DATA NOTE

Plastid genome of *Chenopodium petiolare* from Trujillo, Peru

Flavio Aliaga[1,2,3*], Mario Zapata-Cruz[4] and Silvia Ana Valverde-Zavaleta[4]

Abstract

Objectives The Peruvian Andean region is an important center for plant domestication. However, to date, there have been few genetic studies on native grain, which limits our understanding of their genetic diversity and the development of new genetic studies for their breeding. Herein, we revealed the plastid genome of *Chenopodium petiolare* to expand our knowledge of its molecular markers, evolutionary studies, and conservation genetics.

Data description Total genomic DNA was extracted from fresh leaves (voucher: USM < PER > :MHN333570). The DNA was sequenced using Illumina Novaseq 6000 (Macrogen Inc., Seoul, Republic of Korea) and reads 152,064 bp in length, with a large single-copy region of 83,520 bp and small single-copy region of 18,108 bp were obtained. These reads were separated by a pair of inverted repeat regions (IR) of 25,218 bp, and the overall guanine and cytosine (GC) was 37.24%. The plastid genome contains 130 genes (111 genes were unique and 19 genes were found duplicated in each IR region), including 86 protein-coding genes, 36 transfer RNA-coding genes, eight ribosomal RNA-coding genes, and 25 genes with introns (21 genes with one intron and four genes with two introns). The phylogenetic tree reconstructed based on single-copy orthologous genes and maximum likelihood analysis indicated that *Chenopodium petiolare* is most closely related to *Chenopodium quinoa*.

Keywords Plastid genome, Chloroplast genome, Chenopodiaceae, *Chenopodium petiolare*, Lomas del Cerro Campana, Trujillo, Peru

Objective

Chenopodium petiolare Kunth is a native grain of the Andean region, this annual herb grows in the Peruvian Andean formations at altitudes of 200–3,900 m.a.s.l., and its grains are small and black with high concentration of saponins [1, 2]. It is a diploid species with a small number of chromosomes ($2n = 2x = 18$) belonging to the Chenopodiaceae family. Its outstanding features are drought stress tolerance and resistance to diseases [1, 3]. *Chenopodium petiolare* has multiple uses including being used as cattle feed, in cooking local dishes such as quispiño (dark muffin), and in traditional medicine mainly for bone fractures [1].

The plastid genome has a quadripartite structure: a large single-copy (LSC) of 80–90 kilobase pairs (kb), a small single-copy (SSC) of 16–27 kb, and two sets of inverted repeats (IRs) of 20–28 kb, with 110–130 unique genes, including protein-coding genes, transfer RNA (tRNA), and ribosomal RNA (rRNA) [4, 5]. In recent years, declining genome sequencing costs resulted in more than 790 complete plant genomes of different

*Correspondence:
Flavio Aliaga
flavio.aliaga@upn.edu.pe
[1] Grupo de Investigación en Ecología Evolutiva, Protección de Cultivos, Remediación Ambiental, y Biotecnología (EPROBIO), Universidad Privada del Norte, Trujillo 13011, Peru
[2] Dirección de Investigación, Innovación y Responsabilidad Social, Universidad Privada del Norte, Trujillo 13009, Peru
[3] Capítulo de Ingeniería Agronómica, Consejo Departamental de La Libertad (CDLL), Colegio de Ingenieros del Perú (CIP), Trujillo 13008, Peru
[4] Plant Science Laboratory (PSL), Trujillo 13009, Peru

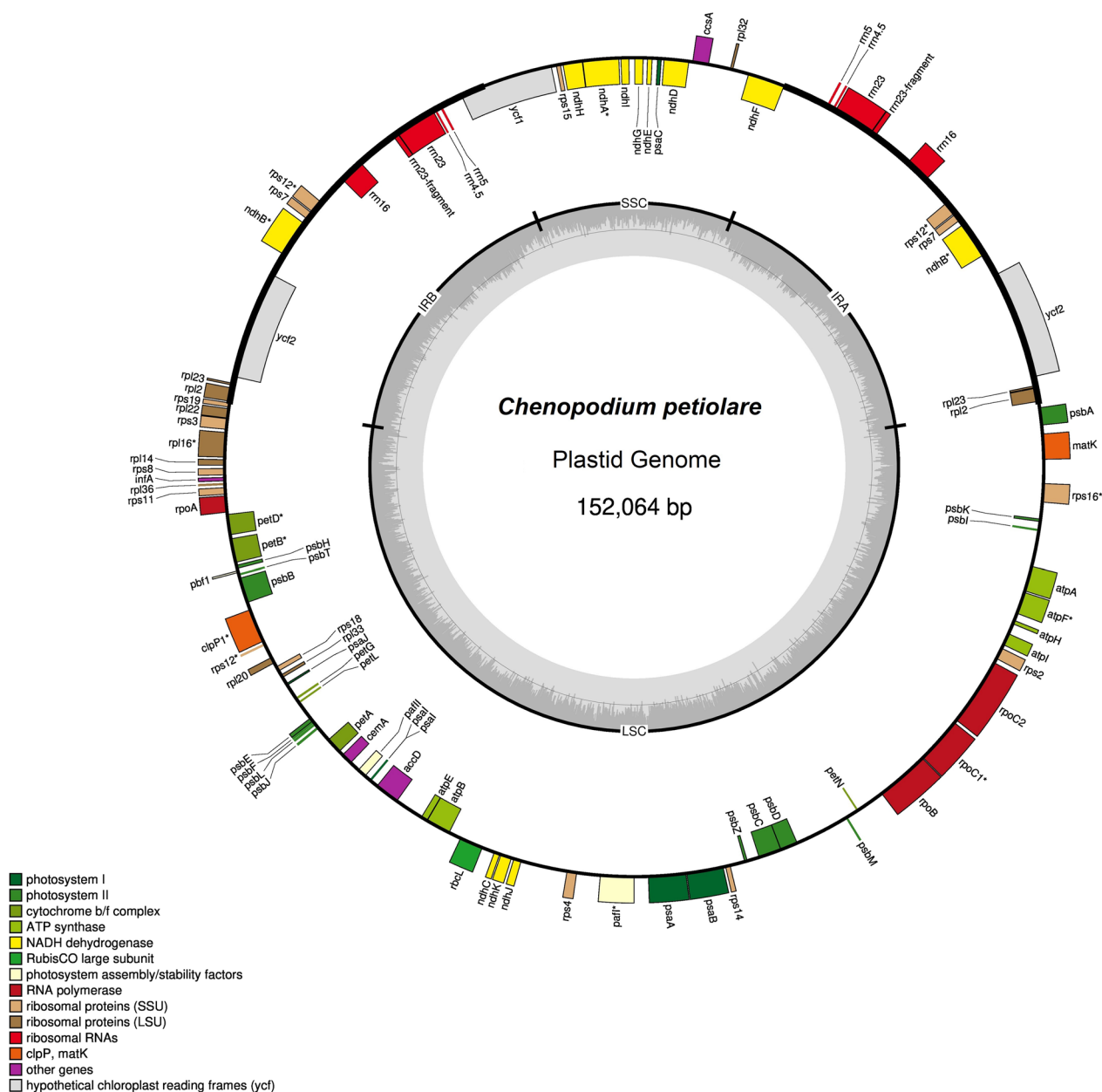Aliaga *et al. BMC Research Notes* (2024) 17:69

Page 2 of 5

Fig. 1 Circular map of *Chenopodium petiolare* chloroplast genome. The thick lines indicate the IR1 and IR2 regions, which separate the large single-copy (LSC) and small single-copy (SSC) regions. Genes marked inside the circle are transcribed clockwise, and genes marked outside the circle are transcribed counterclockwise. Genes are color-coded based on their function, shown at the bottom left. The inner circle indicates the inverted boundaries and guanine and cytosine (GC) content

species becoming available [6, 7]. Recently, some *Chenopodium* plastid genomes such as *Chenopodium acuminatum* [8], *Chenopodium album* [9], *Chenopodium quinoa* [10], *Chenopodium ficifolium* [11], became publicly available. However, despite the few genetic data available, we have only begun to investigate the genomics of native grains of great importance for plant breeding programs. In the present study, we report the first plastid genome sequence submitted for an isolate of *Chenopodium petiolare*, which will expand our knowledge about its plant molecular breeding, molecular markers, evolutionary studies, and conservation genetics.
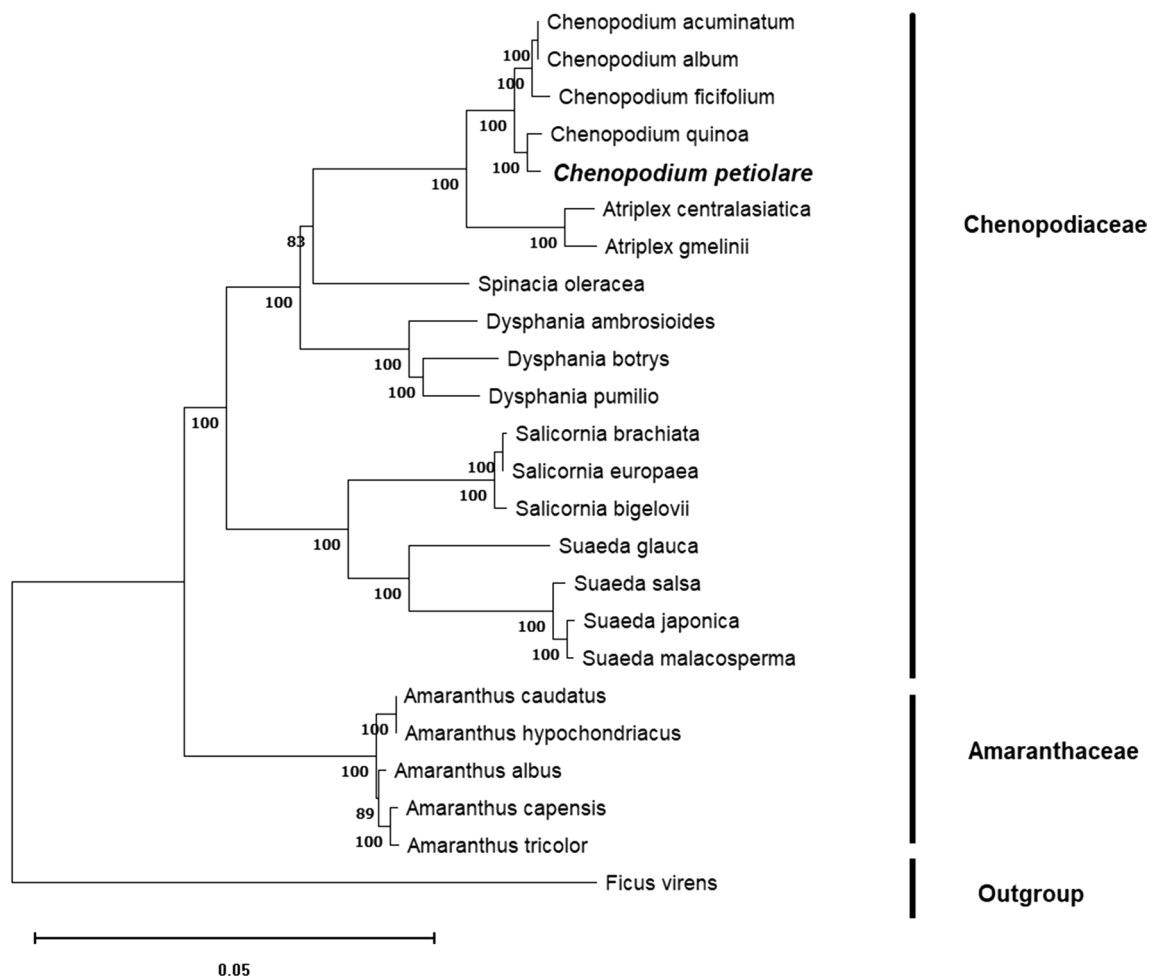
Aliaga *et al. BMC Research Notes*   (2024) 17:69

Page 3 of 5

Fig. 2 Phylogenetic tree of 24 plastid genomes. Maximum likelihood analysis based on single-copy orthologous protein. Bootstrap values on the branches were calculated from 1000 replicates

Table 1 Overview of data files/data sets

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data file 1	Herbarium specimen voucher of *Chenopodium petiolare* Kunth (USM < PER > :333,570)	*Picture file (.jpg)*	*Figshare* https://doi.org/10.6084/m9.figshare.23574303.v1 [25]
Data file 2	Figure 1 Circular map of *Chenopodium petiolare* plastid genome	*Picture file (.jpg)*	*Figshare* https://doi.org/10.6084/m9.figshare.23574270.v1 [26]
Data file 3	Plastid genome features of the *Chenopodium petiolare*	*Document file (.docx)*	*Figshare* https://doi.org/10.6084/m9.figshare.23574306.v1 [27]
Data file 4	Genes present in the plastid genome of *Chenopodium petiolare*	*Document file (.docx)*	*Figshare* https://doi.org/10.6084/m9.figshare.23574312.v1 [28]
Data file 5	Figure 2 Phylogenetic tree of 24 plastid genomes	*Picture file (.jpg)*	*Figshare* https://doi.org/10.6084/m9.figshare.23574327.v1 [29]
Data set 1	*Chenopodium petiolare* chloroplast, complete genome	*Fasta file (.fasta)*	GenBank from NCBI repository under the accession number OQ957163 (https://identifiers.org/ncbi/insdc:OQ957163) [30]

Data description

Total genomic DNA was extracted from approximately 100 mg of fresh leaves (from voucher number USM < PER >:MHN333570) (Data file 1) using a cetyltrimethyl ammonium bromide (CTAB) protocol [12]. Genomic DNA quality was assessed using a fluorometry-based Qubit (Thermo Fisher Scientific, USA) coupled to a Broad Range Assay kit (Thermo Fisher Scientific, USA). High-quality DNA (230/260 and 260/280 ratios > 1.8) was normalized (20 ng/μL) to examine its integrity using 1% (*w*/*v*) agarose gel electrophoresis. Qualified DNA was fragmented, and the TruSeq Nano DNA kit (Illumina, San Diego, CA, USA) was used to construct an Illumina paired-end (PE) library. PE sequencing (2×150 bp) was performed using the Illumina NovaSeq 6000 platform (Macrogen, Inc., Seoul, Republic of Korea) [13]. All adapters and low-quality reads were removed using the FastQC [14] and Cutadapt [15] programs. PE reads (2×150 bp) were evaluated for quality using QUAST [16] analysis, and subsequent steps used clean data. Then, clean reads obtained were assembled into a circular contig using NOVOPlasty (version.4.3) [17], with *C. quinoa* (NC_034949) as the reference. Data can be accessed from NCBI GenBank under the accession number OQ957163 [30]. The plastid genome was annotated using the Dual Organellar GenoMe Annotator GeSeq [18] and CpGA-VAS2 [19]. A circular genome map was constructed using OGDRAW (version 1.3.1) [20] (Fig. 1). The plastid genome encoded 130 genes, of which 111 were unique, and 19 were duplicated in the inverted repeat (IR) region. The chloroplast genome contained 86 protein-coding genes, 36 tRNA-coding genes, eight rRNA-coding genes, and 25 genes with introns (21 genes with one intron and four genes with two introns), as shown in Data file 3.

The plastome contained 111 unique genes, of which there were 28 tRNA genes, four rRNA genes, and 79 protein-coding genes. The latter comprised 21 ribosomal subunit genes (nine large subunits and 12 small subunit), four DNA-directed RNA polymerase genes, 45 genes were involved in photosynthesis (11 encoded subunits of the NADH oxidoreductase, seven for photosystem I, 14 for photosystem II, six for the cytochrome b6/f complex, six for different subunits of ATP synthase, and one for the large chain of ribulose biphosphate carboxylase), eight genes were involved in different functions, and one gene was of unknown function (Data file 4). Phylogenetic analysis reconstruction was performed using 24 complete chloroplast genome sequences to infer the phylogenetic relationships among *Chenopodium* species, and *Ficus virens* was used as an outgroup (Fig. 2). Single-copy orthologous genes were identified using the Orthofinder pipeline (version 2.2.6) [21]. For each gene family, the nucleotide sequences were aligned using the L-INS-i algorithm in MAFFT (version 7.453) [22]. A phylogenetic tree based on maximum likelihood (ML) was constructed using RAxML (version 8.2.12) [23] with the GTR CAT model. A phylogenetic ML tree was reconstructed and edited using MEGA (version 11) [24] with 1000 replicates. The phylogenetic tree illustrated that *Chenopodium petiolare* is closely related to *Chenopodium quinoa* [10].

Limitations

This study used leaf samples of *Chenopodium petiolare* from the Lomas del Cerro Campana Private Conservation Area in Trujillo, Peru. Administratively, this process takes longer than necessary to obtain the corresponding access permit for plant sample collection.

Abbreviations
LSC	Large single-copy
SSC	Small single-copy
IR	Inverted repeat
tRNA	Transfer RNA
rRNA	Ribosomal RNA

Author contributions
FA conceived and designed the experiments. FA, MZ-C and SAV-Z performed the experiments, analyzed the data and wrote the manuscript. FA prepared the figures and tables. FA, MZ-C and SAV-Z corrected and proofread the manuscript. All authors read and approved the final manuscript.

Availability of data and materials
The data described in this Data note can be freely and openly accessed on GenBank of NCBI repository under the accession number OQ957163, and figshare. Please see Table 1 and references list [25–30] for details and links to the data.

Declarations

Ethics approval and consent to participate
Not applicable.

Aliaga *et al. BMC Research Notes*   (2024) 17:69

Page 5 of 5

References

1. Mujica A, Jacobsen S. La Quinua (Chenopodium quinoa Willd.) y sus parientes silvestres. In: Moraes R, Øllgaard B, Kvist L, Borchsenius F, Balslev H, editors. Botánica Económica de los Andes Centrales. La Paz: Universidad Mayor de San Andrés; 2006. p. 453–6.
2. Tropicos. Missouri Botanical Garden. 2024. https://www.tropicos.org/collection/1924364. Accessed 29 Jan 2024.
3. Romero M, Mujica A, Pineda E, Ccamapaza Y, Zavalla N. Genetic identity based on simple sequence repeat (SSR) markers for Quinoa (Chenopodium quinoa Willd.). Cienc Investig Agrar. 2019;46:166–78.
4. Ozeki H, Umesono K, Inokuchi H, Kohchi T, Ohyama K. The chloroplast genome of plants: a unique origin. Genome. 1989;31:169–74.
5. Wang W, Lanfear R. Long-reads reveal that the chloroplast genome exists in two distinct versions in most plants. Genome Biol Evol. 2019;11:3372–81.
6. Marks RA, Hotaling S, Frandsen PB, VanBuren R. Representation and participation across 20 years of plant genome sequencing. Nat Plants. 2021;7:1571–8.
7. Sun Y, Shang L, Zhu QH, Fan L, Guo L. Twenty years of plant genome sequencing: achievements and challenges. Trends Plant Sci. 2022;27:391–401.
8. Wariss HM, Qu XJ. The complete chloroplast genome of Chenopodium acuminatum Willd. (Amaranthaceae). Mitochondrial DNA B Resour. 2021;6:174–5.
9. Devi RJ, Thongam B. Complete chloroplast genome sequence of *Chenopodium album* from Northeastern India. Genome Announc. 2017;5:e01150-e1217.
10. Gao MZ, Dong YH, Valcárcel V, Ren ZM, Li YL. Complete chloroplast genome of the grain *Chenopodium quinoa Willd.*, an important economical and dietary plant. Mitochondrial DNA B Resour. 2021;6:40–2.
11. Yongsung K, Youngjae C, Jongsun P. The complete chloroplast genome of *Chenopodium ficifolium* Sm. (Amaranthaceae). Mitochondrial DNA B Resour. 2019;4:872–3.
12. Doyle J. DNA Protocols for Plants. In: Hewitt GM, Johnston AWB, Young JPW, editors. Molecular Techniques in Taxonomy. Berlin: Springer; 1991. p. 283–93.
13. Modi A, Vai S, Caramelli D. Lari M (2021) The illumina sequencing protocol and the novaseq 6000 system. In: Mengoni A, Bacci G, Fondi M, editors. Bacterial Pangenomics: methods and protocols. New York: Springer; 2021. p. 15–42.
14. Wingett SW, Andrews S. FastQ screen: a tool for multi-genome mapping and quality control. F1000Res. 2018;7:1–5.
15. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17:10–2.
16. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29:1072–5.
17. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 2017;45:e18.
18. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, et al. GeSeq - versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 2017;45:W6-11.
19. Shi L, Chen H, Jiang M, Wang L, Wu X, Huang L, et al. CPGAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic Acids Res. 2019;47:W65-73.
20. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 2019;47:W59-64.
21. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20:1–14.
22. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
23. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3.
24. Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. Mol Biol Evol. 2021;38:3022–7.
25. Aliaga F, Zapata-Cruz M, Valverde-Zavaleta SA. Herbarium specimen voucher of *Chenopodium petiolare* Kunth (USM<PER>:333570). figshare. 2023. https://doi.org/10.6084/m9.figshare.23574303.v1
26. Aliaga F, Zapata-Cruz M, Valverde-Zavaleta SA. Circular map of *Chenopodium petiolare* chloroplast genome. figshare. 2023. https://doi.org/10.6084/m9.figshare.23574270.v1
27. Aliaga F, Zapata-Cruz M, Valverde-Zavaleta SA. Chloroplast genome features of the *Chenopodium petiolare*. figshare. 2023. https://doi.org/10.6084/m9.figshare.23574306.v1
28. Aliaga F, Zapata-Cruz M, Valverde-Zavaleta SA. List gene on *Chenopodium petiolare* choroplast genome. figshare. 2023. https://doi.org/10.6084/m9.figshare.23574312.v1.
29. Aliaga F, Zapata-Cruz M, Valverde-Zavaleta SA. Phylogenetic tree of 24 chloroplast genome. figshare. 2023. https://doi.org/10.6084/m9.figshare.23574327.v1
30. Genbank of National Center for Biotechnology Information (NCBI). 2023. https://identifiers.org/ncbi/insdc:OQ957163. Accessed 29 Jan 2024.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.