas quelantes chamadas fitossideróforos, que são liberadas no solo para capturar Fe^{3+} . O complexo Fe^{3+} -fitossideróforo é então importado por transportadores específicos como *Yellow Stripe-Like* (YSL) (Eide et al., 1996; Henriques et al., 2002; Marschner; Romheld; Kissel, 1986; Robinson et al., 1999; Vert et al., 2002, 2009).

Dentro da Estratégia I, transportadores ZIP/IRT são peças centrais. Estruturalmente, ZIPs são proteínas transmembrana com múltiplas hélices que formam um canal para a passagem de íons bivalentes (Guerinot, 2000; Vatansever; Özyiğit; FiLiZ, 2016). Possuem motivos conservados de coordenação metálica na região transmembrana e alças citosólicas ricas em histidinas, que contribuem para seletividade e estabilidade da ligação ao metal (Zhang et al., 2017). Pequenas mudanças nesses resíduos críticos podem alterar afinidade por ferro, zinco ou

manganês, mostrando como variações estruturais impactam a função (Ajeesh Krishna et al., 2020; Zhang et al., 2017). Estudos comparativos indicam que diferentes linhagens de plantas modificaram motivos estruturais de ZIPs e IRTs de forma paralela, sugerindo que capacidades similares de captação de ferro podem ter evoluído mais de uma vez ao longo da história das plantas verdes (Guerinot, 2000; Hall, 2003; Milner et al., 2013; Vatansever; Özyiğit; FiLiZ, 2016). Isso reforça a ideia de que pressões ambientais intensas, como solos pobres em ferro, favorecem múltiplas soluções moleculares equivalentes, mas evolutivamente independentes.

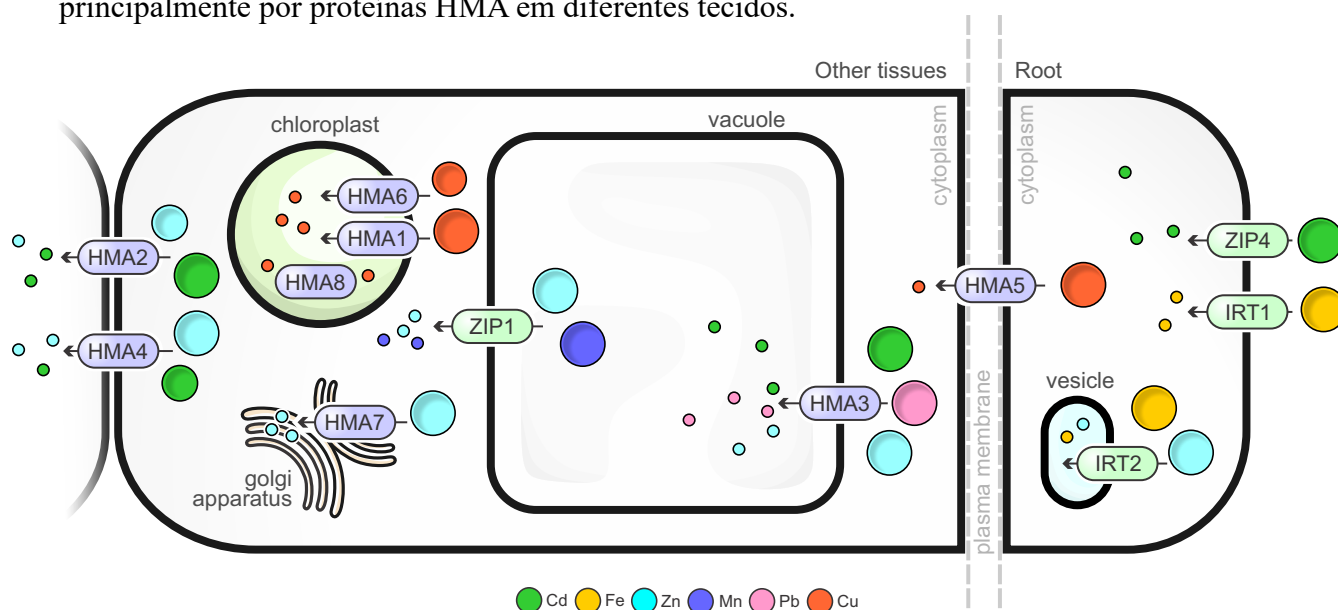
Após a absorção radicular, a distribuição interna de metais depende de sistemas ativos de transporte, entre os quais as P-tipo ATPases, também chamadas de HMAs, são fundamentais. Essas proteínas promovem o transporte direcionado de metais como Cu, Zn, cádmio (Cd), Co e chumbo (Pb) através de membranas utilizando energia derivada da hidrólise de ATP (Bandaranayake; Yoder, 2013). Estruturalmente, HMAs possuem um domínio citosólico ATPase responsável por fornecer energia, um domínio de ligação a metais formado por sítios de coordenação específicos e múltiplas hélices transmembrana que constituem o canal de translocação (Palmgren; Harper, 1999). A modulação desses domínios pode alterar tanto a especificidade pelo metal quanto o local de atuação da proteína, como transporte para o vacúolo, plastídios, citoplasma ou membrana plasmática (Hall, 2003).

A diversidade funcional dentro da família HMA reflete especializações evolutivas que atendem a necessidades fisiológicas distintas. Algumas HMAs participam da detoxificação, movendo metais tóxicos para organelas de armazenamento, enquanto outras garantem o fornecimento de cobre para plastídios ou zinco para proteínas estruturais e enzimáticas (Arnesano et al., 2002; Huffman; O'Halloran, 2001). Como ocorre com ZIPs, HMAs apresentam padrões de duplicação e especialização que variam entre linhagens vegetais, sugerindo que a pressão seletiva por homeostase metálica moldou repetidamente a evolução dessa família (Sánchez-Pérez et al., 2025). A estrutura modular das HMAs facilita esse processo, pois mutações em regiões regulatórias ou em domínios específicos podem gerar novas funções sem comprometer o mecanismo central de transporte.

Em uma escala evolutiva mais ampla, análises filogenômicas indicam que as HMAs fazem parte de uma linhagem antiga de ATPases metálicas cuja diversificação remonta aos primeiros eventos da evolução dos Archaeplastida (Hanikenne; Baurain, 2014). Estudos comparativos sugerem que a separação entre bombas especializadas no transporte de metais monovalentes e divalentes ocorreu muito cedo, possivelmente após duplicações gênicas ancestrais que precederam a diversificação das algas verdes e das plantas terrestres (Chan et al.,

2010; Hanikenne; Baurain, 2014). Posteriormente, eventos adicionais de duplicação e perda gênica em diferentes linhagens contribuíram para a expansão funcional observada em angiospermas, resultando em subgrupos com distribuição taxonômica distinta e adaptações específicas relacionadas à compartimentalização celular e à fisiologia dos plastídios (Hanikenne et al., 2008; Ó Lochlainn et al., 2011; Craciun et al., 2012). Esse histórico evolutivo reforça a ideia de que a diversidade funcional atual das HMAs não surgiu de maneira abrupta, mas foi construída gradualmente por meio da combinação de herança ancestral, eventos de endossimbiose e duplicações específicas de linhagem, criando um repertório modular capaz de sustentar diferentes estratégias de homeostase metálica ao longo da história das plantas.

Figura 5 – Homeostase de metais em plantas por proteínas HMA e ZIP. Modelo esquemático da homeostase de metais em células vegetais, ilustrando, por exemplo, a captação de ferro (Fe^{2+}) na raiz por transportadores da família IRT/ZIP e o transporte para compartimentos subcelulares, incluindo Complexo de Golgi, vacúolo e cloroplasto, mediado principalmente por proteínas HMA em diferentes tecidos.



Fonte: Autor da Pesquisa (2026), adaptado de Bashir et al. (2016); Wu et al. (2021); Khan et al. (2024).

Sob uma perspectiva comparativa mais ampla, a evolução de ZIP/IRT e de HMAs exemplifica como plantas solucionaram, ao longo do tempo, os desafios relacionados à aquisição e ao uso de metais essenciais. Tanto a existência de duas (ou mais) estratégias independentes de captação de ferro quanto a diversificação funcional de famílias de transportadores mostram que a nutrição mineral é um eixo de forte inovação evolutiva. Assim como visto na Seção I para vias de defesa e regulação por RNAs e até mesmo estilos de vidas de plantas parasitas, a história dessas proteínas revela que plantas ajustam suas redes

moleculares em resposta a limitações ambientais, reutilizando módulos ancestrais ou inovando quando necessário.

Nesta seção, são reunidos conceitos que aprofundam a compreensão da evolução e do funcionamento dos sistemas responsáveis pela aquisição e pela homeostase de metais em plantas. Discutimos as estratégias de captação de ferro e a diversidade das famílias ZIP e IRT, destacando que transportadores do tipo IRT1 emergiram independentemente em diferentes linhagens de plantas, constituindo um caso claro de convergência evolutiva. Examinamos também a família HMA, cuja arquitetura modular e variedade funcional ilustram como plantas ajustam mecanismos de transporte interno para equilibrar nutrição e evitar toxicidade.

4.2. CAPÍTULO 4 – FERROUS IRON UPTAKE VIA IRT1/ZIP EVOLVED AT LEAST TWICE IN GREEN PLANTS

Este Capítulo 4 apresenta o artigo “*Ferrous iron uptake via IRT1/ZIP evolved at least twice in green plants*”, publicado na revista *New Phytologist* (fator de impacto 8.1; DOI: 10.1111/nph.19036), no qual sou primeiro autor. O trabalho investiga a evolução dos transportadores *Iron-Regulated Transporters* (IRT) 1, um dos principais componentes da Estratégia I de aquisição de ferro em plantas, e demonstra que a capacidade de captar Fe^{2+} emergiu independentemente em diferentes linhagens de plantas verdes. Por meio de análises filogenômicas, predições funcionais e mapeamento de domínios críticos de coordenação metálica, o estudo revela que transportadores com propriedades equivalentes às de IRT1 surgiram por trajetórias evolutivas distintas, configurando um caso claro de convergência funcional. Além disso, o artigo contextualiza a distribuição de outros membros da família ZIP ao longo da história evolutiva das plantas, destacando como mudanças em regiões estruturais específicas, especialmente nas transmembranas e em alças citosólicas ricas em histidina, contribuíram para a aquisição ou perda de afinidade por ferro. Esses resultados fornecem um arcabouço evolutivo sólido para interpretar as estratégias de nutrição mineral em diferentes grupos vegetais e complementam a discussão geral desta Seção II, que aborda a diversidade estrutural e funcional de transportadores essenciais à homeostase de metais.

Viewpoints

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

Summary

Iron (Fe) is essential for virtually all organisms, being irreplaceable because of its electrochemical properties that enable many biochemical processes, including photosynthesis. Besides its abundance, Fe is generally found in the poorly soluble form of ferric iron (Fe^{3+}), while most plants uptake the soluble form Fe^{2+} . The model angiosperm *Arabidopsis thaliana*, for example, captures Fe through a mechanism that lowers rhizosphere pH through proton pumping that increases Fe^{3+} solubility, which is then reduced by a membrane-bound reductase and transported into the cell by the zinc-regulated, iron-regulated transporter-like protein (ZIP) family protein AtIRT1. ZIP proteins are transmembrane transporters of divalent metals such as Fe^{2+} , Zn^{2+} , Mn^{2+} , and Cd^{2+} . In this work, we investigated the evolution of functional homologs of *IRON-REGULATED TRANSPORTER 1*/ZIP in the supergroup Archaeplastida (Viridiplantae + Rhodophyta + Glaucophyta) using 51 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 by ZIP proteins evolved independently at least twice throughout green plant evolution. Our results indicate that the archetypical *IRON-REGULATED TRANSPORTER* (IRT) proteins from angiosperms likely emerged before the origin of land plants during early streptophyte algae terrestrialization, a process that required the evolution of Fe acquisition in terrestrial subaerial settings.

Introduction

Iron (Fe) is a fundamental element for virtually all cellular organisms. In plants, Fe deficiency results in serious physiological problems, such as failures in photosynthetic electron transport and sulfur and nitrogen metabolism, resulting in impaired growth (Crichton, 2016). Iron is not readily bioavailable, despite being one of the most abundant metals in Earth's crust. In soil, Fe is predominantly found as ferric iron (Fe^{3+}), which is poorly bioavailable to plants. In opposition, Fe^{2+} (ferrous iron) can be readily absorbed by plant roots. In freshwater environments, Fe can be found dissolved in greater abundance than in marine environments, where its limitation impacts photosynthetic productivity (Blomqvist *et al.*, 2004; Frey & Reed, 2012).

The great physiological importance that Fe has for plants and other living organisms is possibly explained by the role of Fe-based catalysts in early life, before the origin of extant cellular domains (Camprubi *et al.*, 2017). The decrease in Fe bioavailability was ultimately caused by the emergence of oxygenic photosynthesis in Cyanobacteria, which caused the transition of Earth's atmosphere from weakly reducing into oxidizing, a phenomenon known as the Great Oxidation Event (GOE) that happened between 2.4 and 2.2 billion years ago (Scott *et al.*, 2008; Wade *et al.*, 2021). The GOE started hundreds of millions of years after the emergence of oxygenic photosynthesis (Oliver *et al.*, 2021), most likely because initially the produced oxygen reacted with reduced compounds that acted as oxygen sinks preventing its accumulation in the atmosphere (Anbar *et al.*, 2007; Garcia-Pichel *et al.*, 2019). Throughout evolutionary time, organisms acquired specific mechanisms for Fe uptake that can sense and respond to soluble Fe in the environment. Land plants (Embryophyta) evolved at least two different Fe capture strategies. Based on Fe reduction, Strategy I is carried out by most plants, and Strategy II, based on chelation, seems to be specific to the Poaceae family (Römheld & Marschner, 1986; Connorton *et al.*, 2017).

In *Arabidopsis thaliana* Strategy I (Fig. 1a), root cells pump protons to the surrounding soil through H-ATPase 2 (AHA₂), decreasing the pH of the rhizosphere and increasing the ferric Fe hydroxide ($\text{Fe}(\text{OH})_3$) solubility. Ferric iron is then reduced by ferric reduction oxidase 2 (FRO₂) reductase to Fe^{2+} and transported into cells by the zinc-regulated, iron-regulated transporter-like protein (ZIP) transporter *IRON-REGULATED TRANSPORTER 1* (IRT1; Robinson *et al.*, 1999; Henriques *et al.*, 2002; Varotto *et al.*, 2002; Vert *et al.*, 2002; Santi & Schmidt, 2009). Low availability of Fe induces the synthesis of AHA₂, FRO₂, and IRT1 proteins that form complexes at the plasma membrane of the root epidermis cells (Martín-Barranco *et al.*, 2020). The acidification and reduction by AHA₂–FRO₂ occur rather poorly in alkaline soils, so this step can also be performed using Fe-mobilizing coumarins containing a catechol group such as fraxetin, sideretin, and esculetin, which have the ability to bind, solubilize, and reduce Fe from insoluble sources (Kai *et al.*, 2006; Rajniak *et al.*, 2018; Siwinska *et al.*, 2018; Tsai *et al.*, 2018; Robe *et al.*, 2021). The Feruloyl-CoA 6'-Hydroxylase 1 enzyme is responsible for the conversion of feruloyl-CoA to 6'-hydroxyferuloyl-CoA, which is the common precursor for the synthesis of coumarins secreted in the rhizosphere by ABCG transporter PLEIOTROPIC DRUG RESISTANCE 9 (Rodríguez-Celma *et al.*, 2013; Fourcroy *et al.*, 2014; Schmid *et al.*, 2014). Before their secretion into the rhizosphere, coumarins are deglycosylated by β -glucosidases (Clemens & Weber, 2016; Connorton *et al.*, 2017).

In Strategy II, root cells secrete phytosiderophores into the soil, which chelate $\text{Fe}(\text{OH})_3$ (Takagi *et al.*, 1984). Phytosiderophores are synthesized from S-adenosyl-methionine, generating compounds of

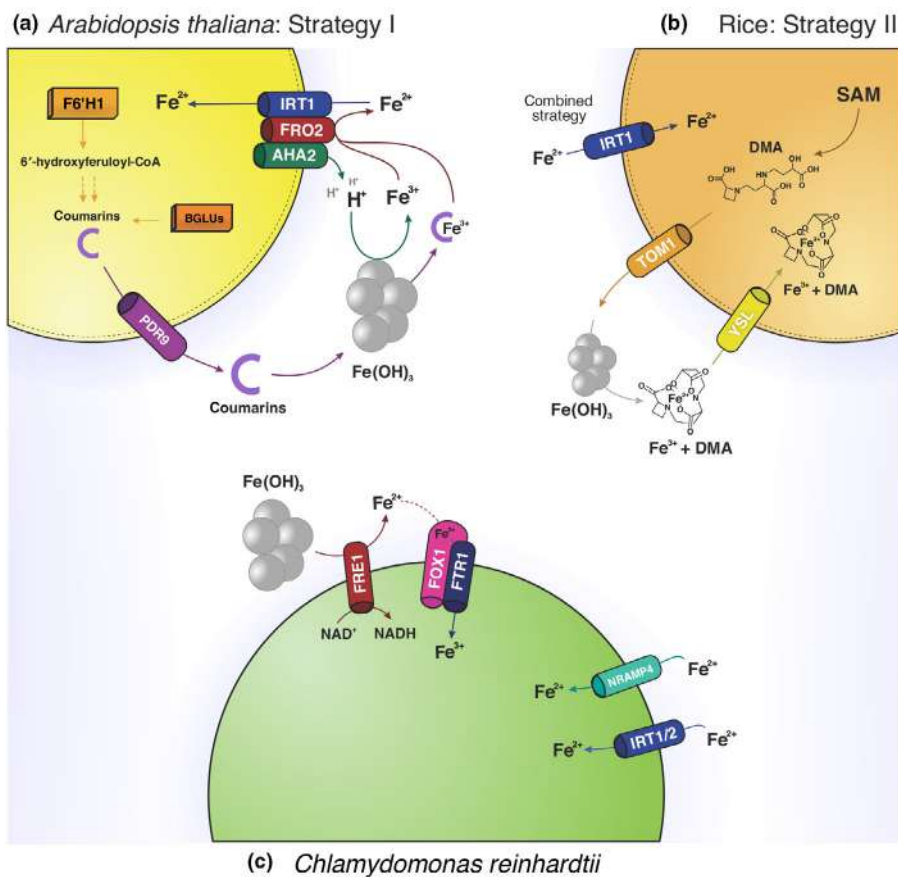


Fig. 1 Iron (Fe) uptake strategies in (a) *Arabidopsis thaliana* (Strategy I), (b) rice (Strategy II and *IRON-REGULATED TRANSPORTER 1* (IRT1)-dependent Fe²⁺ uptake), and (c) *Chlamydomonas reinhardtii*. Cylindrical shapes indicate membrane proteins involved in different steps of Fe capture. Blue shapes represent Fe transporters, and red shapes represent reductases. Orange boxes inside cells represent enzymes involved in the biosynthesis of coumarins. The arrows represent the synthesis and flux of coumarins and phytosiderophores, and the flux of iron and hydrogen ions. Arrow colors match the proteins involved in each step. Double dashed arrows between 6'-hydroxyferuloyl-CoA and coumarins indicate the involvement of multiple enzymes.

the mugineic acid family, such as 2'-deoxymugineic acid, that are secreted via transporter of mugineic acid family phytosiderophores 1 (Takagi *et al.*, 1984; Mori & Nishizawa, 1987; Cheng *et al.*, 2007; Nozoye *et al.*, 2011; Bashir *et al.*, 2017). Phytosiderophores form complexes with Fe³⁺ that are transported into the cell by Yellow Stripe 1-like proteins (Fig. 1b; Inoue *et al.*, 2009; Lee *et al.*, 2009). Species of the rice genus (*Oryza*) can combine Strategy II with direct uptake of Fe²⁺ via IRT1, in what is called the combined strategy (Fig. 1b). This is probably an adaptation of roots to flooded areas, where the lower O₂ levels causes Fe to be reduced to Fe²⁺ (Connorton *et al.*, 2017; Wairich *et al.*, 2019; Martín-Barranco *et al.*, 2021).

In the model chlorophyte *Chlamydomonas reinhardtii*, that was originally isolated from damp soils (Harris, 2001), Fe absorption occurs by reduction through FRE1 reductase (ferric/cupric reductase transmembrane component 1), followed by reoxidation by the putative multicopper oxidase FOX1. Iron is then transported into the cell by the FTR1 permease (high-affinity iron permease 1). The FOX1–FTR1 complex is a high-affinity Fe acquisition system that acquires Fe under very low concentrations. *Chlamydomonas* also directly captures Fe²⁺ through the ZIP transporters CrIRT1 and CrIRT2, and the metal ion transporter CrNRAMP4 (Natural Resistance-Associated Macrophage Protein 4; Fig. 1c; Allen *et al.*, 2007; Blaby-Haas & Merchant, 2012; Glaesener *et al.*, 2013; Kroh & Pilon, 2020; Martín-Barranco *et al.*, 2021).

IRON-REGULATED TRANSPORTER 1 is the first described and most well-known member of the large ZIP family of membrane transporters, identified from an *A. thaliana* cDNA capable of

complementing the Fe uptake function in yeast mutants (Eide *et al.*, 1996). Zinc-regulated, iron-regulated transporter-like proteins can transport a broad range of metal other than Fe, such as zinc (Zn) and manganese (Mn). In fact, in addition to *Arabidopsis* IRT1, the ZIP family is named after the yeast high-affinity Zn transporter zinc-regulated transporter 1 (Eide *et al.*, 1996; Zhao & Eide, 1996). Although there are not many studies on the physiological roles of ZIP genes in plants, complementation studies of yeast metal uptake mutants and *in planta* data indicate that ZIP proteins vary in metal transport ability. Zinc-regulated, iron-regulated transporter-like proteins specificity varies between Zn, Mn, or Fe, and some are able to transport more than one of these metals (Milner *et al.*, 2013; Zheng *et al.*, 2018; Pita-Barbosa *et al.*, 2019; Lee *et al.*, 2021). However, not all the members of the ZIP family are involved in plasma membrane metal uptake. The *Arabidopsis* AtIRT2 is an example of a ZIP protein that participates in Fe homeostasis by transporting Fe²⁺ into intracellular vesicles of root cells (Vert *et al.*, 2009). *IRON-REGULATED TRANSPORTER 1* from rice that functions in Fe uptake in roots is encoded by a gene closely related to IRT1 from *A. thaliana* suggesting that the IRT1-based Fe²⁺ uptake mechanism might be ancestral to angiosperms (Ricachenevsky & Sperotto, 2014).

Freshwater and terrestrial/subaerial environments have different Fe distribution and bioavailability. 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 Fe deficiency. Given that an IRT1-based mechanism for Fe^{2+} uptake is present in angiosperms and in chlorophyte *Chlamydomonas*, it was assumed that this mechanism shares a common evolutionary origin in green plants. Based on these questions, this study used a comparative genomics approach encompassing 51 high-quality genomes to better understand the origin and diversification of IRT1/ZIP proteins in Archaeplastida, the supergroup of photosynthetic eukaryotes that comprises the land plants, green and red algae, and the small group of the glaucophytes (Supporting Information Table S1; Fig. S1).

IRT1/ZIP gene family is present in all Archaeplastida lineages and expanded in land plants

In our analyses, we found 484 homolog ZIP proteins, each corresponding to a unique genomic *locus*, distributed across the 51 species of archaeplastidians in our dataset (Table S2; Methods S1). We detected IRT1/ZIP homologs not only in all lineages of green plants but also in glaucophytes and rhodophytes, which indicates that genes encoding IRT1/ZIP proteins in land plants were likely inherited vertically from the last common ancestor of Archaeplastida (Fig. 2a). Zinc-regulated, iron-regulated transporter-like proteins are present in all domains of life, so most likely ZIP genes were present in early eukaryotes before the divergence of archaeplastidians (Lehtovirta-Morley *et al.*, 2017). Our data suggest that archaeplastidial algae typically have < 10 ZIP genes (Fig. 2a). We separated the algae by habitat (marine, freshwater, or terrestrial/subaerial) and noticed a tendency of freshwater algae to display a higher number of ZIP genes when compared to terrestrial or oceanic (Fig. S2a), which could indicate an evolutionary advantage to have an increased repertoire of metal ion transporters. In freshwater environments, algae adopt specific approaches to Fe absorption, such as those described in the chlorophyte *Chlamydomonas* (Fig. 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 (Hoffmann *et al.*, 2012; Behnke & LaRoche, 2020). Understanding the availability of Fe in water is essential to uncover how the Fe capture mechanism in the first photosynthetic eukaryotes may have developed initially.

Among archaeplastidians, land plants generally have a significantly greater number of ZIP genes (Fig. 2a). However, we observed significant variability in the ZIP gene repertoire between land plant species, ranging from *c.* 10 to 30 genes per genome (Fig. 2a). Our data indicate that vascular plants (Tracheophyta) tend to have more ZIP genes than bryophytes (Bryophyta; Fig. 2a). It indicates that tracheophytes might have experienced increased fitness provided by the fixation of duplicated 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 the protein transporters.

The copy number of a gene family might be influenced by whole-genome duplications (WGD), making it difficult to compare deeply divergent organisms like green plants, red algae, and glaucophytes, that have experienced a variety of independent rounds of WGDs. 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. This enables us to detect whether ZIP genes are enriched in each lineage whole set of protein-coding genes. Results indicate that chlorophytes tend to dedicate a significantly greater fraction of their entire gene set to encode ZIP proteins when compared to red algae, charophytes, and land plants (Fig. 2b). The median frequency observed in our dataset of land plants, charophytes, and red algae is around four ZIP genes for each 10 000 protein-coding genes, while chlorophytes show a median around six ZIP genes for each 10 000 protein-coding genes (Fig. 2b). Bryophytes and tracheophytes have similar genomic frequency of ZIP genes (Fig. 2b), indicating that the increased number of ZIP genes observed in tracheophyte genomes (Fig. 2a) is coupled with an overall increase in the total number of protein-coding gene in this group. Although we accumulated knowledge in angiosperms such as *Arabidopsis* and rice, the molecular mechanisms of Fe acquisition are less well understood outside this group. Our results suggest, however, that ZIP genes might be important players in bryophytes. Experiments with the liverwort *Marchantia polymorpha* showed evidence for the use of Strategy I to capture Fe^{2+} through members of the ZIP family such as MpZIP3 (Mapoly0169s0010) and MpZIP5 (Mapoly0088s0041; Lo *et al.*, 2016). We did not observe a significant difference in ZIP genomic frequency regarding algal habitat (Fig. S2b), suggesting that no ZIP-specific gene amplification trend can be linked to the habitat in archaeplastidial algae. We also compared the number and genomic frequency of ZIP homologs between terrestrial and aquatic land plants. No significant difference was detected regarding the number of ZIP genes per genome or the genomic frequency of ZIP genes between terrestrial and aquatic embryophytes, suggesting that no specific trend in ZIP gene amplification can be linked to the habitat (Fig. S2c,d). Among chlorophytes, algae from the Class Chlorophyceae tend to have more ZIP genes per genome (Fig. S2e), although no differences regarding the relative genomic frequency of ZIP genes were found between the Chlorophyta Classes Chlorophyceae, Trebouxiophyceae, and Mamiellophyceae (Fig. S2f).

We then wondered whether the variation in the 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. Our results show a general trend for all Viridiplantae groups. Species with more protein-coding genes tend to have more ZIP copies (Fig. 2c), indicating that independent expansions in the whole gene space of different taxa involved the fixation of duplicated copies of ZIP genes. Thus, the expanded set of ZIP observed in some species might be the result of gene fixation after lineage-specific WGD events (Fig. 2c). The correlation seems to be weaker in red algae and the single nuclear complete genome available for glaucophytes prevent any further conclusion. Our data support the hypothesis that expansions in genome size *per se* might not explain the observed

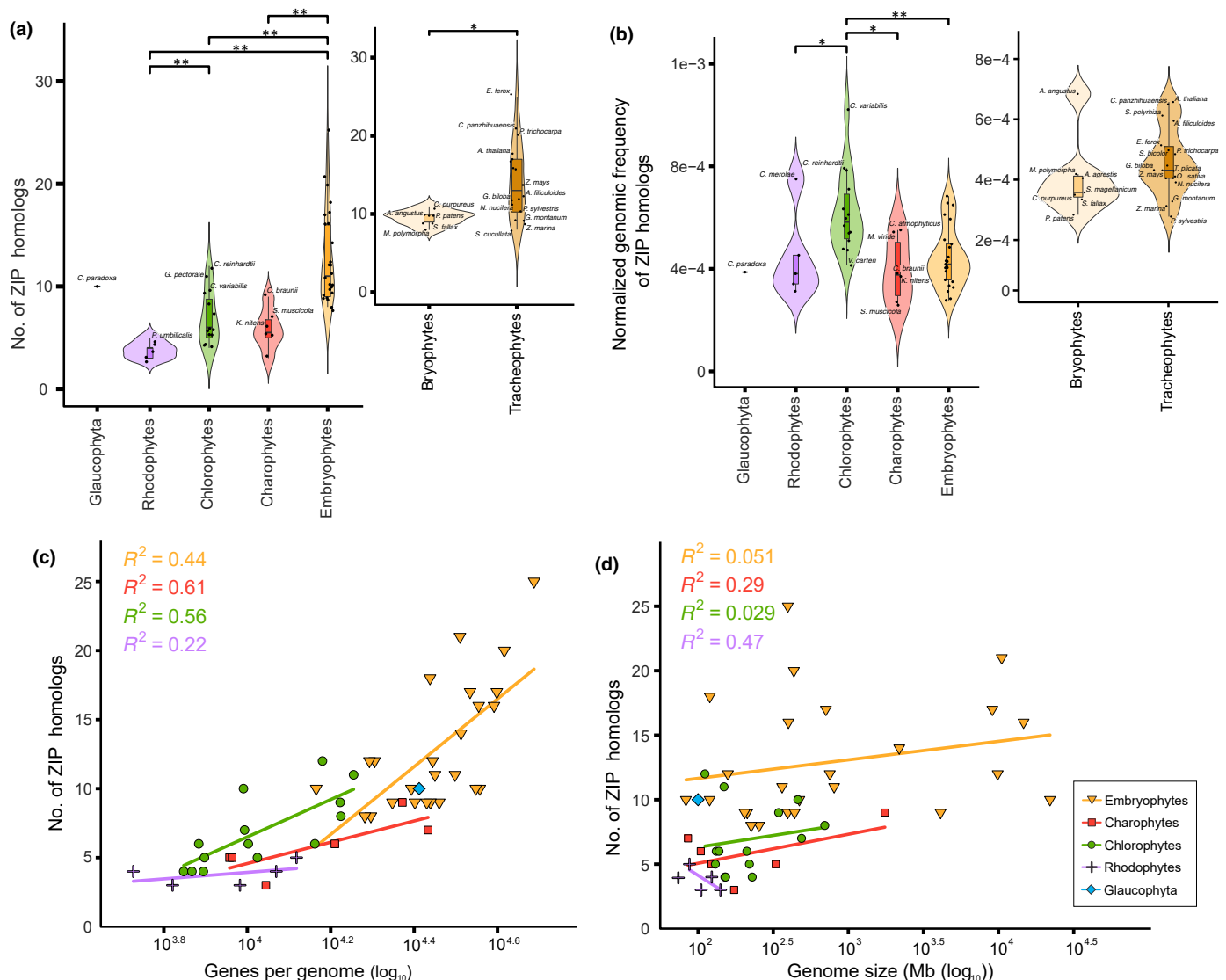


Fig. 2 Distribution of zinc-regulated, iron-regulated transporter-like protein (ZIP) genes across archaeplastidians. Violin plots with integrated boxplot diagram showing: (a) distribution of the number of ZIP homologs found per genome for diverse taxonomic groups; (b) distribution of the relative genomic frequency of ZIP genes per genome (normalized by the total number of protein-coding genes per genome). Brackets indicate significant differences between medians with asterisks representing the *P*-values (Wilcoxon test: **, *P* < 0.01; *, *P* < 0.05). The lower and upper boundaries of the boxplot represent the 1st and 3rd quartiles with the median represented as a line across the box. Boxplot whiskers represent 1.5 times the interquartile range above the upper quartile and below the lower quartile. Scatter plots showing the correlations between the number of ZIP genes per genome and (c) the total number of protein-coding genes per genome and (d) genome size. Linear regression lines are shown for each taxa with matching colors.

variation in ZIP genes between or within groups of green plants (Fig. 2d). However, we detected a moderate negative correlation between the number of ZIP genes and the genome size in red algae (Fig. 2d).

Ferrous iron uptake via IRT1/ZIP proteins evolved at least twice in the kingdom of green plants (Viridiplantae)

To better understand the phylogenetic relationship between IRT1/ZIP proteins, we constructed a phylogenetic tree (Figs 3, S3) using the 484 unique ZIP genes identified across our sample of 51 genomes. The resulting phylogenetic tree can be subdivided into

two deeply divergent major clades, herein called X and Y (Fig. 3a). Interestingly, there was no clustering among the IRT1 proteins from land plants and IRT1/2 from the chlorophyte *Chlamydomonas*. The well-characterized IRT1/ZIP proteins involved in Fe²⁺ uptake from the angiosperms *Arabidopsis* and rice, and from the liverwort *Marchantia* (MpZip3 and MpZIP5) all belong to Clade X, while *Chlamydomonas* IRT1 and IRT2 (herein called CrIRT1 and CrIRT2, respectively) belong to Clade Y (Fig. 3a). MpZIP3 is phylogenetically closer to the archetypal angiosperm IRT proteins from *Arabidopsis* and rice in the subclade X1, while MpZIP5 belongs to the subclade X2 that is more distantly related to *Arabidopsis* and rice IRTs (Fig. 3a). It indicates that divergent ZIP proteins might be recruited into Strategy I in land plants, possibly

due to low specificity to metal transport and fast changes in regulatory networks that may include new ZIP genes in the early transcriptional response to environmental Fe depletion. We also performed a sequence similarity network (SSN) analysis, that is based in pairwise protein identity, with the same 484 protein sequences that yielded similar results, confirming the distant relationship between Clades X and Y, and recovering the same structure of subclades (Fig. 3b). These results suggest that the mechanism of ferrous Fe acquisition via IRT1/ZIP proteins evolved at least twice in green plants from deeply divergent ancestral genes of the ZIP family.

Angiosperm subclade Y1 proteins AtZTP29 (AT3G20870) and OsZIP13 (LOC_Os02g10230) from *Arabidopsis* and rice, respectively, and subclade Y2 AtZIP13 (AT3G08650) from *Arabidopsis* are most likely Zn transporters (Wang *et al.*, 2010; Ivanov & Bauer, 2017; Zheng *et al.*, 2018). It suggests that *Chlamydomonas* CrIRT1/2 are more closely related to Zn transporters proteins from angiosperms than to their archetypical Fe²⁺ uptake IRT proteins (Figs 3, S3). The *Arabidopsis* proteins from Clade X AtZIP1

(AT3G12750; subclade X1) and AtZIP2 (AT5G59520; subclade X2) are involved in the uptake of Zn or Mn, or both (Milner *et al.*, 2013; Zheng *et al.*, 2018). OsZIP5 (LOC_Os05g39560) and OsZIP9 (LOC_Os05g39540) proteins from subclade X1 function in Zn uptake from the soil in rice roots (Huang *et al.*, 2020; Tan *et al.*, 2020; Yang *et al.*, 2020). The closely related subclade X1 proteins AtIRT3 (AT1G60960), AtZIP4 (AT1G10970), and AtZIP9 (AT4G33020) from *Arabidopsis* and OsZIP7 (LOC_Os05g10940) from rice are involved in Zn xylem loading and root to shoot translocation (Ricachenevsky *et al.*, 2018; Tan *et al.*, 2019; Gindri *et al.*, 2020; Lee *et al.*, 2021). This suggests that the IRT1 from angiosperms like *Arabidopsis* and rice are more closely related to proteins acting in Zn or Mn transport than to the functional homologous in Fe²⁺ uptake CrIRT1 and CrIRT2 from *Chlamydomonas*.

These observations indicate that even deeply divergent ZIP proteins can evolve to fulfill the functional role of IRT1 in Fe²⁺ uptake first discovered in *Arabidopsis*. Zinc-regulated, iron-regulated transporter-like proteins might have a fast evolutionary

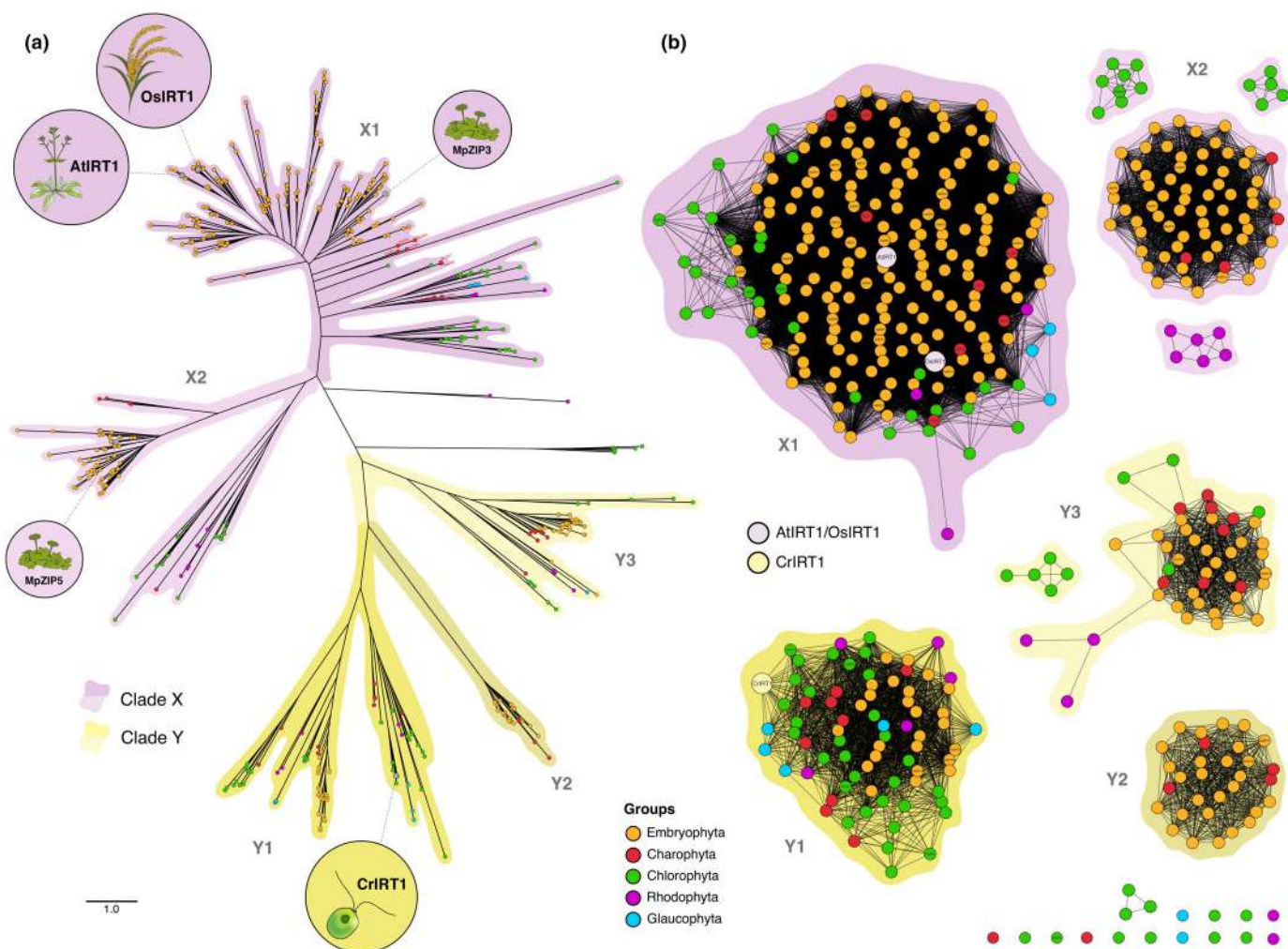


Fig. 3 Maximum-likelihood and sequence similarity network analyses of the *IRON-REGULATED TRANSPORTER 1* (IRT1)/ zinc-regulated, iron-regulated transporter-like protein (ZIP) family in Archaeplastida. (a) Maximum-likelihood phylogenetic tree highlighting Clade X and Clade Y and their subclades X1, X2, Y1, Y2, and Y3. Proteins discussed in the text are highlighted. We also performed a Bayesian phylogenetic analysis using the same alignment that yielded similar results (Supporting Information Fig. S4). See Methods S1 for a brief description. (b) Sequence similarity network (SSN) showing clustering of proteins in agreement with the phylogenetic analysis. Color scheme reflects the taxonomic classification of species.

turnover between different metal acquisition and transport mechanisms due to an overall low metal selectivity coupled with transcriptional network rewiring of ZIP genes (Nocedal & Johnson, 2015).

IRT1/ZIP protein involved in Fe²⁺ uptake have high degree of sequence plasticity

Previous works have demonstrated that the amino acid residues aspartic acid (D) at positions 108 and 144, serine (S) at position 206, and histidine (H) at position 232 of the *Arabidopsis* IRT1 are critical for Fe transport (Fig. 4a; Rogers *et al.*, 2000; Quintana *et al.*, 2022). If those residues are substituted with alanine (A), the protein loses its ability to transport Fe. These residues are also essential for the secondary transport of Zn in addition to glutamic acid (E) at position 111, which abolishes the transport of Zn when substituted with an A without affecting the transport of Fe (Fig. 4a; Rogers *et al.*, 2000). These results suggest that small changes in ZIP proteins can affect metal selectivity. However, if we align the CrIRT1 amino acid sequence with IRT1 from *Arabidopsis* and rice (Fig. 4b), no conservation in three of the four critical positions is observed, suggesting that IRT proteins might exhibit some degree of sequence plasticity. Whether the presence of an A in position 111 in the CrIRT1 is related to Zn transport activity should be better investigated with functional studies (Fig. 4b).

Our results suggest that Strategy I of land plants evolved based on Clade X IRT/ZIP 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 subclade X1: g37664 (CbZIP1) and g42085 (CbZIP2) from *Chara braunii* (Charophyceae), kfl00040_0120 (KnZIP1), and kfl00394_0040 (KnZIP2) from *Klebsormidium nitens* (Klebsormidiophyceae), SM000007S20791 (SmZIP1) and SM000022S07262 (SmZIP2) from *Spirogloea muscicola* (Zygnematophyceae), and Chrsp27S04275 (CaZIP1) and Chrsp157S02258 (CaZIP2) from *Chlorokybus atmophyticus* (Chlorokybophyceae; Fig. 4c). We observed D108 position residue conservation in distinct charophyte lineages related to *Arabidopsis* IRT1. However, most charophyte IRT1-related proteins have an E residue in position 144 (Fig. 4c). Interestingly, E and D residues are both acidic amino acids, with similar structure and chemical properties, indicating that the substitution D144E might not affect Fe transport. We also observed conservation in the essential amino acids S206 and H232, which are reported to be critical for its *in planta* function of high-affinity Fe²⁺ transport activity (Quintana *et al.*, 2022), in all analyzed charophyte proteins (Fig. 4c). These results suggest that the charophyte proteins from subclade X1 closely related to IRT1 from *Arabidopsis* and rice are compatible with Fe²⁺ acquisition in Strategy I and its role in Fe homeostasis. It opens the possibility that Strategy I might have evolved during early land colonization by early charophytes before the origin of land plants. Functional studies on the physiology of Fe acquisition by charophytes, especially with terrestrial/subaerial 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 (AtZTP29, AtZIP13, OsZIP13, OsZIP16, and OsZIP11), together with IRT1 from *Arabidopsis* and rice from Clade X (Fig. 4d). No apparent sequence conservation was observed in the critical residues, except for H232, suggesting that Clade Y and Clade X proteins might have relevant amino acid sequence differences and still be able to function in the transport of the same metal ions.

Distantly related IRT1/ZIP proteins involved in Fe²⁺ uptake are structurally convergent

To further support the hypothesis that divergent ZIP proteins fulfill the role of Fe²⁺ uptake performed by IRT1 in angiosperms, we modeled the three-dimensional structure of the known ferrous iron uptake proteins from land plants (AtIRT1, OsIRT1, and MpZIP3 from subclade X1 and MpZIP5 from subclade X2) and chlorophyte *Chlamydomonas* (CrIRT1 from subclade Y1). We also included in the analysis the Clade X charophyte proteins CbZIP1, KnZIP1, SmZIP1, and CaZIP1 from subclade X1 and the Clade Y putative Zn transporters OsZIP13 and AtZTP29 from subclade Y1 (Table S3). The predicted structures were aligned in pairs, allowing us to estimate the structural similarity between proteins.

The pairwise root-mean-square deviation (RMSD) values suggest that Clade X proteins are structurally conserved (Fig. 5a). The angiosperm proteins AtIRT1 and OsIRT1 have highly similar predicted structures, with an RMSD value of 0.53 Å (lower values indicate higher structural similarity). Interestingly, the charophyte proteins from subclade X1 have a high degree of structural similarity with AtIRT1, with pairwise RMSD values ranging between 0.79 and 1.68 Å (Fig. 5a,b), suggesting that they might be functionally conserved. The IRT/ZIP proteins from the soil charophytes *Klebsormidium* (KnZIP1) and *Spirogloea* (SmZIP1) showed remarkable structural conservation with AtIRT1 (RMSD = 0.79 and 1.08 Å, respectively), suggesting that Strategy I Fe uptake via IRT1 might have evolved in the Streptophyta lineage before the origin of land plants, during the adaptation of early charophytes to the Fe acquisition in terrestrial settings.

Surprisingly, the predicted structure of the Fe transporter CrIRT1 from the chlorophyte *Chlamydomonas* (subclade Y1) showed more similarity to structures from the Clade X, such as the Fe transporters OsIRT1 (RMSD = 2.2 Å) and AtIRT1 (RMSD = 3.8 Å; Fig. 5a,c), than to the phylogenetically related putative Zn transporters OsZIP13 (RMSD = 4.3 Å) and AtZTP29 (RMSD = 4.6 Å) from its own subclade Y1 (Fig. 5a,d). These results indicate that, although CrIRT1 and AtIRT1/OsIRT1 are only distantly related in terms of ZIP family phylogeny, as revealed by their primary sequence of amino acids (Fig. 3), they evolved to some degree of structure convergence which might be related to the Fe²⁺ transport activity. While the predicted structure of the Fe transporter MpZIP3 (subclade X1) from the liverwort *Marchantia* is highly similar to other IRT1/ZIP proteins from the subclade X1 (RMSD ranging from 0.7 to 1.6 Å), MpZIP5 (subclade X2) structure shows a more moderate structural conservation with proteins from both clades Y (RMSD with AtZTP29 = 2.2 Å from subclade Y1) and X (RMSD with AtIRT1 = 2.3 Å from subclade X1; Fig. 5a,e). These results

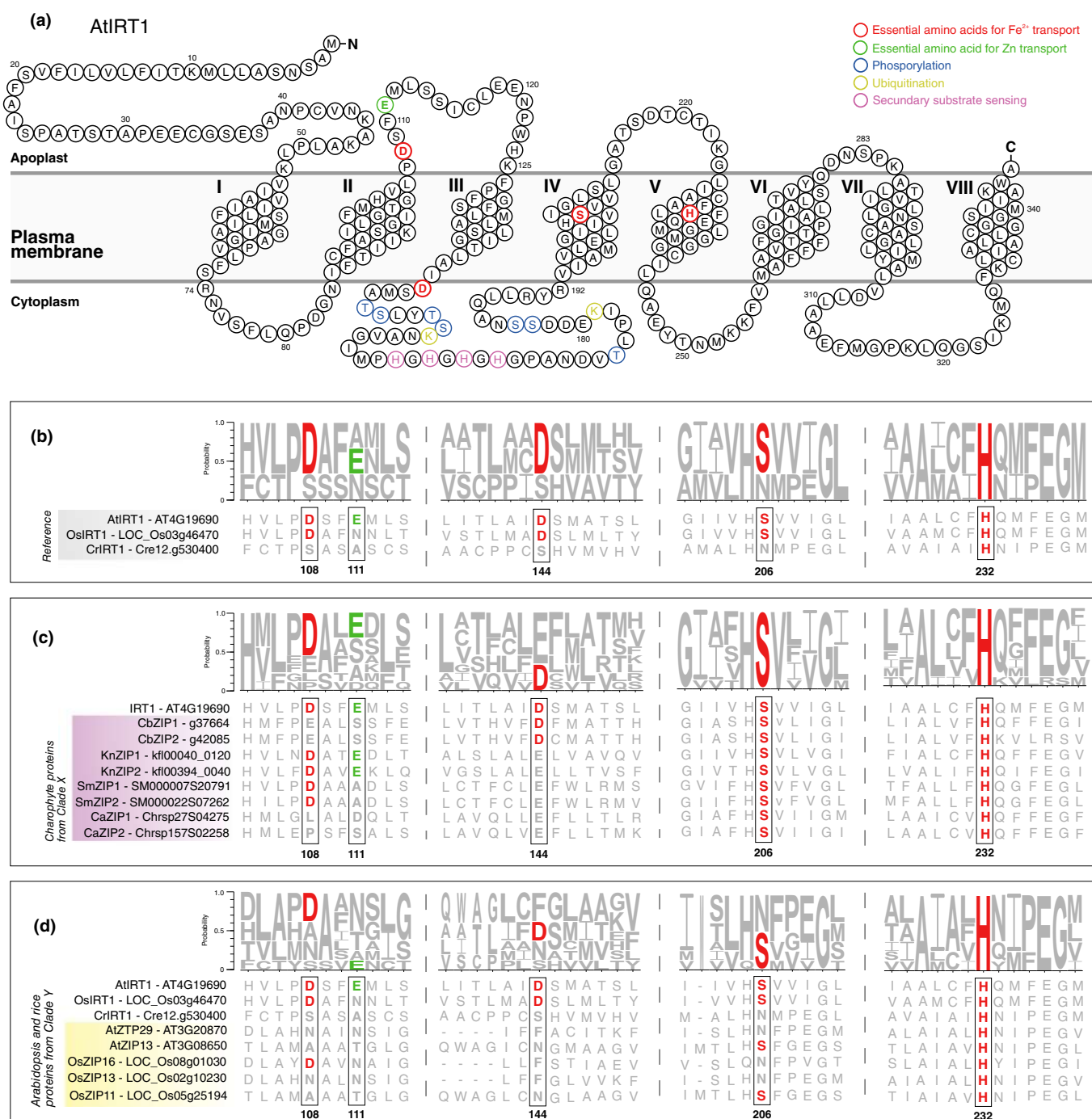


Fig. 4 Multiple sequence alignments and logo sequence comparison of AtIRT1, OsIRT1, and CrIRT1 with their closest related proteins. (a) Topological model of AtIRT1 highlighting the essential amino acid residues for Fe²⁺ transport (red), Zn transport (green), cytosolic phosphorylation (blue) and ubiquitination (yellow), and H residues acting in secondary substrate sensing (purple). (b–d) Alignments showing the Asp-106, Asp-144, Ser-206, and His-232 (red) essential amino acids for Fe transport and Glu-111 (green) essential residue for Zn transport. The size of residues within a stack indicates the relative frequency of each amino acid per position of the selected proteins. Protein amino acid numbering is related to AtIRT1. (b) Alignment between *IRON-REGULATED TRANSPORTER 1* (IRT1) sequences from *Arabidopsis thaliana*, rice, and *Chlamydomonas reinhardtii*. (c) Charophyte ZIP proteins from Clade X aligned with the AtIRT1. (d) Arabidopsis and rice ZIP proteins from Clade Y aligned with AtIRT1, OsIRT1, and CrIRT1.

indicate that ZIP proteins involved in the acquisition of Fe²⁺ can tolerate some degree of structural variations and still be functional.

All Clade X and Y protein models showed a cytosolic loop of variable length (Fig. 5b–e) that is consistently found between transmembrane helices III and IV, as seen for AtIRT1 (Fig. 4a;

Table S4). CrIRT1 is the only protein analyzed where the cytosolic loop is found between transmembrane helices V and VI (Table S4). Specific amino acids in this loop are target of post-translational protein modifications in *Arabidopsis* (Fig. 4a), being involved in protein ubiquitination, internalization, and degradation, as part of

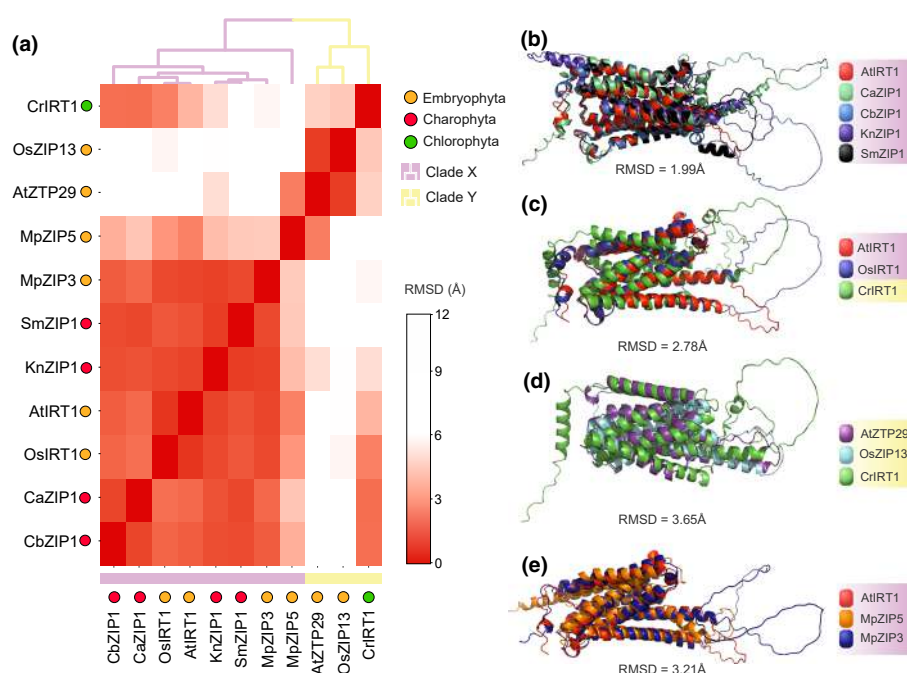


Fig. 5 IRON-REGULATED TRANSPORTER 1 (IRT1)/zinc-regulated, iron-regulated transporter-like protein (ZIP) three-dimensional structure analysis. (a) Heatmap showing the structural similarity of pairwise alignments between selected IRT1/ZIP proteins from Clade X and Clade Y. The structural similarity is indicated by the root-mean-square deviation (RMSD) (Å) value, where the lower value (red) indicates more structural similarity. (b–e) Aligned structural models of (b) CrIRT1, AtIRT1, and OsIRT1; (c) AtIRT1 and its closest related charophyte ZIP proteins from subclade X1; (d) CrIRT1, AtZTP29, and OsZIP13 from subclade Y1; and (e) MpZIP5 from subclade X2, MpZIP3, and AtIRT1 from subclade X1.

the mechanism of control of Fe homeostasis (Barberon *et al.*, 2011; Quintana *et al.*, 2022). We were unable to find apparent conservation of such amino acids between the modeled IRT1/ZIP proteins, indicating that the loop sequence might be variable (Table S4). In fact, the amino acid sequences of the loop portion are significantly more variable than the rest of the protein (Fig. S5), suggesting that this regulatory part of the protein might evolve fast between IRT1/ZIP homologs. Our results indicate that post-translational protein modifications controlling IRT1/ZIP protein turnover are likely variable between taxa.

Concluding remarks and future perspectives

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 critical amino acids for Fe transport in the *Arabidopsis* IRT1 protein seem to have evolved before the split of land plants from streptophytic algae. In fact, structural modeling suggests that putative IRT1 proteins from charophytes are highly conserved with the archetypical angiosperm IRT1 counterparts, indicating that the IRT1-dependent Fe uptake typically found among land plants might have originated in early terrestrial charophytes before the origin of the first embryophytes. Functional data on Fe acquisition of soil charophytes (such as *Klebsormidium*, *Spirogloea*, and *Mesotaenium*) might help understand how land plants' ancestors were able to first obtain Fe in terrestrial settings with subaerial exposure and direct contact with the substrate. On the contrary, IRT1/2 from the chlorophyte *Chlamydomonas* has a divergent protein sequence regarding key residues known in *Arabidopsis*, suggesting that multiple primary amino acid sequences of ZIP proteins might be fitted to transport Fe^{2+} . This might be explained, at least partially, by the evolution of

a structural convergence between distantly related IRT1/ZIP proteins.

The phylogenetic diversity of IRT1/ZIP proteins used in Fe uptake by green plants indicate that network 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 metal transporters, suggesting that much evolutionary diversity is still to be found in Fe uptake in less studied groups such as charophytes, chlorophytes, and red algae.

Acknowledgements

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) – Finance Code 001, that provided a master's and PhD scholarship to WFCR. We want to thank the anonymous referees that helped us improve this manuscript.

Author contributions

L-ED-B designed the research. WFCR carried out the sequence data collection, analyses, and visualizations. ABPL contributed to protein three-dimensional modeling and structural analyses. L-ED-B, WFCR, FKR and JEL contributed to data interpretation and manuscript structure. WFCR and L-ED-B wrote the manuscript. All authors participated in manuscript editing and approved the final version.

Competing interests

None declared.

ORCID

Luiz-Eduardo Del-Bem  <https://orcid.org/0000-0001-8472-4476>

Joni Esrom Lima  <https://orcid.org/0000-0001-5436-6081>

Ayrton Breno P. Lisboa  <https://orcid.org/0000-0002-4344-1100>

Felipe Klein Ricachenevsky  <https://orcid.org/0000-0001-5429-3759>

Wenderson Felipe Costa Rodrigues  <https://orcid.org/0000-0002-6167-0900>

Data availability

All data used in this study are publicly available and can be accessed through the links provided in Table S1.

Wenderson Felipe Costa Rodrigues^{1,2,3} ,
Ayrton Breno P. Lisboa^{1,3} , **Joni Esrom Lima**⁴ ,
Felipe Klein Ricachenevsky^{5,6}  and
Luiz-Eduardo Del-Bem^{1,3,4,7*} 

¹Graduate Program in Bioinformatics, Institute of Biological Sciences (ICB), Federal University of Minas Gerais (UFMG), 31270-901 Belo Horizonte, Brazil;

²Graduate Program in Plant Biology, Department of Botany, Institute of Biological Sciences (ICB), Federal University of Minas Gerais (UFMG), 31270-901 Belo Horizonte, Brazil;

³Del-Bem Lab, Department of Botany, Institute of Biological Sciences (ICB), Federal University of Minas Gerais (UFMG), 31270-901 Belo Horizonte, Brazil;

⁴Department of Botany, Institute of Biological Sciences (ICB), Federal University of Minas Gerais (UFMG), 31270-901 Belo Horizonte, Brazil;

⁵Department of Botany, Institute of Biosciences (IB), Federal University of Rio Grande do Sul (UFRGS), 91501-900 Porto Alegre, Brazil;

⁶Graduate Program in Cellular and Molecular Biology, Center for Biotechnology, Federal University of Rio Grande do Sul (UFRGS), 91501-900 Porto Alegre, Brazil;

⁷MSU-DOE Plant Research Laboratory, Michigan State University, East Lansing, MI 48824, USA
(*Author for correspondence: email delbem@ufmg.br)

References

- Allen MD, del Campo JA, Kropat J, Merchant SS. 2007. *FEA1*, *FEA2*, and *FRE1*, encoding two homologous secreted proteins and a candidate ferriredutase, are expressed coordinately with *FOX1* and *FTR1* in iron-deficient *Chlamydomonas reinhardtii*. *Eukaryotic Cell* 6: 1841–1852.
- Anbar AD, Duan Y, Lyons TW, Arnold GL, Kendall B, Creaser RA, Kaufman AJ, Gordon GW, Scott C, Garvin J *et al.* 2007. A whiff of oxygen before the great oxidation event? *Science* 317: 1903–1906.
- Barberon M, Zelazny E, Robert S, Conéjéro G, Curie C, Friml J, Vert G. 2011. Monoubiquitin-dependent endocytosis of the IRON-REGULATED TRANSPORTER 1 (IRT1) transporter controls iron uptake in plants. *Proceedings of the National Academy of Sciences, USA* 108: E450–8.
- Bashir K, Nozoye T, Nagasaka S, Rasheed S, Miyauchi N, Seki M, Nakanishi H, Nishizawa NK. 2017. Paralogs and mutants show that one DMA synthase functions in iron homeostasis in rice. *Journal of Experimental Botany* 68: 1785–1795.
- Behnke J, LaRoche J. 2020. Iron uptake proteins in algae and the role of Iron Starvation-Induced Proteins (ISIPs). *European Journal of Phycology* 55: 339–360.
- Blaby-Haas CE, Merchant SS. 2012. The ins and outs of algal metal transport. *Biochimica et Biophysica Acta (BBA) – Molecular Cell Research* 1823: 1531–1552.
- Blomqvist S, Gunnars A, Elmgren R. 2004. Why the limiting nutrient differs between temperate coastal seas and freshwater lakes: a matter of salt. *Limnology and Oceanography* 49: 2236–2241.
- Camprubi E, Jordan SF, Vasiliadou R, Lane N. 2017. Iron catalysis at the origin of life. *IUBMB Life* 69: 373–381.
- Cheng L, Wang F, Shou H, Huang F, Zheng L, He F, Li J, Zhao F-J, Ueno D, Ma JF *et al.* 2007. Mutation in nicotianamine aminotransferase stimulated the Fe(II) acquisition system and led to iron accumulation in rice. *Plant Physiology* 145: 1647–1657.
- Clemens S, Weber M. 2016. The essential role of coumarin secretion for Fe acquisition from alkaline soil. *Plant Signaling & Behavior* 11: e1114197.
- Connorton JM, Balk J, Rodriguez-Celma J. 2017. Iron homeostasis in plants – a brief overview. *Metallomics* 9: 813–823.
- Crichton RR. 2016. *Iron metabolism: from molecular mechanisms to clinical consequences*. Chichester, West Sussex, UK: John Wiley & Sons.
- Eide D, Broderius M, Fett J, Guerinot ML. 1996. A novel iron-regulated metal transporter from plants identified by functional expression in yeast. *Proceedings of the National Academy of Sciences, USA* 93: 5624–5628.
- Fourcroy P, Sisó-Terraza P, Sudre D, Savirón M, Rey G, Gaymard F, Abadía A, Abadía J, Álvarez-Fernández A, Briat J. 2014. Involvement of the ABCG 37 transporter in secretion of scopoletin and derivatives by *Arabidopsis* roots in response to iron deficiency. *New Phytologist* 201: 155–167.
- Frey PA, Reed GH. 2012. The ubiquity of iron. *ACS Chemical Biology* 7: 1477–1481.
- García-Pichel F, Lombard J, Soule T, Dunaj S, Wu SH, Wojciechowski MF. 2019. Timing the evolutionary advent of Cyanobacteria and the later great oxidation event using gene phylogenies of a sunscreen. *mBio* 10: e00561-19.
- Gindri RG, Navarro BB, da Cruz Dias PV, Tarouco CP, Nicoloso FT, Brunetto G, Berghetti ÁLP, da Silva LOS, Fett JP, Menguer PK *et al.* 2020. Physiological responses of rice (*Oryza sativa* L.) *oszip7* loss-of-function plants exposed to varying Zn concentrations. *Physiology and Molecular Biology of Plants* 26: 1349–1359.
- Glaesener AG, Merchant SS, Blaby-Haas CE. 2013. Iron economy in *Chlamydomonas reinhardtii*. *Frontiers in Plant Science* 4: 337.
- Harris EH. 2001. *Chlamydomonas* as a model organism. *Annual Review of Plant Physiology and Plant Molecular Biology* 52: 363–406.
- Henriques R, Jásik J, Klein M, Martinoia E, Feller U, Schell J, Pais MS, Koncz C. 2002. Knock-out of *Arabidopsis* metal transporter gene IRT1 results in iron deficiency accompanied by cell differentiation defects. *Plant Molecular Biology* 50: 587–597.
- Hoffmann L, Breitbarth E, Boyd P, Hunter K. 2012. Influence of ocean warming and acidification on trace metal biogeochemistry. *Marine Ecology Progress Series* 470: 191–205.
- Huang S, Sasaki A, Yamaji N, Okada H, Mitani-Ueno N, Ma JF. 2020. The ZIP transporter family member OsZIP9 contributes to root zinc uptake in rice under zinc-limited conditions. *Plant Physiology* 183: 1224–1234.
- Inoue H, Kobayashi T, Nozoye T, Takahashi M, Kakei Y, Suzuki K, Nakazono M, Nakanishi H, Mori S, Nishizawa NK. 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. *Journal of Biological Chemistry* 284: 3470–3479.
- Ivanov R, Bauer P. 2017. Sequence and coexpression analysis of iron-regulated ZIP transporter genes reveals crossing points between iron acquisition strategies in green algae and land plants. *Plant and Soil* 418: 61–73.
- Kai K, Shimizu B, Mizutani M, Watanabe K, Sakata K. 2006. Accumulation of coumarins in *Arabidopsis thaliana*. *Phytochemistry* 67: 379–386.

- Kroh GE, Pilon M. 2020. Regulation of iron homeostasis and use in chloroplasts. *International Journal of Molecular Sciences* 21: 3395.
- Lee S, Chiecko JC, Kim SA, Walker EL, Lee Y, Guerinot ML, An G. 2009. Disruption of *OsYSL15* leads to iron inefficiency in rice plants. *Plant Physiology* 150: 786–800.
- Lee S, Lee J, Ricachenevsky FK, Punshon T, Tappero R, Salt DE, Guerinot ML. 2021. Redundant roles of four ZIP family members in zinc homeostasis and seed development in *Arabidopsis thaliana*. *The Plant Journal* 108: 1162–1173.
- Lehtovirta-Morley L, Alsarraf M, Wilson D. 2017. Pan-domain analysis of ZIP zinc transporters. *International Journal of Molecular Sciences* 18: 2631.
- Lo J, Tsednee M, Lo Y, Yang S, Hu J, Ishizaki K, Kohchi T, Lee D, Yeh K. 2016. Evolutionary analysis of iron (Fe) acquisition system in *Marchantia polymorpha*. *New Phytologist* 211: 569–583.
- Martín-Barranco A, Spielmann J, Dubeaux G, Vert G, Zelazny E. 2020. Dynamic control of the high-affinity iron uptake complex in root epidermal cells. *Plant Physiology* 184: 1236–1250.
- Martín-Barranco A, Thomine S, Vert G, Zelazny E. 2021. A quick journey into the diversity of iron uptake strategies in photosynthetic organisms. *Plant Signaling & Behavior* 16: 1975088.
- Milner MJ, Seamon J, Craft E, Kochian LV. 2013. Transport properties of members of the ZIP family in plants and their role in Zn and Mn homeostasis. *Journal of Experimental Botany* 64: 369–381.
- Mori S, Nishizawa N. 1987. Methionine as a dominant precursor of phytosiderophores in Gramineae plants. *Plant and Cell Physiology* 28: 1081–1092.
- Nocedal I, Johnson AD. 2015. How transcription networks evolve and produce biological novelty. *Cold Spring Harbor Symposia on Quantitative Biology* 80: 265–274.
- Nozoye T, Nagasaka S, Kobayashi T, Takahashi M, Sato Y, Uozumi N, Nakanishi H, Nishizawa NK. 2011. Phytosiderophore efflux transporters are crucial for iron acquisition in graminaceous plants. *Journal of Biological Chemistry* 286: 5446–5454.
- Oliver T, Sánchez-Baracaldo P, Larkum AW, Rutherford AW, Cardona T. 2021. Time-resolved comparative molecular evolution of oxygenic photosynthesis. *Biochimica et Biophysica Acta (BBA) – Bioenergetics* 1862: 148400.
- Pita-Barbosa A, Ricachenevsky FK, Wilson M, Dottorini T, Salt DE. 2019. Transcriptional plasticity buffers genetic variation in zinc homeostasis. *Scientific Reports* 9: 19482.
- Quintana J, Bernal M, Scholle M, Holländer-Czytko H, Nguyen NT, Piotrowski M, Mendoza-Cózatl DG, Haydon MJ, Krämer U. 2022. Root-to-shoot iron partitioning in *Arabidopsis* requires IRON-REGULATED TRANSPORTER1 (IRT1) protein but not its iron(II) transport function. *The Plant Journal* 109: 992–1013.
- Rajniak J, Giehl RFH, Chang E, Murgia I, von Wiren N, Sattely ES. 2018. Biosynthesis of redox-active metabolites in response to iron deficiency in plants. *Nature Chemical Biology* 14: 442–450.
- Ricachenevsky FK, Punshon T, Lee S, Oliveira BHN, Trenz TS, dos Maraschin FS, Hindt MN, Danku J, Salt DE, Fett JP *et al.* 2018. Elemental profiling of rice FOX lines leads to characterization of a new Zn plasma membrane transporter, OsZIP7. *Frontiers in Plant Science* 9: 865.
- Ricachenevsky FK, Sperotto RA. 2014. There and back again, or always there? The evolution of rice combined strategy for Fe uptake. *Frontiers in Plant Science* 5: 189.
- Robe K, Izquierdo E, Vignols F, Rouached H, Dubos C. 2021. The coumarins: secondary metabolites playing a primary role in plant nutrition and health. *Trends in Plant Science* 26: 248–259.
- Robinson NJ, Procter CM, Connolly EL, Guerinot ML. 1999. A ferric-chelate reductase for iron uptake from soils. *Nature* 397: 694–697.
- Rodríguez-Celma J, Lin W-D, Fu G-M, Abadia J, Lopez-Millan A-F, Schmidt W. 2013. Mutually exclusive alterations in secondary metabolism are critical for the uptake of insoluble iron compounds by *Arabidopsis* and *Medicago truncatula*. *Plant Physiology* 162: 1473–1485.
- Rogers EE, Eide DJ, Guerinot ML. 2000. Altered selectivity in an *Arabidopsis* metal transporter. *Proceedings of the National Academy of Sciences, USA* 97: 12356–12360.
- Römhelt V, Marschner H. 1986. Evidence for a specific uptake system for iron phytosiderophores in roots of grasses. *Plant Physiology* 80: 175–180.
- Santi S, Schmidt W. 2009. Dissecting iron deficiency-induced proton extrusion in *Arabidopsis* roots. *New Phytologist* 183: 1072–1084.
- Schmid NB, Giehl RFH, Doll S, Mock H-P, Strehmel N, Scheel D, Kong X, Hider RC, von Wiren N. 2014. Feruloyl-CoA 6'-Hydroxylase1-dependent coumarins mediate iron acquisition from alkaline substrates in *Arabidopsis*. *Plant Physiology* 164: 160–172.
- Scott C, Lyons TW, Bekker A, Shen Y, Poulton SW, Chu X, Anbar AD. 2008. Tracing the stepwise oxygenation of the Proterozoic ocean. *Nature* 452: 456–459.
- Siwinska J, Siatkowska K, Olry A, Grosjean J, Hehn A, Bourgaud F, Meharg AA, Carey M, Lojkowska E, Ichnatowicz A. 2018. Scopoletin 8-hydroxylase: a novel enzyme involved in coumarin biosynthesis and iron-deficiency responses in *Arabidopsis*. *Journal of Experimental Botany* 69: 1735–1748.
- Takagi S, Nomoto K, Takemoto T. 1984. Physiological aspect of mugineic acid, a possible phytosiderophore of graminaceous plants. *Journal of Plant Nutrition* 7: 469–477.
- Tan L, Qu M, Zhu Y, Peng C, Wang J, Gao D, Chen C. 2020. ZINC TRANSPORTER5 and ZINC TRANSPORTER9 function synergistically in zinc/cadmium uptake. *Plant Physiology* 183: 1235–1249.
- Tan L, Zhu Y, Fan T, Peng C, Wang J, Sun L, Chen C. 2019. OsZIP7 functions in xylem loading in roots and inter-vascular transfer in nodes to deliver Zn/Cd to grain in rice. *Biochemical and Biophysical Research Communications* 512: 112–118.
- Tsai H-H, Rodríguez-Celma J, Lan P, Wu Y-C, Vélez-Bermúdez IC, Schmidt W. 2018. Scopoletin 8-hydroxylase-mediated fraxetin production is crucial for iron mobilization. *Plant Physiology* 177: 194–207.
- Varotto C, Maiwald D, Pesaresi P, Jahns P, Salamini F, Leister D. 2002. The metal ion transporter IRT1 is necessary for iron homeostasis and efficient photosynthesis in *Arabidopsis thaliana*: Knock-out of IRT genes. *The Plant Journal* 31: 589–599.
- Vert G, Barberon M, Zelazny E, Séguéla M, Briat J-F, Curie C. 2009. *Arabidopsis* IRT2 cooperates with the high-affinity iron uptake system to maintain iron homeostasis in root epidermal cells. *Planta* 229: 1171–1179.
- Vert G, Grotz N, Dédaldéchamp F, Gaymard F, Guerinot ML, Briat J-F, Curie C. 2002. IRT1, an *Arabidopsis* transporter essential for iron uptake from the soil and for plant growth. *Plant Cell* 14: 1223–1233.
- Wade J, Byrne DJ, Ballentine CJ, Drakesmith H. 2021. Temporal variation of planetary iron as a driver of evolution. *Proceedings of the National Academy of Sciences, USA* 118: e2109865118.
- Wairich A, de Oliveira BHN, Arend EB, Duarte GL, Ponte LR, Sperotto RA, Ricachenevsky FK, Fett JP. 2019. The combined strategy for iron uptake is not exclusive to domesticated rice (*Oryza sativa*). *Scientific Reports* 9: 16144.
- Wang M, Xu Q, Yu J, Yuan M. 2010. The putative *Arabidopsis* zinc transporter ZTP29 is involved in the response to salt stress. *Plant Molecular Biology* 73: 467–479.
- Yang M, Li Y, Liu Z, Tian J, Liang L, Qiu Y, Wang G, Du Q, Cheng D, Cai H *et al.* 2020. A high activity zinc transporter OsZIP9 mediates zinc uptake in rice. *The Plant Journal* 103: 1695–1709.
- Zhao H, Eide D. 1996. The yeast ZRT1 gene encodes the zinc transporter protein of a high-affinity uptake system induced by zinc limitation. *Proceedings of the National Academy of Sciences, USA* 93: 2454–2458.
- Zheng X, Chen L, Li X. 2018. *Arabidopsis* and rice showed a distinct pattern in ZIPs genes expression profile in response to Cd stress. *Botanical Studies* 59: 22.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Phylogeny of Archaeplastida.

Fig. S2 Distribution of ZIP genes in algal and plant species of different classes and habitats.

Fig. S3 Expanded maximum-likelihood phylogenetic tree of ZIP family in Archaeplastida.

Fig. S4 Bayesian phylogenetic tree of ZIP family in Archaeplastida.

Fig. S5 Pairwise p-distances of the primary amino acid sequences of the modeled proteins.

Methods S1 Methods for identification of IRT1/ZIP, sequence alignment, phylogenetic and sequence similarity network analyses, and protein structural analyses.

Table S1 List of complete predicted proteome databases used in this study.

Table S2 Table showing the number of ZIP family homologs found in each species of our dataset.

Table S3 IRT1/ZIP used in the structural modeling.

Table S4 Cytosolic loops identified in the modeled IRT1/ZIP.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

Key words: Archaeplastida, comparative genomics, iron uptake, IRT1, ZIP transporters.

Received, 14 September 2022; accepted, 29 November 2022.

New Phytologist Supporting Information

Article Title:

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

Authors:

Wenderson Felipe Costa Rodrigues, Ayrton Breno P. Lisboa, Joni Esrom Lima, Felipe Klein Ricachenevsky, Luiz-Eduardo Del-Bem.

Article acceptance date: 29 November 2022

Short legends for Supporting Information

Supplemental Methods 1. Methods for identification of IRT1/ZIP proteins, sequence alignment, phylogenetic and sequence similarity network (SSN) analyses, and protein structural analyses.

Supplemental Figure 1. Phylogeny of Archaeplastida.

Supplemental Figure 2. Distribution of ZIP genes in algal and plant species of different classes and habitats.

Supplemental Figure 3. Expanded maximum-likelihood phylogenetic tree of ZIP family in Archaeplastida.

Supplemental Figure 4. Bayesian phylogenetic tree of ZIP family in Archaeplastida.

Supplemental Figure 5. Pairwise p-distances of the primary amino acid sequences of the modeled proteins.

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

Supplemental Table 2. Table showing the number of ZIP family homologs found in each species of our dataset.

Supplemental Table 3. IRT1/ZIP proteins used in the structural modeling.

Supplemental Table 4. Cytosolic loops identified in the modeled IRT1/ZIP proteins.

Supplemental Methods 1. Methods for identification of IRT1/ZIP proteins, sequence alignment, phylogenetic and sequence similarity network (SSN) analyses, and protein structural analyses.

Datasets

The HMM (hidden Markov model) profile for Zip family (PF02535) was obtained from the Pfam database v35.0 (<https://pfam.xfam.org/>) (Mistry *et al.*, 2021). The complete predicted proteomes of the analyzed organisms were collected from the databases detailed in Supplemental Table 1.

Identification of ZIP family homologs

We performed a HMMER v3.2.1 (Eddy, 2011) search with the ZIP HMM profile PF02535 to identify ZIP protein homologs in our dataset (e-value cutoff $< E^{-10}$). Functional well-characterized IRT1/ZIP proteins such as AtIRT1 and OsIRT1 contain only a single ZIP protein domain. To verify the basic structure of the proteins retrieved, they were scanned by PfamScan (<https://www.ebi.ac.uk/Tools/pfa/pfamscan/>) against Pfam-A.hmm database (v35.0) with an e-value cutoff of E^{-10} to search for the presence of additional protein domains. 484 out of 498 proteins (~97.2%) had only a single ZIP protein domain identified and were selected for the subsequent analyses. Six proteins (~1.2%) showed at least one positive hit with an additional domain other than ZIP (ChlNC64A_1.52390 and ChlNC64A_1.59363 from *Chlorella variabilis*, Coc_C169_1.49037 from *Coccomyxa subellipsoidea*, Picsp_1.460 from *Picochlorum soloecismus*, Zosma03g34380 from *Zostera marina*, and Cypan_014662 from *Cycas panzhihuaensis*) and eight proteins (~1.6%) that had a duplicate ZIP domain (Cyapar1_16293_g16035 from *Cyanophora paradoxa*, Gonpec1.5823 from *Gonium pectorale*, Cre13.g576050 from *Chlamydomonas reinhardtii*, Vocar.0001s0160 from *Volvox carteri*, kfl00274_0050 from *Klebsormidium nitens*, SM000001S04682 and SM000010S04341 from *Spirogloea muscicola*, and Cypan_025071 from *Cycas panzhihuaensis*) were discarded. We kept only the longest protein isoform with full functional domain encoded by each unique gene. Through an in-house script written in Perl, we discarded partial sequences or pseudogenes by eliminating protein sequences shorter than 138 aa, which represents 40% of the length of the well-characterized IRT1 protein from *Arabidopsis thaliana*. Also, we collapsed 100% redundant sequences with *skipredundant* from EMBOSS v6.6.0 package (Rice *et al.*, 2000).

All subsequent analyses and plots were performed with the 484 filtered protein sequences using Rstudio v2022.07.2 (RStudio Team, 2020). The relative genomic frequency of *ZIP* genes per genome was calculated by dividing the number of *ZIP* homologs found in each genome by the total number of protein-coding genes in the correspondent genome. 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 v7.47 (<https://mafft.cbrc.jp/alignment/software/>) (Katoh, 2002) with the following parameters -*-thread 10 --reorder --leavegappyregion --maxiterate 1000 --retree 1 --localpair*. The alignment output containing 4.056 positions (including all gaps) was exported in Phylip format (*--phylipout*) for subsequent phylogenetic analyses. We also aligned our sequences using MUSCLE in MEGA 11 software (Kumar *et al.*, 2018) to analyze the conservation of amino acids between proteins and calculate the genetic *p*-distances shown in Supplemental Figure 5. The amino acid conservation plots shown in Figure 4 were generated by WebLogo 3 (<http://weblogo.threeplusone.com/>) (Crooks *et al.*, 2004), 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 (Nguyen *et al.*, 2015). The amino acid substitution model was selected using the SMS method with Akaike Information Criterion (AIC) as the selection criterion. The best-fit model according to AIC was WAG+F+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 (Guindon *et al.*, 2010) with 1000 replicates. We also performed a Bayesian phylogenetic analysis, with PhyloBayes v4.1 under the CAT-GRT model (Lartillot *et al.*, 2009). Both phylogenetic trees were rooted at the midpoint and plotted using FigTree v1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>).

Sequence Similarity Network (SSN)

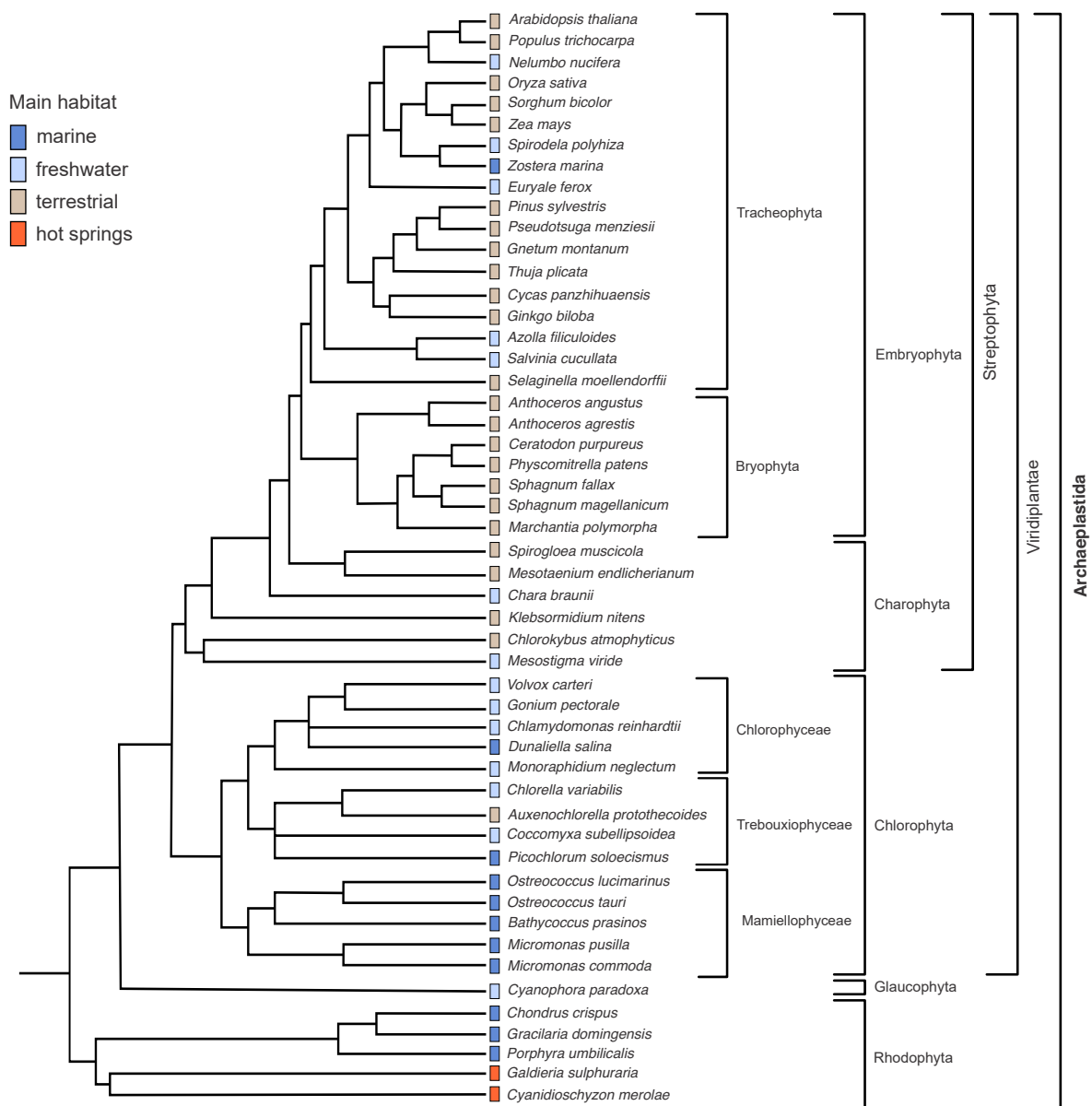
To corroborate our phylogenetic results, we performed a network analysis based on pairwise sequence similarity. We generated networks with an alignment score threshold $\geq 25\%$ sequence identity in the Enzymatic Similarity Tool (EFI-EST tool) (<https://efi.igb.illinois.edu/efi-est/>) (Zallot *et al.*, 2019). After that, we visualize and edit the network with Cytoscape v3.8.2 software (Shannon *et al.*, 2003), keeping only edges connecting proteins with identity $\geq 30\%$.

Protein three-dimensional structure modeling and analyses

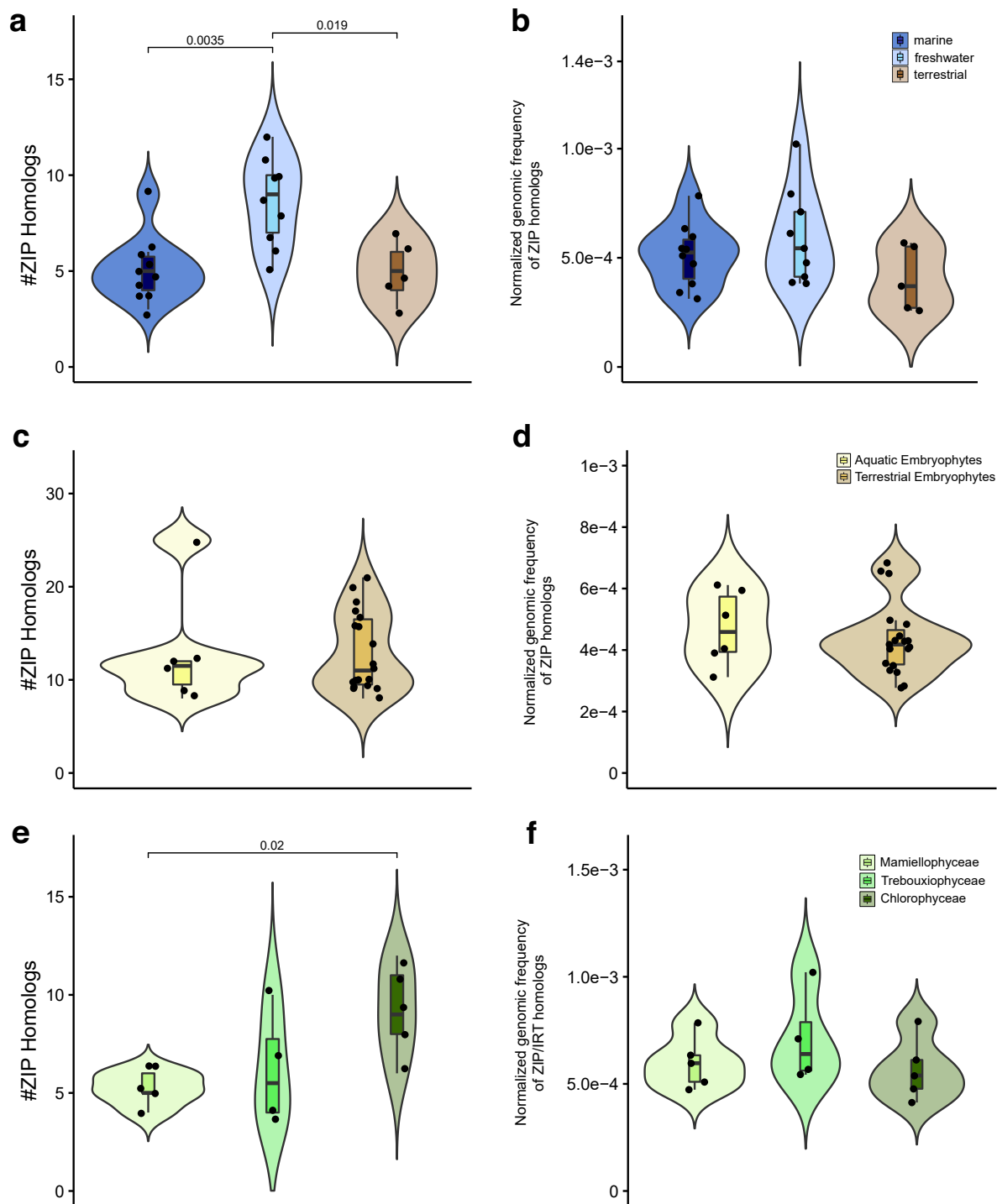
Signal peptide was predicted before three-dimensional structure modeling using Phobius (Kall *et al.*, 2007) and if a signal peptide was predicted, the correspondent sequence was removed prior to modeling (Supplemental Table 3). Protein structure modeling was performed using AlphaFold (Jumper *et al.*, 2021), with default parameters. PDB files underwent sidechain energy minimization on FoldIt Standalone (Kleffner *et al.*, 2017), which uses RosettaCM algorithms (Kuhlman *et al.*, 2003), to resolve sidechain atomic clashes on the models. PDB files were then visualized on PyMOL (Schrödinger, 2015).

To assess the root-mean-square deviation (RMSD) of alpha carbon (C α) atomic positions, we aligned the three-dimensional models with RaptorX Structure Alignment Server DeepAlign (Wang *et al.*, 2013) for three or more structures. For pairwise alignment, structures were aligned on PyMOL.

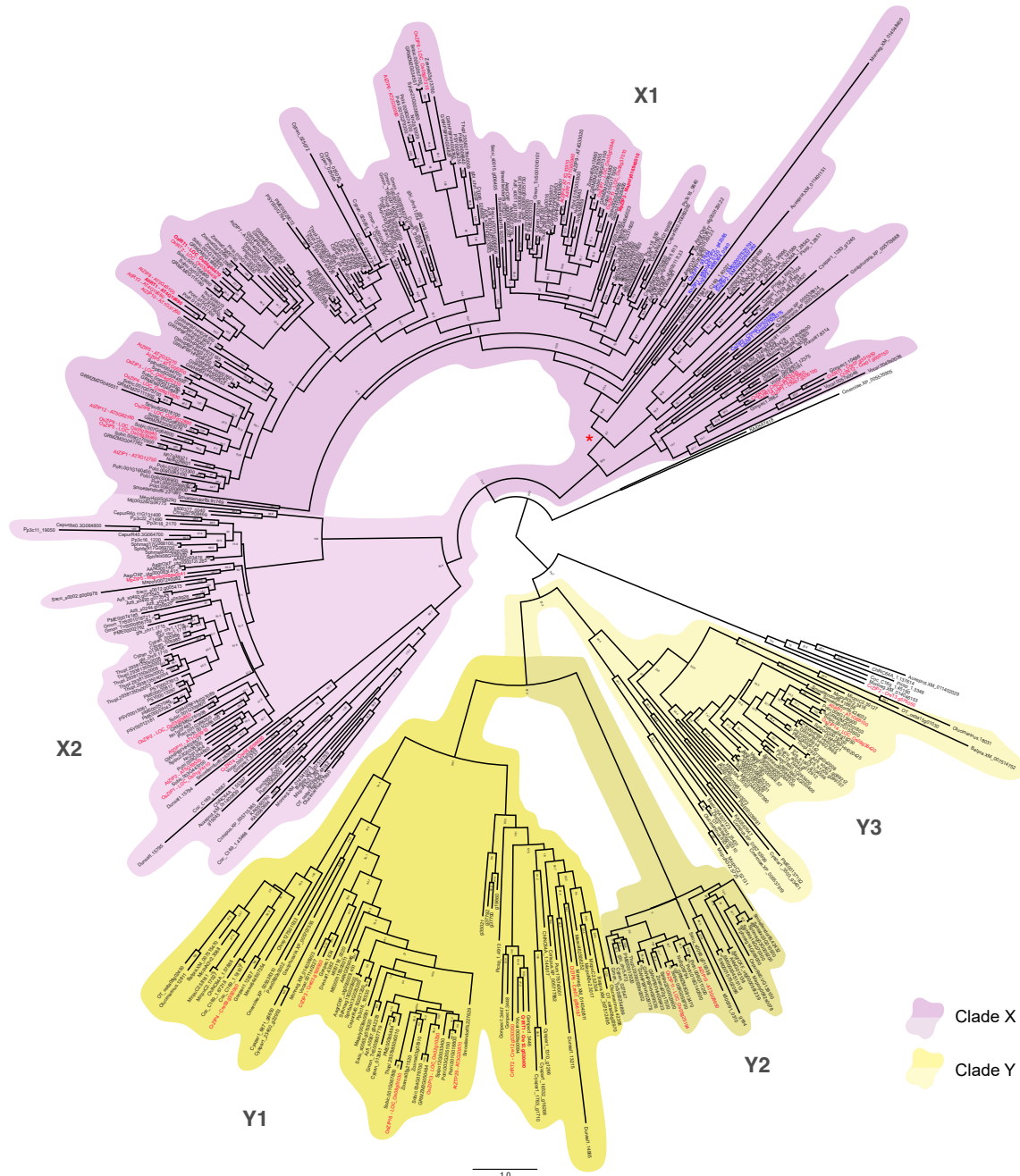
Supplemental Figures 1-5



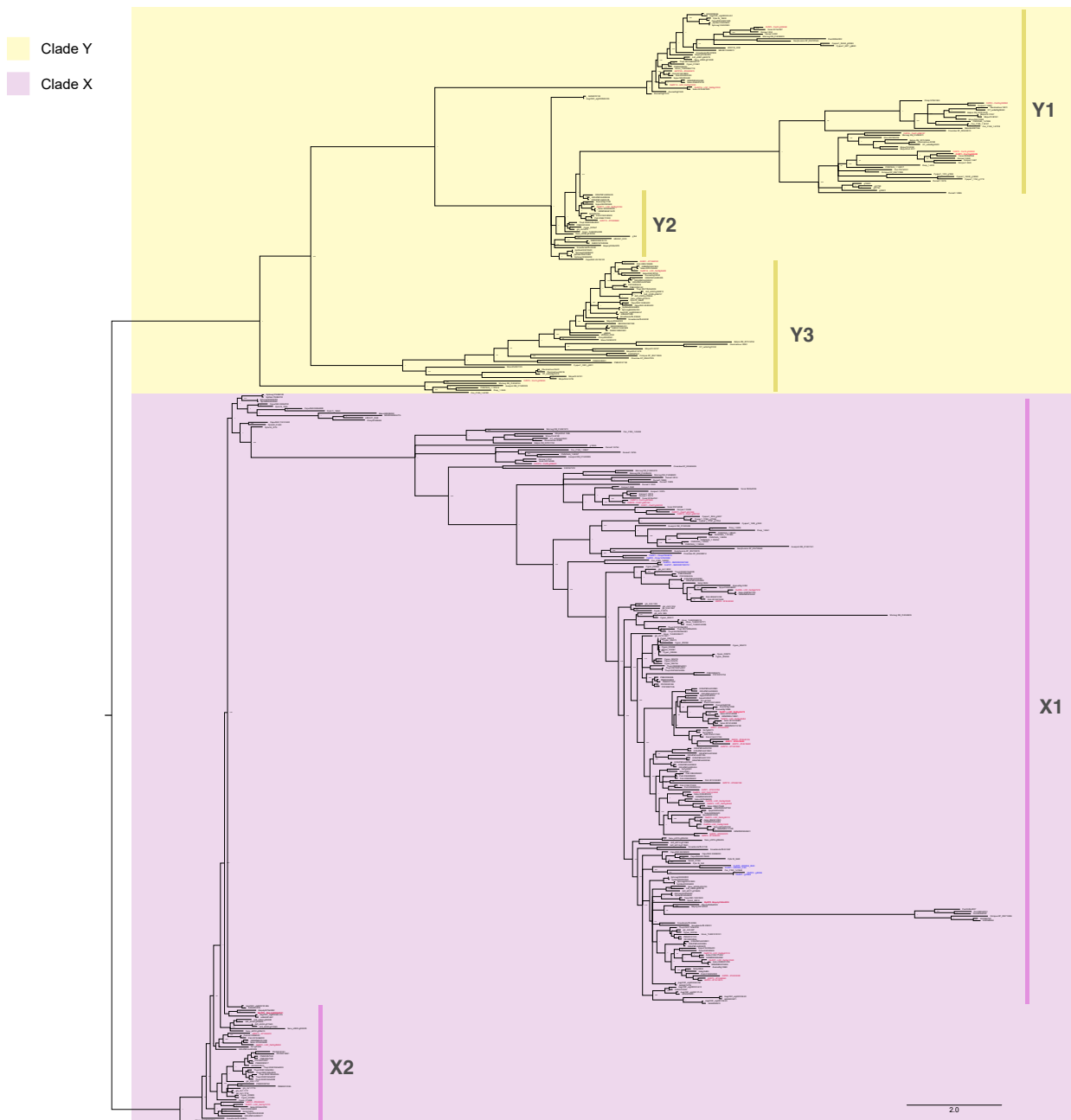
Supplemental Figure 1. Phylogeny of Archaeplastida. Faded colors (orange, red, green, purple, and blue) represent the main archaeplastidian groups analyzed in this work. Topology was based on Lemieux et al. (2015); Cole et al. (2019a,b); Li et al. (2020); Wang et al. (2021).



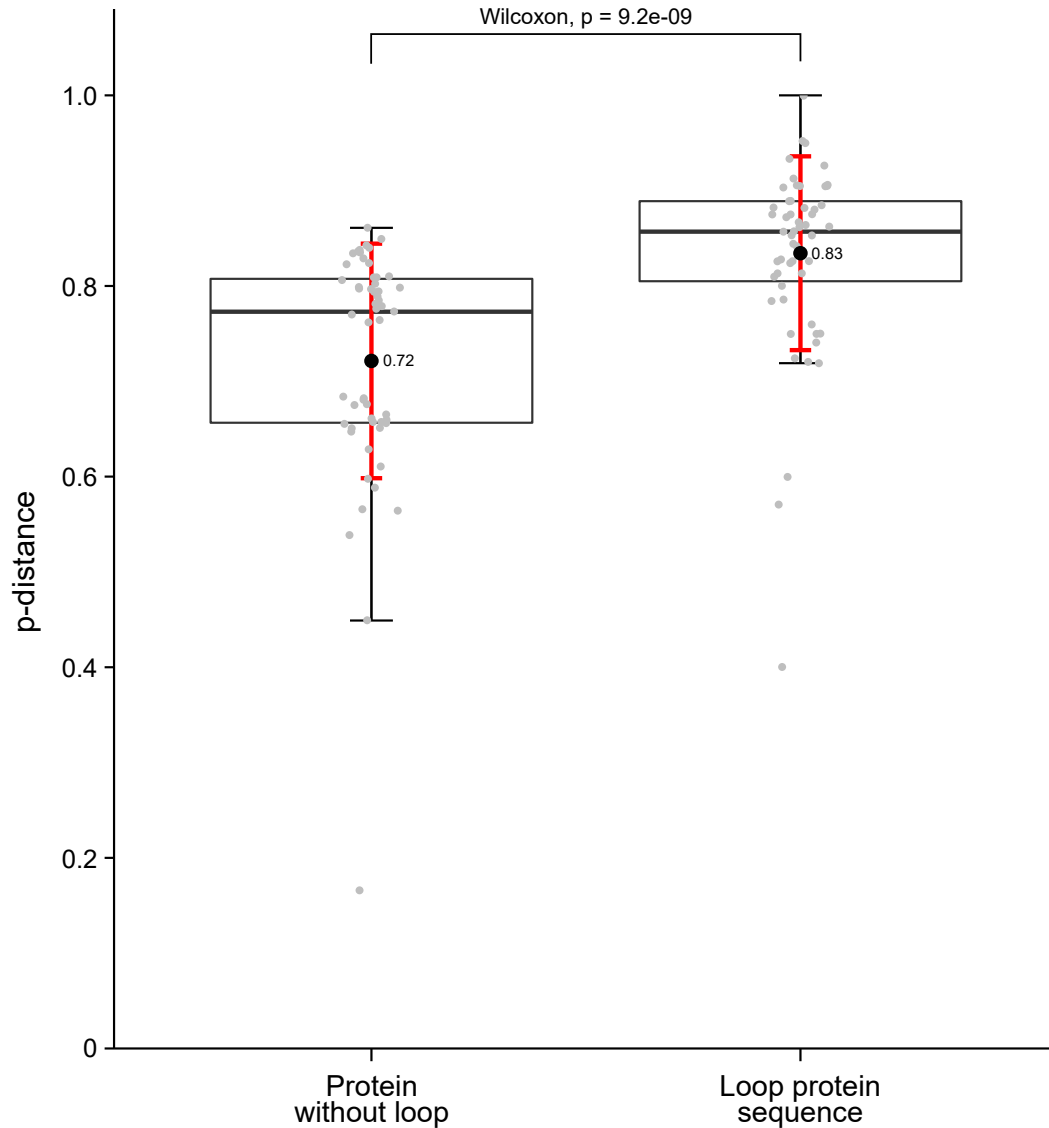
Supplemental Figure 2. Distribution of ZIP genes in algae and plant species of different classes and habitats. Plots showing the global distribution of ZIP homologs, and the frequency of proteins found normalized by the number of genes per genome by (a,b) algae species in different habitats, (c,d) terrestrial and aquatic embryophytes species and (e,f) different Chlorophyta class.



Supplemental Figure 3. Expanded maximum-likelihood phylogenetic tree of ZIP family in Archaeplastida. Expanded version of the phylogenetic tree shown in Figure 3a displaying complete protein IDs and branch support values. ZIP proteins from *Arabidopsis thaliana*, *Oryza sativa* and *Chlamydomonas reinhardtii* are highlighted in red. All ZIP proteins from charophytes used in our logo analysis in Figure 4 and three-dimensional structural modeling in Figure 5 are highlighted in blue. The red * represents the node that integrates embryophyte and charophyte proteins that exhibit the conserved amino acid residues essential for Fe transport shown in Figure 4b-d.



Supplemental Figure 4. Bayesian phylogenetic tree of ZIP family in Archaeplastida. Bayesian phylogenetic tree was done with PhyloBayes v4.1. The main clades and subclades identified in our analyzes are highlighted. ZIP proteins from *Arabidopsis thaliana*, *Oryza sativa* and *Chlamydomonas reinhardtii* are highlighted in red. All ZIP proteins from charophytes used in our logo analysis in Figure 4 and three-dimensional structural modeling in Figure 5 are highlighted in blue.



Supplemental Figure 5. Pairwise p-distances of the primary amino acid sequences of the modeled proteins. Boxplots showing the pairwise distances calculated by p-distance (number of differences/number of aligned positions) between protein cytosolic loops and the rest of the protein sequences. Briefly, p-distance = 0.0 indicates total conservation and p-distance = 1.0 indicates no sequence conservation. Analysis indicates that the amino acid sequences of the loop portion of the IRT1/ZIP proteins are more variable than the rest of the proteins. The black dots indicate the means and red bars are the standard deviation.

Supplemental Tables 1-4

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

Species	Group	Link
<i>Arabidopsis thaliana</i>	Tracheophyta	https://phytozome-next.jgi.doe.gov/info/Athaliana_TAIR10
<i>Oryza sativa</i>	Tracheophyta	https://phytozome-next.jgi.doe.gov/info/Osativa_v7_0
<i>Populus trichocarpa</i>	Tracheophyta	https://phytozome-next.jgi.doe.gov/info/Ptrichocarpa_v4_1
<i>Sorghum bicolor</i>	Tracheophyta	https://phytozome-next.jgi.doe.gov/info/Sbicolor_v3_1_1
<i>Selaginella moellendorffii</i>	Tracheophyta	https://phytozome-next.jgi.doe.gov/info/Smoellendorffii_v1_0
<i>Zea mays</i>	Tracheophyta	https://phytozome-next.jgi.doe.gov/info/Zmays_RefGen_V4
<i>Zostera marina</i>	Tracheophyta	https://phytozome-next.jgi.doe.gov/info/Zmarina_v3_1
<i>Spirodela polyrhiza</i>	Tracheophyta	https://phytozome-next.jgi.doe.gov/info/Spolyrhiza_v2
<i>Nelumbo nucifera</i>	Tracheophyta	https://figshare.com/ndownloader/files/26149538
<i>Euryale ferox</i>	Tracheophyta	https://download.cncb.ac.cn/gwh/Plants/Euryale_ferox_Eferox_GWHBFHH00000000/GWHBFHH000000000.Protein.faa.gz
<i>Salvinia cucullata</i>	Tracheophyta	https://www.fembase.org/ftp/Salvinia_cucullata/Salvinia_asm_v1.2/Salvinia_cucullata.protein.highconfidence_v1.2.fasta
<i>Azolla filiculoides</i>	Tracheophyta	https://www.fembase.org/ftp/Azolla_filiculoides/Azolla_asm_v1.1/Azolla_filiculoides.protein.highconfidence_v1.1.fasta
<i>Ginkgo biloba</i>	Tracheophyta	https://figshare.com/articles/dataset/annotation_of_Ginkgo_biloba/14759223
<i>Cycas panzhihuaensis</i>	Tracheophyta	https://db.cngb.org/codeplot/datasets/PwRfGHfPs5qG3gE
<i>Pinus sylvestris</i>	Tracheophyta	ftp://ftp.psb.ugent.be/pub/plaza/plaza_gymno_01/Fasta/proteome.psy.csv.gz
<i>Thuja plicata</i>	Tracheophyta	https://phytozome-next.jgi.doe.gov/info/Tplicata_v3_1
<i>Gnetum montanum</i>	Tracheophyta	https://datadryad.org/stash/dataset/doi:10.5061/dryad.0vm37
<i>Pseudotsuga menziesii</i>	Tracheophyta	ftp://ftp.psb.ugent.be/pub/plaza/plaza_gymno_01/Fasta/proteome.pme.csv.gz
<i>Anthoceros agrestis</i> [Oxford]	Bryophyta	https://www.hornworts.uzh.ch/static/download/a_agr_oxford.zip
<i>Anthoceros angustus</i>	Bryophyta	https://datadryad.org/stash/dataset/doi:10.5061/dryad.msbec2ftv
<i>Marchantia polymorpha</i>	Bryophyta	https://phytozome-next.jgi.doe.gov/info/Mpolymorpha_v3_1
<i>Sphagnum fallax</i>	Bryophyta	https://phytozome-next.jgi.doe.gov/info/Sfallax_v1_1
<i>Physcomitrella patens</i>	Bryophyta	https://phytozome-next.jgi.doe.gov/info/Ppatens_v3_3
<i>Ceratodon purpureus</i>	Bryophyta	https://phytozome-next.jgi.doe.gov/info/CpurpureusGG1_v1_1

<i>Sphagnum magellanicum</i>	Bryophyta	https://phytozome-next.jgi.doe.gov/info/Smagellanicum_v1_1
<i>Chara braunii</i>	Charophyta	https://bioinformatics.psb.ugent.be/gdb/Chara_braunii/chbra_iso_noTE_23546_pep.fasta
<i>Klebsormidium nitens</i>	Charophyta	http://www.plantmorphogenesis.bio.titech.ac.jp/~algae_genome_project/klebsormidium/kf_download/160614_klebsormidium_v1.1_AA.fasta
<i>Mesotaenium endlicherianum</i>	Charophyta	https://ndownloader.figshare.com/files/17819138
<i>Chlorokybus atmophyticus</i>	Charophyta	ftp://ftp.cngb.org/pub/CNSA/data1/CNP0000228/CNS0021447/CNA0002353/scaffold_Chlorokybus_atmophyticus.pep.gz
<i>Mesostigma viride</i>	Charophyta	ftp://ftp.cngb.org/pub/CNSA/data1/CNP0000228/CNS0021438/CNA0002352/scaffold_Mesostigma_viride.pep.gz
<i>Spirogloea muscicola</i>	Charophyta	https://ndownloader.figshare.com/files/17819147
<i>Auxenochlorella protothecoides</i>	Chlorophyta	https://genome.jgi.doe.gov/portal/Auxeprot1/download/Auxeprot1_GeneCatalog_proteins_20170909.aa.fasta.gz
<i>Bathycoccus prasinos</i>	Chlorophyta	https://genome.jgi.doe.gov/portal/Batpra1/download/Batpra1_GeneCatalog_proteins_20180426.aa.fasta.gz
<i>Chlorella variabilis</i>	Chlorophyta	https://genome.jgi.doe.gov/portal/ChINC64A_1/download/Chlorella_NC64A.all_proteins.fasta.gz
<i>Coccomyxa subellipsoidea</i>	Chlorophyta	https://genome.jgi.doe.gov/portal/Coc_C169_1/download/Coccomyxa_C169_v2_all_proteins.fasta.gz
<i>Chlamydomonas reinhardtii</i>	Chlorophyta	https://genome.jgi.doe.gov/portal/pages/dynamicOrganismDownload.js?organism=Phytozome#
<i>Dunaliella salina</i>	Chlorophyta	https://genome.jgi.doe.gov/portal/Dunsal1/download/Dunsal1_GeneCatalog_proteins_20180508.aa.fasta.gz
<i>Gonium pectorale</i>	Chlorophyta	https://genome.jgi.doe.gov/portal/Gonpec1/download/Gonpec1_GeneCatalog_proteins_20180501.aa.fasta.gz
<i>Micromonas commoda</i>	Chlorophyta	https://genome.jgi.doe.gov/portal/MicpuN3v2/download/MicpuN3v2_GeneCatalog_proteins_20160404.aa.fasta.gz
<i>Monoraphidium neglectum</i>	Chlorophyta	https://genome.jgi.doe.gov/portal/Monneg1/download/Monneg1_GeneCatalog_proteins_20170920.aa.fasta.gz
<i>Micromonas pusilla</i>	Chlorophyta	https://genome.jgi.doe.gov/portal/MicpuC2/download/MicromonasCCMP1545.allModels.aa.fasta.gz
<i>Ostreococcus lucimarinus</i>	Chlorophyta	https://phytozome-next.jgi.doe.gov/info/Olucimarinus_v2_0
<i>Ostreococcus tauri</i>	Chlorophyta	https://genome.jgi.doe.gov/portal/Ostta4221_3/download/Ostta4221_3_GeneCatalog_proteins_20161028.aa.fasta.gz
<i>Picochlorum soloecismus</i>	Chlorophyta	https://genome.jgi.doe.gov/portal/Picsp_1/download/Picsp_1_GeneCatalog_proteins_20170909.aa.fasta.gz
<i>Volvox carteri</i>	Chlorophyta	https://phytozome-next.jgi.doe.gov/info/Vcarteri_v2_1
<i>Chondrus crispus</i>	Rhodophyta	https://greenhouse.lanl.gov/greenhouse/organisms/
<i>Cyanidioschyzon merolae</i>	Rhodophyta	https://greenhouse.lanl.gov/greenhouse/organisms/
<i>Galdieria sulphuraria</i>	Rhodophyta	https://greenhouse.lanl.gov/greenhouse/organisms/
<i>Porphyra umbilicalis</i>	Rhodophyta	https://phytozome-next.jgi.doe.gov/info/Pumbilicalis_v1_5
<i>Gracilaria domingensis</i>	Rhodophyta	https://www.ncbi.nlm.nih.gov/genome/107867
<i>Cyanophora paradoxa</i>	Glaucophyta	http://cyanophora.rutgers.edu/cyanophora_v2018

Supplemental Table 2. Table showing the number of ZIP family homologs found by species in our dataset.

Species	Group	# of ZIP homologs
<i>Arabidopsis thaliana</i>	Embryophyta	18
<i>Oryza sativa</i>	Embryophyta	16
<i>Populus trichocarpa</i>	Embryophyta	20
<i>Sorghum bicolor</i>	Embryophyta	17
<i>Selaginella moellendorffii</i>	Embryophyta	9
<i>Zea mays</i>	Embryophyta	14
<i>Zostera marina</i>	Embryophyta	9
<i>Spirodela polyrhiza</i>	Embryophyta	12
<i>Nelumbo nucifera</i>	Embryophyta	11
<i>Euryale ferox</i>	Embryophyta	25
<i>Salvinia cucullata</i>	Embryophyta	8
<i>Azola filiculoides</i>	Embryophyta	12
<i>Ginkgo biloba</i>	Embryophyta	12
<i>Cycas panzhihuaensis</i>	Embryophyta	21
<i>Pinus sylvestris</i>	Embryophyta	10
<i>Thuja plicata</i>	Embryophyta	17
<i>Gnetum montanum</i>	Embryophyta	9
<i>Pseudotsuga menziesii</i>	Embryophyta	16
<i>Anthoceros agrestis</i> [Oxford]	Embryophyta	10
<i>Anthoceros angustus</i>	Embryophyta	10
<i>Marchantia polymorpha</i>	Embryophyta	8
<i>Sphagnum fallax</i>	Embryophyta	9
<i>Physcomitrella patens</i>	Embryophyta	10
<i>Ceratodon purpureus</i>	Embryophyta	11
<i>Sphagnum magellanicum</i>	Embryophyta	9
<i>Chara braunii</i>	Charophyta	9
<i>Klebsormidium nitens</i>	Charophyta	6
<i>Mesotaenium endlicherianum</i>	Charophyta	3
<i>Chlorokybus atmophyticus</i>	Charophyta	5

<i>Mesostigma viride</i>	Charophyta	5
<i>Spirogloea muscicola</i>	Charophyta	7
<i>Auxenochlorella protothecoides</i>	Chlorophyta	4
<i>Bathycoccus prasinos</i>	Chlorophyta	4
<i>Chlorella variabilis</i>	Chlorophyta	10
<i>Coccomyxa subellipsoidea</i>	Chlorophyta	7
<i>Chlamydomonas reinhardtii</i>	Chlorophyta	12
<i>Dunaliella salina</i>	Chlorophyta	9
<i>Gonium pectorale</i>	Chlorophyta	11
<i>Micromonas commoda</i>	Chlorophyta	6
<i>Monoraphidium neglectum</i>	Chlorophyta	8
<i>Micromonas pusilla</i>	Chlorophyta	5
<i>Ostreococcus lucimarinus</i>	Chlorophyta	6
<i>Ostreococcus tauri</i>	Chlorophyta	5
<i>Picochlorum soloecismus</i>	Chlorophyta	4
<i>Volvox carteri</i>	Chlorophyta	6
<i>Chondrus crispus</i>	Rhodophyta	3
<i>Cyanidioschyzon merolae</i>	Rhodophyta	4
<i>Galdieria sulphuraria</i>	Rhodophyta	3
<i>Porphyra umbilicalis</i>	Rhodophyta	5
<i>Gracilaria domingensis</i>	Rhodophyta	4
<i>Cyanophora paradoxa</i>	Glaucophyta	10
Total		484

Supplemental Table 3. IRT1/ZIP Proteins used in the structural modeling. If predicted, signal peptides were removed from the final protein model.

Protein name	Organism	Protein ID	Accession code	Signal peptide (aa position)	Protein length (aa)
AtIRT1	<i>Arabidopsis thaliana</i>	AT4G19690	Q38856 ^A	1-27	347
CrIRT1	<i>Chlamydomonas reinhardtii</i>	Cre12.g530400	A0A2K3D4S5 ^A	1-28	367
OsIRT1	<i>Oryza sativa</i>	LOC_Os03g46470	Q75HB1 ^A	1-33	374
CbZIP1	<i>Chara braunii</i>	g37664	A0A388LNG6 ^A	NA	375
KnZIP1	<i>Klebsormidium nitens</i>	kfl00040_0120	GAQ79869.1 ^B	NA	340
SmZIP1	<i>Spirogloea muscicola</i>	SM000007S20791	NA	NA	319
CaZIP1	<i>Chlorokybus atmophyticus</i>	Chrsp27S04275	NA	NA	381
AtZTP29	<i>Arabidopsis thaliana</i>	AT3G20870	Q940Q3 ^A	1-22	276
OsZIP13	<i>Oryza sativa</i>	LOC_Os02g10230	EEE56500.1 ^B	NA	276
MpZIP3	<i>Marchantia polymorpha</i>	Mapoly0169s0010	A0A2R6W308 ^A	NA	381
MpZIP5	<i>Marchantia polymorpha</i>	Mapoly0088s0041	A0A0E3DAH5 ^A	1-32	344

^A: UNIPROT accession code.

^B: GENBANK accession code.

Supplemental Table 4. Cytosolic loops identified in the modeled IRT1/ZIP proteins.

Protein name	Loop length	Loop between helices	Loop sequence
AtIRT1	158-180	3-4	GIMPHGHGHGHGPANDVTLPKE
CrIRT1	87-166	5-6	HGSMLASGSSFCTPSASASCSSCCCTGGEAHID EKSAAACPPCSHVMVHVAPQPVASGDGEDVMAGGGGK QAGGRAGSSA
OsIRT1	163-205	3-4	RSKPRPSSGGDVAAVADHGESPQGHRRHGHGHGHGHGMAVAKP
CbZIP1	179-205	3-4	GAGAMLAVTAMGKTPGAEGKGDVVVDC
KaZIP1	130-182	3-4	GHHHPDHVHVHGLAHQGVAPSVDSGDAHPPAGSVKSLDVEAG SVAKESTFGTP
SmZIP1	119-150	3-4	YGLPTQASLEDSDEVQGGLSMGDDLEGAKANP
CaZIP1	164-200	3-4	LAQDVHQPTVPSTSTTTTITSSAAPT LVHRGHAHGAG
AtZTP29	96-106	2-3	FIPEPTLGPSTDGKRRKKNGD
OsZIP13	86-108	3-4	FIPEPTVVPTADAGKKQTDDDGS
MpZIP3	130-209	3-4	SHGHTEDYEKLPTRKSI L DSDSTHEPLLQNSSTSLAVAEPVHKTN GSQHHEKHTERSHTHTPNGGADEEAGHNGNHLVLA
MpZIP5	157-188	3-4	TGKESAITGAVGPNDHEAGHMVQATLIHTASM

References

- Cole TCH, Bachelier JB, Hilger HH. 2019a.** *Tracheophyte phylogeny poster - Vascular plants: systematics and characteristics*. PeerJ Preprints.
- Cole TCH, Hilger HH, Stevens P. 2019b.** *Angiosperm phylogeny poster (APP) – Flowering plant systematics, 2019*. PeerJ Preprints.
- Crooks GE, Hon G, Chandonia J-M, Brenner SE. 2004.** WebLogo: A Sequence Logo Generator: Figure 1. *Genome Research* **14**: 1188–1190.
- Eddy SR. 2011.** Accelerated Profile HMM Searches (WR Pearson, Ed.). *PLoS Computational Biology* **7**: e1002195.
- Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010.** New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology* **59**: 307–321.
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, *et al.* 2021.** Highly accurate protein structure prediction with AlphaFold. *Nature* **596**: 583–589.
- Kall L, Krogh A, Sonnhammer ELL. 2007.** Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. *Nucleic Acids Research* **35**: W429–W432.
- Katoh K. 2002.** MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**: 3059–3066.
- Kleffner R, Flatten J, Leaver-Fay A, Baker D, Siegel JB, Khatib F, Cooper S. 2017.** Foldit Standalone: a video game-derived protein structure manipulation interface using Rosetta. *Bioinformatics* **33**: 2765–2767.
- Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D. 2003.** Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. *Science* **302**: 1364–1368.
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018.** MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms (FU Battistuzzi, Ed.). *Molecular Biology and Evolution* **35**: 1547–1549.
- Lartillot N, Lepage T, Blanquart S. 2009.** PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**: 2286–2288.
- Lemieux C, Vincent AT, Labarre A, Otis C, Turmel M. 2015.** Chloroplast phylogenomic analysis of chlorophyte green algae identifies a novel lineage sister to the Sphaeropleales (Chlorophyceae). *BMC Evolutionary Biology* **15**: 264.
- Li L, Wang S, Wang H, Sahu SK, Marin B, Li H, Xu Y, Liang H, Li Z, Cheng S, *et al.* 2020.** The genome of *Prasinoderma coloniale* unveils the existence of a third phylum within green plants. *Nature Ecology & Evolution* **4**: 1220–1231.

Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, et al. 2021. Pfam: The protein families database in 2021. *Nucleic Acids Research* **49**: D412–D419.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* **32**: 268–274.

Rice P, Longden I, Bleasby A. 2000. EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics* **16**: 276–277.

RStudio Team. 2020. *RStudio: Integrated Development Environment for R*. Boston, MA: RStudio, PBC.

Schrödinger L. 2015. The {PyMOL} Molecular Graphics System, Version~1.8.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research* **13**: 2498–2504.

Wang S, Liang H, Xu Y, Li L, Wang H, Sahu DN, Petersen M, Melkonian M, Sahu SK, Liu H. 2021. Genome-wide analyses across Viridiplantae reveal the origin and diversification of small RNA pathway-related genes. *Communications Biology* **4**: 412.

Wang S, Ma J, Peng J, Xu J. 2013. Protein structure alignment beyond spatial proximity. *Scientific Reports* **3**: 1448.

Zallot R, Oberg N, Gerlt JA. 2019. The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. *Biochemistry* **58**: 4169–4182.

4.3. CAPÍTULO 5 – COMPREHENSIVE CHARACTERIZATION OF HMA TRANSPORTERS IN COMMON BEAN

Neste capítulo é apresentado o artigo “*Comprehensive characterization of HMA transporters in common bean: Tissue-specific expression and response to metal stress*”, publicado na revista *Crop Science* (fator de impacto 1.9; DOI: 10.1002/csc2.70188), do qual sou primeiro autor. O estudo realiza uma caracterização abrangente da família de HMAs em feijão (*Phaseolus vulgaris*), integrando análises genômicas, estruturais e transcriptômicas para compreender o papel desses transportadores na homeostase de metais e na resposta a condições de toxicidade. O trabalho identifica e descreve os genes PvHMA quanto a seus motivos conservados, topologia transmembrana e predições de localização subcelular, além de organizá-los em subgrupos funcionais com base em filogenias. A análise de expressão revela padrões específicos de tecidos e respostas regulatórias a estresses por metais como Cu e Zn, sugerindo especializações fisiológicas entre os diferentes membros da família. Esses resultados contribuem para o entendimento da organização estrutural e funcional das HMAs em plantas cultivadas e fornecem subsídios para abordagens futuras em biofortificação, resposta ao estresse e estudos de nutrição mineral.

ORIGINAL ARTICLE

Genomics, Molecular Genetics & Biotechnology

Comprehensive characterization of HMA transporters in common bean: Tissue-specific expression and response to metal stress

Wenderson Felipe Costa Rodrigues¹ | Laura Oliveira Pires¹ | Lucas Roberto Oliveira² |
Juliane K. Ishida^{1,2,3} 

¹Department of Botany, Institute of Biological Sciences (ICB), Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil

²Center for Nuclear Energy in Agriculture (CENA), University of São Paulo (USP), Piracicaba, Brazil

³Department of Genetics, Luiz de Queiroz College of Agriculture (ESALQ), University of São Paulo (USP), Piracicaba, São Paulo, Brazil

Correspondence

Juliane K. Ishida, Department of Genetics, Luiz de Queiroz College of Agriculture (ESALQ), University of São Paulo (USP), Avenida Pádua Dias, 11, Piracicaba, SP, Brazil.

Email: julianeishida@usp.br

Assigned to Associate Editor Alvaro Sanz-Saez.

Funding information

Fundação de Amparo à Pesquisa do Estado de Minas Gerais, Grant/Award Number: APQ-02786-21; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Grant/Award Number: Finance Code 001; Instituto Serrapilheira, Grant/Award Number: R-2111-40087; Science Without Borders Program from the Brazilian Federal Government

Abstract

Plant growth and development require tightly regulated concentration of heavy metal ions, which function as essential nutrients. In this context, the P1B-heavy metal ATPase (HMA), also known as the HMA family, is critical for mediating metal ion uptake, root-to-shoot translocation, and vacuolar sequestration in plants. This study presents a comprehensive characterization of 11 HMA genes in *Phaseolus vulgaris* L. (*PvHMA*). Phylogenetic analysis classified the *PvHMA*s genes into six distinct clusters, supported by conserved gene structure and motif distributions. *PvHMA1*, -2, -5, -6, -7, and -11 exhibited widespread expression across multiple tissues and harbored a diverse array of cis-regulatory elements in their promoter, suggesting multiple roles in plant growth and development. In contrast, *PvHMA3* and *PvHMA8* displayed tissue-specific expression patterns, being predominantly expressed in roots and leaves, respectively. Under zinc stress, *PvHMA1*, localized in chloroplasts, showed marked upregulation in shoot tissue. Notably, this transcriptional response was not observed under copper exposure, despite the high structural similarity between *PvHMA1* and its *Arabidopsis thaliana* homolog, *AtHMA1* (where *AtHMA* is *Arabidopsis thaliana* HMA). *PvHMA2*, an ortholog of the *A. thaliana* HMA2/4, exhibited increased sensitivity to cobalt stress. Additionally, *PvHMA5* and -11 were differentially expressed in shoots in response to zinc treatment. Collectively, these findings provide a detailed overview of the HMA family in *P. vulgaris* and reveal a complex regulatory network of transporters involved in heavy metal homeostasis, with implications for plant nutrition, development, and stress responses.

Abbreviations: AtHMA, *Arabidopsis thaliana* HMA; HM, heavy metal; HMA, Heavy Metal ATPase; OsHMA, *Oryza sativa* HMA; PvHMA, *Phaseolus vulgaris* L. HMA; RMSD, Root-mean-square deviation; SOD, superoxide dismutase; TM, transmembrane.

Wenderson Felipe Costa Rodrigues and Laura Oliveira Pires contributed equally to this work.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Crop Science* published by Wiley Periodicals LLC on behalf of Crop Science Society of America.

Plain Language Summary

Plants need certain metal ions, like zinc and copper, to grow properly. Special proteins called heavy metal ATPases (HMAs) help move these metals around the plant and store them safely. In this study, we looked at *HMA* genes in common bean plants (*Phaseolus vulgaris*). We grouped the genes into six families and found that some are active in many plant parts, while others work mainly in roots or leaves. When the plants were exposed to zinc, some genes, like *PvHMA1* (where *PvHMA* is *Phaseolus vulgaris* L. HMA) and *PvHMA2*, changed their activity, especially in shoots. These changes did not happen under copper stress. Our results show that *HMA* genes help plants manage metal nutrients and respond to stress. This research could help improve plant health and nutrition.

1 | INTRODUCTION

Heavy metals (HMs) are generally defined as transition and post-transition elements with high density (Hawkes, 1997). This broad category includes more than 20 elements, some of which are highly toxic, such as mercury (Hg), arsenic (As), lead (Pb), cadmium (Cd), cobalt (Co), copper (Cu), manganese (Mn), molybdenum (Mo), nickel (Ni), and zinc (Zn). At elevated concentration, these elements interfere with key cellular functions by inactivating metabolic enzymes, disrupting membrane integrity, and inducing excessive production of reactive oxygen species (Rascio & Navari-Izzo, 2011). However, when present at optimal levels, several HMs function as micronutrients contributing to proper plant growth and development (Alloway, 2013). The intensive and prolonged use of chemical fertilizers in crop cultivation has led to a significant accumulation of HMs in the soil (Atafar et al., 2010), resulting in environmental contamination (Mortvedt, 1996), negative impacts on human health (Mortvedt, 1996), and alteration in the structure and function of soil microbial communities (Navarrete et al., 2017). Understanding the molecular mechanisms that regulate the uptake, translocation, and sequestration of HMs in plants is critical for defining the thresholds at which these elements transition from being beneficial to toxic.

Among the proteins involved in HM homeostasis, members of the P-type ATPase superfamily play a central role. These transporters utilize ATP hydrolysis to mediate the active transport of cations across biological membranes against their electrochemical gradients (Hanikenne & Baurain, 2014). The “P-type” designations refer to the transient formation of a phosphorylated intermediate during the transport cycle (Pedersen & Carafoli, 1987). P-type ATPases are conserved across all domains of life and participate in a wide range of cellular processes (Farley, 2012). Based on substrate specificity, they are classified into five major subfamilies (P1–P5) (Palmgren & Axelsen, 1998). For instance, the H⁺ATPase in

plants and fungi belongs to the P3A-type, whereas Na⁺/K⁺ and Ca²⁺ATPases in animals fall under the P2C/D- and P2A/B-types, respectively (Kühlbrandt, 2004). The P1B-type ATPases, also known as heavy metal ATPase (HMA), form a distinct subgroup specialized in the transport of transition metals. These proteins are characterized by the presence of eight transmembrane (TM) domains, including a highly conserved CPx motif (cysteine-proline-cysteine, cysteine-proline-histidine, or cysteine-proline-serine) within the sixth TM domain, which is responsible for metal ion binding (Hanikenne & Baurain, 2014). Additionally, HMAs possess five conserved signature motifs—DKTGT, GDGxNDxP, PxxK, HP, and S/TGE—that are critical for ATP hydrolysis and metal transport (Argüello et al., 2007). Understanding the function and regulation of HMA transporters is essential for advancing strategies in phytoremediation, crop biofortification, and tolerance to metal stress.

Plant HM transporters play a crucial role in the tolerance of HMs and in maintaining the balance of metal ions inside the plant (Williams et al., 2000). The HMA family of transporters can be categorized into two subgroups: one that binds to monovalent ions such as Cu⁺ and Ag⁺, and another that is linked with divalent ions including Zn²⁺, Cd²⁺, Pb²⁺, and Co²⁺ (Axelsen & Palmgren, 2001). The *Arabidopsis thaliana* genome contains eight members (Baxter et al., 2003). Among them, *AtHMA2* (where *AtHMA* is *Arabidopsis thaliana* HMA) and *AtHMA4* are expressed in vascular tissues and produce proteins involved in metal translocation through the xylem (Hussain et al., 2004; Sinclair et al., 2007). In the related species *Arabidopsis halleri*, constitutive expression of *HMA4* is directly associated with the species' ability to tolerate and hyperaccumulate Zn and Cd (Nouet et al., 2015). Enhanced metal tolerance in *A. thaliana* can be achieved by overexpressing *AtHMA3*, a tonoplast-localized transporter that facilitates vacuolar sequestration of toxic metals (Morel et al., 2009). Transgenic lines overexpressing *AtHMA3* exhibit enhanced tolerance to Cd, Zn, Pb, and Co, while mutants

lacking this transporter show increased sensitivity, particularly to Cd and Zn stress (Morel et al., 2009). Similar findings have been reported in other species; for example, transgenic tobacco expressing *SaHMA3h* from *Sedum alfredii* displayed improved cadmium tolerance and accumulation (Z. Zhang et al., 2016). In *Brassica rapa*, variations in leaf cadmium content were linked to differential expression of *BjHMA3* (L. Zhang et al., 2019). Additionally, ectopic expression of *OsHMA3* (where *OsHMA* is *Oryza sativa* HMA) in rice resulted in greater Cd tolerance and reduced Cd levels in aerial tissues and grains, while increasing Cd accumulation in the roots (Ueno et al., 2010). The AtHMA1, localized in the chloroplast envelope, is known for its role in copper translocation (Seigneurin-Berny et al., 2006). This transporter is essential for delivering Cu to superoxide dismutase (SOD), a key enzyme that mitigates oxidative stress generated during photosynthesis (Seigneurin-Berny et al., 2006). Under Zn (II) excess, AtHMA1 also contributes to detoxification by removing Zn from plastids (Kim et al., 2009). In addition to Cu, AtHMA1 has been shown to interact with calcium, functioning in association with calcium ATPase (Moreno et al., 2008). AtHMA5–8 are predicted to mediate the transporter of Cu and silver (Ag). AtHMA6 (AtPAA1) and AtHMA8 (AtPAA2) are both involved in chloroplast Cu homeostasis. AtHMA6 is located in the plastid envelope, while AtHMA8 is situated in the thylakoid membranes (Abdel-Ghany et al., 2005; Shikanai et al., 2003). AtHMA5 is primarily involved in Cu detoxification in roots (Andrés-Colás et al., 2006). Lastly, AtHMA7, also known as RAN1, delivers Cu to post-Golgi compartments and is essential for maturation of ethylene receptors (Hirayama et al., 1999). The *Oryza sativa* (rice) genome encodes nine HMA genes, with OsHMA1–3 belonging to the Zn/Co/Cd/Pb group and OsHMA4–9 belonging to the Cu/Ag group (Baxter et al., 2003). Among these, OsHMA2 plays a key role in the translocation of Zn/Cd from root cells to vascular tissues, thereby promoting their movement from roots to shoots (Sato-Nagasawa et al., 2012; Takahashi et al., 2012). OsHMA3 transports and stores Cd ions in vacuoles within root cells, while OsHMA9 facilitates the export of Cu and Zn ions from the cells to the apoplastic region (Lee et al., 2007).

Phaseolus vulgaris L., commonly known as the common bean, represents the most significant grain legume for direct human consumption globally (Food and Agriculture Organization [FAO], www.fao.org/faostat). Cultivated extensively, it constitutes a staple food for millions of individuals, particularly in Latin America and Africa, where it serves as a primary source of dietary sustenance (FAO, www.fao.org/faostat). The crop's importance is multifaceted, encompassing both nutritional and agronomic dimensions. Nutritionally, common beans provide a cost-effective source of high-quality protein, complex carbohydrates, dietary fiber, and essential micronutrients, notably iron (Fe) and zinc (Murube et al., 2021), and thus play a critical role in public health strate-

gies addressing malnutrition. Its widespread consumption and nutrient density render it a suitable candidate for biofortification programs (Sperotto & Ricachenevsky, 2017) aimed at mitigating hidden hunger, particularly micronutrient deficiencies such as iron deficiency anemia, which affects billions worldwide. From an agricultural perspective, common beans contribute to sustainable farming systems through symbiotic associations with nitrogen-fixing *Rhizobia*, enhancing soil fertility and reducing reliance on synthetic fertilizers (Abd-Alla et al., 2023). Understanding the molecular components involved in HM homeostasis in *P. vulgaris* is fundamental to improving our knowledge of how micronutrients are taken up from the soil and allocated to edible plant tissues. This insight has the potential to inform strategies aimed at enhancing nutritional content and developing biofortified cultivars. Although members of the HMA gene family have been identified and characterized in several plant species, including *Glycine max* (soybean) (Fang et al., 2016), *Populus trichocarpa* (D. Li et al., 2015), *Brassica napus* (N. Li et al., 2018), maize, *Triticum aestivum* (Batool et al., 2023), *Medicago truncatula* (Ma et al., 2021), and sorghum (Zhiguo et al., 2018), a systematic and comprehensive analysis of these genes in common beans is still lacking. In this study, we performed a genome-wide analysis of 11 putative HMA homologs in *P. vulgaris* L. (*PvHMA*). We examined their gene structure, phylogenetic relationships, conserved motifs, and cis-regulatory elements. Furthermore, we analyzed publicly available gene expression datasets to evaluate the transcriptional profiles of *PvHMA* genes during different developmental stages and under metal stress conditions. Structural modeling was also employed to explore the three-dimensional (3D) conformation of selected *PvHMA* proteins. Together, these findings provide a foundational framework for future functional studies on HMA transporters in common beans, with implications for improving nutrient use efficiency, stress resilience, and biofortification strategies.

2 | MATERIALS AND METHODS

2.1 | Plant material and growth conditions

Seeds of *P. vulgaris* cultivar Carioca IAC-Alvorada were germinated in a climate-controlled growth chamber at 25°C–28°C with a 12-h light/12-h dark photoperiod. Plants were grown in saturated vermiculate for 14 days. Following this period, the substrate was carefully removed, and the roots were washed thoroughly. To assess the effects of HM stress, roots were immersed in aqueous solutions containing divalent metal salts at the following concentrations: 50 µM cobalt chloride (CoCl₂), 100 µM copper sulfate (CuSO₄), 150 µM nickel chloride (NiCl₂), or 200 µM zinc sulfate (ZnSO₄). Control plants were treated with demineralized water under the same

conditions. Treatments were applied for two exposure periods, 2 and 18 h, to simulate both short- and long-term stress conditions. Following exposure, roots and shoots tissues from at least two plants per treatment were harvested separately and immediately flash-frozen in liquid nitrogen. All samples were stored at -80°C until further processing. Three independent biological replicates were performed for each treatment condition.

2.2 | Identification and phylogenetic tree construction of HMAs

The AtHMAs protein sequences were retrieved from the *A. thaliana* genome (TAIR10) (Supporting Information S1) via the Phytozome database (<http://www.phytozome.net/>) and subsequently used as queries in BLASTP (Protein-Protein Basic Local Alignment Search Tool) searches against the *P. vulgaris* genome (G19833) proteome, also accessed through the Phytozome platform. PvHMA homologs were identified based on a minimum similarity threshold of 50%. For each identified gene, the corresponding genomic DNA, coding DNA sequence (CDS), and protein sequences were retrieved. Additional parameters such as protein length, molecular weight (mw), and isoelectric point (pI) were calculated using the Compute pI/Mw (Bjellqvist et al., 1993). Subcellular localization predictions were performed using the TargetP1.1 Server (Emanuelsson et al., 2007) with default parameters. Gene structure was performed using the online tool GSDS 2.0 (<http://gsds.cbi.pku.edu.cn>) (Hu et al., 2015). Conserved motifs among PvHMA proteins were identified using the MEME Suite (Bailey et al., 2009), under default settings. To investigate evolutionary relationships, full-length HMA protein sequences from *P. vulgaris*, *A. thaliana*, *O. sativa* (Baxter et al., 2003), and *G. max* (Fang et al., 2016) were aligned using the CLUSTALW algorithm (Thompson et al., 2002). Phylogenetic trees were constructed with MEGA7 software (Kumar et al., 2016) using the maximum likelihood method with 10,000 bootstrap replications. Hierarchical clustering of PvHMA protein sequences was performed using the GenePattern 2.0 platform (Reich et al., 2006). Pearson correlation coefficients (absolute value) were used as the similarity metric, and the pairwise average-linkage method was applied for clustering. Protein domain comparisons were performed using the SMART database version 7.0 (Letunic et al., 2021), and sequence relationships were constructed with MAFFT (Kato & Standley, 2013).

2.3 | Promoter region analysis and gene expression profiling

The whole-genome sequence of *P. vulgaris* was obtained from the Phytozome database (<https://phytozome.jgi.doe.gov>).

Putative promoter regions, defined as 1 kb upstream of the translation start site (ATG), were extracted for each PvHMA gene using the GENEIOUS software (www.geneious.com). Cis-regulatory elements within these promoter sequences were identified by PlantCARE (Lescot et al., 2002). Organ-specific expression data for the *PvHMA* genes were retrieved from the Phytozome platform, accessed through user login credentials registered in the JGI genome portal. The *P. vulgaris* transcriptome dataset (*P. vulgaris*_442_ver2.1) was downloaded for transcript-level analysis. All graphical elements and figures were created and refined using Adobe Illustrator (Adobe Systems Inc.).

2.4 | RNA isolation and quantitative real-time PCR (RT-qPCR)

Total RNA was extracted from root and shoot tissues using the TRIzol reagent (ThermoFisher Scientific, cat. No. 15596026), following the manufacturer's instructions. To remove residual genomic DNA contamination, RNA samples were treated with DNase I Amplification Grade (ThermoFisher Scientific, cat. No. 1806015). First-strand cDNA synthesis was performed using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Code 4368814), according to the manufacturer's instructions. The qRT-PCR was carried out using the SyBR Green PCR Master Mix (Applied Biosystems, Code 439155) on a real-time PCR detection system. Relative gene expression levels were calculated using the standard curve method. Primer sequences used in the study are listed in Supporting Information S2. As detailed in the figure legends, gene expression quantification was based on three independent biological replicates, except for Shoot 2 h and Root 18 h samples treated with Zn and Co for *PvHMA6*, where only two biological replicates were used due to limited plant material. Each biological replicate consisted of pooled tissues from at least two individual plants and was analyzed with three technical replicates. To determine statistical significance among treatments, *p*-values were calculated using the Student's *t*-test (two-sample assuming equal variances) in Microsoft Excel. Differences were considered significant at $p < 0.05$ and marginally significant at $p < 0.1$.

2.5 | Protein three-dimensional structure prediction of PvHMA

Prior to 3D structure modeling, signal peptides were predicted in SignalP v6.0 (Teufel et al., 2022) and, if identified, removed from the original sequence. However, no PvHMA showed signal peptide via SignalP. Three-dimensional struc-

ture modeling was done via ColabFold (Mirdita et al., 2022), which easily enables structure prediction using AlphaFold2 software (Jumper et al., 2021). Predicted structures were relaxed using Amber force fields and complex prediction (Eastman et al., 2017), where five models were generated, and the highest ranked conformation was selected based on its Predicted Local Distance Difference Test (pLDDT) confidence score. We used PyMOL (pymol.org) to visualize the structures and to know the root-mean-square deviation (RMSD) value of alpha carbon ($C\alpha$) atomic positions with the alignment plugin. The TM topology was predicted using DeepTMHMM (Hallgren et al., 2022) and later identified in the modeled proteins.

3 | RESULTS

3.1 | Identification and classification of HMA genes in the *P. vulgaris* genome

To identify members of the HMA gene family in *P. vulgaris*, BLASTP searches were performed against the common bean genome using the *A. thaliana* HMA protein sequence as queries via the Phytozome database. A total of 11 putative *PvHMA* genes were identified. Conserved motif organization and phylogenetic relationships among the predicted proteins were visualized simultaneously (Figure 1A). Key conserved motifs characteristic of the HMA family, including DKTGT, GDGxNDxP, PxxK, S/TGE, HP, and CPx/SPC, were identified across the sequences (Figure 1B), consistent with previously reported HMA features (Argüello et al., 2007). *PvHMA* sequences contain structural features similar to those found in their homologs in *A. thaliana*, which include the ATPase and hydrolase protein families. A detailed analysis revealed the presence of several conserved domains: the HMA domain (PF00403), the E1-E2 ATPase domain (PF00122), and the hydrolase domain (PF00702) (Supporting Information S3).

Phylogenetic clustering revealed subgroups with distinct motif profiles. Specifically, *PvHMA*2/3, *PvHMA*6/8, *PvHMA*4/7, and *PvHMA*5/9/10 each shared a set of conserved motifs within their respective clades, whereas *PvHMA*1 and *PvHMA*11 exhibited divergent motif compositions (Figure 1A). To investigate the gene structure of *PvHMA* family members, intron/exon organization was analyzed by aligning genomic sequences with their corresponding coding sequences (CDS) (Figure 2). This analysis revealed significant variation in intron size and distribution across different phylogenetic branches. Additional structural features were examined, including gene, transcript, and CDS lengths (Supporting Information S4). The *PvHMA*1 and -6 had longer genomic sequences (16,431 and 24,584 bp, respectively), suggesting the presence of large intronic regions (Supporting

Information S4). In contrast, transcript lengths across the family were relatively uniform, ranging from 2865 to 3941 bp (Supporting Information S4), supporting the hypothesis of extensive intronic expansion in these two genes. Predicted protein lengths varied from 826 to 1187 aa, with corresponding molecular weights ranging from 90 to 130.9 kDa, and the *pI* between 5.26 and 8.69 (Supporting Information S4). Subcellular localization was predicted using TargetP 1.1 to infer the potential biological role of each transporter. *PvHMA*1, -6, and -8 were predicted to localize to the chloroplast, while *PvHMA*5 and -10 were associated with the mitochondrial membrane (Supporting Information S4). *PvHMA*2 was predicted to be directed toward the secretory pathway, whereas *PvHMA*3, -4, -7, -9, and -11 lacked identifiable targeting signals (Supporting Information S4). Taken together, the variation in gene structure, conserved motifs, predicted physicochemical properties, and subcellular localization patterns suggests that *PvHMA* transporters play diverse roles in metal homeostasis and possibly participate in distinct molecular and physiological processes within the common beans.

3.2 | Structural modeling reveals potential biological functions of *PvHMA* proteins

To gain further insight into the structural and functional differentiation of *PvHMA* proteins, we selected *PvHMA*1 and *PvHMA*5 for comparative 3D modeling, as they were predicted to localize to different subcellular compartments (chloroplast and mitochondrion, respectively). Despite *PvHMA*5 having a longer amino acid sequence, both proteins shared similar topologies as predicted by DeepTMHMM, featuring eight TM domains and substantial orientation within the interior of their respective organelles (Figure 3A; Supporting Information S5). Protein structure prediction using AlphaFold2 (via the ColabFold platform) generated high-confidence 3D models for both proteins, revealing the organization of TM helices, conserved domains, and relative orientations of intra- and extra-organellar regions (Figure 3B; Supporting Information S6). Both *PvHMA*1 and *PvHMA*5 contained the conserved CPx motif within TM6, a signature of HMA transporters. However, in *PvHMA*1, the canonical cysteine residue was replaced by serine—an alteration also observed in its *A. thaliana* homolog, *AtHMA*1 (Figure 3C). Structural superposition of *PvHMA*1 and *AtHMA*1 revealed a high degree of similarity, with an RMSD of 0.69 Å (Figure 3D; see also Figure 1A). In contrast, structural alignment between *PvHMA*1 and *PvHMA*5 resulted in a higher RMSD value of 1.91 Å (Figure 3E), indicating substantial divergence despite shared membrane topology. This structural differentiation may reflect distinct metal specificity or transport mechanisms.

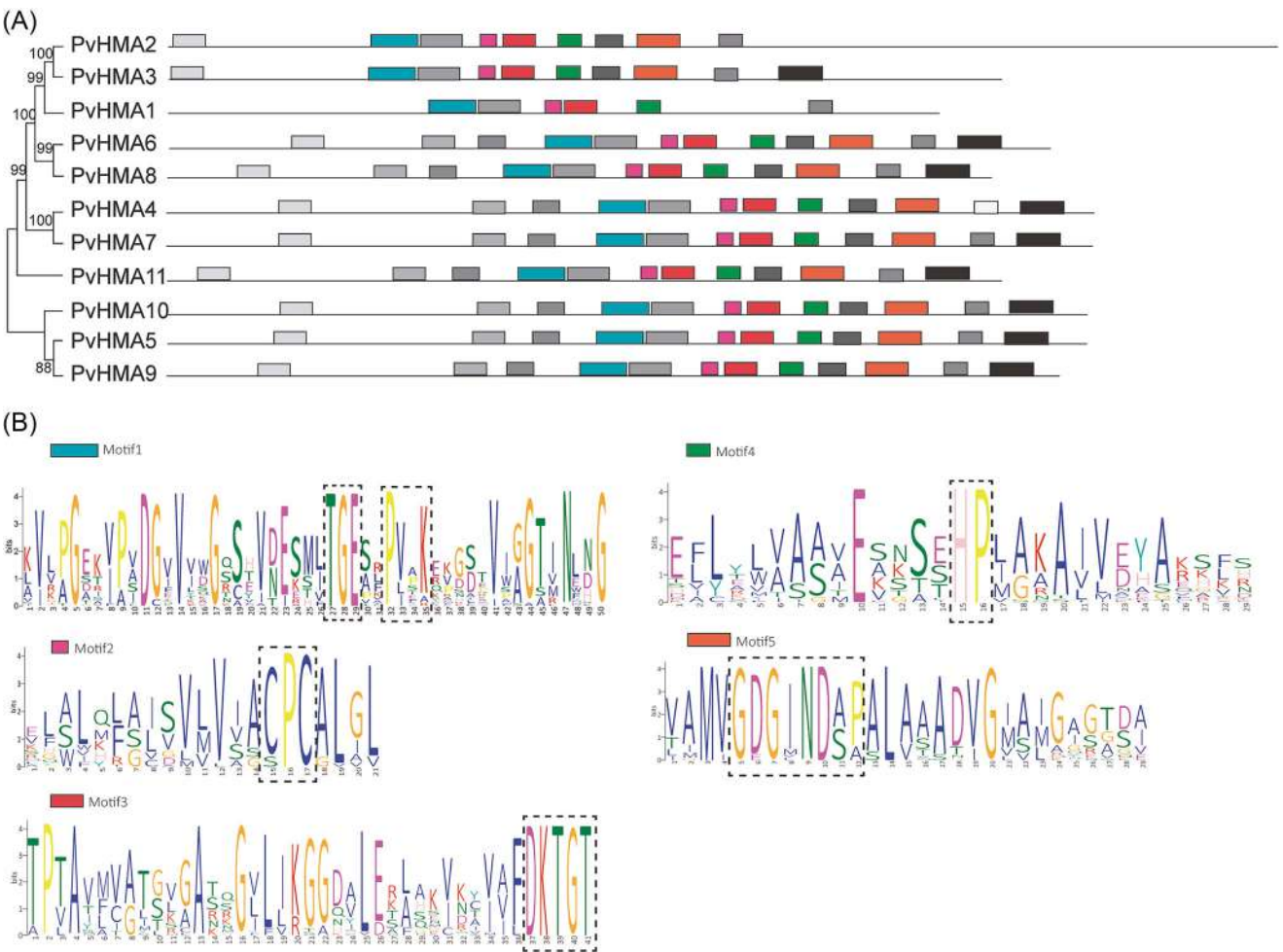


FIGURE 1 Structural characteristics of the heavy metal ATPase (HMA) family in *Phaseolus vulgaris*. (A) The left panel shows a phylogenetic tree constructed from aligned protein sequences using the maximum likelihood method. Bootstrap support values (10,000 replicates) are indicated at internal nodes. The right panel displays conserved motifs identified by the MEME tool, with each colored rectangle representing a distinct motif. Protein sequences are organized according to their phylogenetic groups. (B) Dashed lines highlight motifs that contain conserved sequences characteristic of the HMA family.

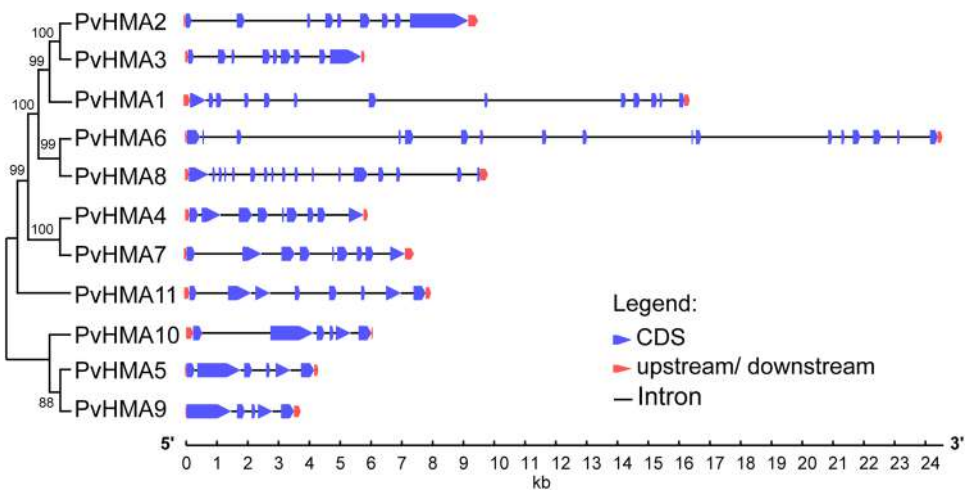


FIGURE 2 Schematic structure of *PvHMA* genes in *Phaseolus vulgaris*. Schematic representation of the exon-intron organization. Black lines indicate introns, blue arrows denote exons, and red arrows represent the 5' and 3' untranslated regions (UTRs). The phylogenetic tree was constructed using the maximum likelihood method, with bootstrap support values (10,000 replicates) shown at internal nodes.

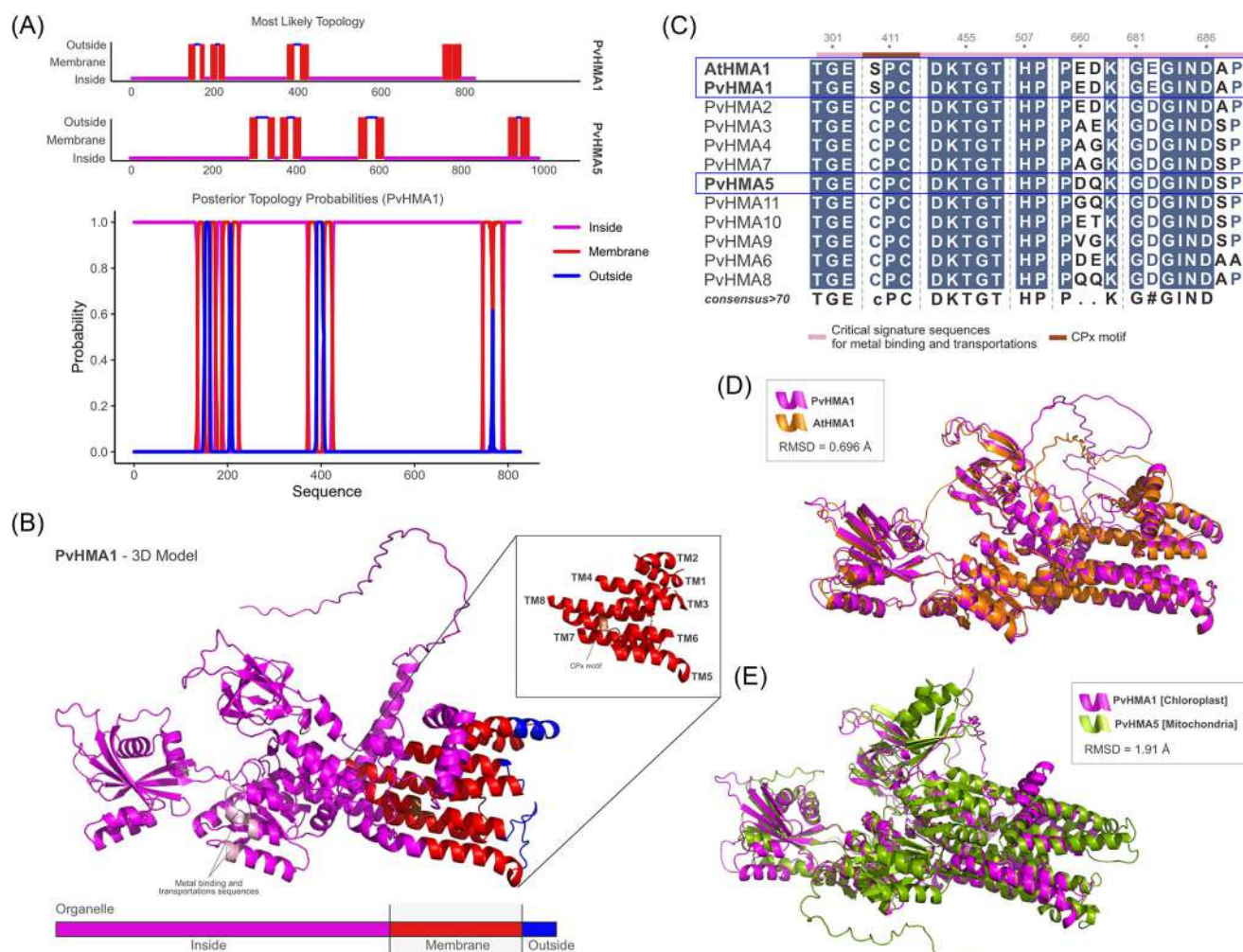


FIGURE 3 Three-dimensional (3D) modeling and structural analysis of PvHMA proteins. (A) Predicted topology of PvHMA1 and PvHMA5, with a probability graph illustrating the transmembrane regions of the PvHMA1 sequence. (B) 3D structure of PvHMA1 based on DeepTMHMM predictions, highlighting eight transmembrane domains and the CPx motif (indicated by a square). (C) Sequence alignment of *Arabidopsis thaliana* HMA1 (AtHMA1) with all PvHMA proteins, emphasizing metal-binding sites and conserved motifs. The blue rectangle marks PvHMA1/AtHMA1 and PvHMA5, which localize to different organelles. (D–E) Structural comparison of 3D models: (D) PvHMA1 versus AtHMA1, and (E) PvHMA1 (chloroplast-localized) versus PvHMA5 (mitochondria-localized). Structural similarity is indicated by root-mean-square deviation (RMSD) (Å) values. HMA, heavy metal ATPase.

3.3 | Comparative phylogenetic analysis of HMA genes in common bean, rice, soybean, and *A. thaliana*

To investigate the evolutionary relationships among HMA transporters across plant species, a phylogenetic tree was constructed using the full-length protein sequences of 11 *P. vulgaris* HMAs (PvHMA1–11), 20 soybean HMAs (GmHMA1–20), eight *A. thaliana* HMAs (AtHMA1–8), and nine rice HMAs (OsHMA1–9). The resulting phylogenetic tree revealed the formation of six major subfamilies, consistent with previous classifications (Fang et al., 2016; Williams & Mills, 2005) (Figure 4). Subfamilies I and II are associated with the transport of Zn, Cd, Pb, and Co. These groups included PvHMA1, -2, and -3, as well as AtHMA1–4 from

A. thaliana, OsHMA1–3 from *O. sativa*, and five soybean homologs (GmHMA5, -13, -14, -18, and -19). The subfamily I contained PvHMA1, AtHMA1, and two soybean sequences (GmHMA5 and 19), suggesting a conserved functional group. The subfamily II positioned PvHMA2 close to AtHMA2, -3, and -4, while PvHMA3 clustered with GmHMA12 (Figure 4). The subfamilies III through VI are primarily associated with Cu transport and potentially Ag, as previously reported (Williams & Mills, 2005). This broader group included PvHMA4, -5, -6, -7, -8, -9, -10, and -11, as well as AtHMA5–8, OsHMA4–9, and 15 soybean HMA genes (Figure 4). Within subfamily III, PvHMA5, -10, and -11 grouped closely with AtHMA5, while PvHMA9 clustered with OsHMA4 and GmHMA3 and -17. Subfamily IV included PvHMA4 and PvHMA7 alongside AtHMA7 and

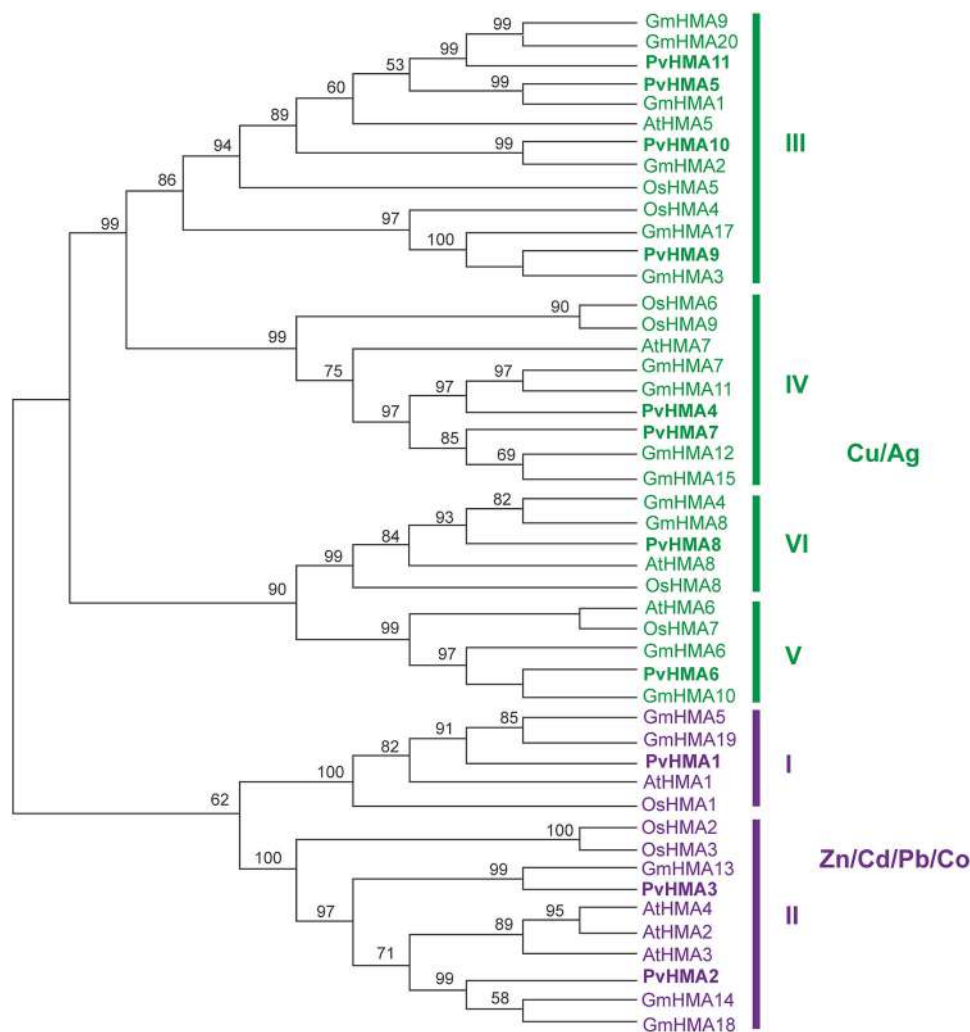


FIGURE 4 Phylogenetic relationships among heavy metal ATPase (HMA) protein sequences. An unrooted phylogenetic tree was constructed using the maximum likelihood method based on predicted HMA protein sequences. Bootstrap values (10,000 replicates) are shown at internal nodes. Sequences are color-coded by the functional group: green for the Cu/Ag clade and purple for the Zn/Cd/Pb/Co clade. *Phaseolus vulgaris* HMA sequences (PvHMA1 to PvHMA11) are highlighted in bold. The soybean genes are identified as GmHMA1–20, *Arabidopsis thaliana* sequences as AtHMA1–7, and rice sequences as OsHMA1–9. Corresponding ID information is provided in Supporting Information S1.

OsHMA6 and -9. Subfamilies V and VI were more distinct, with PvHMA6 closely related to AtHMA6, and PvHMA8 aligning with AtHMA8 and OsHMA8 (Figure 4). These phylogenetic relationships suggest functional conservation and diversification of HMA genes across dicot and monocot lineages, with PvHMA members represented in all six subfamilies, indicating a broad evolutionary distribution and potential functional specialization.

3.4 | Tissue-specific expression patterns and cis-regulatory element analysis of PvHMA genes

To gain further insights into the potential molecular roles of *PvHMA* genes, we examined their *in silico* expression

profiles across various tissues using transcriptome data available. Among all genes analyzed, *PvHMA9* and *PvHMA10* exhibited the lowest transcriptional activity across tissues (Figure 5). In contrast, *PvHMA3* and *PvHMA4* showed preferential expression in mature pods and roots, while *PvHMA8* displayed a predominantly leaf-specific expression pattern. *PvHMA2* and *PvHMA5* were more highly expressed in stems, fruits, and roots. Broad expression across all tissues was observed for *PvHMA1*, *PvHMA6*, *PvHMA7*, *PvHMA8*, and *PvHMA11*, with particularly high transcript levels for *PvHMA1*, *PvHMA7*, and *PvHMA8*, and moderate expression for *PvHMA6* and *PvHMA11* (Figure 5).

To support the expression data, we performed a promoter analysis of the 1000 bp upstream regions of the *PvHMA* genes, identifying putative cis-regulatory elements using the PlantCARE database. For interpretative clarity,

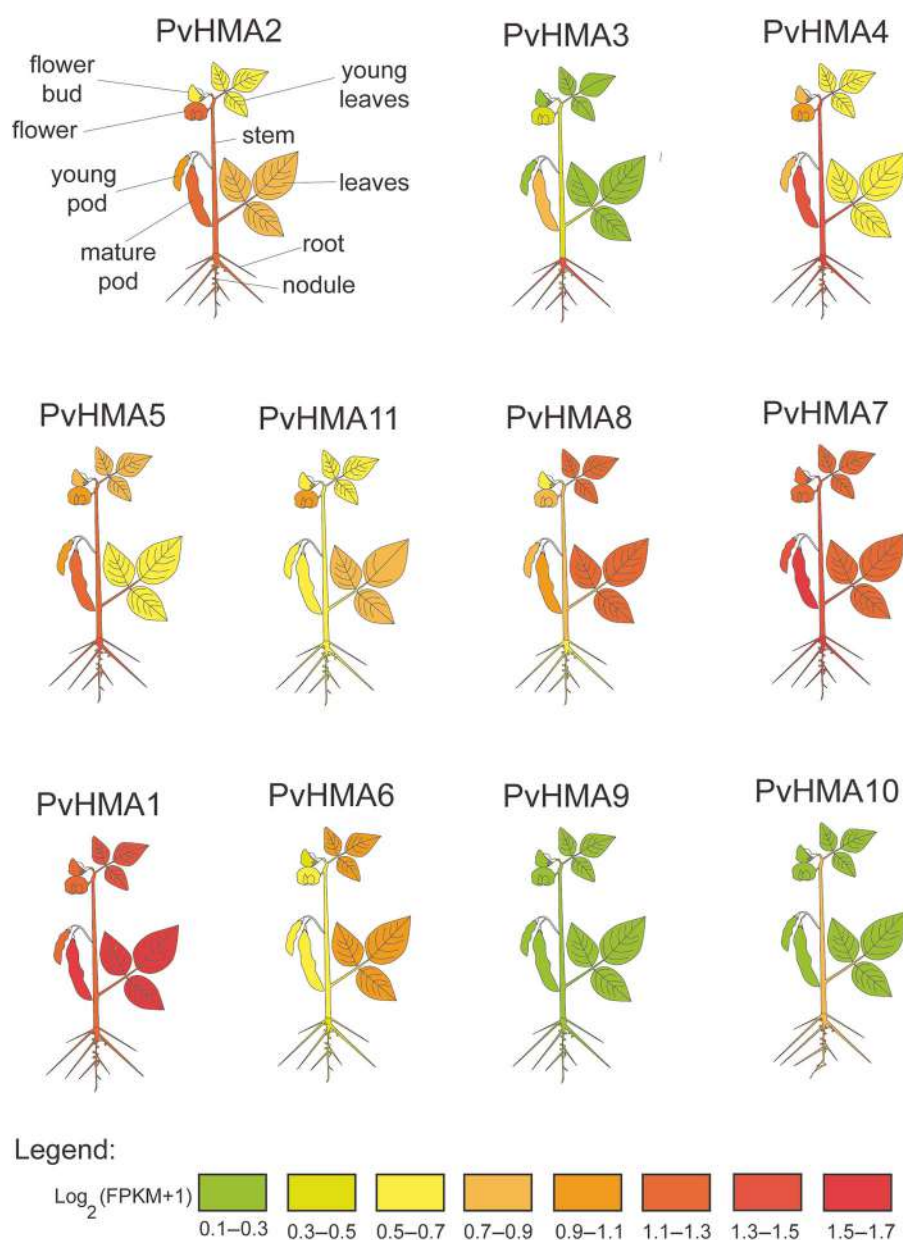


FIGURE 5 Organ-specific expression patterns of *PvHMA* gene during development. RNA-seq data from the Phytozome database (www.phytozome.net) were mapped to *PvHMA* genes to assess their expression across nine organs: flower bud, mature flower, young pod, mature pod, young leaves, mature leaves, stem, root, and nodules. Expression levels are visualized as a heatmap, with color intensity representing $\log_2$ fragments per kilobase million (FPKM) values—ranging from green (low expression) to red (high expression). HMA, heavy metal ATPase.

the elements were categorized into five functional groups: (i) DNA/protein-binding, (ii) biotic/abiotic stress-responsive, (iii) light-responsive, (iv) hormone/metabolism-related, and (v) others. Consistent with their low expression, *PvHMA9* and *PvHMA10* contained the fewest regulatory elements (five and six, respectively) (Supporting Information S7). Notably, *PvHMA3*, which is root-specific, lacked light-responsive elements, while *PvHMA8*, which is highly expressed in leaves, contained the highest number of light-associated elements, suggesting light-dependent transcriptional regulation. Promoters of *PvHMA7*, *PvHMA8*, and *PvHMA11* contained

elements from all five functional categories, supporting their broad expression and potential involvement in diverse physiological processes. Due to the location of *PvHMA6* at the 5' end of scaffold_403, promoter sequence data could not be retrieved, preventing regulatory analysis for this gene.

In summary, *PvHMA1*, *PvHMA2*, *PvHMA5*, *PvHMA6*, *PvHMA7*, and *PvHMA11* exhibit broad expression across multiple tissues, indicating roles in general HM homeostasis. In contrast, *PvHMA3*, *PvHMA4*, *PvHMA8*, and *PvHMA10* show more restricted, tissue-specific expression patterns, pointing to specialized functions in metal transport and reg-

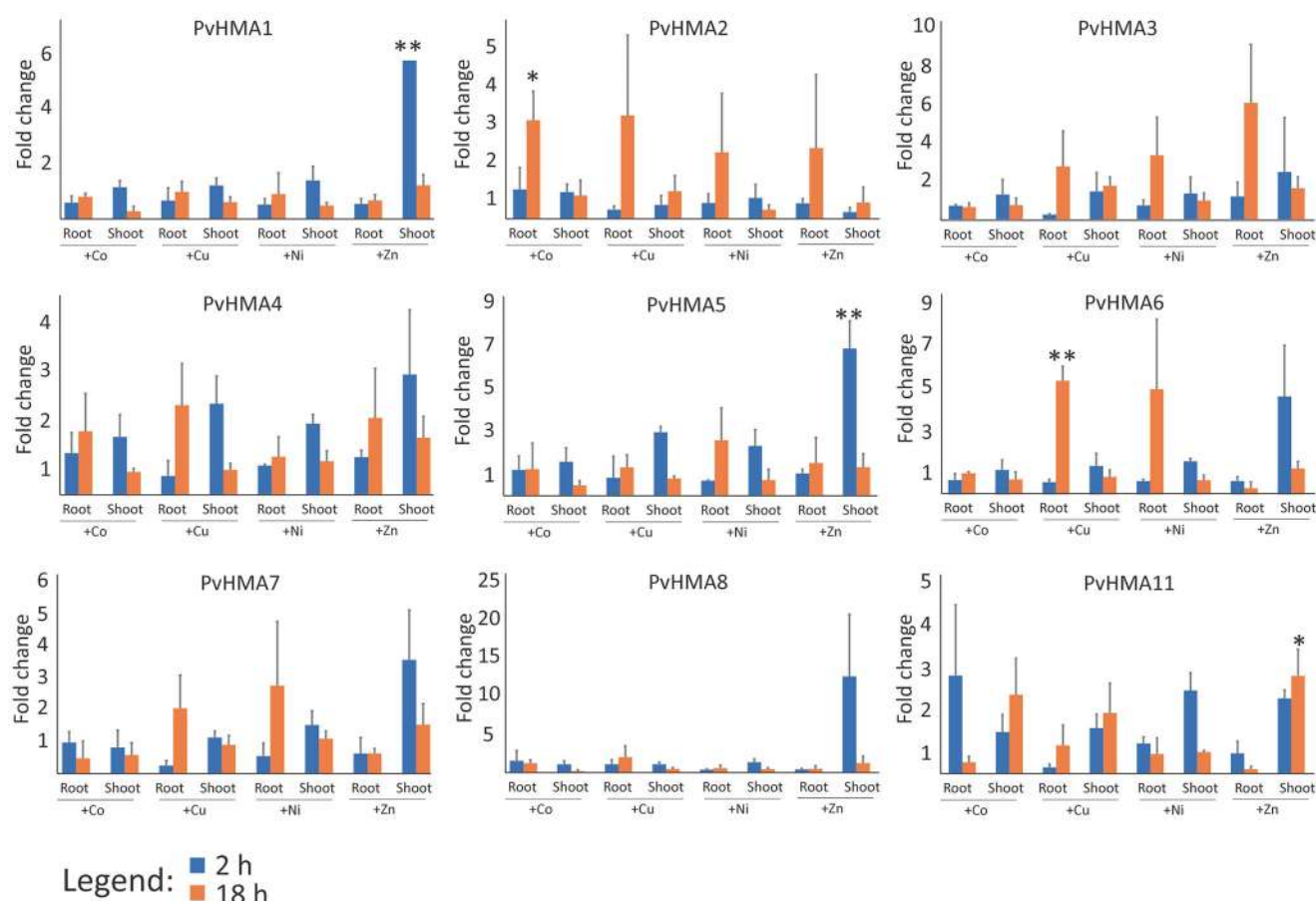


FIGURE 6 Expression of *PvHMA* genes in response to heavy metal exposure. Transcript levels of *PvHMA* genes were measured by quantitative real-time PCR (RT-qPCR) in roots and aerial parts of plants treated with Co (50 μ M), Cu (100 μ M), Ni (150 μ M), or Zn (200 μ M) for 2 h (blue bars) and 18 h (orange bars). Expression values were normalized to a constitutive reference gene, and results represent the mean of two to three biological replicates (see Section 2), each with three technical repetitions. At least two plants were used per biological replicate. Error bars indicate standard error of the mean. Statistical significance was determined using a *t*-test: * $p < 0.1$; ** $p < 0.05$. HMA, heavy metal ATPase.

ulation. These expression trends closely align with the type and abundance of cis-regulatory elements identified in their promoters.

3.5 | Expression profiles of *PvHMA* genes in response to heavy metal stresses

To assess both early-transient and delayed gene inductions, the transcriptional response of *PvHMA* genes to HM exposure was measured at 2 and 18 h after treatment. The RT-qPCR was performed on root and shoot tissues following treatments with Zn^{2+} , Cu^{2+} , Co^{2+} , and Ni^{2+} . Of the 11 *PvHMA* genes analyzed, nine were expressed under at least one metal treatment, while *PvHMA9* and *PvHMA10* showed no detectable expression across all conditions (Figure 6). Significant transcriptional induction of *PvHMA1* and *PvHMA5* was observed in shoots after 2 h of Zn^{2+} exposure, with expression levels markedly higher than untreated controls; however, this induction diminished by 18 h. Conversely, *PvHMA11* exhibited a

progressive and sustained increase in expression throughout zinc treatment, indicating a delayed yet persistent response (Figure 6). Under Co^{2+} treatment, *PvHMA2* was specifically upregulated in roots, suggesting involvement in cobalt transport or detoxification mechanisms. Cu^{2+} exposure elicited elevated expression of *PvHMA6*, predominantly in shoots. Notably, none of the *PvHMA* genes responded transcriptionally to Ni^{2+} under the tested conditions. These results demonstrate that *PvHMA* genes are differentially expressed depending on the specific HM and tissue type, indicating that individual transporters have specialized roles, contributing to metal-specific homeostasis and detoxification mechanisms in *P. vulgaris*.

4 | DISCUSSION

The common bean plays a pivotal role in global food security, serving as a major dietary protein and micronutrient source, especially in low- and middle-income countries. Effi-

cient metal homeostasis is essential for its productivity and nutritional value. HMAs are integral membrane transporters involved in the uptake, transport, sequestration, and detoxification of essential and nonessential metals in plants. While these proteins have been characterized in several model and crop species (Fang et al., 2016; Zhiguo et al., 2018), their diversity and function in *P. vulgaris* remained largely unexplored. Broad phylogenetic comparisons revealed that *P. vulgaris* HMA genes group into six distinct clades, consistent with the established HMA subfamily classification in other angiosperms (Baxter et al., 2003; Tehseen et al., 2010). The high conservation of domain structure and motif composition across these groups supports the notion that the ancestral angiosperm genome contained at least six core HMA genes (Axelsen & Palmgren, 2001; Hanikenne & Baurain, 2014).

The HMA1 transporter has been shown to respond to Zn stress in both *A. thaliana* and rice (Kim et al., 2009). Localized at the chloroplast envelope, it protects the organelle from Zn overaccumulation by translocating the metal to the cytoplasm (Kim et al., 2009). HMA1 plays a crucial role in the protective function of Cu/Zn SOD enzyme in above-ground tissues by providing essential metal cofactors for Cu/Zn SOD activity (Boutigny et al., 2014; Kim et al., 2009). Consistent with previous findings, the common bean homolog of HMA1 contained a chloroplast-targeting peptide signal (Supporting Information S4). Its function in Zn homeostasis in chlorophyllous tissues is supported by its differential expression upon Zn stress in shoots and reproductive organs (Figures 5 and 6) and by the presence of light-responsive cis-elements in its promoter region (Supporting Information S7). While HMA1 has been shown to transport Cu in *A. thaliana* actively (Seigneurin-Berny et al., 2006), our analysis did not reveal an expression response to excessive Cu stress in *PvHMA1* (Figure 6). Despite this, *PvHMA1* shares similarities with *AtHMA1* regarding subcellular localization (Supporting Information S4), protein structure (Figure 3D), and Zn transport (Figure 6). However, their roles in Cu homeostasis may differ. We observed a basal expression of *PvHMA1* in roots that were not differentially expressed by metal exposure (Figure 6). Similar findings have been reported in other plants (D. Li et al., 2015; Seigneurin-Berny et al., 2006), suggesting a parallel function of HMA1 in non-green plastids.

In *A. thaliana*, the HMA2 and HMA4 exhibit redundant functions, sharing similar expression patterns upon Cd and Zn overexposure and predominantly localizing in cells surrounding root vascular vessels (Eren & Argüello, 2004; Hussain et al., 2004). Both transporters facilitate the translocation of Zn from root pericycle cells into the vascular system for delivery to above-ground organs (Hanikenne & Bouché, 2023). Our phylogenetic analysis revealed that a single gene represents the closest homologs of *AtHMA2* and *AtHMA4* in the common bean, *PvHMA2* (Figure 4). Like *AtHMA2/4*,

PvHMA2 is associated with the secretory pathway (Supporting Information S4), translocating metals from the cytoplasm to the extracellular environment. *PvHMA2* exhibited strong expression in vascular tissues and roots (Figure 5) with an overall tendency to be upregulated upon metal treatments (Figure 6). In *A. thaliana*, high levels of *AtHMA2* mRNA were detected in plants treated with various metals, including Ag, Co, Cu, Cd, Mn, and Ni. However, only Zn and Cd were shown to increase the metal-dependent phosphorylation and ATPase activity of HMA2 (Eren & Argüello, 2004). Interestingly, in common beans, *PvHMA2* expression was detected by 40 nm ZnO nanoparticle treatment (da Cruz et al., 2019). These findings suggest that *PvHMA2* plays a role in Zn homeostasis similar to *AtHMA2/4*, although further studies are needed to clarify how different metal species influence its ATPase activity and transport function. Altogether, our phylogenetic and transcriptional data support the hypothesis that *PvHMA2* shares conserved biological functions with its Arabidopsis homologs.

Cu is an essential micronutrient for numerous physiological processes in plants, with roots serving as the primary site for its uptake. Inadequate Cu availability disrupts normal root and leaf development, reduces chlorophyll synthesis, and impairs photosynthetic performance (Hong et al., 2015). Our results indicate that *PvHMA6* plays a role in Cu homeostasis in *P. vulgaris*. This gene displayed strong expression in root tissues and moderate expression in photosynthetic organs (Figure 5), and its transcription was significantly induced following Cu exposure (Figure 6). These patterns suggest a functional role in both Cu uptake from the soil and internal redistribution. In support of this, *PvHMA6* similar to its closest homologs in *A. thaliana* (*AtHMA8*) (Figure 4) and *M. truncatula* (*MtHMA4*) (Ma et al., 2021) are chloroplast-localized transporters (Supporting Information S4). In addition, *AtHMA8* and *MtHMA4* are reported to respond to Cu stress (Ma et al., 2021; Sautron et al., 2016). *AtHMA8*, in particular, has been shown to supply Cu to chloroplastic enzymes such as Cu/Zn SOD and plastocyanin, which are crucial for oxidative stress defense and the electron transport chain during photosynthesis (Abdel-Ghany et al., 2005). Taken together, these findings support a dual role for *PvHMA6* in facilitating Cu acquisition at the root level and in supplying Cu to chloroplasts to sustain photosynthetic function.

Our investigation indicates that *PvHMA5* and *PvHMA11* are involved in Zn homeostasis in common beans. Both genes responded to HM exposure, with *PvHMA5* showing rapid increase in transcript expression within 2 hours, while *PvHMA11* displayed a delayed response, becoming evident after 18 h (Figure 6). Based on their high sequence similarity to *GmHMA1/9* in soybean and predicted localization signals, *PvHMA5* and *PvHMA11* are likely targeted to the mitochondrial membrane (Fang et al., 2016; Supporting Information S4). These homologous transporters in soybean are

known to respond to Cd, prompting investigation into whether PvHMA5 and PvHMA11 share this sensitivity. Interestingly, under our experimental conditions, neither gene responded to Cd, suggesting that these transporters may have diverged functionally in *P. vulgaris*. Unlike their closest homolog in *A. thaliana*, AtHMA5—a plasma membrane Cu transporter that is not responsive to Cd, Zn, or Fe (Andrés-Colás et al., 2006)—PvHMA5 and PvHMA11 appear to be specifically responsive to Zn. These findings suggest that PvHMA5 and PvHMA11 have evolved distinct roles in Zn regulation in common beans, differing from their counterparts in both soybean and *A. thaliana*.

5 | CONCLUSION

Our findings offered a comprehensive perspective on the HMA gene family in common beans. The 11 transporters observed in this study were categorized into six distinct clusters, each exhibiting a consistent gene structure and distribution of motifs. The expression studies provide initial insights into the role of HMA in the absorption, storage, and transit of metals in the roots, stems, and leaves of the common bean. Our findings will help to understand the biological function of the individuals involved in metal balance during the growth of *P. vulgaris*.

AUTHOR CONTRIBUTIONS

Wenderson Felipe Costa Rodrigues: Data curation; formal analysis; investigation; methodology; validation; visualization; writing—original draft; writing—review and editing. **Laura Oliveira Pires:** Data curation; formal analysis; investigation; methodology; validation; visualization; writing—original draft; writing—review and editing. **Lucas Roberto Oliveira:** Data curation; formal analysis; investigation; methodology; writing—original draft. **Juliane K. Ishida:** Conceptualization; data curation; formal analysis; funding acquisition; methodology; project administration; resources; supervision; validation; visualization; writing—original draft; writing—review and editing.

ACKNOWLEDGMENTS

This study was funded by the Serrapilheira Institute (R-2111-40087), the Research Support Foundation of the State of Minas Gerais (FAPEMIG) (APQ-02786-21), and the Science Without Borders Program from the Brazilian Federal Government (88887.092433/2015-00). Additionally, this study received partial funding from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brazil (CAPES), Finance Code 001.

The Article Processing Charge for the publication of this research was funded by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) (ROR identifier: 00x0ma614).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

ORCID

Juliane K. Ishida  <https://orcid.org/0000-0002-0956-7578>

REFERENCES

- Abd-Alla, M. H., Al-Amri, S. M., & El-Enany, A.-W. E. (2023). Enhancing rhizobium–legume symbiosis and reducing nitrogen fertilizer use are potential options for mitigating climate change. *Agriculture*, 13(11), 2092. <https://doi.org/10.3390/agriculture13112092>
- Abdel-Ghany, S. E., Müller-Moulé, P., Niyogi, K. K., Pilon, M., & Shikanai, T. (2005). Two P-type ATPases are required for copper delivery in *Arabidopsis thaliana* chloroplasts. *The Plant Cell*, 17, 1233–1251. <https://doi.org/10.1105/tpc.104.030452>
- Alloway, B. J. (2013). Heavy metals and metalloids as micronutrients for plants and animals. In B. J. Alloway (Ed.), *Heavy metals in soils: Trace metals and metalloids in soils and their bioavailability* (pp. 195–209). Springer.
- Andrés-Colás, N., Sancenón, V., Rodríguez-Navarro, S., Mayo, S., Thiele, D. J., Ecker, J. R., Puig, S., & Peñarrubia, L. (2006). The *Arabidopsis* heavy metal P-type ATPase HMA5 interacts with metal-chaperones and functions in copper detoxification of roots. *The Plant Journal*, 45, 225–236. <https://doi.org/10.1111/j.1365-313X.2005.02601.x>
- Argüello, J. M., Eren, E., & González-Guerrero, M. (2007). The structure and function of heavy metal transport P_{1B}-ATPases. *Biometals*, 20, 233–248. <https://doi.org/10.1007/s10534-006-9055-6>
- Atafar, Z., Mesdaghinia, A., Nouri, J., Homae, M., Yunesian, M., Ahmadi Moghaddam, M., & Mahvi, A. H. (2010). Effect of fertilizer application on soil heavy metal concentration. *Environmental Monitoring and Assessment*, 160, 83–89. <https://doi.org/10.1007/s10661-008-0659-x>
- Axelsen, K. B., & Palmgren, M. G. (2001). Inventory of the superfamily of P-type ion pumps in *Arabidopsis*. *Plant Physiology*, 126, 696–706. <https://doi.org/10.1104/pp.126.2.696>
- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., & Noble, W. S. (2009). MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Research*, 37, W202–W208. <https://doi.org/10.1093/nar/gkp335>
- Batool, T. S., Aslam, R., Gul, A., Paracha, R. Z., Ilyas, M., De Abreu, K., Munir, F., Amir, R., & Williams, L. E. (2023). Genome-wide analysis of heavy metal ATPases (HMAs) in Poaceae species and their potential role against copper stress in *Triticum aestivum*. *Scientific Reports*, 13, 7551. <https://doi.org/10.1038/s41598-023-32023-7>
- Baxter, I., Tchieu, J., Sussman, M. R., Boutry, M., Palmgren, M. G., Gribskov, M., Harper, J. F., & Axelsen, K. B. (2003). Genomic comparison of P-type ATPase ion pumps in *Arabidopsis* and rice. *Plant Physiology*, 132, 618–628. <https://doi.org/10.1104/pp.103.021923>
- Bjellqvist, B., Hughes, G. J., Pasquali, C., Paquet, N., Ravier, F., Sanchez, J.-C., Frutiger, S., & Hochstrasser, D. (1993). The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis*, 14, 1023–1031. <https://doi.org/10.1002/elps.11501401163>
- Boutigny, S., Sautron, E., Finazzi, G., Rivasseau, C., Frelet-Barrand, A., Pilon, M., Rolland, N., & Seigneurin-Berny, D. (2014). HMA1 and PAA1, two chloroplast-envelope P_{1B}-ATPases, play distinct roles in chloroplast copper homeostasis. *Journal of Experimental Botany*, 65, 1529–1540. <https://doi.org/10.1093/jxb/eru020>

- Da Cruz, T. N. M., Savassa, S. M., Montanha, G. S., Ishida, J. K., De Almeida, E., Tsai, S. M., Lavres Junior, J., & Pereira De Carvalho, H. W. (2019). A new glance on root-to-shoot in vivo zinc transport and time-dependent physiological effects of ZnSO_4 and ZnO nanoparticles on plants. *Scientific Reports*, 9, Article 10416. <https://doi.org/10.1038/s41598-019-46796-3>
- Eastman, P., Swails, J., Chodera, J. D., McGibbon, R. T., Zhao, Y., Beauchamp, K. A., Wang, L. P., Simmonett, A. C., Harrigan, M. P., Stern, C. D., Wiewiora, R. P., Brooks, B. R., & Pande, V. S. (2017). OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Computational Biology*, 13, e1005659. <https://doi.org/10.1371/journal.pcbi.1005659>
- Emanuelsson, O., Brunak, S., von Heijne, G., & Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols*, 2, 953–971. <https://doi.org/10.1038/nprot.2007.131>
- Eren, E., & Argüello, J. M. (2004). Arabidopsis HMA2, a divalent heavy metal-transporting $\text{P}_{1\text{B}}$ -type ATPase, is involved in cytoplasmic Zn^{2+} homeostasis. *Plant Physiology*, 136, 3712–3723. <https://doi.org/10.1104/pp.104.046292>
- Fang, X., Wang, L., Deng, X., Wang, P., Ma, Q., Nian, H., Wang, Y., & Yang, C. (2016). Genome-wide characterization of soybean $\text{P}_{1\text{B}}$ -ATPases gene family provides functional implications in cadmium responses. *BMC Genomics*, 17, Article 376. <https://doi.org/10.1186/s12864-016-2730-2>
- Farley, R. A. (2012). Active ion transport by ATP-driven ion pumps. In N. Sperelakis (Ed.), *Cell physiology source book* (pp. 167–177). Academic Press.
- Hallgren, J., Tsigirgos, K. D., Damgaard Pedersen, M., Juan, J., Armenteros, A., Marcatili, P., Nielsen, H., Krogh, A., & Winther, O. (2022). DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. *BioRxiv*. <https://doi.org/10.1101/2022.04.08.487609>
- Hanikenne, M., & Baurain, D. (2014). Origin and evolution of metal P-type ATPases in Plantae (Archaeplastida). *Frontiers in Plant Science*, 4, Article 544. <https://doi.org/10.3389/fpls.2013.00544>
- Hanikenne, M., & Bouché, F. (2023). Iron and zinc homeostasis in plants: A matter of trade-offs. *Journal of Experimental Botany*, 74(18), 5426–5430. <https://doi.org/10.1093/jxb/erad304>
- Hawkes, S. J. (1997). What is a heavy metal? *Journal of Chemical Education*, 74, 1374. <https://doi.org/10.1021/ed074p1374>
- Hirayama, T., Kieber, J. J., Hirayama, N., Kogan, M., Guzman, P., Nourizadeh, S., Alonso, J. M., Dailey, W. P., Dancis, A., & Ecker, J. R. (1999). RESPONSIVE-TO-ANTAGONIST1, a Menkes/Wilson disease-related copper transporter, is required for ethylene signaling in *Arabidopsis*. *Cell*, 97, 383–393. [https://doi.org/10.1016/S0092-8674\(00\)80747-3](https://doi.org/10.1016/S0092-8674(00)80747-3)
- Hong, J., Rico, C. M., Zhao, L. J., Adeleye, A. S., Keller, A. A., Peralta-Videa, J. R., & Gardea-Torresdey, J. L. (2015). Toxic effects of copper-based nanoparticles or compounds to lettuce (*Lactuca sativa*) and alfalfa (*Medicago sativa*). *Environmental Science. Processes & Impacts*, 17, 177–185.
- Hu, B., Jin, J., Guo, A.-Y., Zhang, H., Luo, J., & Gao, G. (2015). GSDS 2.0: An upgraded gene feature visualization server. *Bioinformatics*, 31, 1296–1297.
- Hussain, D., Haydon, M. J., Wang, Y., Wong, E., Sherson, S. M., Young, J., Camakaris, J., Harper, J. F., & Cobbett, C. S. (2004). P-type ATPase heavy metal transporters with roles in essential zinc homeostasis in Arabidopsis. *The Plant cell*, 16, 1327–1339. <https://doi.org/10.1105/tpc.020487>
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596, 583–589. <https://doi.org/10.1038/s41586-021-03819-2>
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772–780. <https://doi.org/10.1093/molbev/mst010>
- Kim, Y.-Y., Choi, H., Segami, S., Cho, H.-T., Martinoia, E., Maeshima, M., & Lee, Y. (2009). AtHMA1 contributes to the detoxification of excess Zn(II) in Arabidopsis. *The Plant Journal*, 58, 737–753. <https://doi.org/10.1111/j.1365-3113X.2009.03818.x>
- Kühlbrandt, W. (2004). Biology, structure and mechanism of P-type ATPases. *Nature Reviews Molecular Cell Biology*, 5, 282–295. <https://doi.org/10.1038/nrm1354>
- Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 33, 1870–1874. <https://doi.org/10.1093/molbev/msw054>
- Lee, S., Kim, Y.-Y., Lee, Y., & An, G. (2007). Rice $\text{P}_{1\text{B}}$ -type heavy-metal ATPase, OsHMA9, is a metal efflux protein. *Plant Physiology*, 145, 831–842. <https://doi.org/10.1104/pp.107.102236>
- Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., Rouzé, P., & Rombauts, S. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Research*, 30, 325–327. <https://doi.org/10.1093/nar/30.1.325>
- Letunic, I., Khedkar, S., & Bork, P. (2021). SMART: Recent updates, new developments and status in 2020. *Nucleic Acids Research*, 49(D1), D458–D460. <https://doi.org/10.1093/nar/gkaa937>
- Li, D., Xu, X., Hu, X., Liu, Q., Wang, Z., Zhang, H., Wang, H., Wei, M., Wang, H., Liu, H., & Li, C. (2015). Genome-wide analysis and heavy metal-induced expression profiling of the HMA gene family in *Populus trichocarpa*. *Frontiers in Plant Science*, 6, Article 1149. <https://doi.org/10.3389/fpls.2015.01149>
- Li, N., Xiao, H., Sun, J., Wang, S., Wang, J., Chang, P., Zhou, X., Lei, B., Lu, K., Luo, F., Shi, X., & Li, J. (2018). Genome-wide analysis and expression profiling of the HMA gene family in *Brassica napus* under Cd stress. *Plant and Soil*, 426, 365–381. <https://doi.org/10.1007/s11104-018-3637-2>
- Ma, Y., Wei, N., Wang, Q., Liu, Z., & Liu, W. (2021). Genome-wide identification and characterization of the heavy metal ATPase (HMA) gene family in *Medicago truncatula* under copper stress. *International Journal of Biological Macromolecules*, 193, 893–902. <https://doi.org/10.1016/j.ijbiomac.2021.10.197>
- Mirdita, M., Schütze, K., Moriawaki, Y., Heo, L., Ovchinnikov, S., & Steinegger, M. (2022). ColabFold: Making protein folding accessible to all. *Nature Methods*, 19(6), 679–682. <https://doi.org/10.1038/s41592-022-01488-1>
- Morel, M., Crouzet, J., Gravot, A., Auroy, P., Leonhardt, N., Vavasseur, A., & Richaud, P. (2009). AtHMA3, a $\text{P}_{1\text{B}}$ -ATPase allowing Cd/Zn/Co/Pb vacuolar storage in Arabidopsis. *Plant physiology*, 149, 894–904. <https://doi.org/10.1104/pp.108.130294>
- Moreno, I., Norambuena, L., Maturana, D., Toro, M., Vergara, C., Orellana, A., Zurita-Silva, A., & Ordenes, V. R. (2008). AtHMA1

- is a thapsigargin-sensitive Ca^{2+} /heavy metal pump. *The Journal of Biological Chemistry*, 283, 9633–9641. <https://doi.org/10.1074/jbc.M800736200>
- Mortvedt, J. J. (1996). Heavy metal contaminants in inorganic and organic fertilizers. *Fertilizer Research*, 43, 55–61. <https://doi.org/10.1007/BF00747683>
- Murube, E., Beleggia, R., Pacetti, D., Nartea, A., Frascarelli, G., Lanzavecchia, G., Bellucci, E., Nanni, L., Gioia, T., Marciello, U., Esposito, S., Foresi, G., Logozzo, G., Frega, G. N., Bitocchi, E., & Papa, R. (2021). Characterization of nutritional quality traits of a common bean germplasm collection. *Foods*, 10(7), 1572. <https://doi.org/10.3390/foods10071572>
- Navarrete, A. A., Mellis, E. V., Escalas, A., Lemos, L. N., Junior, J. L., Quaggio, J. A., Zhou, J., & Tsai, S. M. (2017). Zinc concentration affects the functional groups of microbial communities in sugarcane-cultivated soil. *Agriculture, Ecosystems & Environment*, 236, 187–197. <https://doi.org/10.1016/j.agee.2016.12.009>
- Nouet, C., Charlier, J.-B., Carnol, M., Bosman, B., Farnir, F., Motte, P., & Hanikenne, M. (2015). Functional analysis of the three HMA4 copies of the metal hyperaccumulator *Arabidopsis halleri*. *Journal of Experimental Botany*, 66, 5783–5795. <https://doi.org/10.1093/jxb/erv280>
- Palmgren, M. G., & Axelsen, K. B. (1998). Evolution of P-type ATPases. *Biochimica et Biophysica Acta*, 1365, 37–45. [https://doi.org/10.1016/S0005-2728\(98\)00041-3](https://doi.org/10.1016/S0005-2728(98)00041-3)
- Pedersen, P. L., & Carafoli, E. (1987). Ion motive ATPases. I. Ubiquity, properties, and significance to cell function. *Trends in Biochemical Sciences*, 12, 146–150. [https://doi.org/10.1016/0968-0004\(87\)90071-5](https://doi.org/10.1016/0968-0004(87)90071-5)
- Rascio, N., & Navari-Izzo, F. (2011). Heavy metal hyperaccumulating plants: How and why do they do it? And what makes them so interesting? *Plant Science*, 180, 169–181. <https://doi.org/10.1016/j.plantsci.2010.08.016>
- Reich, M., Liefeld, T., Gould, J., Lerner, J., Tamayo, P., & Mesirov, J. P. (2006). GenePattern 2.0. *Nature Genetics*, 38, 500–501. <https://doi.org/10.1038/ng0506-500>
- Satoh-Nagasawa, N., Mori, M., Nakazawa, N., Kawamoto, T., Nagato, Y., Sakurai, K., Takahashi, H., Watanabe, A., & Akagi, H. (2012). Mutations in rice (*Oryza sativa*) heavy metal ATPase 2 (OSHMA2) restrict the translocation of zinc and cadmium. *Plant and Cell Physiology*, 53, 213–224. <https://doi.org/10.1093/pcp/pcr166>
- Sautron, E., Giustini, C., Dang, T., Moyet, L., Salvi, D., Crouzy, S., Rolland, N., Catty, P., & Seigneurin-Berny, D. (2016). Identification of two conserved residues involved in copper release from chloroplast PIB-1-ATPases. *Journal of Biological Chemistry*, 291(38), 20136–20148. <https://doi.org/10.1074/jbc.M115.706978>
- Seigneurin-Berny, D., Gravot, A., Auroy, P., Mazard, C., Kraut, A., Finazzi, G., Grunwald, D., Rappaport, F., Vavasseur, A., Joyard, J., Richaud, P., & Rolland, N. (2006). HMA1, a new Cu-ATPase of the chloroplast envelope, is essential for growth under adverse light conditions. *The Journal of biological chemistry*, 281, 2882–2892. <https://doi.org/10.1074/jbc.M508333200>
- Shikanai, T., Müller-Moulé, P., Munekage, Y., Niyogi, K. K., & Pilon, M. (2003). PAA1, a P-type ATPase of Arabidopsis, functions in copper transport in chloroplasts. *The Plant cell*, 15, 1333–1346. <https://doi.org/10.1105/tpc.011817>
- Sinclair, S. A., Sherson, S. M., Jarvis, R., Camakaris, J., & Cobbett, C. S. (2007). The use of the zinc-fluorophore, Zinpyr-1, in the study of zinc homeostasis in *Arabidopsis* roots. *New Phytologist*, 174, 39–45. <https://doi.org/10.1111/j.1469-8137.2007.02030.x>
- Sperotto, R. A., & Ricachenevsky, F. K. (2017). Common bean Fe bio-fortification using model species' lessons. *Frontiers in Plant Science*, 8, Article 2187. <https://doi.org/10.3389/fpls.2017.02187>
- Takahashi, R., Bashir, K., Ishimaru, Y., Nishizawa, N. K., & Nakanishi, H. (2012). The role of heavy-metal ATPases, HMAs, in zinc and cadmium transport in rice. *Plant Signaling & Behavior*, 7, 1605–1607. <https://doi.org/10.4161/psb.22454>
- Tehseen, M., Cairns, N., Sherson, S., & Cobbett, C. S. (2010). Metallochaperone-like genes in *Arabidopsis thaliana*. *Metallomics*, 2, 556–564. <https://doi.org/10.1039/c003484c>
- Teufel, F., Almagro Armenteros, J. J., Johansen, A. R., Gíslason, M. H., Pihl, S. I., Tsirigos, K. D., Winther, O., Brunak, S., von Heijne, G., & Nielsen, H. (2022). SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nature Biotechnology*, 40(7), 1023–1025. <https://doi.org/10.1038/s41587-021-01156-3>
- Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2002). *Multiple sequence alignment using ClustalW and ClustalX*. Current Protocols in Bioinformatics.
- Ueno, D., Yamaji, N., Kono, I., Huang, C. F., Ando, T., Yano, M., & Ma, J. F. (2010). Gene limiting cadmium accumulation in rice. *Proceedings of the National Academy of Sciences of the United States of America*, 107(38), 16500–16505. <https://doi.org/10.1073/pnas.1005396107>
- Williams, L. E., & Mills, R. F. (2005). P_{1B}-ATPases—An ancient family of transition metal pumps with diverse functions in plants. *Trends in Plant Science*, 10, 491–502. <https://doi.org/10.1016/j.tplants.2005.08.008>
- Williams, L. E., Pittman, J. K., & Hall, J. L. (2000). Emerging mechanisms for heavy metal transport in plants. *Biochimica et Biophysica Acta (BBA)—Biomembranes*, 1465, 104–126. [https://doi.org/10.1016/S0005-2736\(00\)00133-4](https://doi.org/10.1016/S0005-2736(00)00133-4)
- Zhang, L., Wu, J., Tang, Z., Huang, X. Y., Wang, X., Salt, D. E., & Zhao, F. J. (2019). Variation in the *BrHMA3* coding region controls natural variation in cadmium accumulation in *Brassica rapa* vegetables. *Journal of Experimental Botany*, 70(20), 5865–5878. <https://doi.org/10.1093/jxb/erz310>
- Zhang, Z., Yu, Q., Du, H., Ai, W., Yao, X., Mendoza-Cózatl, D. G., & Qiu, B. (2016). Enhanced cadmium efflux and root-to-shoot translocation are conserved in the hyperaccumulator *Sedum alfredii* (Crassulaceae family). *FEBS Letters*, 590(12), 1757–1764. <https://doi.org/10.1002/1873-3468.12225>
- Zhiguo, E., Tingting, L., Chen, C., & Lei, W. (2018). Genome-wide survey and expression analysis of P_{1B}-ATPases in rice, maize and sorghum. *Rice Science*, 25, 208–217. <https://doi.org/10.1016/j.rsci.2018.06.004>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Rodrigues, W. F. C., Pires, L. O., Oliveira, L. R., & Ishida, J. K. (2025). Comprehensive characterization of HMA transporters in common bean: Tissue-specific expression and response to metal stress. *Crop Science*, 65, e70188. <https://doi.org/10.1002/csc2.70188>

4.4. DISCUSSÃO

A compreensão integrada dos processos evolutivos e funcionais que sustentam a aquisição e a homeostase de metais em plantas revela um panorama complexo no qual inovação evolutiva profunda, especialização funcional e plasticidade estrutural convergem para moldar a fisiologia mineral em diferentes linhagens vegetais. A análise da evolução dos transportadores ZIP/IRT mostra que a capacidade de capturar ferro reduzido não segue uma trajetória linear ou singular na história das plantas verdes, mas emerge repetidamente em linhagens distintas (Rodrigues et al., 2023; Cao; Zhang; Zou, 2025). Essa convergência funcional indica que a limitação de ferro no ambiente constituiu uma pressão seletiva consistente ao longo da evolução, capaz de promover soluções moleculares independentes baseadas em rearranjos estruturais semelhantes. Essa constatação reforça a ideia de que sistemas de aquisição de nutrientes essenciais podem surgir ou ser remodelados diversas vezes quando demandas ecológicas permanecem constantes.

Ao considerar a família HMA, observa-se um padrão evolutivo distinto, caracterizado não por convergência, mas por diversificação funcional dentro de um arcabouço estrutural altamente conservado (Hanikenne; Baurain, 2014). A arquitetura modular das HMAs, que combina domínio citosólico ATPase, sítios de coordenação metálica e múltiplas hélices transmembrana, permite a evolução de especializações fisiológicas sem comprometer o mecanismo central de transporte (Sánchez-Pérez et al., 2025). Essa plasticidade explica por que diferentes HMAs podem atuar na mobilização intracelular de metais essenciais, na detoxificação de elementos tóxicos e no abastecimento de organelas específicas. Em culturas agrícolas, como o feijão, essa diversidade funcional se manifesta em padrões de expressão distintos entre tecidos e em respostas específicas ao estresse metálico (Li; Zhang; Shi, 2024; Shao et al., 2025; Rodrigues et al., 2025), refletindo a importância dessas proteínas para o crescimento, produtividade e tolerância a ambientes desafiadores.

A comparação entre os dois sistemas discutidos nesta seção revela e pressupõe uma complementaridade conceitual importante. Enquanto a aquisição de ferro depende de inovações evolutivas que surgiram em momentos distintos e independentemente em diferentes linhagens, a manutenção interna de metais parece requerer redes altamente integradas de transporte e compartimentalização que evoluíram principalmente por especialização funcional de famílias ancestrais. Essa distinção hipotetiza a existência de dois níveis evolutivos nos processos de nutrição mineral: um nível macroevolutivo, no qual surgem mecanismos de captação completamente novos para enfrentar condições ambientais restritivas, e um nível

microevolutivo, no qual proteínas já estabelecidas são refinadas e adaptadas a demandas celulares específicas, ambos os casos já vistos e discutidos na Seção I sob diferentes cenários.

Sob uma perspectiva mais ampla, os sistemas discutidos reforçam um tema recorrente na evolução das plantas: a capacidade de reutilizar módulos moleculares conservados e adaptá-los a novos contextos fisiológicos. Assim como vias de defesa foram modificadas e, em alguns casos, coaptadas para sustentar o parasitismo, proteínas de transporte metálico também foram reorganizadas ao longo do tempo para garantir eficiência nutricional em ambientes heterogêneos. Nesse sentido, estudos recentes sobre a emergência do metabolismo destacam que inovações bioquímicas frequentemente surgem pela reutilização e reorganização de módulos proteicos antigos associados ao metabolismo redox e ao manuseio de metais, reforçando que a plasticidade estrutural ligada a cofatores metálicos permanece um motor fundamental de inovação evolutiva (Nanda et al., 2025). Essa continuidade entre padrões ancestrais e especializações modernas ajuda a explicar por que diferentes linhagens vegetais foram capazes de desenvolver, de forma independente ou derivada, mecanismos eficientes para captar, distribuir e regular elementos essenciais à vida, passos que devem ter sido fundamentais no seu sucesso adaptativo em todo planeta.

4.5. CONCLUSÃO

A Seção II sintetiza os principais avanços na compreensão da evolução e da funcionalidade dos transportadores envolvidos na aquisição e na homeostase de metais em plantas, demonstrando que esses sistemas são moldados por uma combinação de inovação evolutiva profunda e especialização funcional refinada. A análise da evolução dos transportadores IRT mostra que mecanismos essenciais para a captação de ferro podem surgir várias vezes ao longo da história das plantas, reforçando a relevância das pressões seletivas associadas à disponibilidade limitada desse elemento no ambiente. Em contraste, a diversificação das HMAs reflete a expansão de uma família conservada cuja arquitetura modular permitiu múltiplas especializações, proporcionando às plantas a capacidade de transportar e manejar diferentes metais de forma regulada e eficiente.

Esses dois eixos mostram que sistemas de nutrição mineral não são estáticos, mas dinâmicos, sendo continuamente moldados pela interação entre demandas fisiológicas, condições ambientais e potencial evolutivo das famílias proteicas envolvidas. Assim, a integração das abordagens evolutivas e funcionais discutidas nesta seção oferece uma visão abrangente sobre como plantas desenvolveram e refinaram mecanismos para assegurar sua nutrição mineral, destacando a importância central desses sistemas para a sobrevivência, a adaptação e o desempenho de espécies vegetais em ambientes diversos.

A integração desses resultados com os temas discutidos na Seção I reforça que tanto os sistemas de defesa quanto os mecanismos de aquisição e transporte de metais integram um mesmo conjunto de pressões evolutivas que moldaram a diversificação das plantas ao longo de sua história (resumidos na Figura 1 da Introdução Geral). A emergência de vias capazes de reconhecer sinais químicos, modular respostas imunes ou captar nutrientes essenciais representa soluções paralelas desenvolvidas em resposta a desafios enfrentados desde a conquista do ambiente terrestre, marcada por solos pobres em nutrientes, elevada competição e novos tipos de interações bióticas. A compreensão conjunta desses processos permite reconhecer que a adaptação das plantas envolveu a remodelação de módulos moleculares ancestrais em múltiplas direções, seja para sustentar parasitismo, defesa ou eficiência nutricional. Assim, a leitura integrada das duas seções revela que inovação molecular e evolutiva são narrativas complementares que constituem pilares centrais na forma como plantas estabeleceram sua diversidade atual e se ajustaram a ambientes cada vez mais complexos.

REFERÊNCIAS

- AJEESH KRISHNA, T. P. *et al.* Structure, Function, Regulation and Phylogenetic Relationship of ZIP Family Transporters of Plants. **Frontiers in Plant Science**, v. 11, 27 maio 2020.
- ALONSO, José Ramón. **Parasite plant check**. Disponível em: <<https://mappingignorance.org/2016/08/22/parasite-plant-check/>>. Acesso em: 28 fev. 2026.
- ARNESANO, Fabio *et al.* Metallochaperones and Metal-Transporting ATPases: A Comparative Analysis of Sequences and Structures. **Genome Research**, v. 12, n. 2, p. 255–271, 1 fev. 2002.
- BANDARANAYAKE, Pradeepa C. G.; YODER, John I. Haustorium Initiation and Early Development. In: JOEL, Daniel M.; GRESSEL, Jonathan; MUSSELMAN, Lytton J. (Orgs.). **Parasitic Orobanchaceae**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. p. 61–74.
- BASHIR, Khurram *et al.* Regulating Subcellular Metal Homeostasis: The Key to Crop Improvement. **Frontiers in Plant Science**, v. 7, 5 ago. 2016.
- BEHNKE, Joerg; LAROCHE, Julie. Iron uptake proteins in algae and the role of Iron Starvation-Induced Proteins (ISIPs). **European Journal of Phycology**, v. 55, n. 3, p. 339–360, 2 jul. 2020.
- BOWMAN, John L.; BRIGINSHAW, Liam N.; FLORENT, Stevie N. Evolution and co-option of developmental regulatory networks in early land plants. In: **Current Topics in Developmental Biology**. [S.l.]: Elsevier, 2019. v. 131 p. 35–53.
- CAI, Liming *et al.* Deeply Altered Genome Architecture in the Endoparasitic Flowering Plant *Sapria himalayana* Griff. (Rafflesiaceae). **Current Biology**, v. 31, n. 5, p. 1002–1011.e9, mar. 2021.
- CAO, Zhenqiu; ZHANG, Hongjiu; ZOU, Zhengting. Language models reveal a complex sequence basis for adaptive convergent evolution of protein functions. **Proceedings of the National Academy of Sciences**, v. 122, n. 39, p. e2418254122, 30 set. 2025.
- CHAN, Henry *et al.* The P-Type ATPase Superfamily. **Microbial Physiology**, v. 19, n. 1–2, p. 5–104, 2010.
- CHANG, Mayland; LYNN, David G. The haustorium and the chemistry of host recognition in parasitic angiosperms. **Journal of Chemical Ecology**, v. 12, n. 2, p. 561–579, fev. 1986.
- CHEN, Qingshuai *et al.* From “Dark Matter” to “Star”: Insight Into the Regulation Mechanisms of Plant Functional Long Non-Coding RNAs. **Frontiers in Plant Science**, v. 12, p. 650926, 7 jun. 2021.
- CLARK, James W.; DONOGHUE, Philip C. J. Whole-Genome Duplication and Plant Macroevolution. **Trends in Plant Science**, v. 23, n. 10, p. 933–945, out. 2018.
- CLARKE, Christopher R. *et al.* Molecular Dialog Between Parasitic Plants and Their Hosts. **Annual Review of Phytopathology**, v. 57, n. 1, p. 279–299, 25 ago. 2019.
- CONNORTON, James M.; BALK, Janneke; RODRÍGUEZ-CELMA, Jorge. Iron homeostasis in plants – a brief overview. **Metallomics**, v. 9, n. 7, p. 813–823, 1 jul. 2017.
- COSGROVE, Daniel J. Loosening of plant cell walls by expansins. **Nature**, v. 407, n. 6802, p. 321–326, set. 2000.

CRACIUN, Adrian R. *et al.* Variation in HMA4 gene copy number and expression among *Nocca caerulescens* populations presenting different levels of Cd tolerance and accumulation. **Journal of Experimental Botany**, v. 63, n. 11, p. 4179–4189, 28 jun. 2012.

CUI, Jun. Editorial: lncRNA in plants: function, mechanisms and applications. **Frontiers in Plant Science**, v. 14, p. 1238185, 28 jun. 2023.

CUI, Songkui *et al.* Ethylene signaling mediates host invasion by parasitic plants. **Science Advances**, v. 6, n. 44, p. eabc2385, 30 out. 2020.

DUBEY, Neeraj K. *et al.* Enhanced Host-Parasite Resistance Based on Down-Regulation of *Phelipanche aegyptiaca* Target Genes Is Likely by Mobile Small RNA. **Frontiers in Plant Science**, v. 8, p. 1574, 12 set. 2017.

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.

GONZÁLEZ PLAZA, Juan José. Small RNAs as Fundamental Players in the Transference of Information During Bacterial Infectious Diseases. **Frontiers in Molecular Biosciences**, v. 7, p. 101, 16 jun. 2020.

GOYET, Vincent *et al.* Haustorium Inducing Factors for Parasitic Orobanchaceae. **Frontiers in Plant Science**, v. 10, p. 1056, 29 ago. 2019.

GUERINOT, Mary Lou. The ZIP family of metal transporters. **Biochimica et Biophysica Acta (BBA) - Biomembranes**, v. 1465, n. 1–2, p. 190–198, maio 2000.

HALL, J. L. Transition metal transporters in plants. **Journal of Experimental Botany**, v. 54, n. 393, p. 2601–2613, 29 out. 2003.

HANIKENNE, Marc *et al.* Evolution of metal hyperaccumulation required cis-regulatory changes and triplication of HMA4. **Nature**, v. 453, n. 7193, p. 391–395, maio 2008.

HANIKENNE, Marc; BAURAIN, Denis. Origin and evolution of metal P-type ATPases in Plantae (Archaeplastida). **Frontiers in Plant Science**, v. 4, 2014.

HENRIQUES, Rossana *et al.* Knock-out of Arabidopsis metal transporter gene IRT1 results in iron deficiency accompanied by cell differentiation defects. **Plant Molecular Biology**, v. 50, n. 4/5, p. 587–597, 2002.

HONAAS, Loren A. *et al.* Functional genomics of a generalist parasitic plant: Laser microdissection of host-parasite interface reveals host-specific patterns of parasite gene expression. **BMC Plant Biology**, v. 13, n. 1, p. 9, dez. 2013.

HUFFMAN, David L.; O'HALLORAN, Thomas V. Function, Structure, and Mechanism of Intracellular Copper Trafficking Proteins. **Annual Review of Biochemistry**, v. 70, n. 1, p. 677–701, jun. 2001.

ICHIHASHI, Yasunori *et al.* Transcriptomics exposes the uniqueness of parasitic plants. **Briefings in Functional Genomics**, v. 14, n. 4, p. 275–282, jul. 2015.

ISHIDA, Julianne K. *et al.* Local Auxin Biosynthesis Mediated by a YUCCA Flavin Monooxygenase Regulates Haustorium Development in the Parasitic Plant *Phtheirospermum japonicum*. **The Plant Cell**, v. 28, n. 8, p. 1795–1814, ago. 2016.

- JHU, Min-Yao; SINHA, Neelima R. *Cuscuta* species: Model organisms for haustorium development in stem holoparasitic plants. **Frontiers in Plant Science**, v. 13, p. 1086384, 12 dez. 2022.
- JONES, Jonathan D. G.; STASKAWICZ, Brian J.; DANGL, Jeffery L. The plant immune system: From discovery to deployment. **Cell**, v. 187, n. 9, p. 2095–2116, abr. 2024.
- KADO, Tomoyuki; INNAN, Hideki. Horizontal Gene Transfer in Five Parasite Plant Species in Orobanchaceae. **Genome Biology and Evolution**, v. 10, n. 12, p. 3196–3210, 1 dez. 2018.
- KELLY, C. K. Resource choice in *Cuscuta europaea*. **Proceedings of the National Academy of Sciences**, v. 89, n. 24, p. 12194–12197, 15 dez. 1992.
- KELLY, Colleen K. Plant Foraging: A Marginal Value Model and Coiling Response in *Cuscuta Subinclusa*. **Ecology**, v. 71, n. 5, p. 1916–1925, out. 1990.
- KHAN, Noor Muhammad *et al.* Genome-wide identification of heavy-metal ATPases genes in *Areca catechu*: investigating their functionality under heavy metal exposure. **BMC Plant Biology**, v. 24, n. 1, p. 484, 31 maio 2024.
- KOBAYASHI, Takanori; NISHIZAWA, Naoko K. Iron Uptake, Translocation, and Regulation in Higher Plants. **Annual Review of Plant Biology**, v. 63, n. Volume 63, 2012, p. 131–152, 2 jun. 2012.
- KOKLA, Anna; MELNYK, Charles W. Developing a thief: Haustoria formation in parasitic plants. **Developmental Biology**, v. 442, n. 1, p. 53–59, out. 2018.
- LANDIS, Jacob B. *et al.* Impact of whole-genome duplication events on diversification rates in angiosperms. **American Journal of Botany**, v. 105, n. 3, p. 348–363, mar. 2018.
- LAOHAVISIT, Anuphon *et al.* Quinone perception in plants via leucine-rich-repeat receptor-like kinases. **Nature**, v. 587, n. 7832, p. 92–97, 5 nov. 2020.
- LENTON, Timothy M. *et al.* First plants cooled the Ordovician. **Nature Geoscience**, v. 5, n. 2, p. 86–89, fev. 2012.
- LORETI, Elena; PERATA, Pierdomenico. Mobile plant MICRORNAs allow communication within and between organisms. **New Phytologist**, v. 235, n. 6, p. 2176–2182, set. 2022.
- MA, Yitong *et al.* Genome-wide identification and characterization of the heavy metal ATPase (HMA) gene family in *Medicago truncatula* under copper stress. **International Journal of Biological Macromolecules**, v. 193, p. 893–902, dez. 2021.
- MARSCHNER, H.; RÖMHELD, V. Strategies of plants for acquisition of iron. **Plant and Soil**, v. 165, n. 2, p. 261–274, 1 jun. 1994.
- MARSCHNER, H.; ROMHELD, V.; KISSEL, M. Different strategies in higher plants in mobilization and uptake of iron. **Journal of Plant Nutrition**, v. 9, n. 3, p. 695–713, mar. 1986.
- MCNEAL, Joel R. *et al.* Phylogeny and origins of holoparasitism in Orobanchaceae. **American Journal of Botany**, v. 100, n. 5, p. 971–983, maio 2013.
- MEHRA, Aradhana; FARAGO, Margaret E. Metal ions and plant nutrition. **Plants and the chemical elements. Biochemistry, uptake, tolerance and toxicity**, p. 32–66, 1994.

- MILNER, Matthew 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.
- MOLINA, Jeanmaire *et al.* Possible Loss of the Chloroplast Genome in the Parasitic Flowering Plant *Rafflesia lagascae* (Rafflesiaceae). **Molecular Biology and Evolution**, v. 31, n. 4, p. 793–803, abr. 2014.
- MOREL, F. M. M.; PRICE, N. M. The Biogeochemical Cycles of Trace Metals in the Oceans. **Science**, v. 300, n. 5621, p. 944–947, 9 maio 2003.
- MORRIS, Jennifer L. *et al.* The timescale of early land plant evolution. **Proceedings of the National Academy of Sciences**, v. 115, n. 10, 6 mar. 2018.
- NANDA, Vikas *et al.* On the emergence of metabolism: the evolution of proteins that powered life. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 380, n. 1931, p. 20240090, 7 ago. 2025.
- NGOU, Bruno Pok Man *et al.* Evolutionary trajectory of pattern recognition receptors in plants. **Nature Communications**, v. 15, n. 1, p. 308, 1 fev. 2024.
- NGOU, Bruno Pok Man; DING, Pingtao; JONES, Jonathan D. G. Thirty years of resistance: Zig-zag through the plant immune system. **The Plant Cell**, v. 34, n. 5, p. 1447–1478, 26 abr. 2022.
- NICKRENT, Daniel L. Parasitic angiosperms: How often and how many? **TAXON**, v. 69, n. 1, p. 5–27, 2020.
- Ó LOCHLAINN, Seosamh *et al.* Tandem Quadruplication of HMA4 in the Zinc (Zn) and Cadmium (Cd) Hyperaccumulator *Noccaea caerulea*. **PLoS ONE**, v. 6, n. 3, p. e17814, 10 mar. 2011.
- PALMGREN, Michael Gjedde; HARPER, Jeffrey F. Pumping with plant P-type ATPases. **Journal of Experimental Botany**, v. 50, n. Special_Issue, p. 883–893, jun. 1999.
- RANJAN, Aashish *et al.* De Novo Assembly and Characterization of the Transcriptome of the Parasitic Weed Dodder Identifies Genes Associated with Plant Parasitism. **Plant Physiology**, v. 166, n. 3, p. 1186–1199, 5 nov. 2014.
- RAVEN, J. A.; EDWARDS, D. Roots: evolutionary origins and biogeochemical significance. **Journal of Experimental Botany**, v. 52, n. suppl_1, p. 381–401, mar. 2001.
- ROBINSON, Nigel J. *et al.* A ferric-chelate reductase for iron uptake from soils. **Nature**, v. 397, n. 6721, p. 694–697, fev. 1999.
- RODRIGUES, Wenderson Felipe Costa *et al.* Ferrous iron uptake via IRT1 / ZIP evolved at least twice in green plants. **New Phytologist**, p. nph.18661, 10 jan. 2023.
- RUNO, Steven; KURIA, Eric K. Habits of a highly successful cereal killer, *Striga*. **PLOS Pathogens**, v. 14, n. 1, p. e1006731, 11 jan. 2018.
- RUNO, Steven; WICKE, Susann; THOROGOOD, Chris. Parasitic plants are models for examining global food security, biodiversity loss and host–parasite evolution, in a changing world. **PLANTS, PEOPLE, PLANET**, v. 7, n. 2, p. 303–307, mar. 2025.
- SABOT, F.; SCHULMAN, A. H. Parasitism and the retrotransposon life cycle in plants: a hitchhiker’s guide to the genome. **Heredity**, v. 97, n. 6, p. 381–388, dez. 2006.

- SÁNCHEZ-PÉREZ, Charles R. *et al.* Heavy Metal-Associated Proteins in Plants: Genome-Wide Identification and Functional Insights from Angiosperms and Ancestral Plants. **Journal of Molecular Evolution**, v. 93, n. 4, p. 527–542, ago. 2025.
- SHAHID, Saima *et al.* MicroRNAs from the parasitic plant *Cuscuta campestris* target host messenger RNAs. **Nature**, v. 553, n. 7686, p. 82–85, jan. 2018.
- SHAO, Leilei *et al.* Genome-wide survey of the HMA gene family in wheat (*Triticum aestivum* L.) and its potential role in cadmium stress. **BMC Genomics**, v. 26, n. 1, p. 619, 1 jul. 2025.
- SHIMIZU, Kohki; AOKI, Koh. Development of Parasitic Organs of a Stem Holoparasitic Plant in Genus *Cuscuta*. **Frontiers in Plant Science**, v. 10, p. 1435, 12 nov. 2019.
- SMITH, Aaron T.; SMITH, Kyle P.; ROSENZWEIG, Amy C. Diversity of the metal-transporting P1B-type ATPases. **JBIC Journal of Biological Inorganic Chemistry**, v. 19, n. 6, p. 947–960, 1 ago. 2014.
- TAGLIABUE, Alessandro *et al.* The integral role of iron in ocean biogeochemistry. **Nature**, v. 543, n. 7643, p. 51–59, mar. 2017.
- THOROGOOD, Chris J. *et al.* Endoparasitic plants and fungi show evolutionary convergence across phylogenetic divisions. **New Phytologist**, v. 232, n. 3, p. 1159–1167, nov. 2021.
- TOH, Shigeo *et al.* Structure-function analysis identifies highly sensitive strigolactone receptors in *Striga*. **Science**, v. 350, n. 6257, p. 203–207, 9 out. 2015.
- TOMILOV, Alexey A. *et al.* Localized Hormone Fluxes and Early Haustorium Development in the Hemiparasitic Plant *Triphysaria versicolor*. **Plant Physiology**, v. 138, n. 3, p. 1469–1480, 1 jul. 2005.
- VATANSEVER, Recep; ÖZYIĞIT, Ibrahim Ilker; FILİZ, Ertuğrul. Comparative and phylogenetic analysis of zinc transporter genes/proteins in plants. **TURKISH JOURNAL OF BIOLOGY**, v. 40, p. 600–611, 2016.
- VERT, Grégory *et al.* Arabidopsis IRT2 cooperates with the high-affinity iron uptake system to maintain iron homeostasis in root epidermal cells. **Planta**, v. 229, n. 6, p. 1171–1179, maio 2009.
- VERT, Grégory *et al.* IRT1, an Arabidopsis Transporter Essential for Iron Uptake from the Soil and for Plant Growth. **The Plant Cell**, v. 14, n. 6, p. 1223–1233, jun. 2002.
- WESTWOOD, James H. *et al.* The evolution of parasitism in plants. **Trends in Plant Science**, v. 15, n. 4, p. 227–235, abr. 2010.
- WU, Danxia *et al.* The Mechanism of Metal Homeostasis in Plants: A New View on the Synergistic Regulation Pathway of Membrane Proteins, Lipids and Metal Ions. **Membranes**, v. 11, n. 12, p. 984, 15 dez. 2021.
- WU, Feihua *et al.* Hydrogen peroxide sensor HPCA1 is an LRR receptor kinase in Arabidopsis. **Nature**, v. 578, n. 7796, p. 577–581, 27 fev. 2020.
- WU, Yuguo *et al.* Bidirectional lncRNA Transfer between *Cuscuta* Parasites and Their Host Plant. **International Journal of Molecular Sciences**, v. 23, n. 1, p. 561, 5 jan. 2022.
- XI, Zhenxiang *et al.* Horizontal transfer of expressed genes in a parasitic flowering plant. **BMC Genomics**, v. 13, n. 1, p. 227, dez. 2012.

XI, Zhenxiang *et al.* Massive Mitochondrial Gene Transfer in a Parasitic Flowering Plant Clade. **PLoS Genetics**, v. 9, n. 2, p. e1003265, 14 fev. 2013.

XU, Yuxing *et al.* Comparative genomics of orobanchaceous species with different parasitic lifestyles reveals the origin and stepwise evolution of plant parasitism. **Molecular Plant**, v. 15, n. 8, p. 1384–1399, ago. 2022.

YANG, Wenjing *et al.* Epigenetic modifications: Allusive clues of lncRNA functions in plants. **Computational and Structural Biotechnology Journal**, v. 21, p. 1989–1994, 2023.

YANG, Zhenzhen *et al.* Comparative Transcriptome Analyses Reveal Core Parasitism Genes and Suggest Gene Duplication and Repurposing as Sources of Structural Novelty. **Molecular Biology and Evolution**, v. 32, n. 3, p. 767–790, 1 mar. 2015.

YANG, Zhenzhen *et al.* Convergent horizontal gene transfer and cross-talk of mobile nucleic acids in parasitic plants. **Nature Plants**, v. 5, n. 9, p. 991–1001, 22 jul. 2019.

YOSHIDA, Satoko *et al.* Genome Sequence of *Striga asiatica* Provides Insight into the Evolution of Plant Parasitism. **Current Biology**, v. 29, n. 18, p. 3041–3052.e4, set. 2019.

YOSHIDA, Satoko *et al.* The Haustorium, a Specialized Invasive Organ in Parasitic Plants. **Annual Review of Plant Biology**, v. 67, n. 1, p. 643–667, 29 abr. 2016.

YOSHIDA, Satoko; KEE, Yee Jia. Large-scale sequencing paves the way for genomic and genetic analyses in parasitic plants. **Current Opinion in Biotechnology**, Food Biotechnology - Plant Biotechnology. v. 70, p. 248–254, 1 ago. 2021.

YU, Xiao-Qian *et al.* PTI-ETI synergistic signal mechanisms in plant immunity. **Plant Biotechnology Journal**, v. 22, n. 8, p. 2113–2128, ago. 2024.

YUAN, Minhang *et al.* PTI-ETI crosstalk: an integrative view of plant immunity. **Current Opinion in Plant Biology**, v. 62, p. 102030, ago. 2021.

ZHANG, Tuo *et al.* Crystal structures of a ZIP zinc transporter reveal a binuclear metal center in the transport pathway. **Science Advances**, v. 3, n. 8, p. e1700344, 4 ago. 2017.

5. ANEXOS – PRODUÇÃO ACADÊMICA

Nesta seção são apresentados a produção acadêmica durante o período do doutorado da autora (2022-2026) assim como os prêmios obtidos.

ANEXO A – TRABALHOS PUBLICADOS (ARTIGOS)

RODRIGUES, WENDERSON FELIPE COSTA; PIRES, LAURA OLIVEIRA; OLIVEIRA, LUCAS ROBERTO; ISHIDA, JULIANE K. *Comprehensive characterization of HMA transporters in common bean: Tissue-specific expression and response to metal stress. CROP SCIENCE*, v. 65, p. e70188, 2025.

KFOURY, BEATRIZ; **RODRIGUES, WENDERSON FELIPE COSTA**; KIM, SANG-JIN; BRANDIZZI, FEDERICA; DEL-BEM, LUIZ-EDUARDO. *Multiple horizontal gene transfer events have shaped plant glycosyl hydrolase diversity and function. NEW PHYTOLOGIST*, v. 242, p. 809-824, 2024.

RODRIGUES, WENDERSON FELIPE COSTA; LISBOA, AYRTON BRENO P.; LIMA, JONI ESROM; RICACHENEVSKY, FELIPE KLEIN; DEL-BEM, LUIZ-EDUARDO. *Ferrous iron uptake via / evolved at least twice in green plants. NEW PHYTOLOGIST*, v. 237, p. 699-1066, 2023.

SANTOS, THIAGO M.; LISBOA, AYRTON B. P.; **RODRIGUES, WENDERSON**; GOMES, HELENA; ABRAHÃO, JÔNATAS; DEL-BEM, LUIZ-EDUARDO. *Human variation in the protein receptor ACE2 affects its binding affinity to SARS-CoV-2 in a variant-dependent manner. JOURNAL OF BIOMOLECULAR STRUCTURE AND DYNAMICS*, v. 41, p. 2947-2955, 2023.

ANEXO B – TRABALHOS PUBLICADOS (LIVROS/CAPÍTULOS)

PIRES, L. O.; **RODRIGUES, W. F. C.**; ISHIDA, J. K.; DEL-BEM, LUIZ-EDUARDO. *Origin and evolutionary history of defense-related proteins in plants*. In: Santosh Kumar Upadhyay. (Org.). *Defense-Related Proteins in Plants*. 1ed. Amsterdã (Holanda): Elsevier, 2024, v. 1.

OLIVEIRA, M. F.; SANTOS, P. O.; **RODRIGUES, W. F. C.**; PIRES, L. O. Anais do XV Workshop do Programa de Pós-graduação em Biologia Vegetal da Universidade Federal de Minas Gerais. 1. ed. Belo Horizonte: UFMG, 2023. v. 1. 71p.

ANEXO C – MANUSCRITOS EM FASE DE SUBMISSÃO

LAURA O. PIRES*; WENDERSON F. C. RODRIGUES*; JULIANE K. ISHIDA. *Comparative Genomic and Transcriptomic Analyses Suggest the Evolutionary Convergency of UMAMIT transporter in Parasitic Plants*. A ser submetido para revista *PLANT BIOLOGY* (fator de impacto 3.8) em 2026.

WENDERSON F. C. RODRIGUES; LAURA O. PIRES; LUIZ-EDUARDO DEL-BEM; JULIANE K. ISHIDA. *Investigating lncRNAs at different levels of parasitism in parasitic plants*. A ser submetido para revista *PHYSIOLOGIA PLANTARUM* (fator de impacto 3.6) em 2026.

WENDERSON F. C. RODRIGUES; LUIZ-EDUARDO DEL-BEM; PEDRO J. BALLESTER; JULIANE K. ISHIDA. *AI-guided two-stage screening identifies putative haustorium non-inducers / inhibitors molecules in parasitic plants*. A ser submetido em 2026.

AYRTON B. P. LISBOA*; WENDERSON F. C. RODRIGUES*; MATHEUS L. C. BONACCORSI DE CAMPOS; EMERSON G. COLAÇO; MARIA HELENA S. GOLDMAN; LUIZ-EDUARDO DEL-BEM. *SCII unveiled: tracing its deep evolutionary origins through horizontal gene transfer analyses*. A ser submetido para revista *NEW PHYTOLOGIST* (fator de impacto 8.1) em 2026.

AYRTON B. P. LISBOA; VICTOR S. PYLRO; WENDERSON F. C. RODRIGUES; LUIZ-EDUARDO DEL-BEM. *Seasonal rainfall regulates soil charophyte algae abundance: insights into the origin of soils and land plants*. A ser submetido para revista *NEW PHYTOLOGIST* (fator de impacto 8.1) em 2026.

MATHEUS L. C. B. DE CAMPOS*; WENDERSON F. C. RODRIGUES*; HELENA F. GOMES; FELIPE K. RICACHENEVSKY; JONI E. LIMA; LUIZ-EDUARDO DEL-BEM. *Plant iron acquisition strategy II evolved through multiple horizontal gene transfer events*. A ser submetido para revista *PNAS* (fator de impacto 9.1) em 2026.

ANDRIELE WAIRICH; LUCAS R. PONTE; PALOMA K. MENGUER; GUILHERME L. DUARTE; EZEQUIEL B. AREND; MARCOS BÖNES; MATHEUS L. C. B. DE CAMPOS; WENDERSON F. C. RODRIGUES; JANETTE P. FETT; LUIZ-EDUARDO DEL-BEM; RAUL A. SPEROTTO; FELIPE RICACHENEVSKY. *The wild rice species *Oryza rhizomatis* constitutively accumulates sodium in leaves and tolerates salt stress*. Submetido para revista *PLANTA* (fator de impacto 3.8) em 2026.

* Autores dividem primeira autoria.

ANEXO D – PRÊMIOS E INDICAÇÕES

Best Poster Award (*Phylogeny and Evolution*) for the work entitled "Origin, distribution and evolution of the DOMON protein superfamily in plant lineage". **WENDERSON FELIPE COSTA RODRIGUES**, RODOLFO A. MANIERO, JONI ESROM LIMA, RICARDO F. H. GIEHL, LUIZ EDUARDO VIEIRA DEL-BEM. 20th International Conference of the Brazilian Association for Bioinformatics and Computational Biology (X-Meeting 2024). **Associação Brasileira de Bioinformática e Biologia Computacional - AB3C**.

Honorable Mention (*Phylogeny and Evolution*) for the work entitled "The independent origins of PT and YSL subfamilies of plant OPT iron transporters: an horizontal gene transference case". MATHEUS LEWI CRUZ BONACCORSI DE CAMPOS; **WENDERSON FELIPE COSTA RODRIGUES**; HELENA FERREIRA GOMES; JONI ESROM LIMA; FELIPE KLEIN RICACHENEVSKY; LUIZ EDUARDO VIEIRA DEL-BEM. 20th International Conference of the Brazilian Association for Bioinformatics and Computational Biology (X-Meeting 2024). **Associação Brasileira de Bioinformática e Biologia Computacional - AB3C**.

Nominated for the Alcides Carvalho Award (Genetics and Plant Breeding) for the work entitled "Computational identification and comparison of long non-coding RNA in parasitic plants families (Orobanchaceae and Convolvulaceae)". **Wenderson Felipe Costa Rodrigues**, Laura Oliveira Pires, Luiz Eduardo Vieira Del-Bem, Juliane Karine Ishida. 68th Brazilian Congress of Genetics (2023). **Sociedade Brasileira de Genética**.