

GENOME SEQUENCING

The complete genome sequences of 56 *Erythroxylum* species

Dawson M White¹, Lyndel Meinhard², Bryan Bailey², Stacy Pirro³

¹ Negaunee Integrative Research Center, Field Museum, ² ARS/USDA, ³ Biodiversity, Iridian Genomes

Keywords: erythroxylum, genomes

<https://doi.org/10.56179/001c.40336>

Biodiversity Genomes

We present the whole genome sequences of 56 wild *Erythroxylum* species from Africa, China, and the American tropics. Deep Illumina sequencing was performed on a single leaf of each voucher. We de novo assembled sequence reads and then identified and used conserved regions across all preassemblies join contigs in a finishing step. The raw and assembled data is publicly available via Genbank.

Introduction

The coca genus, *Erythroxylum*, represents a pantropical clade of small trees and shrubs mostly known by the cocaine-alkaloid bearing coca crops (*Erythroxylum coca* & *E. novogranatense*). Within the species-rich *Erythroxylum* genus, about 205 species (75%) occur in the American tropics, with centers of diversity in eastern Brazil, Colombia, and Venezuela. Another center of diversity is Madagascar, with about 33 species (12% of *Erythroxylum*; (Daly 2004; White 2019).

Many species of *Erythroxylum* are utilized in traditional medicine, most conspicuously *E. monogynum* of India as well as the coca crops – which have been cultivated and used for more than 9,000 years (Dillehay et al. 2010; White et al. 2020). A recent review by Lv and colleagues (Lv et al. 2022) reports our state of knowledge on the natural products and bioactivity of the group. The diversity of tropane alkaloids, diterpenes, triterpenes, flavonoids, and derivatives found in *Erythroxylum* provide a rich source of bioactive compounds for medicinal applications.

Dating back to Darwin (Darwin 1877), *Erythroxylum* has also served as a case study of the diversity and biology of heterostyly. All species in the group are distylous, with the exception of 11 dioecious species in the Caribbean region (Fuentes Marrero 2018). The genomic basis of distyly in *Erythroxylum* remains unstudied.

These genome assemblies add to a growing repository of plant biology resources that will assist investigations focused on plant ecology, evolution, and natural products.

Methods

Leaf tissue from individual wild trees was collected for this study. DNA extraction was performed using the Qiagen DNAeasy genomic extraction kit 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 × 150 bp format. The resulting fastq files were trimmed of adapter/primer sequence and low-quality regions with Trimmomatic v0.33 (Bolger, Lohse, and Usadel 2014). The trimmed sequence was assembled by SPAdes v2.5 (Bankevich et al. 2012) followed by a finishing step using Zanfona (Kieras, O’Neill, and Pirro 2021).

Results

Data Availability

Raw read data and assembled genome sequences are available via Genbank.

taxname	genome_size	N50	SRA_acc	genome_acc
<i>Erythroxyllum acuminatum</i>	520,944,951	26.6 MB	SRR16518260	JANKOD000000000
<i>Erythroxyllum alaternifolium</i>	1,140,126,797	54.9 MB	SRR19167553	JANIIA000000000
<i>Erythroxyllum anguifugum</i>	1,398,821,111	61.1 MB	SRR19071309	JANIPN000000000
<i>Erythroxyllum areolatum</i>	1,181,226,222	63.9 MB	SRR19067407	JANIJH000000000
<i>Erythroxyllum baracoense</i>	674,802,793	32.4 MB	SRR20897294	JAOTAY000000000
<i>Erythroxyllum bequaertii</i>	1,069,376,509	66.2 MB	SRR19169396	JANIJF000000000
<i>Erythroxyllum brevipes</i>	934,221,715	62.1 MB	SRR16641246	JANIHY000000000
<i>Erythroxyllum carthagenense</i>	1,260,613,813	66.4 MB	SRR16134036	JANZLD000000000
<i>Erythroxyllum cassinoides</i>	1,151,892,187	66.6 MB	SRR19077077	JANIPM000000000
<i>Erythroxyllum cataractarum</i>	1,016,125,520	69.3 MB	SRR13024521	JANJEY000000000
<i>Erythroxyllum citrifolium</i>	1,079,479,909	61.8 MB	SRR17913416	JANKOE000000000
<i>Erythroxyllum confusum</i>	1,475,026,490	94.5 MB	SRR17999774	JANZXO00000000
<i>Erythroxyllum coriaceum</i>	1,397,282,891	60.7 MB	SRR17999775	JANQAP000000000
<i>Erythroxyllum densum</i>	961,004,938	34.7 MB	SRR16133662	JANPWT000000000
<i>Erythroxyllum divaricatum</i>	883,246,698	33.4 MB	SRR19579570	JANTPM000000000
<i>Erythroxyllum echinodendron</i>	767,751,986	32.0 MB	SRR19170120	JANZXR000000000
<i>Erythroxyllum engleri</i>	592,837,144	30.0 MB	SRR17999777	JANZLJ000000000
<i>Erythroxyllum fimbriatum</i>	623,511,794	25.1 MB	SRR17839727	JANZLH000000000
<i>Erythroxyllum flavicans</i>	692,607,313	32.8 MB	SRR17853861	JANZXP000000000
<i>Erythroxyllum foetidum</i>	616,347,371	36.4 MB	SRR16132259	JANZLF000000000
<i>Erythroxyllum glaucum</i>	679,154,896	34.3 MB	SRR17834921	JANZLK000000000
<i>Erythroxyllum gonocladus</i>	524,222,889	24.5 MB	SRR17999776	JANZLY000000000
<i>Erythroxyllum gracilipes</i>	675,766,074	35.3 MB	SRR13004464	JANZLC000000000
<i>Erythroxyllum guanchezii</i>	729,565,920	31.0 MB	SRR20897173	JAOTAX000000000
<i>Erythroxyllum haughtii</i>	657,627,051	34.1 MB	SRR19169397	JANSVB000000000
<i>Erythroxyllum havanense</i>	823,408,479	30.7 MB	SRR16128326	JANTNU000000000
<i>Erythroxyllum hondense</i>	991,084,625	65.4 MB	SRR19579446	JANKYF000000000
<i>Erythroxyllum incrassatum</i>	698,894,607	27.8 MB	SRR16511100	JANVCL000000000

taxname	genome_size	N50	SRA_acc	genome_acc
Erythroxyllum kapplerianum	838,545,021	30.4 MB	SRR19612725	JANSVIO000000000
Erythroxyllum kunthianum	1,325,754,177	50.8 MB	SRR19612843	JANIMX000000000
Erythroxyllum lineolatum	500,963,852	31.2 MB	SRR19612731	JAOQMO000000000
Erythroxyllum longipes	725,526,181	33.6 MB	SRR20971339	JAOQMS000000000
Erythroxyllum macrophyllum	678,289,703	22.1 MB	SRR17839751	JAOQMK000000000
Erythroxyllum mucronatum	712,302,707	23.3 MB	SRR20897490	JAOSBP000000000
Erythroxyllum orinocense	682,537,204	28.9 MB	SRR19076290	JANZLX000000000
Erythroxyllum oxycarpum	651,926,524	28.3 MB	SRR20971318	JAOTAZ000000000
Erythroxyllum panamense	572,741,246	34.2 MB	SRR17839679	JAOTIZ000000000
Erythroxyllum pedicellare	772,611,573	33.9 MB	SRR19612799	JAOQMP000000000
Erythroxyllum pelleterianum	574,749,498	28.0 MB	SRR20971396	JAOQMR000000000
Erythroxyllum platyclados	505,405,536	19.5 MB	SRR19624006	JAOQMM000000000
Erythroxyllum plowmanianum	930,928,887	29.8 MB	SRR19071311	JANVCQ000000000
Erythroxyllum raimondii	507,564,953	25.4 MB	SRR19622179	JAOQMN000000000
Erythroxyllum reticulatum	933,378,646	32.7 MB	SRR19169603	JANSVA000000000
Erythroxyllum roigii	624,502,766	26.1 MB	SRR16638488	JANZXV000000000
Erythroxyllum roraimae	659,978,004	27.6 MB	SRR19579568	JAOSBL000000000
Erythroxyllum rotundifolium	884,866,973	30.0 MB	SRR20912047	JAOSBO000000000
Erythroxyllum rufum	1,026,915,537	66.4 MB	SRR19169398	JANIHZ000000000
Erythroxyllum savannarum	545,546,056	27.3 MB	SRR19612969	JAOTAW000000000
Erythroxyllum shatona	692,895,338	36.7 MB	SRR19579569	JAOSBK000000000
Erythroxyllum squamatum	552,878,497	30.0 MB	SRR20971441	JAOTBA000000000
Erythroxyllum suave	858,558,633	31.2 MB	SRR20912045	JAOSBN000000000
Erythroxyllum subrotundum	548,290,955	26.5 MB	SRR19616481	JAOQMQ000000000
Erythroxyllum tenue	639,465,533	26.9 MB	SRR19580238	JAOSBM000000000
Erythroxyllum ulei	1,042,824,208	62.9 MB	SRR16633785	JANITX000000000
Erythroxyllum urbanii	763,890,317	28.6 MB	SRR19626866	JAOQML000000000
Erythroxyllum williamsii	1,013,529,337	66.2 MB	SRR16133663	JANITW000000000

.....

Funding

Funding was provided by Iridian Genomes, grant # IRGEN_RG_2021-1345 Genomic Studies of Eukaryotic Taxa. Dawson White is supported by the Grainger Bioinformatics Center at the Field Museum and an NSF Postdoctoral Fellowship in Biology, award 2010821.

Submitted: October 18, 2022 EST, Accepted: November 04, 2022 EST

REFERENCES

- Bankevich, Anton, Sergey Nurk, Dmitry Antipov, Alexey A. Gurevich, Mikhail Dvorkin, Alexander S. Kulikov, Valery M. Lesin, et al. 2012. “SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing.” *Journal of Computational Biology* 19 (5): 455–77. <https://doi.org/10.1089/cmb.2012.0021>.
- Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. “Trimmomatic: A Flexible Trimmer for Illumina Sequence Data.” *Bioinformatics* 30 (15): 2114–20. <https://doi.org/10.1093/bioinformatics/btu170>.
- Daly, D.C. 2004. “Erythroxylaceae (Coca Family).” In *Flowering Plants of the Neotropics*, edited by N. Smith, S.A. Mori, A. Henderson, D.W. Stevenson, and S.V. Heald, 143–45. Princeton: Princeton University Press.
- Darwin, C. 1877. *The Different Forms of Flowers on Plants of the Same Species*. Cambridge, UK: Cambridge University Press.
- Dillehay, Tom D., Jack Rossen, Donald Ugent, Anathasios Karathanasis, Víctor Vásquez, and Patricia J. Netherly. 2010. “Early Holocene Coca Chewing in Northern Peru.” *Antiquity* 84 (326): 939–53. <https://doi.org/10.1017/s0003598x00067004>.
- Fuentes Marrero, I.M. 2018. *Morfología floral y polimorfismos sexuales en el género Erythroxylum (Erythroxylaceae) en Cuba*.
- Kieras, M., K. O’Neill, and S. Pirro. 2021. “Zanfona, a Genome Assembly Finishing Tool for Paired-End Illumina Reads.” 2021. <https://github.com/zanfona734/zanfona>.
- Ly, Y., T. Tian, Y.-J. Wang, J.-P. Huang, and S.-X. Huang. 2022. “Advances in Chemistry and Bioactivity of the Genus *Erythroxylum*.” *Nat Prod Bioprospect* 12: 15.
- White, Dawson M. 2019. “Biogeography, Diversification, and Domestication in the Coca Family (Erythroxylaceae).”
- White, Dawson M, Jen-Pan Huang, Orlando Adolfo Jara-Muñoz, Santiago Madriñán, Richard H Ree, and Roberta J Mason-Gamer. 2020. “The Origins of Coca: Museum Genomics Reveals Multiple Independent Domestications from Progenitor *Erythroxylum Gracilipes*.” *Systematic Biology* 70 (1): 1–13. <https://doi.org/10.1093/sysbio/syaa074>.