Reference genome assembly of a tetraploid accession of the tuber crop *Tropaeolum tuberosum*

Daniela Ruiz-Mateus[1, †], Ina Scheffler[2,†], María del Pilar Márquez-Cardona[1], Thomas Greb[2,3], Wilson Terán[1] and Pascal Hunziker[2,3*]

[1]Biología de Plantas y Sistemas Productivos, Departamento de Biología, Pontificia Universidad Javeriana, Bogotá, Colombia

[2]Centre for Organismal Studies (COS), Heidelberg University, Germany

[3]Cluster of Excellence GreenRobust, Heidelberg University, Germany

[†]Authors contributed equally

**Author for correspondence:* Pascal Hunziker, Centre for Organismal Studies (COS), Heidelberg University, Im Neuenheimer Feld 360, 69120 Heidelberg, Germany, phone: +49-(0)6221-54-5785, email: pascal.hunziker@cos.uni-heidelberg.de

Abstract

Tropaeolum tuberosum is a tuber-forming crop native to the Andes, valued for its nutritional content, pest resistance and adaptation to high-altitude environments. Despite its contribution to food security and sustainable agriculture, it remains an underutilized species with scarce genomic resources, limiting genetic and functional studies. To address this gap, we generated a reference genome assembly for a European *ex situ* tetraploid accession of *T. tuberosum* using PacBio HiFi sequencing. The assembly spans 1.3 Gb in 1,805 contigs (contigs N50 = 32.2 Mb, longest contig = 60 Mb) and recovers 87% of the estimated genome size. We assessed assembly completeness and accuracy using Benchmarking Universal Single-Copy Orthologs, which detected 98.5% complete genes (21.7% single-copy, 76.8% duplicated), 0.7% fragmented, and 0.7% missing, demonstrating near-complete gene space recovery consistent with a high-quality tetraploid genome. To evaluate transferability, we resequenced a field-collected Colombian genotype of economic relevance using Oxford Nanopore technology. Read mapping showed 99.7% of primary alignments to the reference (weighted mean coverage = 16.4×, 96.1% at ≥5×), confirming broad sequence conservation between accessions and validating the suitability of the *ex situ* reference as an anchor for *in situ* germplasm. Repetitive elements accounted for 71.3% of the genome. Using ANNEVO, we annotated 56,354 high-confidence protein-coding genes, achieving 98.3% complete BUSCOs, a PSAURON score of 97.2, and 90.5% taxonomic consistency with the rosid lineage (OMArk). This high-quality reference genome establishes a foundational resource for comparative genomics, population genomics and functional analyses on *T. tuberosum*, supporting future breeding and conservation of this important food resource.

Keywords: PacBio HiFi, underutilized, orphan crop, cubio, mashua, isaño

Significance statement

T. tuberosum is an underutilized Andean tuber crop that combines highly nutritious tubers with remarkable disease resistance and UV tolerance, supporting Andean smallholder livelihoods under harsh high-altitude conditions. Beyond its role as a food crop, it is cultivated widely as an ornamental due to its abundant, brightly colored flowers closely resembling those of its relative nasturtium. Despite this dual relevance, genomic resources have been entirely lacking, limiting efforts to understand its domestication history, adaptation, and potential as a climate-resilient alternative to potato with higher nutraceutical properties. Our reference genome assembly provides the first genome-scale resource for *T. tuberosum*, creating a foundation for dissecting polyploid genome organization, key agronomic and nutritional traits, and interactions with biotic and abiotic stress. By enabling evolution, domestication, comparative genomics, breeding, conservation, and functional studies, this resource unlocks *T. tuberosum* as a model for resilient, diversified cropping systems globally.

Introduction

Tropaeolum tuberosum Ruiz & Pav. (2n = 4x = 52), also known as *mashua, cubio* or *isaño*, is a tuber-forming crop traditionally cultivated by smallholder farmers in local communities in the Andes mountains. It thrives under harsh conditions – including strong diurnal temperature shifts and nutrient-poor soils – producing nutrient-rich tubers with high levels of carbohydrates, proteins, vitamins and diverse secondary metabolites with antioxidant, anti-inflammatory, diuretic and antimicrobial properties, including glucosinolates (Apaza Ticona et al. 2020; Luera-Quiñones et al. 2025; Coloma et al. 2022). These traits make *T. tuberosum* a promising species for food security as well as resilient and sustainable agriculture.

T. tuberosum belongs to the genus *Tropaeolum* (family Tropaeolaceae, order Brassicales), a Neotropical lineage of roughly 90 herbaceous species that includes ornamental nasturtiums and several tuber-forming taxa (Andersson & Andersson 2000). Its placement in the

Brassicales is of particular biochemical interest: although glucosinolates are characteristic of this order, their diversity outside the Brassicaceae remains poorly characterized , highlighting the value of genomic data from non-brassicaceous taxa (Agerbirk & Koch 2025; Bird et al. 2025).

Like other edible Andean root and tuber crops, *T. tuberosum* exhibits remarkable morphological and genetic diversity (Pissard et al. 2008; Ruíz-Mateus et al. 2025; Fonseca & Márquez-Cardona 2024), yet low commercial value and limited market integration are jeopardizing its conservation as a food resource. In contrast to globally important tuber crops such as potato, its improvement has relied primarily on traditional farmer selection. Thus, despite its agronomic and nutritional potential, *T. tuberosum* remains largely underutilized (Luera-Quiñones et al. 2025).

No reference genome has been available for *T. tuberosum* and genomic resources remain scarce. Here we report a long-read–based, pseudo-haplotype–resolved genome assembly of a tetraploid accession of *T. tuberosum*. While a recent preprint described a genome for *Tropaeolum majus* (Friedhoff et al. 2024), this is the first assembly of *T. tuberosum*, providing a genomic framework for investigating specialized metabolism, tuber development, and environmental adaptation in a non-brassicaceous Brassicales lineage, while laying the foundation for genomics-assisted improvement of this neglected crop.

Results and Discussion

We sequenced and assembled the genome of *Tropaeolum tuberosum* Ruiz & Pav (Tropaeolaceae*, Brassicales*) accession BGHEID007454 from the living collection of the Botanic Garden Heidelberg, Germany (**Figure 1a-b**). Using single-molecule real-time sequencing technology from Pacific Biosciences, a total of 128.2 Gb of PacBio HiFi sequencing data were produced. *K*-mer analysis revealed an estimated monoploid genome size of 418.1 Mb, consistent with an autotetraploid genome architecture and an estimated

heterozygosity rate of 4.7% (**Figure 1c**). A genome survey based on *k*-mer frequency analysis of Illumina data estimated comparable metrics (**Supplementary Fig. S1**). *K*-mer-based genome size estimates (418-497 Mb) were moderately smaller than the monoploid genome size inferred from flow cytometry (2c = 2.48 Gb) (Veselý et al. 2013), which assuming tetraploidy yields a monoploid size of 620 Mb. Such underestimation is commonly observed in repeat-rich polyploid genomes due to compression of repetitive and near-identical homeologous regions. *K*-mer analysis of both HiFi and Illumina data supports an autotetraploid origin of *T. tuberosum*, with aaab forms (2.75%) exceeding aabb forms (1.72%) (**Figure 1c, Supplementary Fig. S1**), in line with the dominant cytological evidence (Johns & Neil Towers 1981).

Genome assembly

Initial genome assembly using hifiasm yielded a draft primary assembly containing 2,189 contigs and a pseudo-haplotype-resolved assembly comprising two haplotypes (hap1, 1.21 Gb and hap2, 0.98 Gb) of unequal size, as expected for autotetraploid genome assembly (**Supplementary Table S1-2**). Because the four homeologous copies do not segregate into two discrete sequence clusters, phasing algorithms partition them based on local heterozygosity patterns. Bandage plots showed numerous small contigs with low depth (**Supplementary Fig. S2**). Coverage distribution was bimodal, with a peak at 30–200× representing the nuclear genome and a tail of high-coverage contigs (>200×, n = 93) (**Supplementary Fig. S3**). 292 contigs showed very low coverage (<10×). Kraken2 classification confirmed that all 292 low-coverage contigs originated from non-target organisms (**Supplementary Table S3**). Contaminants (n = 292), plastid-derived (n = 60), mitochondria-derived (n = 30) and repeat contigs (n = 2) were removed (384 contigs in total, accounting for 17.5% of the original assembly by count).

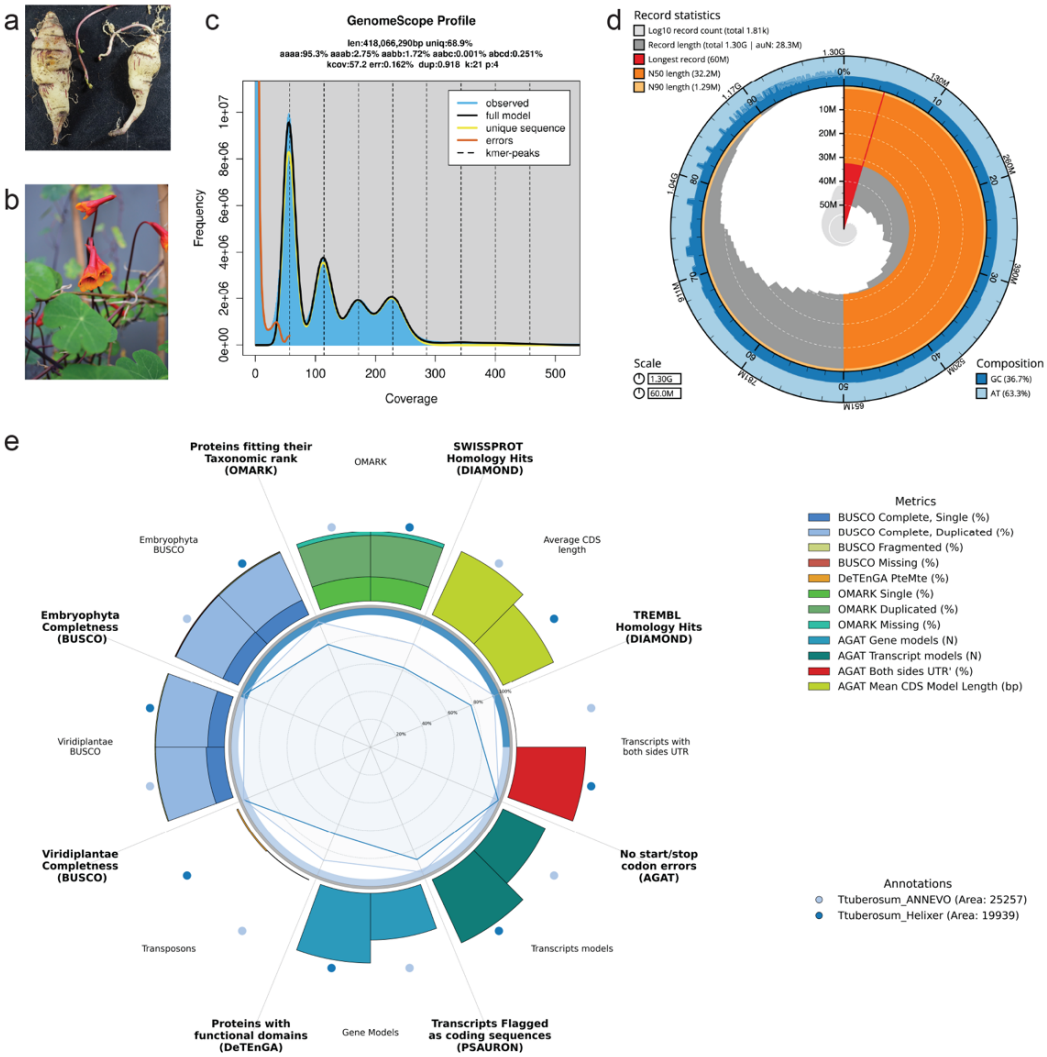

Fig. 1: The *Tropaeolum tuberosum* reference genome. **(a)** Tubers of the BGHEID007454 accession. **(b)** Leaves, stems and flowers of the BGHEID007454 accession. **(c)** Distribution profile of 21-mer analysis**. (d)** Snail plot of the primary reference genome. **(e)** Benchmarking Helixer and ANNEVO annotations.

The final primary assembly comprised 1,805 contigs (N50 of 32.2 Mb; largest = 60 Mb), with 90% of the assembly contained in 84 contigs (**Figure 1d, Table 1, Supplementary Table S1**). The size of 1.3 Gb matches the expected size of two monoploid complements (~1.24 Gb (Veselý et al. 2013)), consistent with a collapsed pseudo-haploid assembly representing most sequence diversity across four homeologs. BUSCO analysis showed 98.5% completeness, 76.8% duplication, and 0.7% fragmentation, with identical results before and after decontamination, confirming no gene loss (**Table 1, Supplementary Table S1**). Merqury analysis yielded a consensus quality value (QV) of 60.4 and *k*-mer completeness of 75.4%

124 (**Table 1, Supplementary Table S4**). The read-only and assembly copy number spectra
125 aligned at ~56× coverage, as expected for a collapsed tetraploid, indicating that reduced
126 multiplicity reflects intended homeolog collapse rather than sequence loss (**Supplementary**
127 **Fig. S4**). Assembly spectra exceeding read-only lines at higher multiplicities confirm
128 comprehensive representation of multicopy sequences (**Supplementary Fig. S5**).

129 To assess the utility of the assembly as a reference for natural populations, Nanopore long
130 reads (~24×) from the field-collected genotype B15, originating from a farm in the department
131 of Boyacá, Colombia, were mapped to the final primary assembly. A mapping rate of 99.7%
132 and ≥98% of the assembly covered at a minimum depth of 5× demonstrate that the assembly
133 captures the genomic content of geographically distant, *in situ* material with high fidelity,
134 supporting its use as a reference for population-level and ecological genomic studies
135 (**Supplementary Table S5**).

136 Of the 1,805 assembly contigs, 1,009 (55.9%) were covered at ≥10× mean depth, and a further
137 288 (16.0%) at 5–10× (**Supplementary Table S5**). The remaining 508 contigs (28.2%)
138 showed low or absent coverage (<5×), collectively spanning only 23.5 Mb — 1.8% of the
139 nuclear assembly by length. Repeat density in this fraction (998 masked bp/kb) was
140 indistinguishable from that of well-covered contigs (952 bp/kb), against a genome-wide
141 nuclear repeat content of 71.3%, indicating that the low-coverage fraction is not an enrichment
142 of repetitive sequence but rather reflects the structural properties of short, repeat-dense
143 contigs that are inherently difficult to anchor by read mapping regardless of genotype.
144 Consistent with this, gene density on low-coverage contigs (164 models/Mb) was comparable
145 to that of well-covered contigs (128 models/Mb), and only 3,446 of 56,354 annotated gene
146 models (6.1%) resided in this fraction (**Supplementary Table S5**). The cross-genotype
147 validation therefore confirms that the assembly provides comprehensive coverage of the genic
148 and single-copy complement of the *T. tuberosum* genome, establishing it as a transferable
149 reference resource for the species.

150 **Table 1:** Statistics for the *Tropaeolum tuberosum* reference genome assembly and
151 annotation.

Primary assembly	
Number of contigs	1,805
Total assembly size	1.30 Gbp
Largest contig	60 Mbp
N50 contig length	32.2 Mbp
L50 contig count	16
N90 contig length	1.3 Mbp
L90 contig count	84
GC content	36.70 %
BUSCO v5 embryophyta_odb10	C:98.5% [S:21.7%, D:76.8%], F:0.7%, M:0.7%, n:1614, E:1.9%
QV	60.4025
K-mer completeness	75.4455
Repeats	
LINE retroelements	54,466 (36.8%)
LTR retroelements	559,198 (33.6%)
DNA transposons	49,786 (1.9%)
Unclassified	680,448 (18.7%)
Small RNA	31,591 (13.2%)
Annotation (ANNEVO)	
Gene Models	56,354
Exons	271,743
BUSCO v5 embryophyta_odb10	C:98.3% [S:24.2%, D:74.2%], F:0.7%, M:0.9%, n:1614
PSAURON score	97.2
OMArk Consistency Results	Cons:90.45% [P:7.27%; F:2.32%], Inco:6.95% [P:1.54%, F:0.34%], Cont:0.00%, Unkn:2.60%
OMArk Completeness Results	malvids-HOGs: 11,704; S:34.71%, D:59.25% [U:59.02%, E:0.23%], M:6.03%
OMArk Species Composition	rosids: 97.40%
Proteins With TREMBL Hits	96.15
Proteins With SWISSPROT Hits	80.39

152

153 Together, the assembly statistics, BUSCO gene completeness, Merqury QV scores, *k*-mer
154 completeness and cross-genotype validation demonstrate that the *T. tuberosum* genome
155 assembly is highly accurate, complete and suitable for downstream analyses. Our primary
156 assembly provides an annotation-grade reference genome sequence that was used for all
157 gene prediction and functional annotation analyses described below.

158 The final haplotype-resolved assembly comprised hap1 (2,022 contigs, N50 = 18.4 Mb) and
159 hap2 (580 contigs, N50 = 26.3 Mb), totalling 1.19 Gb and 0.96 Gb, respectively
160 (**Supplementary Table S2**). Combined, the 2.19 Gb assembly represents ~3.5× the
161 monoploid genome size (620 Mb), recovering ~88% of the expected tetraploid complement
162 (~2.48 Gb); the missing ~12% likely corresponds to highly repetitive or heterochromatic

163 regions. Gene completeness analysis showed slightly lower BUSCO scores per haplotype but
164 98.8% completeness when merged, indicating that unassembled sequence is largely non-
165 genic (**Supplementary Table S2**). Merqury analysis revealed ~70% *k*-mer recovery per
166 haplotype and 95% combined, with consensus QV 60–65, reflecting high sequence accuracy.
167 The *k*-mer spectrum displayed four coverage peaks (~56×, ~112×, ~168×, ~224×), matching
168 the expected 1×–4× pattern for an autotetraploid at ~56× per haploid copy
169 (**Supplementary Fig. S6**). Minimal read-only fractions confirmed near-complete
170 representation of the read *k*-mer space, while balanced hap1- and hap2-specific fractions and
171 a dominant shared component at higher multiplicities (**Supplementary Fig. S7**) are
172 characteristic of autopolyploidy. Distinct haplotype-specific peaks verified successful phasing,
173 and per-haplotype spectra showed that missing *k*-mers reflect sequence assigned to the
174 alternate haplotype rather than gaps (**Supplementary Figs. S8–S9**). Together, these results
175 indicate a high-quality pseudo-haplotype-resolved, partially phased assembly of an
176 autotetraploid genome

177

178 **De novo repeat and gene annotation**

179 Using RepeatModeler, we identified 1,706 repeat families, including 732 structurally defined
180 LTR retrotransposons (**Table 1, Supplementary Table S6**). Repetitive elements accounted
181 for 71.3% of the genome, dominated by Gypsy and Copia LTR retrotransposons. This
182 composition aligns with other eudicot genomes of similar size (Chanderbali et al. 2022), where
183 LTR retrotransposons contribute substantially to genome expansion (Zhang et al. 2020). The
184 final nonredundant repeat library was used for genome-wide masking and annotation.

185 Helixer and ANNEVO predicted 84,059 and 56,354 protein-coding genes, respectively (**Figure**
186 **1e, Table 1, Supplementary Table S7**). Both annotations showed near-complete gene space
187 coverage, with BUSCO completeness scores of 98.1% and 98.3%, respectively. In both cases,
188 the majority of complete BUSCOs were duplicated (~75%), consistent with the retention of

duplicated genes following ancient whole genome duplication events common in flowering plants (Van de Peer et al. 2017). Despite comparable completeness, ANNEVO consistently produced higher-quality gene models across all metrics assessed. ANNEVO achieved a substantially higher ORF accuracy score (PSAURON: 97.2 vs. 87.5), greater taxonomic consistency with the rosid lineage (OMArk: 90.5% vs. 71.5%), and a higher proportion of proteins with significant homology hits in both TrEMBL (96.2% vs. 78.2%) and Swiss-Prot (80.4% vs. 62.0%). Helixer's higher gene count was accompanied by clear signs of inflation: 4,557 overlapping gene models (vs. 59 in ANNEVO), 504 models with missing stop codons, 4.2% of proteins shorter than 50 amino acids (vs. 0.1%), and nearly four times more transposable element-derived transcripts (DeTEnGA PteMte: 1.02% vs. 0.27%). The higher proportion of unknown hits in Helixer's OMArk analysis (19.95% vs. 2.60%) further suggests that a substantial fraction of its additional gene models lack biological support.

Based on these results, the ANNEVO annotation was selected as the primary gene set. The 56,354 high-confidence protein-coding genes predicted by ANNEVO are consistent with the tetraploid genome of *T. tuberosum*. The closely related diploid *Tropaeolum majus* harbors 32,924 protein-coding genes (Friedhoff et al. 2024). Assuming a similar ancestral gene complement, whole genome duplication would be expected to approximately double this number, with subsequent gene loss and fractionation yielding the observed count in *T. tuberosum*.

Assembly and annotation of organelle genomes

The plastid genome was assembled as two complete circular paths of identical length (~150 kb), representing the two orientations of the large single-copy / inverted repeat / small single-copy (LSC/IR/SSC) quadripartite structure characteristic of land plant chloroplasts (**Supplementary Figure S10**). This result is consistent with the well-documented flip-flop recombination of the IR region and confirms assembly completeness.

215 Mitochondrial genome assembly recovered one primary scaffold of 550,896 bp and a complex
216 assembly graph with over 100 repeat patterns (**Supplementary Figure S11**). Such structural
217 complexity is typical of plant mitochondria, which undergo repeat-mediated recombination and
218 exist as dynamic genomic populations. Thirty high-coverage HiFi contigs mapped to this
219 scaffold, confirming their mitochondrial origin and warranting their exclusion from the nuclear
220 assembly. The most extreme outlier (ptg000094l, 1,293×) showed 33 BLASTn hits to the
221 mitochondrial genome, indicating a highly reiterated region with elevated read depth.

222

223 **Conclusions**

224 This first reference genome assembly for *T. tuberosum* achieves high completeness and
225 sequence quality, providing a foundation for comparative genomics, functional analyses and
226 diversity studies. This resource enables investigation of polyploid genome organization,
227 tuber development, and specialized metabolism in *T. tuberosum* and comparative studies
228 across the Tropaeolaceae family.

229

230 **Materials and Methods**

231 DNA in vitro-grown *T. tuberosum* accession BGHEID007454 (Botanic Garden Heidelberg,
232 Germany) was extracted using a CTAB-based protocol (Vaillancourt & Buell 2019). The
233 Colombian B15 morphotype was collected from small producers' farms in Boyacá, Colombia
234 (Fonseca & Márquez-Cardona 2024), and maintained in a conservation plot at San Javier
235 Farm, Cogua, Cundinamarca, Colombia (05°03'37.0" N, 073°56'41.6" W; 2590 m). DNA of
236 B15 was extracted from fresh leaf tissue as described (Kalendar et al. 2021).

237 BGHEID007454 DNA was sequenced on an Illumina NextSeq2000 at the Deep Sequencing
238 Core Facility, Heidelberg University. Long-read sequencing used the SMRTbell Express
239 Template Prep Kit 2.0 on a PacBio Revio by Novogene (https://www.novogene.com/). B15

240 DNA libraries (SQK-LSK114) were sequenced on a PromethION 24 (V14 chemistry, R10 flow
241 cells) by SNP Saurus (http://www.snpsaurus.com). Genome size, heterozygosity, and
242 duplication rate were estimated using jellyfish and GenomeScope2.

243 HiFi reads were assembled with hifiasm. The primary assembly underwent contaminant
244 (Kraken2; n=292) and organelle (BLASTn; n=89) removal. Quality metrics (gfastats, BUSCO,
245 merqury, BlobToolKit) and B15 ONT mapping (minimap2; mosdepth) are detailed in the
246 Supplementary Data.

247 The assembly was soft-masked (RepeatModeler, RepeatMasker). Gene models were
248 predicted ab initio using Helixer and ANNEVO. Annotation quality was benchmarked with
249 GAQET2 including BUSCO, OMArk, PSAURON, DeTEnGA/TEsorter, and DIAMOND (see
250 Supplementary Data).

251 Plastid and mitochondrial genomes were assembled from Illumina reads using GetOrganelle.
252 The plastid was annotated with GeSeq. Full details are in the Supplementary Data.

253

254 **Data Availability:**

255 Raw reads for the BGHEID007454 and B15 genotypes used in this study have been deposited
256 in the European Nucleotide Archive (ENA) at EMBL-EBI under accession numbers
257 PRJEB108055 (https://www.ebi.ac.uk/ena/browser/view/PRJEB108055) and PRJEB109811
258 (https://www.ebi.ac.uk/ena/browser/view/PRJEB109811), respectively. Reference genome
259 assembly (v1.2), haplotype sequences and annotation files have been deposited at FigShare
260 (https://doi.org/10.6084/m9.figshare.31338463).

261

262 **Ethics and Compliance:**

263 The *Tropaeolum tuberosum* accession BGHEID007454 was obtained under a Material
264 Transfer Agreement from the Botanic Garden Heidelberg, Germany. Maintained in European

265 collections since at least 1995 (predating the Nagoya Protocol and EU Regulation 511/2014),
266 it lacks documented collecting event or provenance. The collection and access to the
267 Colombian genetic resource *Tropaeolum tuberosum* B15 were covered under the Access to
268 Genetic Resources and Derived Products Agreement No. 256, "Studies on the Genetic
269 Diversity of Native Crops of Colombia," and its Addendum No. 1, signed between the Ministry
270 of Environment and Sustainable Development of Colombia and Pontificia Universidad
271 Javeriana.

272

273 **Author contributions:**

274 Study conception and design: PH. Sample preparation and experimental work: DRM, IS,
275 MPM, PH. Data analysis: DRM, WT, PH. Manuscript draft: PH. Manuscript final revision and
276 consent: DRM, IS, MPM, TG, WT, PH.

277

278 **Acknowledgements:**

279 We thank the Botanical Garden Heidelberg for plant material (BG HEID 007454), Sandra
280 Schmöckel and David Jarvis for helpful discussions, and Aureliano Bombarely for early
281 ANNEVO testing. We thank David Ibberson (Deep Sequencing Core Facility, Heidelberg
282 University) for technical support and Marcus Koch for commenting on the manuscript. We
283 acknowledge the SDS@hd data storage service supported by MWK and DFG (INST 35/1803-
284 1 FUGG, INST 35/1804-1 LAGG) and computational resources provided by the High
285 Performance and Cloud Computing Group, University of Tübingen, through bwHPC and DFG
286 (Project 455787709, bwForCluster BinAC2). This work was funded by the Klaus Tschira Boost
287 Fund (Klaus Tschira Stiftung/GSO e.V.) awarded to PH, and Fontagro (IDB) under grant "Root
288 to Food: Improvement of Yield in Potatoes and Other Andean Tubers " (ATN/RF-18120-RG)
289 to MPMC and WT.

290 **References**

291 Agerbirk N, Koch MA. 2025. Inconclusive evidence for chain-elongated methionine-derived
292 glucosinolates beyond the family Brassicaceae. Biochem. Syst. Ecol. 122:105033. doi:
293 10.1016/j.bse.2025.105033.

294 Andersson L, Andersson S. 2000. A molecular phylogeny of *Tropaeolaceae* and its systematic
295 implications. Taxon. 49:721–736. doi: 10.2307/1223973.

296 Apaza Ticona LN, Tena Pérez V, Bermejo Benito P. 2020. Local/traditional uses, secondary
297 metabolites and biological activities of Mashua (Tropaeolum tuberosum Ruíz & Pavón). J.
298 Ethnopharmacol. 247:112152. doi: 10.1016/j.jep.2019.112152.

299 Bird KA, Ramos AA, Kliebenstein DJ. 2025. Phylogenetic and genomic mechanisms shaping
300 glucosinolate innovation. Curr. Opin. Plant Biol. 85:102705. doi: 10.1016/j.pbi.2025.102705.

301 Chanderbali AS et al. 2022. Buxus and Tetracentron genomes help resolve eudicot genome
302 history. Nat. Commun. 13:643. doi: 10.1038/s41467-022-28312-w.

303 Coloma A et al. 2022. Characterization of Nutritional and Bioactive Compound in Three
304 Genotypes of Mashua (Tropaeolum tuberosum Ruiz and Pavón) from Different Agroecological
305 Areas in Puno. International Journal of Food Science. 2022:7550987. doi:
306 10.1155/2022/7550987.

307 Fonseca LR, Márquez-Cardona M del P. 2024. Morphological and molecular characterization
308 of cubios (Tropaeolum tuberosum Ruiz & Pavón) collected in two municipalities in Boyaca -
309 Colombia. Univ Sci (Bogota). 29:82–95. doi: 10.11144/Javeriana.SC291.mamc.

310 Friedhoff R, Wittstock U, Pucker B. 2024. Genome sequence of the medicinal plant
311 *Tropaeolum majus* provides insights into flavonoid biosynthesis. BioRxiv. doi:
312 10.1101/2024.11.01.621629.

Johns T, Neil Towers GH. 1981. Isothiocyanates and thioureas in enzyme hydrolysates of Tropaeolum tuberosum. Phytochemistry. 20:2687–2689. doi: 10.1016/0031-9422(81)85268-5.

Kalendar R, Boronnikova S, Seppänen M. 2021. Isolation and Purification of DNA from Complicated Biological Samples. Methods Mol. Biol. 2222:57–67. doi: 10.1007/978-1-0716-0997-2_3.

Luera-Quiñones S, Correa MJ, Moreno-Rojo C, Salvador-Reyes R, Paucar-Menacho LM. 2025. Mashua (Tropaeolum tuberosum Ruiz & Pavón): Nutritional Composition, Bioactive Compounds, and Functional Potential as an Andean Natural Ingredient. Foods. 14. doi: 10.3390/foods14244198.

Pissard A, Arbizu C, Ghislain M, Bertin P. 2008. Influence of Geographical Provenance on the Genetic Structure and Diversity of the Vegetatively Propagated Andean Tuber Crop, Mashua (*Tropaeolum tuberosum*), Highlighted by Intersimple Sequence Repeat Markers and Multivariate Analysis Methods. Int. J. Plant Sci. 169:1248–1260. doi: 10.1086/591979.

Ruíz-Mateus D, Márquez-Cardona M del P, Terán W. 2025. Cubio tuber (*Tropaeolum tuberosum*, Tropaeolaceae) morphological diversity in Colombia and Bolivia. Univ Sci (Bogota). 30:180–193. doi: 10.11144/Javeriana.SC30.ctmd.

Vaillancourt B, Buell CR. 2019. High molecular weight DNA isolation method from diverse plant species for use with Oxford Nanopore sequencing. BioRxiv. doi: 10.1101/783159.

Van de Peer Y, Mizrachi E, Marchal K. 2017. The evolutionary significance of polyploidy. Nat. Rev. Genet. 18:411–424. doi: 10.1038/nrg.2017.26.

Veselý P, Bureš P, Šmarda P. 2013. Nutrient reserves may allow for genome size increase: evidence from comparison of geophytes and their sister non-geophytic relatives. Ann. Bot. 112:1193–1200. doi: 10.1093/aob/mct185.

337 Zhang S-J, Liu L, Yang R, Wang X. 2020. Genome size evolution mediated by gypsy
338 retrotransposons in brassicaceae. Genomics Proteomics Bioinformatics. 18:321–332. doi:
339 10.1016/j.gpb.2018.07.009.