



GENOME NOTE

The complete genome sequences of *Erythroxylum coca* and *Erythroxylum novogranatense*

[version 1; peer review: peer review discontinued]

Dawson M. White¹, Lyndel Meinhardt², Bryan Bailey², Stacy Pirro³

¹Negaunee Integrative Research Center, The Field Museum, Chicago, IL, 60605, USA

²ARS/USDA, Beltsville, MD, 20705, USA

³IRIDIAN GENOMES, Bethesda, MD, 20817, USA

V1 First published: 25 Apr 2022, 11:457
<https://doi.org/10.12688/f1000research.108549.1>

Latest published: 25 Apr 2022, 11:457
<https://doi.org/10.12688/f1000research.108549.1>

Peer review discontinued

At the request of the author(s), this article is no longer under peer review. What does this mean?

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

Abstract

The flowering plant genus *Erythroxylum* contains approximately 280 species, including the economically and socially consequential crops called coca. We present the genome sequences of *Erythroxylum coca* and *E. novogranatense*, two cultigens produced for medicinal and quotidian use in the Andes and Amazon regions of South America, as well as the international cocaine industry. Sequencing was performed on an Illumina X-Ten platform, and reads were assembled by a de novo method followed by finishing via comparison with several species from the same genus. The BioProject, raw and assembled data can be accessed in GenBank for *E. coca* (PRJNA676123; JAJMLV000000000) and *E. novogranatense* (PRJNA675212; JAJKBF000000000).

Keywords

Erythroxylum coca, *Erythroxylum novogranatense*, genome, assembly



This article is included in the **Genomics and Genetics** gateway.

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

Author roles: White DM: Conceptualization, Investigation; Meinhardt L: Resources; Bailey B: Resources; Pirro S: Data Curation

Competing interests: No competing interests were disclosed.

Grant information: Dawson White is supported by an NSF Postdoctoral Fellowship in Biology, award number 2010821.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2022 White DM *et al.* 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: White DM, Meinhardt L, Bailey B and Pirro S. **The complete genome sequences of *Erythroxylum coca* and *Erythroxylum novogranatense* [version 1; peer review: peer review discontinued]** F1000Research 2022, 11:457 <https://doi.org/10.12688/f1000research.108549.1>

First published: 25 Apr 2022, 11:457 <https://doi.org/10.12688/f1000research.108549.1>

Introduction

The leaves of the coca plant have been used as a medicine and mild stimulant in South America for over 8,000 years (Plowman, 1984; Dillehay *et al.*, 2010). In more recent history, few plants have had such far-reaching effects on human health and international relations (Restrepo *et al.*, 2019). Coca crops produce the alkaloid cocaine: a natural insecticide (Nathanson *et al.*, 1993), Western medicine's first local anesthetic, and a controlled narcotic whose supply chains and illicit international markets have caused decades of social disaster.

Coca is classified into two species, *Erythroxylum coca* and *E. novogranatense* (Erythroxylaceae, Malpighiales), each with two taxonomic varieties. These two species are found only in cultivation, having resulted from independent origins of domestication from the wild *E. gracilipes* (White *et al.*, 2021).

The two varieties used in this study, *E. coca* var. *ipadu* Plowman, known as Amazonian coca, and *E. novogranatense* var. *truxillense* (Rusby) Plowman, known as Trujillo coca, are regionally distinct crops. *Erythroxylum coca* var. *ipadu* is cultivated by indigenous groups in the lowland Amazon basin of Colombia, Brazil, and Perú. *Erythroxylum novogranatense* var. *truxillense* is grown primarily in the dry valleys of northwestern Perú and is exported as a flavoring agent of Coca Cola®. These taxa have been crossed to produce improved hybrid varieties for the cocaine market, which are currently grown in southern Colombia and possibly southern Mexico (Casale *et al.*, 2014; Rodriguez, 2015).

Complete genome sequences for *E. coca* var. *ipadu* and *E. novogranatense* var. *truxillense* will provide insight into the origins, evolution, and modern breeding patterns of coca crops, as well as the of the cocaine biosynthesis pathway.

Methods

DNA from each species was provided by USDA/ARS Sustainable Perennial Crops Laboratory for use in this study.

Sequencing libraries were constructed with the Illumina TruSeq kit using standard protocols for the 2×150 bp format. Sequencing was performed on an Illumina X-Ten platform.

Raw, paired-end sequence data was trimmed of adapter sequence and low-quality regions using Trimmomatic (Bolger *et al.*, 2014). Genome preassemblies were constructed using SPAdes (Bankevich *et al.*, 2012), and finished with Zanfona (Kieras *et al.*, 2021).

Results

The results of genome assemblies are as follows:

specimen	accession	genome size	N50
<i>E. coca</i> var. <i>ipadu</i>	JAJMLV000000000	584,059,212	3,914,021
<i>E. novogranatense</i> var. <i>truxillense</i>	JAJKBf000000000	573,212,218	5,183,633

References

- Bankevich A, *et al.*: **SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing.** *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology.* 2012; **19**(5): 455–477. RRID: SCR_000131. [PubMed Abstract](#) | [Publisher Full Text](#)
- Bolger AM, Lohse M, Usadel B: **Trimmomatic: A flexible trimmer for Illumina Sequence Data.** *Bioinformatics.* 2014; **30**(15): 2114–2120. [Publisher Full Text](#)
- Casale JF, Mallette JR, Jones LM: **Chemosystematic identification of fifteen new cocaine-bearing *Erythroxylum* cultigens grown in Colombia for illicit cocaine production.** *Forensic Sci. Int.* 2014; **237**: 30–39. [PubMed Abstract](#) | [Publisher Full Text](#)
- Dillehay TD, Rossen J, Ugent D, *et al.*: **Early Holocene Coca Chewing In Northern Peru.** *Antiquity.* 2010; **84**: 939–953. [Publisher Full Text](#)

Kieras M, Peterson R, O'Neill K, Pirro S: **ZANFONA, a genome finishing process for short read assemblies.** 2021.

[Reference Source](#) | [Reference Source](#)

Nathanson JA, Hunnicutt EJ, Kantham L, *et al.*: **Cocaine as a naturally occurring insecticide.** *Proc. Natl. Acad. Sci. U. S. A.* 1993; **90**: 9645–9648.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Plowman T: **The Ethnobotany of Coca (Erythroxylum spp., Erythroxylaceae).** *Adv. Econ. Bot.* 1984; **1**: 62–111.

Restrepo DA, *et al.*: **Erythroxylum in Focus: An Interdisciplinary Review of an Overlooked Genus.** *Molecules.* 21 Oct. 2019; **24**(20): 3788.

[PubMed Abstract](#) | [Publisher Full Text](#)

Rodríguez Zapata FV: **Genome size and descriptors of leaf morphology as indicators of hybridization in colombian cultigens of coca Erythroxylum spp.** Thesis, Universidad de los Andes. 2015; 42 pp.

White DM, Huang J-P, Orlando AJ-M, *et al.*: **The Origins of Coca: Museum Genomics Reveals Multiple Independent Domestications from Progenitor Erythroxylum gracilipes.** *Syst. Biol.* 2021; **70**: 1–13.

[PubMed Abstract](#) | [Publisher Full Text](#)

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