



DATA NOTE

The complete genome sequence of *Stevia rebaudiana*, the Sweetleaf [version 1; peer review: 2 approved, 1 approved with reservations]

Kathleen O'Neill, Stacy Pirro

IRIDIAN GENOMES, Bethesda, MD, 20817, USA

V1 First published: 21 Jul 2020, 9:751
<https://doi.org/10.12688/f1000research.24396.1>
 Latest published: 21 Jul 2020, 9:751
<https://doi.org/10.12688/f1000research.24396.1>

Abstract

The Sweetleaf (*Stevia rebaudiana*: Asteraceae) is widely grown for use as a sweetener. We present the whole genome sequence and annotation of this species. A total of 146,838,888 paired-end reads consisting of 22.2G bases were obtained by sequencing one leaf from a commercially grown seedling. The reads were assembled by a de-novo method followed by alignment to related species. Annotation was performed via GenMark-ES. The raw and assembled data is publicly available via GenBank: Sequence Read Archive (SRR6792730) and Assembly (GCA_009936405).

Keywords

Stevia rebaudiana, Sweetleaf, genome, assembly, annotation



This article is included in the **Draft Genomes** collection.



This article is included in the **Iridian Genomes** collection.

Open Peer Review

Reviewer Status

	Invited Reviewers		
	1	2	3
version 1 21 Jul 2020	 report	 report	 report
1. Eric Lu Zhang , Hong Kong Baptist University, Kowloon Tong, Hong Kong			
2. Andrew Miller , University of Illinois, Illinois, USA Illinois Natural History Survey, Illinois, USA			
3. Dawson White , Field Museum of Natural History, Chicago, USA			

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Stacy Pirro (info@iridiangenomes.org)

Author roles: **O'Neill K:** Resources; **Pirro S:** Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Project Administration, Validation, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This study was supported by IRIDIAN GENOMES (IRGEN-55670).

Copyright: © 2020 O'Neill K and Pirro S. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: O'Neill K and Pirro S. **The complete genome sequence of *Stevia rebaudiana*, the Sweetleaf [version 1; peer review: 2 approved, 1 approved with reservations]** F1000Research 2020, 9:751 <https://doi.org/10.12688/f1000research.24396.1>

First published: 21 Jul 2020, 9:751 <https://doi.org/10.12688/f1000research.24396.1>

Introduction

The Sweetleaf (*Stevia rebaudiana*: Asteraceae) is cultivated commercially for use as a sweetener. The sweetness is due to various steviol glycosides, primarily stevioside and rebaudioside. These compounds have 200-300X the sweetness of sugar (Abdullateef & Osman, 2012) but have no calories. The market for raw Stevia and derived products is expected to exceed 1B USD by 2021 (International Stevia Council, 2017).

Stevia rebaudiana has been used as a sweetener for centuries in Brazil and Paraguay (Misra *et al.*, 2011). Botanist Moisés Santiago Bertoni first described the plant as growing in eastern Paraguay and noted its use as a sweetener (Bertoni, 1899).

Chemists Bridel and Lavielle isolated the glycosides stevioside and rebaudioside that give the leaves their sweet taste (Bridel & Lavielle, 1931). The chemical structures of the aglycone steviol and its glycoside have been solved (Mosettig & Nes, 1955).

A complete genome sequence for this species will assist with discovering markers for crop yields, disease and drought resistance, and determining the biochemical pathways for the relevant metabolites.

Methods

A single commercially grown *Stevia rebaudiana* plant was used for this study (Behnke Nurseries, Beltsville, MD, USA). DNA extraction was performed on tissue from a single leaf using the Qiagen DNAeasy genomic extraction kit for plants, using the standard process. A paired-end sequencing library was constructed using the Illumina TruSeq kit, according to the manufacturer's

instructions. The library was sequenced on an Illumina Hi-Seq platform in paired-end, 2 × 150bp format.

The resulting fastq files were trimmed of adapter/primer sequence and low-quality regions with Trimmomatic v0.33 (Bolger *et al.*, 2014). The trimmed sequence was assembled by SPAdes v2.5 (Bankevich *et al.*, 2012) followed by a finishing step using RagTag v1.0.0 (Alonge, 2020) to make additional contig joins based on conserved regions in related plant species: *Erigeron canadensis* (GCA_010389155), *Mikania micrantha* (GCA_009363875), and *Helianthus annuus* (GCA_002127325). Default parameters were used for all assembly steps.

Annotation was performed using GeneMark-ES v2.0 (Lomsadze *et al.*, 2005). Annotation was performed fully de novo without a curated training set and default parameters.

Results

The genome assembly yielded a total sequence length of 411,383,069 bp over 55,557 scaffolds with an N50 of 37,276,437. The GeneMark-ES annotation resulted in 24,994 genes.

Data availability

Underlying data

Raw and assembled data is publicly available via GenBank:

Raw genome of *Stevia rebaudiana*, Accession number SRR6792730: <https://www.ncbi.nlm.nih.gov/sra/?term=SRR6792730>

Assembly of *Stevia rebaudiana*, Accession number ASM993640v1: https://www.ncbi.nlm.nih.gov/assembly/GCA_009936405.1/

References

- Abdullateef RA, Osman M: **Studies on effects of pruning on vegetative traits in *Stevia rebaudiana* Bertoni (Compositae).** *Int J Biol.* 2012; 4(1).
[Publisher Full Text](#)
- Alonge M: **Ragtag: Reference-guided genome assembly correction and scaffolding.** GitHub archive. 2020.
[Publisher Full Text](#)
- Bankevich A, Nurk S, Antipov D, *et al.*: **SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing.** *J Comput Biol.* 2012; 19(5): 455–477.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bertoni MS: **Revista de Agronomia de l'Assomption.** 1899.
- Bolger AM, Lohse M, Usadel B: **Trimmomatic: A flexible trimmer for Illumina Sequence Data.** *Bioinformatics.* 2014; 30(15): 2114–20.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bridel M, Lavielle R: **Sur le principe sucre des feuilles de kaa-he-e (stevia rebaudiana B).** *Comptes rendus de l'Académie des Sciences.* 1931; 192: 1123–5.
- International Stevia Council: **Industry data report, September 2017.** 2017.
- Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, *et al.*: **Gene identification in novel eukaryotic genomes by self-training algorithm.** *Nucleic Acids Res.* 2005; 33(20): 6494–6506.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Misra H, Soni M, Silawat N, *et al.*: **Antidiabetic activity of medium-polar extract from the leaves of *Stevia rebaudiana* Bert. (Bertoni) on alloxan-induced diabetic rats.** *J Pharm Bioallied Sci.* 2011; 3(2): 242–8.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mosettig E, Nes WR: **Stevioside. II. The structure of the aglucon.** *J Org Chem.* 1955; 20(7): 884–899.
[Publisher Full Text](#)

Open Peer Review

Current Peer Review Status: ? ✓ ✓

Version 1

Reviewer Report 18 November 2020

<https://doi.org/10.5256/f1000research.26914.r74315>

© 2020 White D. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Dawson White 

Department of Science and Education, Field Museum of Natural History, Chicago, IL, USA

Stevia is an economically important plant with no prior genomic resource for researchers. Standard methods were used for laboratory preparation, sequencing, assembly, and annotation of the genome. The resulting assembly and gene annotations are good based on the provided statistics.

I would like to see that a voucher specimen of the sequenced plant (or a similar individual) at Behnke Nurseries was deposited in a herbarium. This is important for reproducible science and its omission almost causes me to mark that the methods are only "partly" sufficient.

Is the rationale for creating the dataset(s) clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of methods and materials provided to allow replication by others?

Yes

Are the datasets clearly presented in a useable and accessible format?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Plant systematics.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 07 August 2020

<https://doi.org/10.5256/f1000research.26914.r67544>

© 2020 Miller A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Andrew Miller 

¹ University of Illinois, Illinois, USA

² Illinois Natural History Survey, Illinois, USA

The authors present the whole genome sequence of Sweetleaf, the plant where the sweetener, Stevia, is derived. I am surprised someone has not already sequenced this genome - it is certainly timely. The data are well presented and publicly available through NCBI. It is well-written and worthy of indexing. I only have minor comments below:

1. Should the keywords repeat words found in the title?

2. Can more be said in the Results about the genome?

Is the rationale for creating the dataset(s) clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of methods and materials provided to allow replication by others?

Yes

Are the datasets clearly presented in a useable and accessible format?

Yes

Competing Interests: I am a collaborator with Staci Pirro and we may publish a paper together in the future.

Reviewer Expertise: Sanger sequencing; mycology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 07 August 2020

<https://doi.org/10.5256/f1000research.26914.r68815>

© 2020 Zhang E. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Eric Lu Zhang

Department of Computer Science, Hong Kong Baptist University, Kowloon Tong, Hong Kong

The authors sequenced and assembled the genome of Sweetleaf by Illumina short-reads and annotated the genes using de novo prediction. The data could be useful to the field, but some points need to be clarified:

1. Commonly, the first step of genome assembly is to estimate the genome size using k-mer frequency distribution. The authors should add such analysis in the revision.
2. Based on the results in (1), the genome size of *Stevia rebaudiana* is ca. 1.3Gb. I don't think SPAdes is a good choice as it only works well on the small genomes.
3. Gene prediction is too simple and may lose considerable genes. The prediction should include multifaceted information, such as de novo prediction, Homology-based genes prediction *et al.*
4. Add a citation for "standard process" of DNA extraction and the model of sequencing platform, e.g. Hi-Seq 2500.
5. It is better to include a table to show the statistics, such as N50, genome size, the number of genes.
6. Both of the assemblies from short-reads and short-reads+RagTag should be provided, because RagTag may introduce certain bias.

Is the rationale for creating the dataset(s) clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Partly

Are sufficient details of methods and materials provided to allow replication by others?

Partly

Are the datasets clearly presented in a useable and accessible format?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Computational genomics, bioinformatics

I confirm that I have read this submission and believe that I have an appropriate level of

expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research