



Published in final edited form as:

*Biodivers Genomes*. 2022 ; 2022: . doi:10.56179/001c.39776.

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

Dawson White<sup>1</sup>, Lyndel Meinhardt<sup>2</sup>, Bryan Bailey<sup>2</sup>, Stacy Pirro<sup>3</sup>

<sup>1</sup>Grainger Bioinformatics Center, Science and Education Department, Field Museum

<sup>2</sup>ARS/USDA

<sup>3</sup>Iridian Genomes

### Abstract

The flowering plant genus *Erythroxylum* contains approximately 300 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; genome

### 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. 2020).

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, Mallette, and Jones 2014; Rodríguez Zapata 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, Lohse, and Usadel 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	584,053,830	71.4 MB
<i>E. novogranatense</i> var. <i>truxillense</i>	JAJKBF000000000	573,249,677	50.4 MB

## Acknowledgements

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

## Funding

Funding was provided by Iridian Genomes, grant # IRGEN\_RG\_2021-1345 Genomic Studies of Eukaryotic Taxa.

## REFERENCES

- Bankevich Anton, Nurk Sergey, Antipov Dmitry, Gurevich Alexey A., Dvorkin Mikhail, Kulikov Alexander S., Lesin Valery M., et al. 2012. "SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing." *Journal of Computational Biology* 19 (5): 455–77. 10.1089/cmb.2012.0021. [PubMed: 22506599]
- Bolger Anthony M., Lohse Marc, and Usadel Bjoern. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics* 30 (15): 2114–20. 10.1093/bioinformatics/btu170. [PubMed: 24695404]
- Casale John F., Mallette Jennifer R., and Jones Laura M. 2014. "Chemosystematic Identification of Fifteen New Cocaine-Bearing *Erythroxylum* Cultigens Grown in Colombia for Illicit Cocaine

- Production.” *Forensic Science International* 237: 30–39. 10.1016/j.forsciint.2014.01.012. [PubMed: 24561557]
- Dillehay Tom D., Rossen Jack, Ugent Donald, Karathanasis Anathasios, Vásquez Víctor, and Netherly Patricia J.. 2010. “Early Holocene Coca Chewing In Northern Peru.” *Antiquity* 84 (326): 939–53. 10.1017/s0003598x00067004.
- Kieras M, Peterson R, O’Neill K, and Pirro S. 2021. ZANFONA, a Genome Finishing Process for Short Read Assemblies. <https://github.com/zanfona734/zanfona>.
- Nathanson JA, Hunnicutt EJ, Kantham L, and Scavone C. 1993. “Cocaine as a Naturally Occurring Insecticide.” *Proceedings of the National Academy of Sciences of the United States of America* 90 (20): 9645–48. 10.1073/pnas.90.20.9645. [PubMed: 8415755]
- Plowman T 1984. “The Ethnobotany of Coca (*Erythroxylum* Spp., *Erythroxylaceae*.)” Repos. Inst. CEDRO.
- Restrepo David A., Saenz Ernesto, Jara-Muñoz Orlando Adolfo, Calixto-Botía Iván F, Rodríguez-Suárez Sioly, Zuleta Pablo, Chavez Benjamin G., Sanchez Juan a., and D’Auria John C.. 2019. “*Erythroxylum* in Focus: An Interdisciplinary Review of an Overlooked Genus.” *Molecules* (Basel, Switzerland) 24 (20): 3788. 10.3390/molecules24203788.
- Rodríguez Zapata FV 2015. “Genome Size and Descriptors of Leaf Morphology as Indicators of Hybridization in Colombian Cultigens of Coca *Erythroxylum* Spp.” Thesis, Universidad de los Andes.
- White Dawson M., Huang Jen-Pan, Jara-Muñoz Orlando Adolfo, Madriñán Santiago, Ree Richard H., and Mason-Gamer Roberta J.. 2020. “The Origins of Coca: Museum Genomics Reveals Multiple Independent Domestications from Progenitor *Erythroxylum Gracilipes*.” *Systematic Biology* 70 (1): 1–13. 10.1093/sysbio/syaa074.