A draft genome assembly of “Cas” (*Psidium friedrichsthalianum* (O. Berg) Nied.): an indigenous crop of Costa Rica untapped

Mónica Rojas-Gómez · Jose Pablo Jiménez-Madrigal · Maripaz Montero-Vargas · Randall Loaiza-Montoya · Max Chavarría · Esteban Meneses · Eric J. Fuchs

Received: 28 June 2021 / Accepted: 17 October 2021 / Published online: 16 November 2021

© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract *Psidium friedrichsthalianum* (O. Berg) Nied. is a tropical tree species in the Myrtaceae family, natively distributed from southern Mexico to eastern Venezuela and Ecuador and commonly known as “Cas”, ”Costa Rican guava“ or “Sour Guava”. The “Cas” produces a fruit with a rather distinctive acidic flavor and has bioactive compounds with biological potential equal or greater than common Guava; is considered an indigenous crop in Costa Rica with characteristics as a functional food untapped. This species has not been completely domesticated, and can be found in home gardens, paddocks, small groups, and, more recently, in small and medium sized plantations. Also, the plantations of this species do not have technical and scientific support or agronomic promotion from industry, nor are there genetic resources or germplasm readily available to farmers. This limits its commercial development and the implementation of selection or genetic improvement programs. In this study, we present the first draft assembly of the Cas genome using PacBio long reads and the Canu assembly pipeline. Our draft assembly has a total length of 417.64 Mb, with 24 440 contigs and a N50 contig size of 21.3 Kb. Structural annotation resulted in 59 036 gene models. Functional annotation was conducted against the non-redundant set of genes from the KEGG database. Of the 52 422
complete genes models, 15.55% (8 153) presented homology with KEGG orthologs. The genes found in our Cas draft assembly were compared to those found in *Eucalyptus grandis* W. Hill. in the KEGG repository. According to the KEGG pathway assignments, 33 isoforms were annotated as part of the flavonoid biosynthetic pathway. In addition, 19 isoforms were annotated as part of phenylpropanoid biosynthetic pathway. The results of this study provide an overview of the first draft of the Cas genome assembly using PacBio long reads. This new genomic resource represents the basis for exploring the genetic potential of this crop with characteristics as a functional food.

**Keywords** *P. friedrichsthalianum* · Cas · Indigenous crop · Draft genome assembly · Genomic resources · Functional food

**Introduction**

*Psidium friedrichsthalianum* (O. Berg) Nied. is a tropical tree species in the Myrtaceae family, subfamily Myrtoideae, tribe Myrteae (Lucas et al. 2019). This family includes several plant species of economic importance, such as guava (*Psidium guajava* L.), eucalyptus (*Eucalyptus grandis* W. Hill.), wax apple (*Syzygium samarangense* (Blume) Merr. YLM Perry), and pink myrtle (*Rhodomyrtus tomentosa* (Aiton) Hassk) (Grattapaglia et al. 2012; Wilson et al. 2004). *P. friedrichsthalianum* is a medium-sized tree (Fig. 1), natively distributed from southern Mexico, to eastern Venezuela and Ecuador (Govaerts et al. 2008). This tree has perfect allogamous flowers (Barahona and Rivera 1995) and produces a fleshy berry-like fruit, commonly known as "Cas", “Costa Rican guava” or “Sour Guava” (Cuadrado-Silva et al. 2017a). The fruit has a greenish to yellow exocarp and a pulp with a rather distinctive acidic flavor (Cuadrado-Silva et al. 2017a, b; Thaptimthong et al. 2016). The species is diploid (2n = 2x = 22) (Srivastava 1977), with an estimated genome size of 958 Mb (Rojas-Gómez et al. 2021).

Recent studies have shown that fruit of Cas (*P. friedrichsthalianum*) has bioactive compounds and biological potential equal or greater than guava (*P. guajava*) (González et al. 2012; Cuadrado-Silva et al. 2017a). This tree produces phenolic compounds, mainly quercetin aglycones and proanthocyanidins, derived from epi-catechin units. Other compounds such as ellagittannins and benzophenones have also been putatively identified (Rojas-Garbanzo et al. 2021, 2019; Cuadrado-Silva et al. 2017a, b; Granda-dos-Chinchilla et al. 2016; Flores et al. 2013; González et al. 2012). Previous studies have categorized it as a functional food due to its association with positive health effects. Anti-diabetic, anti-plasmodic, hepatoprotective, antioxidant, antimicrobial, anti-inflammatory, and anti-platelet activity are some of the reported bioactivities for leaves and fruit extracts (Cuadrado-Silva et al. 2017b; González et al. 2012;
Flores et al. 2015, 2013; Granados-Chinchilla et al. 2016; Rojas-Garbanzo et al. 2021).

In Costa Rica, the fruit of Cas is consumed mainly as fresh fruit and is used to prepare traditional foods, such as pasteurized soft drinks, ice creams, juices, desserts, and jams. It is sold at grocery stores and farmers markets (León 2000). In Costa Rica, this species has not been completely domesticated, and can be found in home gardens, paddocks, small groups, and, more recently, in small and medium-sized plantations. In general, Cas is considered an indigenous crop in Costa Rica with characteristics as a functional food untapped. The plantations of this species do not have technical and scientific support or agronomic promotion from industry, nor are there genetic resources or germplasm readily available to farmers. This limits its commercial development and the implementation of selection or genetic improvement programs.

Previous work on the assembly of the complete genome of economically important members of the Myrtaceae, such as Psidium guajava L. (Feng et al. 2020), Eucalyptus grandis W. Hill. (Myburg et al. 2014) and Leptospermum scoparium JRForst. & G.Forst (Thrimawithana et al. 2019) has been published. In this study, we present the first draft assembly of the Cas genome (P. friedrichsthalianum) using PACBio long reads and the Canu assembly pipeline (Koren et al. 2017). This new genomic resource represents the basis for exploring the genetic potential of this crop with characteristics as a functional food.

Materials and methods

Sampling permits

All the needed permits for sampling leaf tissue from Cas were requested by Comisión de Biodiversidad-UCR at Universidad de Costa Rica (under permit # 133-2018) and Mr. Alfonso Ruiz, owner of the farm where the sampling was carried out.

Plant material, DNA preparation, and genome sequencing

Plant material used for this study was collected from a small plantation located at Tacacori, Alajuela, Costa Rica (DMS: 10° 03’ 07.3” N–84° 12’ 52.3” W). Leaf tissue was collected from mature trees and transported to the laboratory at room temperature in sterile polypropylene tubes (50 mL) containing silica gel. Genomic DNA was extracted from 4 samples of 50 mg each of dry leaf tissue, using the DNeasy Plant Mini kit (QIAGEN, Inc.). DNA size, quantity, and integrity was assessed with an Agilent’s TapeStation system (model 4200, CA, USA). The sample with the highest molecular weight and DNA integrity was selected for sequencing. SMRTbell library preparation (template size 4–10 Kb) and next-generation sequencing was performed by the North Carolina State University Genomic Sciences Laboratory (Raleigh, NC, USA) using the PacBio Sequel I system.

Genome assembly, gene prediction, and annotation

Raw reads (BAM files) from each SMRTcell were converted to FASTQ format, using bam2fastq, and collated into a single file. These long reads were corrected, trimmed and assembled using Canu v2.0 (Koren et al. 2017), with corrected Error Rate set to 0.155, min Read Length set to 500, min Overlap Length set to 400, and genome size equal to 958 Mbp, based on previous flow cytometry estimates (Rojas-Gómez et al. 2021). Genome assembly quality statistics were estimated with QUAST v5.0.2 (Gurevich et al. 2013). For comparison purposes, the Guava genome sequence (GenBank assembly accession: GCA_016432845.1) was set as a reference for estimating misassemblies, genome fraction, duplication ratio, and potential transposable elements with QUAST. Assembly completeness was assessed with BUSCO v4.1.2, using the Eudicotyledons (odb10) database (Seppey et al. 2019). Structural annotation and gene predictions were obtained with Augustus v3.3.3 (Stanke et al. 2008) and filtered with gFACs v1.1.2 (Caballero and Wegrzyn 2019), excluding incomplete gene models and redundant exons.

Functional annotation was performed with BlastKOALA (Kanehisa et al. 2016) and the non-redundant set of genes from the KEGG database (Kanehisa and Goto 2000; Kanehisa 2019; Kanehisa et al. 2021), using default parameters. Additionally, annotation of the flavonoid biosynthesis pathway was performed using modules M00137 and M00138, from genes annotated in Eucalyptus grandis W.Hill, a member of
the Myrtaceae family and the closest relative for which there is data available.

**Results and discussion**

Genome assembly and annotation

A de novo draft assembly of the Cas genome was produced using 58.7 Gb of PacBio Sequel I raw data and the Canu assembler pipeline. The draft assembly has a total length of 417.64 Mb, with 24,440 contigs and a N50 contig size of 21.3 Kb. The assembly length accounts for 43.6% of the expected genome size (Rojas-Gomez et al. 2020, 2021) and yet it presents 62.6% complete universal single-copy orthologs (Table 1). This may indicate the accumulation of repetitive DNA or transposable elements (TE) throughout the evolutionary history of the species, which contributes to the larger genome size but are hard to sequence and assemble. Partial or complete genome duplication is also a common feature in plant genomes (Wendel et al. 2016). The GC content of the Cas draft assembly was 40.05%, similar to that of *P. guajava* L. (Feng et al. 2021) and *E. grandis* W. Hill. (Myburg et al. 2014), the two most closely related species with sequenced genomes. Furthermore, when comparing the Cas assembly to the latest Guava reference sequence, the aligned genome fraction was 21.57% with a duplication ratio of 1.38. This indicates that a fraction of the Cas assembly shares a strong homology with the Guava reference, hinting at conserved regions in the *Psidium* spp. genomes. Conversely, up to 60 of the missassemblies between the Cas and Guava sequence can be attributed to potential transposable elements (TE) of different lengths, which along with the duplication ratio would help explain the differences in genome size, as previously mentioned.

Structural annotation of the draft assembly resulted in 59,036 gene models. However, after filtering out incomplete genes, overlapping exons, and duplicates, a total of 52,422 genes were retained. This is a larger number of predicted protein-coding genes than in the sequenced genomes of the myrtaceous species, *Leptospermum scoparium* JRForst. & G.Forst (31,220 genes), *Eucalyptus grandis* W. Hill. (36,376 genes) and *Psidium guajava* L. (25,601 genes) (Thrimawithana et al. 2019; Myburg et al. 2014; Feng et al. 2021). In addition, of the remaining gene models, 6,106 correspond to mono-exonic genes, while 46,316 correspond to multi-exonic genes. On average, multi-exonic genes are composed of 5.9 exons and 4.9 introns. Median CDS length was 1,279 nt, but multi-exonic gene size varies greatly with CDS ranging from 209 to 14,373 nt. For mono-exonic genes the average length was 948.6 nt. This is consistent with reports of plants having more protein coding genes of smaller size and with fewer exons per gene than other organisms (Ramirez-Sanchez et al. 2016).

Functional annotation was conducted against the non-redundant set of genes from the KEGG database. Of the 52,422 complete genes models, 15.55% (8,153) presented homology with KEGG orthologs. The majority of the annotated genes correspond with genetic information processing pathways, such as proteins involved in transcription and translation processes, as well as folding, sorting and degradation. The second largest group of annotated genes correspond to environmental information processing pathways, including membrane transport and signal transduction. Finally, the third largest group of annotated genes correspond to carbohydrate metabolism pathways. A complete distribution of the

| Table 1 | Summary statistics of Cas (*P. friedrichsthalianum*) draft genome |
|---------|-------------------------------------------------|
| Estimated genome size (n, Mb)* | 958 |
| Assembly size (Mb) | 417.64 |
| Number of contigs | 24,440 |
| Largest contig (bp) | 182,085 |
| N50 (bp) | 21,274 |
| L50 (bp) | 6,131 |
| GC (%) | 40.05 |
| Number of N’s per 100 Kbp | 0.0 |
| BUSCOs | C: 62.5% [S:57.4%, D:5.1%], F:10.9%, M:26.6%, n:2121 |

*Based on flow cytometry estimates (Rojas-Gomez et al. 2021)
functional categories of the annotated genes is presented in Fig. 2. The remaining gene models, that is the ones without annotation, were classified as hypothetical proteins. Many of these gene models showed homology with other hypothetical proteins reported in the NCBI database from both closely and distantly related species (data not shown). Because there is not reported evidence of their function in vivo these hypothetical proteins remain uncharacterized.

Genes involved in flavonoid biosynthesis

The most important flavonoids in Cas are proanthocyanidins, anthocyanins, and flavonols. All three classes are being intensively investigated for their potential benefit to human health and are the compounds that provide functional food properties. For example, proanthocyanidins are thought to help in maintaining urinary tract health (Howell 2002; Foo et al. 2000), anthocyanins are important as antioxidants (Yan et al. 2002; Sun et al. 2002), and flavonols are implicated in anti-plasmodic, hepatoprotective, anti-inflammatory, and anti-cancer bioactivities, among others (Hollman and Katan 1997; Ren et al. 2003). Thus, considering the importance of the flavonoids, in this study we focus on analyzing the presence of genes involved in its biosynthesis. For this, we obtained the KEGG reference pathway for flavonoid biosynthesis (KEGG modules: M00137 and M00138; KEGG map: map00941, Fig. 3) and phenylpropanoid biosynthesis (KEGG map: map00940, Fig. 4). Then, we compared our Cas draft assembly with annotated genes of *Eucalyptus grandis* W. Hill. in the KEGG repository.

According to the assignments of the KEGG pathway in the flavonoid biosynthetic pathway, 4 isoforms were annotated for Trans-cinnamate 4-monooxygenase (K00487, [EC: 1.14.1491]), 17 isoforms were annotated for Chalcone synthase (K00660, [EC: 2.3.1.74]), 3 isoforms were annotated for Chalcone isomerase (K01859, [EC: 5.5.1.6]), 6 isoforms were annotated for Naringenin 3-dioxygenase (K00475, [EC: 1.14.11.9]), and 2 isoforms were annotated for Flavanone 4-reductase (K13082, [EC: 1.1.1.234])(Fig. 3 and Table S1). In addition, 10 isoforms were annotated for Phenylalanine ammonia lyase (K10775, [EC: 4.3.1.24]) and 9 isoforms were annotated for 4-coumarate-CoA ligase (K01904, [EC: 6.2.1.12]); as part of the phenylpropanoid biosynthetic pathway (Fig. 4 and Table S1). These genes annotated in our study express enzymes that are precursors in the formation of different flavonoids previously identified in Cas, for
example, Naringenin (Flavanone), Quercetin (Flavonol), Gallocatechin (Monomeric flavanol), Epigallocatechin (Monomeric flavanols) and Myricetin (Hexahydroxyflavone) (Flores et al. 2013; Rojas-Garbanzo et al. 2019, 2021). The identification of genes for flavonoid biosynthesis is very important not only to clarify the expression and accumulation of bioactive molecules of Cas, but also for the application of biotechnology to improve its biosynthesis. To date, most of the articles published in Cas have focused on the chemical characterization of its metabolites.

The number of isoforms reported for the genes annotated in the flavonoid and phenylpropanoid biosynthetic pathways could be related to the number of duplications in our Cas assembly; several authors have suggested that the events of partial or complete duplication of plant genomes are common and expected characteristics, as well, may be important in the evolution of genome sizes in plants (Wang et al. 2016; Wendel et al. 2016). While we concede that the number of isoforms reported may be affected by the nature of the annotation methods (in silico prediction) and reference database used, our results are consistent with isoforms reports in the similar biosynthetic pathway of different species. For example: Theobroma cacao L. (Argout et al. 2011), Glycyrrhiza uralsensis (Mochida et al. 2016) and Carthamus tinctorius L. (Chen et al. 2018).

The genes and regulatory mechanisms involved in the flavonoid pathway are of great interest in Cas, as these affect the temporal and spatial flavonoid biosynthesis as well as the specific flavonoids produced. It has been shown that the central pathway for flavonoid biosynthesis is conserved in most plants, but depending on the species, a group of enzymes, such as
isomerases, reductases, hydroxylases and several Fe$^{2+}$/2-oxoglutarate-dependent dioxygenases can modify their basic flavonoid skeleton, generating different subclasses of flavonoids (Martens et al. 2010). Likewise, it has been shown that flavonoid biosynthesis is tissue specific, regulated by development and can be induced by a variety of environmental factors, including light, UV radiation, fungal infection, interaction with microorganisms, wounding, among others (Winkel-Shirley 2001). This study presents the first annotation of genes involved in flavonoid biosynthesis for Cas. These genes can be targeted for manipulation of flavonoid biosynthesis through various means or used as markers for the selection of desirable flavonoid profiles through breeding. For example, in cranberry it has been shown that an important aspect of anthocyanins as antioxidants is the specific aglycone, as well as the glucoside, since this affects both the antioxidant potential and the bioavailability (Satué-Gracia et al. 1997; Wang et al. 1997).

The results of this study provide an overview of the first draft of the Cas genome assembly using PacBio reads. This assembly represents the basis for new genomic and molecular studies in the future of this crop. In particular, it delves into the evolution of secondary metabolites, mainly of the polyphenolic type, which give this crop special characteristics as a functional food.

**Acknowledgements** We thank local Cas farmer, Mr. Alfonso Ruiz, who provided free samples from his farm.

**Author contributions** MR-G: Conceptualization, methodology, investigation, formal analysis, writing—original draft, project administration. JPJ-M: Methodology, formal analysis, writing—review and editing, data curation. MM-V: Methodology, formal analysis, data curation, writing—review and editing. RL-Mo: Review and editing, funding acquisition, project administration. MC: Review and editing, supervision. EM: Review and editing, supervision, funding acquisition. EF: Conceptualization, investigation, writing—review and editing.

**Funding** This work was supported by the Centro Nacional de Alta Tecnología (CeNAT-CONARE), Foundation of Costa Rica (3-006-213777).

**Data availability** The whole-genome sequence of Cas has been deposited at the DDBJ/ENA/GenBank databases under...
BioProject PRJNA725514, under the accession JAGVVM000000000. The version described in this paper is version JAGVVM010000000. Available data includes raw reads, draft assembly sequence, and gene models. https://www.ncbi.nlm.nih.gov/search/all/?term=PRJNA725514

Declarations

Conflict of interest The authors declared that they have no conflict of interest in this work.

References

Argout X, Salse J, Aury JM et al (2011) The genome of Theobroma cacao. Nat Genet 43:101–108. https://doi.org/10.1038/ng.736
Baraona M, Rivera G (1995) Development of jocote (Spondias purpurea L.) and Cas (Psidium friedrichsthalianum Niedz) in the premontano humid forest of Costa Rica. Mesoam Agron 6:23–31
Caballero M, Weгрzyn J (2019) gFACs: gene filtering, analysis, and conversion to unify genome annotations across alignment and gene prediction frameworks. Gen Proteom Bioinform 17(3):305–310. https://doi.org/10.1016/j.gpb.2019.04.002
Chen J, Tang X, Ren C et al (2018) Full-length transcriptome sequences and the identification of putative genes for flavonoid biosynthesis in safflower. BMC Gen 19:548. https://doi.org/10.1186/s12864-018-4946-9
Cuadrado-Silva CT, Pozo-Bayón MA, Osorio C (2017a) Identification of aroma compounds and precursors of sour guava (Psidium friedrichsthalianum Niedz.) following a sensomics approach. Eur Food Res Technol 243:1–10. https://doi.org/10.1007/s00217-016-2716-y
Cuadrado-Silva CT, Pozo-Bayón MÁ, Osorio C (2017) Targeted metabolomic analysis of polyphenols with antioxidant activity in sour Guava (Psidium friedrichsthalianum Ndz.). Fruit Mol 22(1):11. https://doi.org/10.3390/molecules22010011
Feng C, Feng C, Lin X, Liu S, Li Y, Kang M (2021) A chromosome-level genome assembly provides insights into ascorbic acid accumulation and fruit softening in guava (Psidium guajava). Plant Biotechnol J 19(4):717–730. https://doi.org/10.1111/ptb.13498
Flores G, Dastmalchi K, Wu SB, Whalen K, Dabo AJ, Reynertson K (2013) Phenol-rich extract of the Costa Rican guava pulp (Psidium friedrichsthalianum Ndz) with antioxidant and anti-inflammatory activity. Potential COPD Ther Food Chem 14:889–895. https://doi.org/10.1016/j.foodchem.2013.03.025
Flores G, Wu SB, Negrín A, Kennelly EJ (2015) Chemical composition and antioxidant activity of seven cultivars of guava (Psidium guajava) fruits. Food Chem 170:327–335. https://doi.org/10.1016/j.foodchem.2014.08.076
Foo L, Lu Y, Howell AB, Vorsa N (2000) A-Type proanthocyanidin trimers from cranberry that inhibit adherence of uropathogenic P-fimbriated Escherichia coli. J Nat Prod Chem 63(9):1225–1228. https://doi.org/10.1021/np000128u
González E, Vaillant F, Pérez AG, Rojas T (2012) Antioxidant protection mediated by in vitro cells of human erythrocytes by some common tropical fruits. J Nutr Food Sci 2:139. https://doi.org/10.4172/2155-9600.1000139
Goyaerts R, Sobral N, Ashton P, Barrie F, Holst BK, Landrum LL, Matsumoto K, Fernanda Mazine F, Nic Lughadha, Proença C (2008) World checklist of myrtaceae: Kew Publishing, Royal botanic gardens. pp. 1–455
Granados-Chinchilla F, Villegas E, Molina A, Arias C (2016) Composition, chemical fingerprinting and antimicrobial assessment of costa rican cultivated guavas (Psidium friedrichsthalianum (O. Berg) Ndz and Psidium guajava L.) essential oils from leaves and fruits. Nat Prod Chem Res 4:5. https://doi.org/10.4172/2329-6836.1000236
Grattapaglia D, Vaillancourt RE, Shepherd M, Thumma BR, Foley W, Külheim C, Potts BM (2012) Progress in Myrtaceae genetics and genomics: eucalyptus as the pivotal genus. Tree Genet Gen 8:463–508. https://doi.org/10.1007/s11295-012-0491-x
Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. Bioinformatics 29(8):1072–1075. https://doi.org/10.1093/bioinformatics/btt086
Hollman PCH, Katan MB (1997) Absorption, metabolism and health effects of dietary flavonoids in man. Biomed Pharmacother 51:305–310. https://doi.org/10.1016/S0753-3322(97)88045-6
Howell AB (2002) Cranberry proanthocyanidins and the maintenance of urinary tract health. Crit Rev Food Sci Nutr 42(S3):273–278. https://doi.org/10.1080/1040839029351915
Kanehisa M (2019) Toward understanding the origin and evolution of cellular organisms. Protein Sci 28(11):1947–1951. https://doi.org/10.1002/pro.371
Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M (2021) KEGG: integrating viruses and cellular organisms. Nucl Acids Res 49(1):545–551. https://doi.org/10.1093/nar/gkaa970
Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucl Acids Res 28(1):27–30. https://doi.org/10.1093/nar/28.1.27
Kanehisa M, Sato Y, Morishima K (2016) BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol 428:726–731. https://doi.org/10.1016/j.jmb.2015.11.006
Koren S, Walecz BP, Berlin K, Miller JR, Bergman NH, Phil lippy AM (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Gen Res 27(5):722–736. https://doi.org/10.1101/gr.215087.116
León J (2000) Botany of tropical crops. San José, Costa Rica: Ed. Ilca. pp. 445
Lucas EJ, Holst B, Sobral M, Mazine FF, Barnes-Proença CE, Ribeiro da Costa I, Vasconcelos TNC (2019) A new subtribal classification of tribe Myrtea (Myrtaceae). Syst Bot 44(3):560–569. https://doi.org/10.1002/10.1600/03636419X15620113920608
Martens S, Preuss A, Matern U (2010) Multifunctional flavonoid dioxygenases: flavonols and anthocyanin biosynthesis

Springer
in Arabidopsis thaliana L. Phytochemistry 71:1040–1049. https://doi.org/10.1016/j.phytochem.2010.04.016

Mochida K et al (2016) Draft genome assembly and annotation of Glycyrrhiza uralensis, a medicinal legume. Plant J 89(2):181–194. https://doi.org/10.1111/tpj.13385

Myburg A, Grattapaglia D, Tuskan G (2014) El genoma de Eucalyptus grandis. Nature 510:356–362. https://doi.org/10.1038/nature13308

Ren W, Qiao Z, Wang H, Zhu L, Zhang L (2003) Flavonoids: promising anticancer agents. Med Res Rev 23:519–534. https://doi.org/10.1002/med.10033

Rojas-Garbanzo C, Rodríguez L, Pérez AM, Mayorga-Gross AL, Vásquez-Chaves V, Fuentes E, Palomo I (2021) Antiplatelet activity and chemical characterization by UPLC-DAD-ESI-QTOF-MS of the main polyphenols in extracts from Psidium leaves and fruits. Food Res Int 141(1):45. https://doi.org/10.1016/j.foodres.2020.110070

Rojas-Garbanzo C, Wintera J, Montero ML, Zimmermann BF, Schiebera A (2019) Characterization of phytochemicals in Costa Rican guava (Psidium friedrichsthalianum Ndz) fruit and stability of main compounds during juice processing: (U) HPLC-DAD-ESI-QTD-MS. J Food Compos Anal 75:26–42. https://doi.org/10.1016/j.foodres.2020.110070

Rojas-Gómez M, García-Piñeres A, Bolaños-Villegas P, Arrieta-Espinoza G, Fuchs EJ (2021) Genome size and chromosome number of Psidium friedrichsthalianum (O. Berg) Nied (“Cas”) in six populations of Costa Rica Caryology. Int J Cytol, Cytsystematics Cytoqenet 73(3):55–63. https://doi.org/10.13128/caryologia-646

Satué-Gracia MT, Heinonen M, Frankel EN (1997) Anthocyanins as antioxidants on human low-density lipoprotein and lecithin-liposome systems. J Agric Food Chem 45(9):3362–3367. https://doi.org/10.1021/jf970234a

Seppey M, Manni M, Zdobnov EM (2019) BUSCO: Assessing genome assembly and annotation completeness. In: Kolmam M (ed) Gene prediction methods in molecular biology, vol 1962. Humana, New York. https://doi.org/10.1007/978-1-4939-9173-0_14

Srivastava HC (1977) Cytological studies in Psidium friedrichsthalianum Niedenzu. Cytologia 42;395–400. https://doi.org/10.1508/cytologia.42.395

Stanke M, Diekhans M, Baertsch R, Haussler D (2008) Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics 24(5):637–644. https://doi.org/10.1093/bioinformatics/btn013

Sun J, Chu YF, Wu X, Liu RH (2002) Antioxidant and antiproliferative activities of common fruits. J Agric Food Chem 50(25):7449–7454. https://doi.org/10.1021/jf0207530

Thaptimthong T, Kasemsuk T, Sibmooh N, Unchern S (2016) Platelet inhibitory effects of juices from Pachyrhizus erosus L. root and Psidium guajava L. fruit: a randomized controlled trial in healthy volunteers. BMC Complement Altern Med 16(1):269. https://doi.org/10.1186/s12906-016-1255-1

Thrimawithana AH, Jones D, Hilario E, Grierson E, Liachko I, Sullivan S et al (2019) A whole genome assembly of Leptospermum scoparium (Myrtaceae) for Mānuka research. N Z J Crop Hortic Sci. https://doi.org/10.1080/01140671.2019.1657911

Wang H, Cao G, Prior RL (1997) Oxygen radical absorbing capacity of anthocyanins. J Agric Food Chem 45:304–309. https://doi.org/10.1021/jf960421t

Wang K, Huang G, Zhu Y (2016) Transposable elements play an important role during cotton genome evolution and fibre cell development. Sci ChinaLifeSci. 59: 112–121

Wendel J, Jackson S, Blake CM, Wing RA (2016) Evolution of plant genome architecture. Gen Biol 17:37. https://doi.org/10.1186/s13059-016-0908-1

Wilson PG, O’Brien M, Heslewood MM, Quinn C (2004) Relationships within Myrtaceae sensu lato based on a matK phylogeny. Plant Syst Evol 251:3–19. https://doi.org/10.1007/s00606-004-0162-y

Winkel-Shirley B (2001) Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiol 126:485–493. https://doi.org/10.1104/pp.126.2.485

Yan X, Murphy BT, Hammond GB, Vinson JA, Neto CC (2002) Antioxidant activities and antitumor screening of extracts from cranberry fruit (Vaccinium macrocarpon). J Agric Food Chem 50:5844–5849. https://doi.org/10.1021/jf0202234

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