Communication

Genome of *Qualea grandiflora* Mart. (Vochysiaceae) Reveals Multi-Level Aluminium Handling Mechanisms in a Cerrado Hyperaccumulator Species

Laísa Maria de Resende Castro , Christina Cleo Vinson , Sheila Maysa da Cunha Gordo , Natalia Faustino Cury , Michelle de Souza Fayad André , Thomas Christopher Rhys Williams , Luiz Alfredo Rodrigues Pereira *

Posted Date: 11 February 2026

doi: 10.20944/preprints202602.0920.v1

Keywords: aluminium; Cerrado; genomics; *Qualea grandiflora*

Communication

Genome of *Qualea grandiflora* Mart. (Vochysiaceae) Reveals Multi-Level Aluminium Handling Mechanisms in a Cerrado Hyperaccumulator Species

Laísa Maria de Resende Castro [1,3], Christina Cleo Vinson [2], Sheila Maysa da Cunha Gordo [2], Natalia Faustino Cury [1,4], Michelle de Souza Fayad André [1], Thomas Christopher Rhys Williams [1] and Luiz Alfredo Rodrigues Pereira [1]*

[1] Plant Molecular Biology Laboratory, Department of Botany, University of Brasilia, Brasilia, DF, Postal Code 70919-900, Brazil

[2] Federal University of South and Southeast of Pará, Quadra 07, Lote Especial, s/nº, Nova Marabá, Postal code: 68507-590, Marabá, PA, Brazil

[3] Department of Education, Federal Institute of Pará, Campus Altamira, Altamira, Postal Code 68377-630, PA, Brazil

[4] Embrapa Cenargen, Parque Estação Biológica, PqEB, W5 Norte Av., PoBox 02372 – Postal Code 70770-917, Brasília, DF, Brazil

* Corresponding author: larp@unb.br

Abstract

The lack of reference genomes for non-model species hinders our understanding of aluminum (Al) tolerance and accumulation. We present the first high-quality genome assembly of *Qualea grandiflora* Mart. (Vochysiaceae), an Al-accumulating species endemic to the Brazilian Cerrado. Multi-omics analyses (transcriptomic, proteomic, and metabolomic) reveal that Al is essential for its growth and development. Using a paired-end library and ABySS v2.0, we assembled a genome containing 38,034 annotated genes (63.1% "complete"). Functional annotation via SwissProt/KOG and Blast2GO identified 11 gene families linked to Al response, including ALMT, MATE, ABC, and NRAT1. GO analysis further highlighted enriched processes related to Al metabolism, notably SAM synthetase genes upregulated in roots, which are critical for DNA/RNA methylation and cell wall formation. By establishing *Q. grandiflora* as a genomic model for native Al hyperaccumulation species, this study provides a foundational resource for researching detoxification and ecological adaptations in metallophytes. The annotated sequence is available via NCBI (BioProject PRJNA786741), supported by leaf transcriptomic data from PRJNA358394.

Keywords: aluminium; Cerrado; genomics; *Qualea grandiflora*

1. Introduction

Aluminum (Al) is a major abiotic stressor that inhibits plant growth and development by disrupting critical physiological processes, including respiration, photosynthesis, and gene expression [1,2]. In response to Al toxicity, plants activate oxidative stress metabolism to mitigate cellular damage [2,3]. However, certain species exhibit Al tolerance or even dependence, employing mechanisms such as Al complexation with organic acids (*e.g.*, citrate, malate, and oxalate), leading to accumulation exceeding 1 g Al·kg^{-1} dry mass [2,3]. Notably, in the Brazilian Savanna (Cerrado), several species not only accumulate Al but also require it for optimal growth [1,2]. Among these, the Vochysiaceae family is particularly significant, as all its species are Al accumulators [2].

Qualea grandiflora Mart. (Vochysiaceae, Myrtales; APG IV, 2016), commonly known as "*pau-terra*," is a widely distributed Al-accumulating species in the Cerrado. This diploid (2n = 22) [4,5], hermaphroditic tree disperses seeds during the dry season (July–September) [6] and exhibits

 doi:10.20944/preprints202602.0920.v1

remarkable adaptability to varying light conditions, nutrient availability, and soil pH [7], making it a candidate for reforestation of degraded areas. Beyond its ecological role, *Q. grandiflora* has garnered pharmaceutical interest due to its antibacterial [8], antiulcer [9], anticonvulsant, analgesic, and antioxidant properties [10].

Given its high Al accumulation and tolerance, *Q. grandiflora* has been extensively studied. Andrade et al. [11] demonstrated that Al accumulates in chloroplasts without impairing their function. Furthermore, anatomical and physiological studies revealed that exogenous Al does not inhibit seed germination, with ~60% of Al stored in the seed translocated to seedling leaves [12,13]. Histochemical analyses localized Al primarily in cotyledons, suggesting a potential interaction with lipid and protein reserves, which may contribute to this species' ecological success [13]. Transcriptomic analyses of Al-exposed *Q. grandiflora* leaves revealed upregulation of genes linked to cell wall biosynthesis, primary/secondary metabolism, phytohormones (brassinosteroids, salicylic acid), chloroplast biogenesis, defence responses and Al transport, reinforcing its Al-dependent growth [14].

A well-documented Al tolerance mechanism involves Al-activated exudation of organic acids (O.As.; *e.g.*, malate, citrate) into the rhizosphere, mediated by the ALMT (aluminium-activated malate transporter) and MATE (multidrug and toxic compound extrusion) gene families [15,16]. While transcriptomic studies on Al-accumulating species like *Psychotria rubra* (Rubiaceae) have identified Al-responsive genes [17], the absence of comparative analyses under varying soil acidity limits mechanistic insights.

A critical barrier to understanding Al-responsive mechanisms in non-model plants is the lack of reference genomes. Although next-generation sequencing (NGS) has expanded genomic resources for non-model species [18], no genome assembly exists for *Q. grandiflora*. To address this gap, we present the first *de novo* genome assembly of *Q. grandiflora*, enabling the identification of Al-tolerance and accumulation-related genes**.** This work establishes a foundational genomic resource for studying Al adaptation in Cerrado flora and advances our understanding of metallophyte evolution.

2. Results and Discussion

2.1. Genome Sequencing and Assembly

Whole-genome sequencing of *Qualea grandiflora* was conducted using the Illumina HiSeq 2500 platform with 100 bp paired-end libraries, generating approximately 57.7 Gb of raw data. Quality filtering (Q-score >30) yielded 536,103,004 high-quality reads (99.6% retention rate) from an initial 537,656,590 reads. The assembly achieved approximately 81% coverage of the estimated 500 Mb genome (Table 1), representing superior coverage compared to other Al-accumulating Cerrado species like *Caryocar brasiliense*: 45.7%; and *Eugenia dysenterica*: 56.7% [19].

The assembly pipeline compared ABySS 2.0 and DISCOVAR *de novo*, with ABySS 2.0 producing superior results. The final assembly comprised 406,048,254 bp distributed across 277,998 scaffolds, with scaffold lengths ranging up to 600,552 bp (Table 1). Assembly metrics included an N50 of 2,499 bp, L50 of 35,899, and GC content of 36.3%. Repetitive elements constituted 50.17% (203,730,046 bp) of the assembled genome, consistent with patterns observed in other plant genomes (Table 1).

BUSCO assessment against the Embryophyta database (1,440 orthologs) revealed 63.1% complete genes (including 1.3% duplicated), 13.1% fragmented, and 23.8% missing (Table 2). This completeness level meets established standards for non-model plant genomes, where >50% completeness is considered acceptable and more likely to appear in draft assemblies given the challenges of genome size, complexity, and phylogenetic divergence [20,21]. The 63.1% completeness particularly supports the utility of this assembly for gene-centric analyses, as noted in similar non-model plant genome projects [22].

Table 1. *Qualea grandiflora* Mart. genome assembly statistics.

Genome estimate size (Mb)	**500**

 doi:10.20944/preprints202602.0920.v1

Total size (bp)	406,048,254
Number of contigs	5,905,481
Number of scaffolds	277,998
Maximum scaffold length (bp)	600,552
GC (%)	36.33
N50	2,499
N75	806
L50	35,899
L75	114,236
Repetitive elements - Masked repeats (bp)	203,730,046

Table 2. *Qualea grandiflora* Mart. final genome annotation statistics.

Parameter	BUSCO groups	%
Complete BUSCOs (C)	840	58.4%
Complete and single-copy BUSCOs (S)	803	55.8%
Complete and duplicated BUSCOs (D)	37	2.6%
Fragmented BUSCOs (F)	245	17.0%
Missing BUSCOs (M)	355	24.6%
Total BUSCO groups searched	1,440	100.0

Table 3. Estimation of genome characteristics based on 21-mer statistics.

Number of genes	38,034
Average gene length (bp)	1,353.38
Number of exons	114,847
mRNA	37,691
tRNA	343
Number of genes	38,034
Average gene length (bp)	1,353.38
Number of exons	114,847
mRNA	37,691
tRNA	343

2.2. Annotation and Prediction of Qualea grandiflora Genome

The structural annotation of the *Q. grandiflora* genome identified 38,034 genes, including 37,691 protein-coding mRNA genes and 343 tRNA genes, with an average gene length of 1,353.38 bp and a total of 114,847 exons (Tables 3 and S1). To assess gene completeness, protein-coding sequences were evaluated using BUSCO, revealing that 58.4% of genes were complete, 2.6% were duplicated, 17.0% were fragmented, and 24.6% were missing (Table 2).

For functional annotation, predicted proteins were analysed using InterProScan to identify conserved domains and motifs, followed by comparison against the InterPro database. Additionally, a BLAST search was performed against the UniRef90 Viridiplantae protein database to assign putative functions.

Gene Ontology (GO) analysis classified 7,754 genes into Biological Process (BP), 3,686 into Molecular Function (MF), and 3,447 into Cellular Component (CC) categories (Figure 2). Within BP, the most frequent terms include organic substance metabolic process (4.2%), primary metabolic process (4.0%), nitrogen compound metabolic process (3.7%), cellular metabolic process (4.0%), biosynthetic process (1.8%), gene expression (1.6%), and transport (1.3%). The MF category was dominated by protein binding (2.0%), catalytic activity acting on proteins (2.3%), carbohydrate derivative binding (2.4%), oxidoreductase activity (2.6%), hydrolase activity (2.8%), small molecule binding (3.3%), transferase activity (3.5%), ion binding (5.0%), and organic/heterocyclic compound binding (5.6%). For CC, most genes were linked to intracellular anatomical structures (9.0%),

 doi:10.20944/preprints202602.0920.v1

organelles (7.1%), membranes (6.8%), cytoplasm (5.9%), intrinsic membrane components (4.3%), and cell periphery (1.8%). A summary of these GO term distributions is provided in Figure 2.

2.3. Candidate Genes from the Qualea grandiflora in Response to Al

To identify genes that could be associated with aluminium (Al) metabolism in *Q. grandiflora*, we compared known Al-responsive gene sequences from the literature against the *Q. grandiflora* genome using Blast2GO and BLAST/NCBI [23]. Sequences with the highest similarity and lowest e-values were selected for further analysis (Table S2). Several key genes linked to Al tolerance were identified, including those involved in Al transport, organic acid metabolism, plant growth regulation, the methylation cycle, and cell wall synthesis.

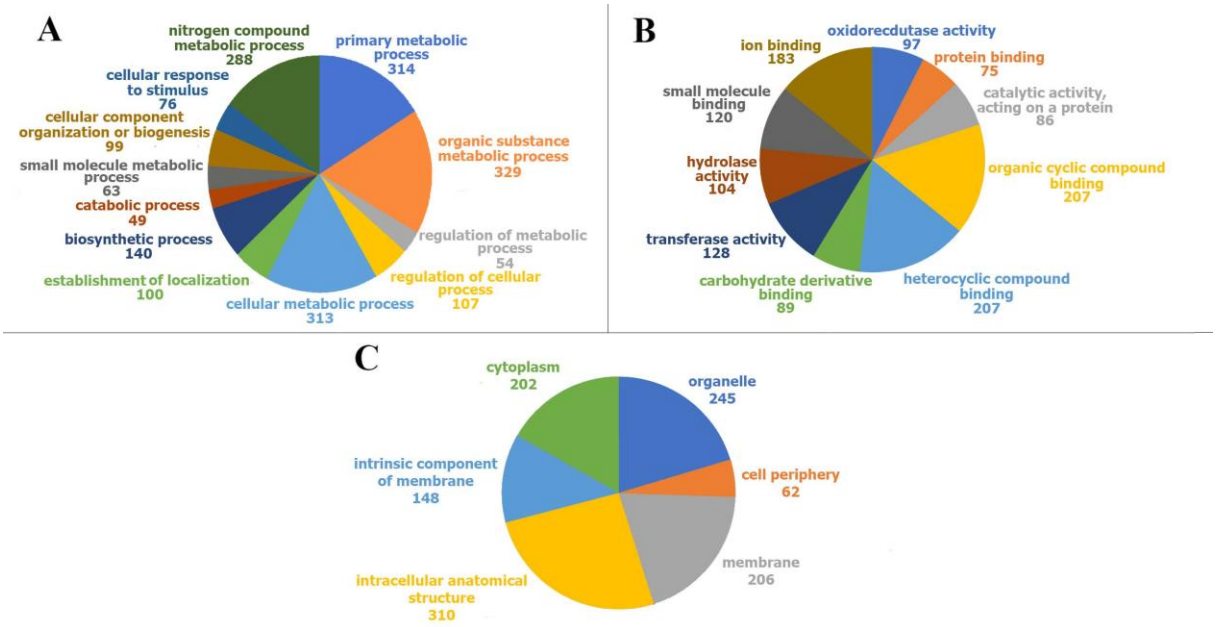

Figure 2. Gene ontology (GO) analysis of gene sequences of *Qualea grandiflora* Mart. (Vochysiaceae). A) *Qualea grandiflora* sequences associated with some Biological Process categories. B) Main molecular function categories that of gene sequences of *Q. grandiflora* according to GO analysis. C) Cellular component categories of some *Q. grandiflora* genes as determined by GO analysis.

Notably, transmembrane transporter genes such as *ALMT* (aluminium-activated malate transporter), *MATE* (multidrug and toxic compound extrusion family), ABC (ATP-binding cassette transporter), and *NRAT* (natural resistance–associated macrophage protein 1) were detected [3,15,24,25]. Transcriptomic analysis revealed that four ABC transporter genes were differentially expressed under Al exposure, with one upregulated in *Q. grandiflora* leaves [14]. In contrast, NRAT family genes showed no differential expression in response to Al [12,26]. Surprisingly, most Al-tolerance-related genes in *Q. grandiflora* were not upregulated under Al stress; instead, many were downregulated in both leaves and roots [12,26]. This suggests that their regulation may depend on additional endogenous or environmental factors beyond Al exposure alone [27].

Genes involved in essential cellular processes were also identified, including *SAMS2_1* and *SAMS2_2* (S-adenosyl methionine synthetase 2), which catalyse the synthesis of *S*-adenosyl methionine (SAM) – a critical methyl group donor for DNA, RNA, protein, and cell wall methylation, as well as ethylene and polyamine biosynthesis [28,29,30]. Intriguingly, *SAMS2* was upregulated in *Q. grandiflora* roots under Al stress [12,26]. Additionally, we identified *SAHH2* (S-adenosylhomocysteine hydrolase 2) and adenosine kinase (ADK; IPR001810), both crucial for maintaining SAM-dependent methylation reactions [31,32,33].

 doi:10.20944/preprints202602.0920.v1

Cell wall-related genes were also prominent, including *PME29*, a pectin methylesterase involved in modulating pectin methylation – a key determinant of cell wall rigidity [32,34,35]. Given that the cell wall is a primary site of Al accumulation and toxicity [3,34,36], the upregulation of PME genes in Al-treated roots [12,26] suggests enhanced cell wall remodelling as a compensatory mechanism.

This study identified Al-responsive genes in *Q. grandiflora*, highlighting potential mechanisms of tolerance and physiological effects of this metal in this species. However, functional characterization of these genes and further investigation into their regulatory dynamics in the presence of Al are needed to fully elucidate their roles.

3. Material and Methods

3.1. Plant Material and DNA Extraction

Qualea grandiflora seeds were collected from a Cerrado stricto sensu area (14° 42' 19.31" S; 47° 40' 11.81" W) in the municipality of Água Fria de Goiás, Goiás, Brazil. Samples from the mother plants were herborized and incorporated into the Herbarium of the University of Brasília (UB); Voucher: 217284.

After germination, leaves from *Q. grandiflora* seedlings were frozen in liquid nitrogen and transported to the Plant Biotechnology Laboratory at the University of Brasília (UnB), for total DNA extraction. Total DNA was extracted using the Qiagen DNeasy Plant Kit (Hilden, Germany). The concentration and quality of the DNA were assessed using a NanoDrop 1000 (Thermo Fisher Scientific, Wilmington, USA), a Bioanalyzer 2100 (Agilent Technologies Inc., Santa Clara, USA), a Qubit fluorometer (Life Technologies, Thermo Fisher Scientific, Wilmington, USA), and a 1% agarose gel (Figure 1).

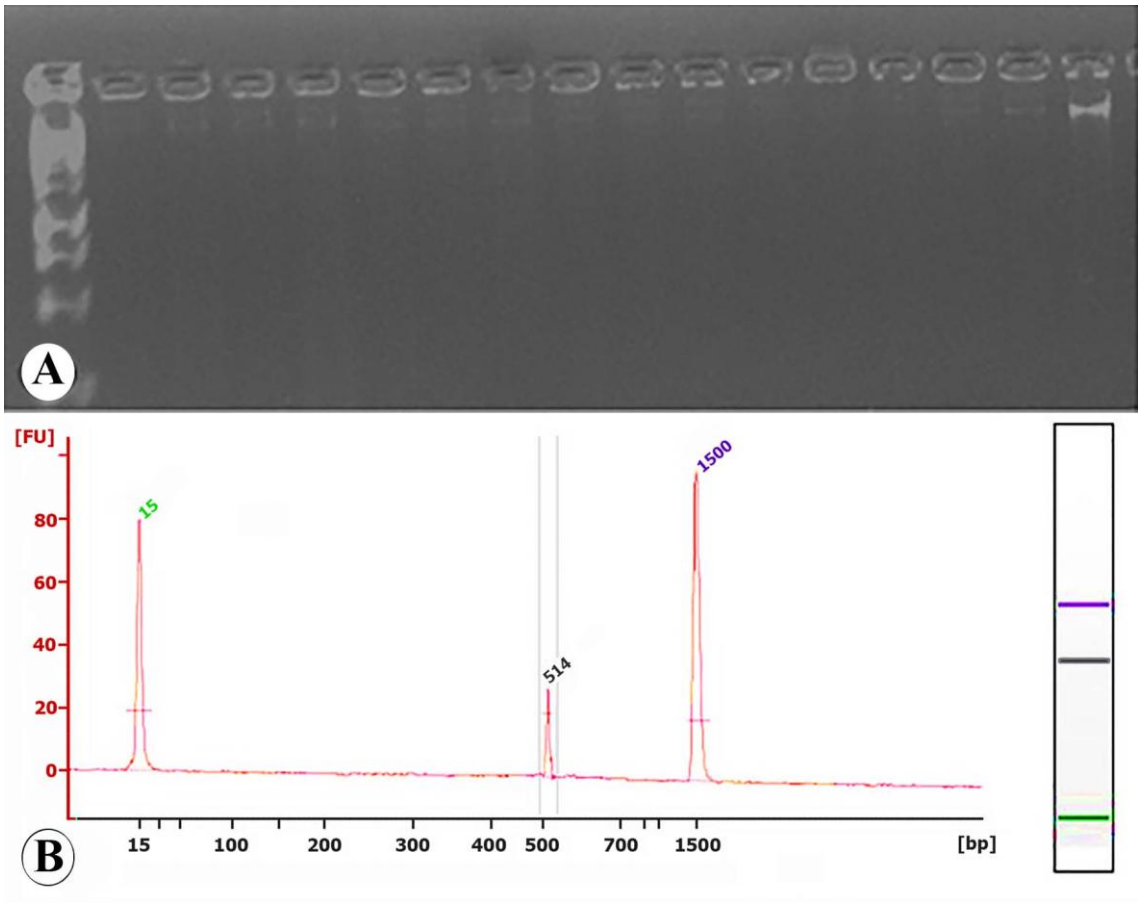

Figure 1. Quality and integrity of the total DNA extracted from *Qualea grandiflora* Mart. A) Agarose gel showing the integrity of *Q. grandiflora* DNA. B) Bioanalyzer data on *Q. grandiflora* DNA quality and integrity (by LACTAD).

 doi:10.20944/preprints202602.0920.v1

3.2. Genome Sequencing and Assembly

Sequencing was performed on an Illumina HiSeq 2500 platform (San Diego, CA, USA), and two paired-end libraries were constructed using the TruSeq DNA Nano Low Throughput Library Prep Kit (Illumina, San Diego, CA, USA), with read sizes ranging from 100 to 300 bp.

Prior to assembly, Illumina reads were filtered and trimmed using Trimmomatic v.0.20 [37] to remove the TruSeq adapters. For the trimmed reads, FastQC (v0.11.3) [38] was used to plot quality scores and sequence length distribution. K-mer analysis was performed in the libraries. Next, genome size estimation was conducted using GenomeScope [39] with Jellyfish v.2.0 [40] to count and build the histogram of K-mers.

The genome was assembled using ABySS v2.0 [41] and DISCOVAR *de novo* [42]. Both assemblers follow the classic De Bruijn graph approach.

To assess the quality of the *de novo* genome assembly, we used BUSCO v3.0.2 [43] against the Embryophyta ortholog database. BUSCO classified the sequences as complete and single-copy (S), complete and duplicated (D), fragmented (F), or missing (M). In addition to assessing the relative gene completeness of the assemblies, the measurement of duplicated hits is useful for evaluating the extent to which the assembler may falsely expand genomic regions, particularly those that are highly polymorphic [43].

3.3. Genome Annotation

For gene prediction and annotation, the Funannotate pipeline v1.8.1 [44], AUGUSTUS v2.5.5 [45] SNAP [46], and GlimmerHMM [47] were used. First, Funannotate cleaned up and sorted the sequences. Then, the RepeatModeler v2.0/RepeatMasker v4.0.1 software identified repetitive elements [48]. Additionally, Funannotate was trained using RNA-Seq data from *Q. grandiflora* (SRA: SRR5248188, SRA: SRR5248189, SRA: SRR5248190, SRA: SRR5248191, SRA: SRR5248192, SRA: SRR5248193) and the Uniprot Myrtales protein sequence database. AUGUSTUS was trained using alignments from the BUSCO dataset [43].

Functional annotations of genes were predicted based on the SwissProt database [49] and the Eukaryotic Orthologous Groups (KOG) protein database [50], using an E-value threshold of <1e-5. Gene classification was performed using InterProScan (v5.20-59.0) for Gene Ontology (GO) terms [51] were evaluated using Blast2GO software [52].

All proteins were assigned based on their similarity to the Pfam database [53] and compared for carbohydrate-active enzyme domains (CAZymes) using GlimmerHMM v3.0.472 [47]; and were also compared to the MEROPS [54] and eggNOG v4.5 [55] databases using BLASTP (version 2.7.1) with an E-value of 1e-5 [56].

The analysis of metabolic pathways was performed by searching the KEGG database (Kyoto Encyclopedia of Genes and Genomes) with an E-value cut-off of 1e-5 using the BlastKOALA mapping platform (KEGG Orthology and Links Annotation) [57]. Functional annotation was based on the following parameters: the Plants taxonomic group and the family eukaryotes database as reference.

3.4. Al-Responsive Genes

Gene sequences associated with Al responses were selected from the literature, combined with BLAST results from NCBI (National Center for Biotechnology Information) (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Batch BLAST similarity searches were performed automatically on the database. We used RNA-Seq data from *Q. grandiflora* (SRA: SRR5248188, SRA: SRR5248189, SRA: SRR5248190, SRA: SRR5248191, SRA: SRR5248192, SRA: SRR5248193).

4. Conclusions and Perspectives

This study is a high-significant genome sequencing of *Q. grandiflora*, achieving a total genome coverage of 81% with excellent metrics, particularly considering the lack of a reference genome and

 doi:10.20944/preprints202602.0920.v1

the non-model status of the species (BioProject PRJNA786741). Additionally, 11 gene families potentially involved in Al response mechanisms were identified in *Q. grandiflora*.

Thus, this study also provides:

The first comprehensive genome resource for an Al-hyperaccumulating Cerrado species

Evidence for both conserved (*e.g.*, transporter families) and unique (*e.g.*, constitutive expression) Al-tolerance strategies

Candidate genes for functional validation, particularly:

PME29-mediated cell wall modification

SAM-dependent methylation pathways

Key unresolved questions to be addressed include:

The role of protein-level regulation in Al metabolic responses in *Q. grandiflora* and other native Al-accumulating species

Ecological implications of constitutive vs. inducible tolerance

Evolutionary origins of Al hyperaccumulation in Vochysiaceae

Supplementary Materials: The following supporting information can be downloaded at: https://....

Author Contributions: Experimental design and execution, L.M.R.C., C.C.V., and L.A.R.P.; In Silico Analyses, C.C.V. and S.M.C.G.; GO terms analysis N.F.C.; Field collection, L.M.R.C., N.F.C. and M.S.F.A.; Writing draft preparation, L.M.R.C. and C.C.V.; Writing review and editing, L.A.R.P. and T.C.R.W.; Research supervision, L.A.R.P.; Funding acquisition, L.A.R.P. and N.F.C.; Project administration, L.A.R.P. All authors have read and agreed to the published version of the manuscript.

Funding: The funding of this research was provided by CAPES, as Doctoral and Post-Doctoral scholarships.

Research grants: FAP-DF (Research Support Foundation of the Federal District), 0193.001622/2017, and 00193-00002089/2023-97.

Data Availability Statement: The data presented in this study are included in the article/Supplementary Materials as well as at https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA786741. Further inquiries can be directed to the corresponding author.

Acknowledgments:. The authors would like to thank the Coordination for the Improvement of Higher-Level Personnel (CAPES), the support of the Research Support Foundation of the Federal District (FAP-DF), the Federal Institute of Para campus Altamira, Altamira, as well as Federal University of South and Southeast of Pará, PA, Brazil for their support of this research.

Conflicts of Interest: The authors declare there is neither conflicts nor competing interests.

References

1. Zaia, M.; Timpone, L.T.; Habermann, G. Do aluminum (Al)-accumulating species from the Brazilian savanna accumulate Al in the roots? *Trees* **2022**, 36, 1677-1685. https://doi.org/10.1007/s00468-022-02301-4
2. Castro, L.M.R.; Vinson, C.C.; Gordo S.M.C.; Cury N.F.; André M.S.F.; Williams T.C.; Pereira L.A.R. Molecular and physiological aspects of plant responses to aluminum: what do we know about Cerrado plants? *Braz. J. Bot.* **2022**, 45, 545-562 https://doi.org/10.1007/s40415-021-00781-1
3. Kochian, L.V.; Piñeros, M.A.; Liu, J.; Magalhães J.V. Plant adaptation to acid soils: the molecular basis for crop aluminum resistance. *Ann. Rev. Plant Biol.* **2015,** *66,* 571-98. https://doi.org/10.1146/annurev-arplant-043014-114822
4. Yamagishi-Costa, J.; Barbosa, A.R.; Shimizu, G.H.; Yamamoto, K.; Forni-Martins, E.R. Chromosome numbers and the systematics of tribe Vochysieae (Vochysiaceae). *Acta Bot. Bras.* **2018**, 32, 314-320. https://doi.org/10.1590/0102-33062017abb0354
5. Potascheff, C.M.; Oddou-Muratorio, S.; Klein, E.K. Figueira, A.; Bressan, E.A.; Oliveira, P.E.; Lander, T.A.; Sebbenn A.M. Stepping stones or stone dead? Fecundity, pollen dispersal and mating patterns of roadside *Qualea grandiflora* Mart. trees. *Conserv. Gen.* **2019**, 20, 1355-1367. https://doi.org/10.1007/s10592-019-01217-w

 doi:10.20944/preprints202602.0920.v1

6. Ferreira, R.A.; Davide, A.C.; Tonetti, O.A.O. Morphology of seeds and seedlings of pau-terra (*Qualea grandiflora* Mart. – Vochysiaceae). *J. Seed Sci.* **2001**, 23, 116-122. http://dx.doi.org/10.17801/0101-3122/rbs.v23n1p116-122
7. Haridasan, M. Aluminium accumulation by some Cerrado native species of central Brazil. *Plant Soil* 65, **1982**, 265–273. https://doi.org/10.1007/BF02374657
8. Ayres, M.C.C.; Escórcio, S.P.; Costa, D.A.; Chaves, M.H.; Júnior, G.M.V.; Cavalheiro, A.J. Chemical constituents from leaves of the *Qualea grandiflora*: attribution of the NMR data of two diastereoisomeric acylated flavonoid glycosids. *Quim. Nova* **2008**, 31, 1481-1484. https://doi.org/10.1590/S0100-40422008000600038
9. Hiruma-Lima, C.A.; Santos, L.C.; Kushima, H.; Pellizzon, C.H.; Silveira, G.G.; Vasconcelos, P.C.P.; Vilegas, W.; Souza Brito, A.R.M. *Qualea grandiflora*, a Brazilian "Cerrado" medicinal plant presents an important antiulcer activity. *J. Ethnopharmacol.* **2006**, 104, 207-214. https://doi.org/10.1016/j.jep.2005.09.002
10. Dousseau, S.; Alvarenga, A.A.; Arantes L.O.; Chaves I.S.; Avelino E.V. Technology of *Qualea grandiflora* Mart. (Vochysiaceae) seeds. *Cerne* **2013**, 19, 93-101. https://doi.org/10.1590/S0104-77602013000100012
11. Andrade, L.R.M.; Barros, L.M.G.; Echevarria G.F.; Velho do Amaral L.I.; Cotta M.G.; Rossatto D.R.; Haridasan M.; Franco A.C. Al-hyperaccumulator Vochysiaceae from the Brazilian Cerrado store aluminum in their chloroplasts without apparent damage. *Environ. Exp. Bot.* **2011**, 70, 37-42. https://doi.org/10.1016/j.envexpbot.2010. 05,013
12. Cury, N.F.; Silva, R.C.C.; Andre, M.S.F.; Fontes, W.; Ricart, C.A.O.; Castro, M.S.; Silveira, C.E.S.; Williams, T.C.R.; de Sousa, M.V.; Pereira, L.A.R. Root proteome and metabolome reveal a high nutritional dependency of aluminium in *Qualea grandiflora* Mart. (Vochysiaceae). *Plant Soil* **2020**, 446, 125-143. https://doi.org/10.1007/s11104-019-04323-3
13. Castro, L.M.R.; Vinson, C.C.; Almeida, A.L.; Cury, N.F.; André, M.S.F.; Williams, T.C.; Pereira, L.A.R. *Qualea grandiflora* Mart. (Vochysiaceae) seed reserves and aluminum: Usage and mobilization during germination and seedling development. *Plant Spec. Biol.* **2024**, 39, 205-219. https://doi.org/10.1111/1442-1984.12469
14. Silva, R.C.C. 2017. Análise do transcriptoma e do metaboloma revelam que *Qualea grandiflora* Mart. possui um metabolismo alumínio-dependente. Doctoral Thesis. Universidade de Brasília. Brazil. http://educapes.capes.gov.br/handle/capes/946323
15. Magalhães, J.V.; Liu, J.; Guimarães, C.T.; Lana, U.G.P.; Alves, V.M.C.; Wang, Y-H.; Schaffert, R.E.; Hoekenga, O.A.; Piñeros, M.A.; Shaff, J.E.; Klein, P.E.; Carneiro, N.P.; Coelho, C.M.; Trick, H.N.; Kochian, LV. A gene in the multidrug and toxic compound extrusion (MATE) family confers aluminum tolerance in sorghum. *Nat. Gen.* **2007,** 39, 1156-1161. https://doi.org/10.1038/ng2074
16. Liu, J.; Magalhaes, J.V.; Shaff, J.; Kochian, L.V. Aluminum-activated citrate and malate transporters from the MATE and ALMT families function independently to confer *Arabidopsis* aluminum tolerance. *Plant J.* **2009,** 57, 389-99. https://doi.org/10.1111/j.1365-313X.2008.03696.x
17. Iguchi, A.; Sanmiya, K.; Watanabe, K. Identification of genes encoding ALMT and MATE transporters as candidate aluminum tolerance genes from a typical acid soil plant, *Psychotria rubra* (Rubiaceae). *Peer J.* **2019**, 7, e7739. https://doi.org/10.7717/peerj.7739
18. Unamba, CIN.; Nag, A.; Sharma, RK. Next generation sequencing technologies: the doorway to the unexplored genomics of non-model plants. *Front. Plant Sci.* **2015**, 6, 1074. https://doi.org/10.3389/fpls.2015.01074
19. Nunes, R.; Gonçalves, A.R.; Telles, M.P.C. Data on the draft genome sequence of *Caryocar brasiliense* Camb. (Caryocaraceae): An important genetic resource from Brazilian savannas. *Data Brief* **2019**, 26, 104543. https://doi.org/10.1016/j.dib.2019.104543
20. Hirsch, C.N.; Buell, C.R. Tapping the promise of genomics in species with complex, non-model genomes. *Ann. Rev. Plant Biol.* **2013**, 64, 89-110. https://doi.org/10.1146/annurev-arplant-050312-120237
21. Seppey, M.; Manni, M.; Zdobnov, EM. BUSCO: Assessing Genome Assembly and Annotation Completeness. In: Kollmar M, eds. *Gene Predict. Meth. Mol. Biol*. **2019**, vol 1962. Humana, New York, NY. https://doi.org/10.1007/978-1-4939-9173-0_14

22. Song, K.; Li, L.; Zhang, G. Coverage recommendation for genotyping analysis of highly heterologous species using next-generation sequencing technology. *Sci. Rep.* **2016**, 6, 35736. https://doi.org/10.1038/srep35736
23. National Center for Biotechnology Information (NCBI). PubChem Element Summary for Atomic Number 13, Aluminum. https://pubchem.ncbi.nlm.nih.gov/element/Aluminum. Accessed Feb. 14, **2022**
24. Sasaki, T.; Yamamoto, Y.; Ezaki, B.; Kasuara M.; Ahn S.J.; Ryan P.; Delhaize E.; Matsumoto H. A wheat gene encoding an aluminum-activated malate transporter. *Plant J.* **2004**, 37, 645-653. https://doi.org/10.1111/j.1365-313x.2003.01991.x
25. Huang, J.; Li, X.; Chen, X.; Guo, Y.; Liang, W.; Wang, H. Genome-wide identification of soybean ABC transporters relate to aluminum toxicity. *Intl. J. Mol. Sci.* **2021**, 22, 6556. https://doi.org/10.3390/ijms22126556
26. Cury, N.F.; Silva, R.C.C.; Fayad, M.; Fontes, W.; Ricart, C.A.O.; Castro, M.S.; Silveira, C.E.S.; Valle de Sousa, M.; Pereira, L.A.R. Proteome dataset of *Qualea grandiflora* Mart. (Vochysiaceae) by LC–MS/MS label-free identification in response to aluminium. *Proteomics* **2019**, 19. https://doi.org/10.1002/pmic.201900148
27. Piñeros, M.A.; Cançado, G.M.; Maron, L.G.; Lyi, S.M.; Menossi, M.; Kochian, L.V. Not all ALMT1-type transporters mediate aluminum-activated organic acid responses: the case of ZmALMT1 - an anion-selective transporter. *Plant J.* **2008,** 53, 352-67. https://doi.org/10.1111/j.1365-313X.2007.03344.x
28. Moffatt, B.A.; Weretilnyk, E.A. Sustaining S-adenosyl-L-methionine-dependent methyltransferase activity in plant cells. *Physiol. Plant.* **2001**, 113, 435-442. https://doi.org/10.1034/j.1399-3054.2001.1130401.x
29. Ezaki, B.; Higashi, A.; Nanba, N.; Nishiuchi, T. An *S*-adenosyl methionine synthetase (SAMS) gene from *Andropogon virginicus* L. confers aluminum stress tolerance and facilitates epigenetic gene regulation in *Arabidopsis thaliana*. *Front. Plant Sci.* **2016,** 7, 1627. https://dx.doi.org/10.3389%2Ffpls.2016.01627
30. He, M-W.; Wang, Y.; Wu, J-Q.; Shu, S.; Sun, J.; Guo, S-R. Isolation and characterization of S-Adenosylmethionine synthase gene from cucumber and responsive to abiotic stress, *Plant Physiol. Biochem.* **2019**, 141, 431-445. https://doi.org/10.1016/j.plaphy.2019.06.006
31. Moffatt, B.A.; Stevens, Y.Y.; Allen, M.S.; Snider, J.D.; Pereira, L.A.R; Todorova, M.I.; Summers, P.S.; Weretilnyk, E.A.; Martin-McCaffrey, L.; Wagner, C. Adenosine kinase deficiency is associated with developmental abnormalities and reduced transmethylation. *Plant Physiol.* **2002,** 128, 812-821. https://doi.org/10.1104/pp.010880
32. Pereira, L.A.R.; Schoor, S.; Goubet, F.; Dupree, P.; Moffatt, B.A. Deficiency of adenosine kinase activity affects the degree of pectin methyl-esterification in cell walls of *Arabidopsis thaliana*. *Planta* **2006,** 224, 1401-1414. https://doi.org/10.1007/s00425-006-0323-z
33. Pereira, L.A.R.; Todorova, M.; Cai, X.; Makaroff, C.A.; Emery, R.J.N.; Moffatt, B.A. Methyl recycling activities are co-ordinately regulated during plant development. *J. Exp. Bot.* **2007**, 58, 1083-1098. https://doi.org/10.1093/jxb/erl275
34. Eticha, D.; Stass, A.; Horst, WJ. Cell-wall pectin and its degree of methylation in the maize root-apex: significance for genotypic differences in aluminium resistance. *Plant Cell Environ.* **2005,** 28, 1410-1420. http://dx.doi.org/10.1111/j.1365-3040.2005.01375.x
35. El-Moneim, D.A.; Contreras, R.; Silva-Navas, J.; Gallego, F.J.; Figueiras, A.M.; Benito, C. Pectin methylesterase gene and aluminum tolerance in *Secale cereale*. *Environ. Exp. Bot.* **2014**, 107, 125-133. http://dx.doi.org/10.1016/j.envexpbot.2014.06.006
36. Ma, X.; An, F.; Wang, L.; Guo, D.; Xie, G.; Liu, Z. Genome-wide identification of aluminum-activated malate transporter (ALMT) gene family in rubber trees (*Hevea brasiliensis*) highlights their involvement in aluminum detoxification. *Forests* **2020**, 11, 142 https://doi.org/10.3390/f11020142
37. Bolger, A.M.; Lohse, M.; Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, 30, 2114-2120. https://doi.org/10.1093/bioinformatics/btu170
38. Andrews, S. FastQC: a quality control tool for high throughput sequence data. **2010,** Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
39. Vurture, G.W.; Sedlazeck, F.J.; Nattestad, M.; Underwood, C.J.; Fang, H.; Gurtowski, J.; Schatz, M.C. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **2017**, 33, 2202–2204, https://doi.org/10.1093/bioinformatics/btx153

 doi:10.20944/preprints202602.0920.v1

40. Marçais, G.; Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **2011**, 27, 764-770. https://doi.org//10.1093/bioinformatics/btr011
41. Jackman, S.D.; Vandervalk, B.P.; Mohamadi, H.; Chu,J.; Yeo, S.; Hammond SA.; Jahesh G.; Khan H.; Coombe L.; Warren R.L.; Birol I. ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. *Gen. Res.* **2017,** 27, 768-777. https://dx.doi.org/10.1101%2Fgr.214346.116
42. Love, R.R.; Weisenfeld, N.I.; Jaffe, D.B.; Besansky, N.I.; Neafsey, D.E. Evaluation of DISCOVAR *de novo* using a mosquito sample for cost-effective short-read genome assembly. *BMC Genomics* **2016,** 17, 187. https://doi.org/10.1186/s12864-016-2531-7
43. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov. EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015,** 31, 3210-3212. https://doi.org//10.1093/bioinformatics/btv351
44. Palmer, J.M.; Stajich, J.E. Funannotate v1.8.1: eukaryotic genome annotation. Zenodo. **2020.** https://doi.org/10.5281/zenodo.4054262
45. Stanke, M.; Keller, O.; Gunduz, I.; Hayes, A.; Waack, S.; Morgenstern, B. AUGUSTUS: ab initio prediction of alternative transcripts. *Nuc. Ac. Res.* **2006**, 34, W435-W439. https://doi.org/10.1093/nar/gkl200
46. Korf, I. Gene finding in novel genomes. BMC *Bioinformatics* **2004**, 5, 59. https://doi.org/10.1186/1471-2105-5-59
47. Cantarel, B.L.; Korf, I.; Robb, S.M.; Parra, G.; Ross, E.; Moore, B.; Holt, C.; Alvarado, A.S.; Yandell, M. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Gen. Res.* **2008**, 18, 188-196. https://doi.org/10.1101/gr.6743907
48. Stanke, M.; Schöffmann, O.; Morgenstern, B.; Waack, S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **2006**, 7, 62. https://doi.org/10.1186/1471-2105-7-62
49. Bairoch, A.; Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL. *Nuc. Ac. Res.* **2000**, 28, 45-8. https://doi.org/10.1093/nar/28.1.45
50. Tatusov, R.L.; Fedorova, N.D.; Jackson, J.D.; Jacobs, A.R.; Kiryutin, B.; Koonin, E.V.; Krylov, D.M.; Mazumder, R.; Mekhedov, S.L.; Nikolskaya, A.N.; Rao, B.S.; Smirnov, S.; Sverdlov, A.V.; Vasudevan, S.; Wolf, Y.I.; Yin, J.J.; Natale, D.A. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **2003**, Epub PMID: 12969510; PMCID: PMC222959. https://doi.org/10.1186/1471-2105-4-41
51. Jones, P.; Binns, D.; Chang, H.Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; Pesseat, S.; Quinn, AF.; Sangrador-Vegas, A.; Scheremetjew, M.; Yong, SY.; Lopez, R.; Hunter, S. InterProScan 5, genome-scale protein function classification. *Bioinformatics* **2014**, 30, 1236-1240. https://doi.org/10.1093/bioinformatics/btu031
52. Conesa, A.; Götz, S.; García-Gómez, J.M.; Terol, J.; Talón, M.; Robles, M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **2005,** 21, 3674-3676. https://doi.org/10.1093/bioinformatics/bti610
53. Finn, R.D.; Bateman, A.; Clements, J.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Heger, A.; Hetherington, K.; Holm, L.; Mistry, J.; Sonnhammer, ELL.; Tate, J.; Punta, M. Pfam: the protein families database. *Nuc. Ac. Res.* **2014,** 42, D222–D230, https://doi.org/10.1093/nar/gkt1223
54. Rawlings, N.D.; Barrett, A.J.; Bateman, A. Using the MEROPS Database for proteolytic enzymes and their inhibitors and substrates. *Curr. Protoc. Bioinfo.* **2014,** 48, 1.25.1-1.25.33. https://doi.org/10.1002/0471250953.bi0125s48
55. Huerta-Cepas, J.; Szklarczyk, D.; Forslund, K.; Cook, H.; Heller, D.; Walter, M.C.; Rattei, T.; Mende, DR.; Sunagawa, S.; Kuhn, M.; Jensen, LJ.; von Mering, C.; Bork, P. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences *Nuc. Ac. Res.* **2016,** 44, D286–D293. https://doi.org/10.1093/nar/gkv1248
56. Huang, L.; Zhang, H.; Peizhi, W.; Entwistle, S.; Li, X.; Yohe, T.; Yi, H.; Yang, Z.; Yin, Y. dbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation. *Nuc. Ac. Res.* **2018,** 46, D516–D521. https://doi.org/10.1093/nar/gkx894

 doi:10.20944/preprints202602.0920.v1

57. Kanehisa, M.; Sato, Y.; Morishima, K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **2016**, 428, 726-731. https://doi.org/10.1016/j.jmb.2015.11.006