Title

Genome assemblies and comparison of two Neotropical spiral gingers: Costus pulverulentus and C. lasius.

Journal

The Journal of heredity, 114(3)

ISSN

0022-1503

Authors

Harenčár, Julia
Vargas, Oscar M
Escalona, Merly
et al.

Publication Date

2023-05-01

DOI

10.1093/jhered/esad018

Peer reviewed

Journal of Heredity, 2023, **114**, 286–293
https://doi.org/10.1093/jhered/esad018
Advance access publication 16 March 2023
Genome Resources

American Genetic Association

OXFORD

Genome Resources

Genome assemblies and comparison of two Neotropical spiral gingers: *Costus pulverulentus* and *C. lasius*

Julia Harenčár[1], Oscar M. Vargas[2], Merly Escalona[3], Douglas W. Schemske[4], Kathleen M. Kay[1]

[1]Ecology and Evolutionary Biology Department, University of California, Santa Cruz, Santa Cruz, CA, United States,
[2]Department of Biological Sciences, California State Polytechnic University, Humboldt, Arcata, CA, United States,
[3]Department of Biomolecular Engineering, University of California, Santa Cruz, Santa Cruz, CA, United States,
[4]Department of Plant Biology, Michigan State University, East Lansing, MI, United States

Address correspondence to J. Harenčár at the address above, or e-mail: jharenca@ucsc.edu.

Corresponding Editor: Rachel Meyer

Abstract

The spiral gingers (*Costus* L.) are a pantropical genus of herbaceous perennial monocots; the Neotropical clade of *Costus* radiated rapidly in the past few million years into over 60 species. The Neotropical spiral gingers have a rich history of evolutionary and ecological research that can motivate and inform modern genetic investigations. Here, we present the first 2 chromosome-level genome assemblies in the genus, for *C. pulverulentus* and *C. lasius*, and briefly compare their synteny. We assembled the *C. pulverulentus* genome from a combination of short-read data, Chicago and Dovetail Hi-C chromatin-proximity sequencing, and alignment with a linkage map. We annotated the genome by mapping a *C. pulverulentus* transcriptome and querying mapped transcripts against a protein database. We assembled the *C. lasius* genome with Pacific Biosciences HiFi long reads and alignment to the *C. pulverulentus* genome. These 2 assemblies are the first published genomes for non-cultivated tropical plants. These genomes solidify the spiral gingers as a model system and will facilitate research on the poorly understood genetic basis of tropical plant diversification.

Spanish Abstract

Costus es un género pantropical de monocotiledóneas herbáceas perennes; el clado neotropical de Costus se diversificó rápidamente en más de 60 especies en los últimos millones de años. Las especies de neotropicales de Costus tienen una rica historia de investigación evolutiva y ecológica que puede motivar e informar las investigaciones genéticas modernas. Aquí, presentamos los dos primeros ensamblajes de genoma a nivel de cromosoma en el género, C. pulverulentus y C. lasius, y comparamos brevemente su sintenía. El genoma de C. pulverulentus lo ensamblamos a partir de una combinación de datos de secuenciación de lectura corta, secuenciación de proximidad de cromatina CHiCago y Hi-C de Dovetail y su alineación con un mapa de ligamiento. Ensamblamos el genoma de C. lasius con lecturas largas HiFi de Pacific Biosciences y alineación con el genoma de C. pulverulentus. Estos dos ensamblajes son los primeros genomas publicados para plantas tropicales no cultivadas. Estos genomas solidifican Costus como un sistema modelo y que facilitarán la investigación sobre la base genética poco conocida de la diversificación de plantas tropicales.

Key words: genome resources, herbaceous, monocot, synteny, tropical plant

Introduction

The tropics encompass the highest plant species diversity (Mutke and Barthlott 2005; Cai et al. 2023), yet there are relatively few tropical plant reference genomes. Whereas high-quality genomic resources exist for some cultivated tropical plants (e.g. Argout et al. 2011—chocolate; Droc et al. 2013—banana; Dereeper et al. 2015—coffee), there are no published high-quality genomes for wild tropical plants. Without reference genomes and broader investigation of tropical plant genetics, we cannot hope to understand the ecology and evolution of tropical plant diversity.

The Neotropical spiral gingers (genus *Costus* L.) are emerging as a model system for tropical plant evolutionary biology. *Costus* is a pantropical genus of perennial understory herbaceous monocots that contains over 80 species (Maas 1972, 1977). The majority (over 60) of these species diverged rapidly and recently (~3 mya) in Central and South America (Vargas et al. 2020). Despite their recent divergence, Neotropical spiral ginger species vary widely in reproductive traits, habitat affinities, and species interactions. Their variable ecology and rapid diversification make spiral gingers an excellent system for studying both speciation and the evolution of morphological, biochemical, and ecological traits.

Evolutionary investigation of the genus dates back to monographic work (Maas 1972, 1977) and early research on *Costus* species interactions and reproductive biology

Received February 22, 2023; Accepted March 15, 2023

(Schemske 1978, 1980, 1981, 1982, 1983; Schemske and Pautler 1984) and has continued since, building a rich foundational understanding of mechanisms of speciation, species interactions, local adaptation, floral evolution, and macroevolutionary patterns of diversification (e.g. Kay 2006; Bartlett and Specht 2010; Chen and Schemske 2015, 2019; Vargas et al. 2020; Ávila-Lovera et al. 2022). *Costus* provides one of the earliest empirical examples of reinforcement, a process which acts to increase reproductive isolation between 2 species occurring in sympatry (Kay and Schemske 2008; Yost and Kay 2009; Surget-Groba and Kay 2013). *Costus* species have specialized pollination interactions with orchid bees and hummingbirds, provide extrafloral nectar rewards to ant defenders, host specialized flower mites and leaf herbivores, and exhibit novel ecophysiological strategies (e.g. Schemske 1980; Kay and Schemske 2003; Kay et al. 2005; García-Robledo et al. 2013; Bizzarri et al. 2022; Harenčár et al. 2022; Kay and Grossenbacher 2022). This richness of ecological and evolutionary research provides an excellent foundation for genomics.

Spiral ginger research also benefits from their unusual tractability compared with other tropical plants. They are generally well described and identifiable by vegetative characteristics, can be maintained in greenhouses, and can be grown from seed or cuttings for use in experiments such as reciprocal transplants. *Costus* species are also tractable for genetic investigation as they are generally diploid with stable ploidy (Maas 1972, 1977), have relatively small genomes (~1 Gb), have publicly available transcriptome data for 5 species (NCBI Acc. PRJNA600282), and have a well-resolved phylogeny estimated from 853 genes (Vargas et al. 2020). Further, *Costus* species are widely interfertile (Kay and Schemske 2008), enabling propagation of hybrids for genetic mapping and focused investigation on the evolutionary role of extrinsic barriers to gene flow. Here, we improve Neotropical *Costus* as a model system with 2 reference genomes spanning the deepest node in the phylogeny.

We generated chromosome-level genome assemblies for *C. pulverulentus* C. Presl and *C. lasius* Loes. (Fig. 1). We first assembled a *C. pulverulentus* genome using a hybrid de novo assembly approach combining short-read shotgun and mate-pair libraries with Hi-C chromatin-proximity and Chicago libraries. We then used a RAD-seq genetic linkage map to pair half-chromosome scaffolds and generate a chromosome-level assembly. We assembled the *C. lasius* genome from PacBio HiFi long-read sequencing data and aligned it with the *C. pulverulentus* genome to assemble chromosome-scale scaffolds. We also estimated both genome sizes with flow cytometry. These genomes, combined with the other genomic resources available for the genus, allow us to leverage existing and ongoing field research to build a nuanced understanding of the almost entirely unexplored genomics of Neotropical plant diversification.

Methods

Biological materials

Costus pulverulentus

The *C. pulverulentus* individual selected for sequencing was grown at the University of California, Santa Cruz (UCSC) from field-collected seed. The seed was collected by DWS on Barro Colorado Island in Panama (9.155, −79.85). A voucher is deposited at the Michigan State University (MSU) Herbarium (Kay 0328 (MSC)). We collected stem meristem and nascent leaf tissue from the live greenhouse individual.

Costus lasius

The *C. lasius* individual selected for sequencing was grown in the UCSC greenhouses from a field-collected cutting. The cutting was collected by KMK from a wild individual growing in El Valle de Antón, Panama (8.633, −80.117). A voucher is deposited at the MSU Herbarium (Kay 0321 (MSC)). We collected stem meristem, nascent leaf, and nascent bud tissue from the greenhouse individual.

Nucleic acid library preparation

Costus pulverulentus

We extracted high molecular weight (HMW) DNA from fresh tissue ground in a bead beater with PVP buffer. We used a modified Qiagen Puregene Gentra kit protocol (S1)

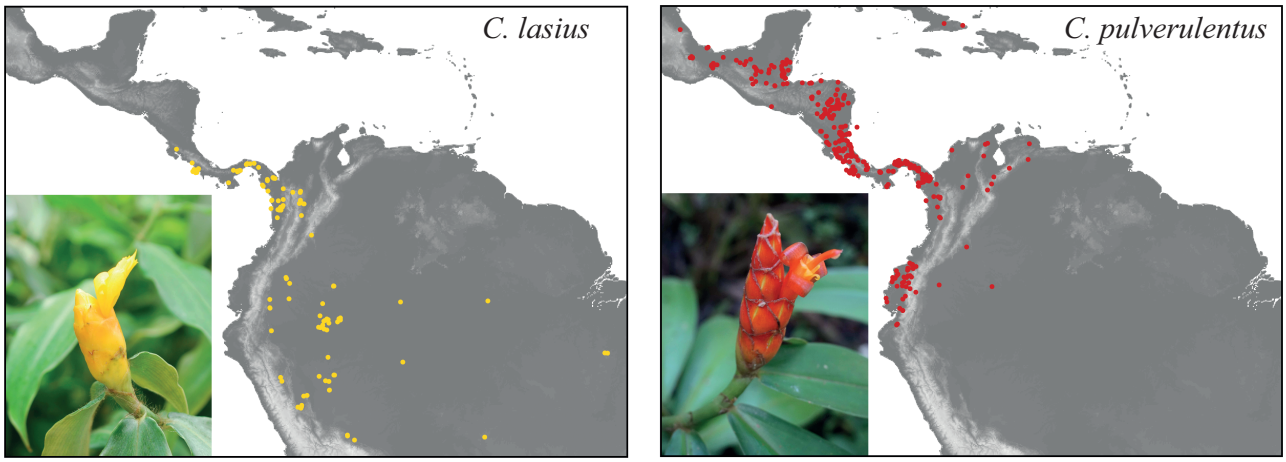

Fig. 1. Geographic distributions and photos of each species. Location data are from the Vargas et al. (2020) supplement and represent expert-verified collection locations. Photos show a terminal inflorescence with a single open flower. Photo credits: *C. lasius* - KMK; *C. pulverulentus* - Rossana Maguiña.

Table 1. Assembly pipeline and software usage.

Purpose	Software and options*	Version
Costus pulverulentus assembly		
De novo assembly of shotgun and mate-pair data	Meraculous-2D	2.2.5.1
Scaffolding with Hi-C and Chicago data	Dovetail Genomics' HiRise	2018
Transcriptome annotation	PASA	2.3.3
Gene annotation	EnTAP	1.9.2
Visualization and scaffold joining	geneious Prime	2022.2.1
Costus lasius assembly		
De novo assembly of PacBio HiFi CCS reads	HiFiasm	0.11
Remove low-coverage, duplicated contigs	purge_dups	1.0
Scaffold based on *C. pulverulentus*	RagTag	2.1.0
Genome quality assessment		
Basic assembly metrics	QUAST (--est-ref-size)	5.0.2
Assembly gap quantification	Asset (det_gaps)	1.0.3
Assembly function; completeness	BUSCO (-l embryophyta)	5.0.0
Assembly completeness	Merqury	2020-01-29
k-mer counting	Meryl (k = 21)	1.0

*Software citations are listed in the text.

for HMW DNA extraction. We assessed DNA concentration with a Qubit Fluorometer (Thermo Fisher Scientific, Waltham, MA) and ran an agarose gel to assess DNA fragment integrity and size. An Illumina DNA TruSeq library was created for regular whole-genome shotgun sequencing. Mate-pair libraries were created with 6 kbp insert sizes using an Illumina Nextera mate-pair library kit. Additional flash-frozen leaf tissue was sent to Dovetail Genomics (Scotts Valley, CA) for Hi-C and Chicago library preparation and sequencing.

Costus lasius

We extracted HMW DNA from 1 g of fresh mixed leaf and bud tissue. We isolated cell nuclei following Workman (2018) and then used the Circulomics Nanobind Plant Nuclei Big DNA kit to extract HMW DNA. We assessed DNA concentration with a Qubit Fluorometer (Thermo Fisher Scientific), purity with 260/280 and 260/230 absorbance ratios from a NanoDrop Spectrophotometer (Thermo Fisher Scientific), and ran a 7% agarose gel at 70 V for over 8 h to assess DNA fragment integrity and size.

DNA sequencing and genome assembly

Costus pulverulentus

We sequenced the whole-genome shotgun libraries on a single lane of an Illumina HiSeq 4000 in paired-end 150 bp mode. We sequenced the 6 kb mate-pair libraries on an Illumina NextSeq 500 in paired-end 150 bp mode. Dovetail Genomics (Scotts Valley, CA) conducted Hi-C and Chicago sequencing.

We conducted initial assembly of the shotgun and 6 kb mate-pair sequencing data with Meraculous-2D (Goltsman et al. 2017). Genome scaffolding was conducted by Dovetail Genomics using their HiRise pipeline (Putnam et al. 2016), combining the initial assembly from shotgun and mate-pair sequencing with Hi-C, and CHiCago library data. This produced 13 large scaffolds representing presumptive full and half-chromosomes. To assemble the 13 largest scaffolds into 9 chromosomes, we mapped RAD-seq markers (Kay and Surget-Groba 2022) against the draft genome using BBmap (Bushnell 2014) and visualized the results with R. We based chromosome naming on the previously published chromosome names (Kay and Surget-Groba 2022), which were based on chromosome centimorgan length from a linkage map.

To create a draft annotation for the *C. pulverulentus* genome, we mapped a transcriptome of *C. pulverulentus* (GenBank accession SRX7544604, Vargas et al. 2020) to the draft genome using PASApipeline (Haas et al. 2003). To search for potential gene functions in our draft annotation, we queried our mapped transcripts against a protein database (GO; Harris et al. 2004) using EnTAP (Hart et al. 2020). Matching protein names and functions were added as annotation notes in the .gff3 annotation file generated with PASA using a custom python script (available in Dryad repository; see "Data availability"). We then imported annotation files to Geneious Prime 2022.2.1 for visualization and, because annotations were added prior to the assembly of large scaffolds into chromosomes, to maintain annotations during final assembly.

All assembly software versions and parameters are listed in Table 1. We conducted additional trimming and removal of mitochondrial sequences per NCBI recommendations upon submission.

Costus lasius

We sent HMW DNA to the Vincent J. Coates Genomics Sequencing Lab (Berkeley, CA) for Pacific Biosciences HiFi library preparation and circular consensus sequencing (CCS) on 2 PacBio Sequel II HiFi SMRT cells.

We assembled the PacBio CCS data into preliminary primary and alternate assemblies with HiFiasm (Cheng et al. 2021). To improve the assemblies, we used purge_dups (Guan et al. 2020) to remove duplicated haplotigs and sequence overlaps from the primary assembly and add them to the alternate assembly. Finally, RagTag (Alonge et al. 2019) was used to align the primary *C. lasius* assembly to the *C. pulverulentus*

assembly and assign corresponding chromosome names. The assembly software versions and parameters are listed in Table 1. We conducted additional trimming and removal of mitochondrial sequences per NCBI recommendations upon submission.

Genome size estimation, quality assessment, and genome comparison

For size estimation, we sent fresh leaf tissue from the sequenced individuals of both species (and 5 other species; S3) to the Flow Cytometry Lab at the Benaroya Research Institute, Seattle, WA. Four replicates were run for each sample. We also calculated the total genome assembly lengths with QUAST (Gurevich et al. 2013).

We also used QUAST to generate contiguity metrics and base pair content. To further evaluate genome completeness and quality, we ran BUSCO (Manni et al. 2021) with the Embryophyta database (embryophyta_odb10; 1,614 genes). To assess base call accuracy (QV) and k-mer completeness, we first generated meryl databases with the shotgun and 6 kb mate-pair sequence data for *C. pulverulentus* and with the PacBio HiFi reads for *C. lasius*. We then ran merqury (Rhie et al. 2020) with the meryl databases and genome assemblies to obtain the QV and k-mer completeness estimates. To determine the number of gaps we first used det_gaps, a submodule from asset (https://github.com/dfguan/asset) that describes the existing gaps in the assemblies.

Finally, to understand genome size discrepancies and determine synteny we compared the *C. pulverulentus* and *C. lasius* assemblies by generating a dotplot with D-GENIES (Cabanettes and Klopp 2018) using Minimap2 (version 2.24; Li 2018) as the aligner and the "some repeats" option.

Results

Costus pulverulentus sequencing, assembly, and annotation

The combined shotgun sequencing runs yielded 615 million PE 150 reads, and the 6 kb mate-pair sequencing yielded 136 million reads. The genome coverage reported by Dovetail Genomics after HiRise Assembly was ~1,959-fold.

The final *C. pulverulentus* genome assembly size (~736 Mbp) is about 300 Mbp smaller than the average flow cytometry estimate (1.036 Gbp). The 9 largest contigs correspond to chromosomes following successful matching of chromosome arms with the linkage map S2. The assembly is composed of 23,271 contigs with an N50 of 63.17 Mbp and the longest contig is 97.28 Mbp. Complete assembly statistics are reported in Table 2.

For the draft annotation, ~93% of our *C. pulverulentus* transcripts (258,934 out of 279,542) mapped to our genome, and about 14% of the mapped transcripts (37,512) matched an entry in the protein database. The transcriptome of *C. pulverulentus* is highly repetitive, representing different protein configurations from single genes. For example, gene 1 (see the .gff annotation files in the Dryad repository listed in "Data availability"), can produce 5 different proteins depending on the inclusion/exclusion of exons.

Costus lasius sequencing and assembly

The combined PacBio sequencing runs yielded 6.3 million HiFi reads representing ~19-fold coverage based on flow cytometry estimated genome size (HiFi read data averaged across both runs: N50 read length 166,084 bp; mean read length 144,929 bp).

The final *C. lasius* genome assembly size (1.28 Gbp) is similar to the average flow cytometry estimate (1.27 Gbp). The 9 largest contigs correspond to homologous chromosomes in the *C. pulverulentus* assembly. The assembly is composed of 278 contigs with an N50 of 119.75 Mb and the longest contig is 163.08 Mb. Complete assembly statistics are reported in Table 2.

Genome comparison

Comparison of the *C. pulverulentus* and *C. lasius* genomes shows that the genomes are largely syntetic (Fig. 2). There is one potential transposition between chromosome 7 in the *C. pulverulentus* genome and chromosome 5 in the *C. lasius* genome. There are also several small potential inversions on chromosomes 3, 8, and 9. However, both the potential transposition and inversions need further validation. The *C. lasius* assembly tends to be longer at chromosome ends and centromeric regions.

Flow cytometry results from a total of 6 species and a hybrid support previous chromosome count data showing that *Costus* are largely diploid (S3; Maas 1972, 1977). Genome sizes ranged from 924 to 1,376 Mbp; the 2 species sequenced here represent relatively small and large genomes at 1,036 Mbp (*C. pulverulentus*) and 1,268 Mpb (*C. lasius*).

Discussion

The 2 genome assemblies described here are the first chromosome-level genomes for the *Costus* genus. The *C. pulverulentus* genome has already facilitated the generation of a well-supported phylogeny (Vargas et al. 2020) and genetic mapping of evolutionarily important floral traits (Kay and Surget-Groba 2022). Both genomes are being used in ongoing investigation of tropical plant evolution, including research on the population genetics of hybridizing species, the genetic basis of habitat isolation, and the extent of introgression across the clade.

Comparison of the *C. lasius* and *C. pulverulentus* genomes clearly demonstrates the advantage of long-read data for genome assembly. The *C. pulverulentus* genome assembly is shorter than the flow cytometry estimate of its size (assembly = 736 Mbp, flow cytometry = 1,036 Mp), potentially due to collapsed assemblies in repetitive regions with short-read data. Many of the contigs from the *C. pulverulentus* genome that were not placed on chromosomes aligned to what are likely centromeric and chromosome end regions in *C. lasius*, which tend to be highly repetitive (Blackburn and Szostak 1984; Lamb et al. 2007). In contrast, long-read data improved the assembly of these regions in *C. lasius*. The synteny plot also hints at why the *C. lasius* genome is longer than the *C. pulverulentus* genome; the majority of breaks in the synteny plot showing where *C. lasius* is longer are also in centromeric and chromosome end regions. This suggests that the larger size of the *C. lasius* genome is due to increases in repetitive regions. The abundance of repetitive DNA is both highly variable even between closely related plants and strongly correlated with genome size (Hancock 2002; Lee and Kim 2014). Differences in repetitive regions, which mostly comprise transposable elements (TEs), can have implications

Table 2. Sequencing and assembly statistics, and accession numbers.

	Species		*C. pulverulentus*	*C. lasius*
BioProjects & Vouchers	NCBI BioProject		PRJNA864859	PRJNA864861
	NCBI BioSample		SAMN13824542	SAMN30087462
	Voucher		Kay 0328 (MSC)	Kay 0321 (MSC)
NCBI genome sequence data	PacBio HiFi reads accession		—	SRR21090603
	Illumina shotgun reads accession		SRR21047455	—
	Illumina mate-pair reads accession		SRR21047454	—
	Genome sequence		JANTFE000000000	JANVAS000000000
	Assembly accession		GCA_027562315.1	GCA_027563935.1
Genome Assembly Quality Metrics	Flow cytometry est. genome size		1,035.59 Mbp	1,268.28 Mbp
	Genome length		735.86 Mbp	1,280.87 Mbp
	# contigs		23,271	278
	Scaffold N50		63,174,634	119,749,194
	Scaffold L50		5	5
	Contig N50		24,502	21,722,320
	Longest contig		97,275,379	163,078,524
	Total gaps		36,438	100
	Gaps per Gbp (#Gaps)		49,518	78
	GC%		39.43	40.75
	Base composition	A	30.32	29.61
		C	19.72	20.38
		G	19.71	20.37
		T	30.25	29.64
	BUSCO	Complete	97.9%	98.7%
		Single	91.9%	91.2%
		Duplicated	6.0%	7.5%
		Fragmented	0.7%	0.4%
		Missing	1.4%	0.9%
		n	1,614	1,614
	Merqury	Base-call QV	80.4	59.1
		k-mer completeness	87.1	90.8

for gene expression and species divergence (Heslop-Harrison 2000; Feschotte and Pritham 2007; Jurka et al. 2007; Heslop-Harrison and Schmidt 2012; Kim et al. 2012; Fedoroff 2013; Lee and Kim 2014). Comparison of the 2 genomes indicates the advantage of long-read data for understanding repetitive regions, enabling investigation of the role of TEs in species distinction.

Genome comparison also reveals that *C. lasius* and *C. pulverulentus* genomes are largely conserved structurally. *C. lasius* and *C. pulverulentus* diverged ~3 mya (Vargas et al. 2020), yet display only a few relatively small potential rearrangements, remaining predominantly syntenic. Some organisms with similar divergence times show similarly conserved genomes. For example, the American bison (*Bison bison bison*) and cattle (*Bos taurus*) diverged ~3.5 mya (Kumar et al. 2017) and have largely conserved chromosomes with no large inversions or transpositions, just insertions and deletions (Oppenheimer et al. 2021). *Drosophila melanogaster* and *D. simulans* diverged ~4 mya (Kumar et al. 2017) and, while their karyotypes are also conserved, they demonstrate more structural differences (especially inversions) which could partially be due to longer divergence and shorter generation time than our focal *Costus* species (Ranz et al. 2007). Shorter generation time may also partially explain greater structural divergence between more closely related organisms, such as an inversion between ecotypes of *Mimulus guttatus* and multiple inversions between *Helianthus annuus* and *H. argophyllus* (diverged ~1.5 mya; Wellenreuther and Bernatchez 2018). Continued sequencing and comparison of genomes like that presented here will enable evaluation of how and when genomes are more versus less structurally conserved.

Several genome assemblies exist for economically important tropical plants, but, unlike *Costus*, these genomes are not well positioned for studying tropical plant diversification. We were only able to find published tropical plant genomes for cultivated species (e.g. Prochnik et al. 2012—cassava; Ming et al. 2012—papaya; Guignon et al. 2016—various). In contrast to these cultivated species, *Costus* is emerging as a model system for investigating the evolution and ecology of tropical plant diversification. These 2 genome assemblies add to growing genomic resources for the genus that include transcriptome data, floral QTL, and a well-resolved phylogeny (Vargas et al. 2020; Kay and

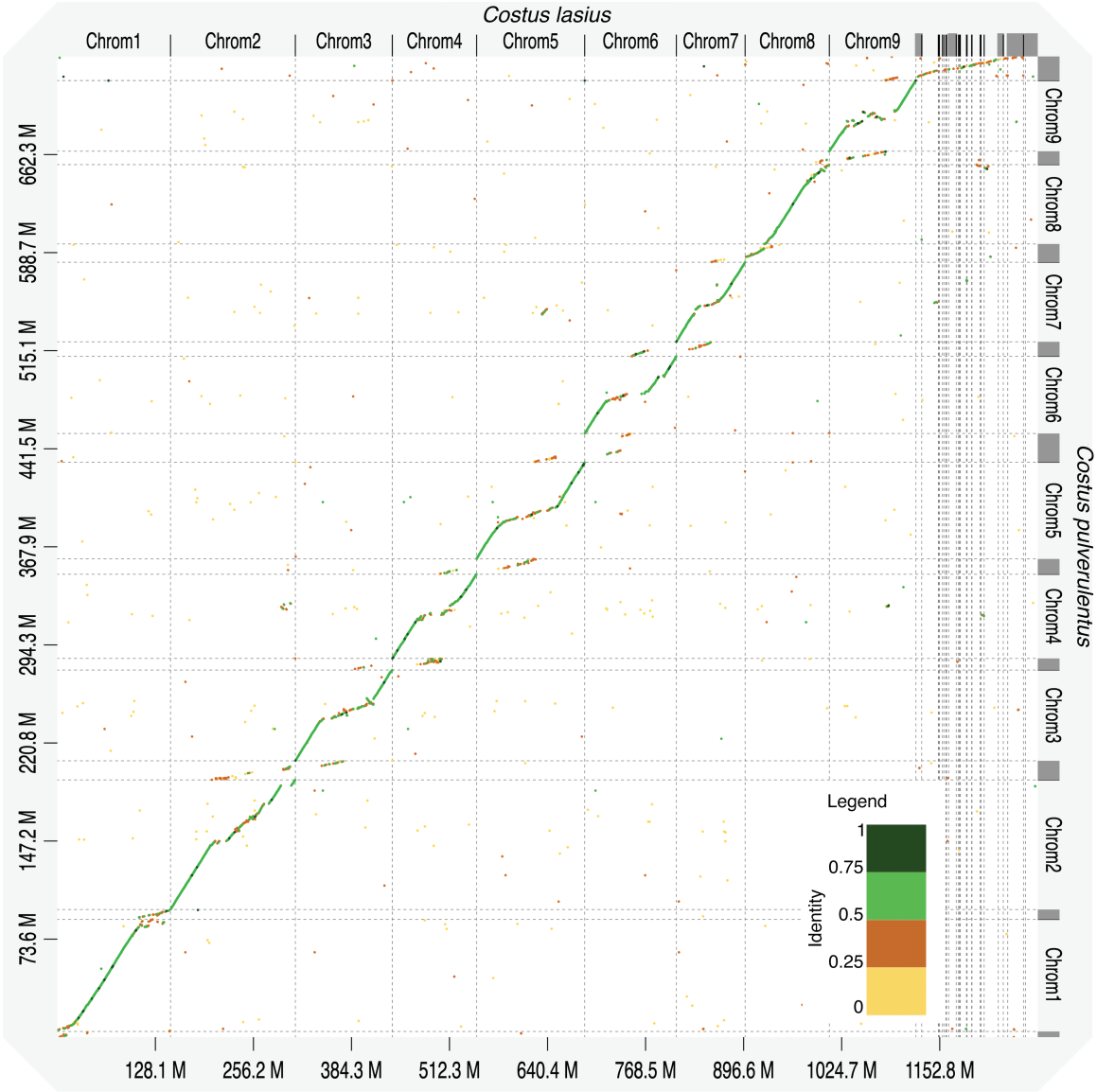

Fig. 2. Syntetic dotplot with the *C. lasius* genome assembly represented by the *x* axis and the *C. pulverulentus* genome assembly represented on the *y* axis. Dots represent matching sequences and where they are found along the genomes. Dot color represents alignment identity, or how well the sequences match between the genomes, with 1 representing a perfect match and 0 representing no match. Dark green represents sequences with alignment scores of 0.75 to 1, light green represents scores from 0.5 to 0.75 alignment identity, etc. We generated the plot with D-GENIES using Minimap 2 for alignment and selecting the "sort" and "hide noise" options. Access the interactive plot here.

Surget-Groba 2022). Further, genomic resources will enable researchers to leverage a rich foundation of natural history, ecological, and evolutionary research in Neotropical *Costus* to understand the genetic basis of coexistence and speciation.

Supplementary material

Supplementary material is available at *Journal of Heredity* online.

S1. Kay HMW extraction protocol.

S2. Linkage map generated with RAD-seq data from a *C. pulverulentus* × *C. scaber* hybrid plotted against the *C. pulverulentus* assembly sequence.

S3. Flow cytometry results.

S4. D-GENIES synteny dotplot interactive offline viewer (https://jharencar.github.io/Costus-interactive-synteny-plot/Clasius_X_Cpulverulentus_interactive_synteny_plot.html).

Acknowledgments

We thank Nedda Saremi for generating the original *C. pulverulentus* assembly that was sent to Dovetail Genomics. We are also grateful to the Dovetail Genomics staff and Vincent J. Coates Genomics Sequencing Lab staff for their dedication to providing quality data. Both of the sequenced individuals were under the long-term care of the UCSC greenhouses; we thank Jim Velzy, Sylvie Childress, and the rest of the greenhouse staff for their excellent plant care. We conducted the majority of assembly and QC after sequencing for *C. lasius*

and after Dovetail scaffolding for *C. pulverulentus* on the UCSC Hummingbird Computational Cluster, and we thank Rion Parsons and the rest of the Hummingbird staff for their support.

Funding

This work was supported by the National Science Foundation Dimensions of Biodiversity grant (DEB-1737889) to KMK; and the John A. Hannah Distinguished Professorship in Plant Biology at MSU held by DWS.

Data availability

Data generated for this study are available under NCBI BioProject PRJNA864859 for *C. pulverulentus* and PRJNA864861 for *C. lasius*. Raw sequencing data for *C. pulverulentus* (NCBI BioSample SAMN13824542) and *C. lasius* (NCBI BioSample SAMN30087462) are deposited in the NCBI Short Read Archive (SRA) under SRR21047455 for *C. pulverulentus* shotgun sequencing data, SRR21047454 for *C. pulverulentus* 6 kp mate-pair sequencing data, and SRR21090603 for *C. lasius* PacBio HiFi sequencing data. The GenBank accession for the *C. pulverulentus* assembly is GCA_027562315.1, and JANTFE000000000 for the genome sequences. The GenBank accession for the *C. lasius* assembly is GCA_027563935.1, and JANVAS000000000 for the genome sequences. *C. pulverulentus* genome annotations are available on Dryad at doi:10.5061/dryad.cfxpnvx94.

References

Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, Lippman ZB, Schatz MC. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. *Genome Biol*. 2019;20:224.

Argout X, Salse J, Aury J-M, Guiltinan MJ, Droc G, Gouzy J, Allegre M, Chaparro C, Legavre T, Maximova SN, et al. The genome of *Theobroma cacao*. *Nat Genet*. 2011;43:101–108.

Ávila-Lovera E, Goldsmith GR, Kay KM, Funk JL. Above- and below-ground functional trait coordination in the neotropical understory genus *Costus*. *AoB PLANTS*. 2022;14:plab073.

Bartlett ME, Specht CD. Evidence for the involvement of GLOBOSA-like gene duplications and expression divergence in the evolution of floral morphology in the Zingiberales. *New Phytol*. 2010;187:521–541.

Bizzarri L, Baer CS, García-Robledo C. DNA barcoding reveals generalization and host overlap in hummingbird flower mites: implications for the mating rendezvous hypothesis. *Am Nat*. 2022;199:576–583.

Blackburn EH, Szostak JW. The molecular structure of centromeres and telomeres. *Annu Rev Biochem*. 1984;53:163–194.

Bushnell B. *BBMap: a fast, accurate, splice-aware aligner*. Berkeley (CA): Lawrence Berkeley National Lab. (LBNL); 2014.

Cabanettes F, Klopp C. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ*. 2018;6:e4958.

Cai L, Kreft H, Taylor A, Denelle P, Schrader J, Essl F, van Kleunen M, Pergl J, Stein A, Winter M, et al. Global models and predictions of plant diversity based on advanced machine learning techniques. *New Phytol*. 2023;237:1432–1445.

Chen GF, Schemske DW. Ecological differentiation and local adaptation in two sister species of Neotropical *Costus* (Costaceae). *Ecology*. 2015;96:440–449.

Chen GF, Schemske DW. Adaptation to seasonal drought in two closely related species of Neotropical *Costus* (Costaceae). *Biotropica*. 2019;51e3:311–315.

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods*. 2021;18:170–175.

Dereeper A, Bocs S, Rouard M, Guignon V, Ravel S, Tranchant-Dubreuil C, Poncet V, Garsmeur O, Lashermes P, Droc G. The coffee genome hub: a resource for coffee genomes. *Nucleic Acids Res*. 2015;43:D1028–D1035.

Droc G, Larivière D, Guignon V, Yahiaoui N, This D, Garsmeur O, Dereeper A, Hamelin C, Argout X, Dufayard J-F, et al. The banana genome hub. *Database*. 2013;2013.

Fedoroff NV. *Plant transposons and genome dynamics in evolution*. Hoboken, New Jersey: John Wiley & Sons; 2013.

Feschotte C, Pritham EJ. DNA transposons and the evolution of eukaryotic genomes. *Annu Rev Genet*. 2007;41:331–368.

García-Robledo C, Erickson DL, Staines CL, Erwin TL, Kress WJ. Tropical plant–herbivore networks: reconstructing species interactions using DNA barcodes. *PLoS One*. 2013;8:e52967.

Goltsman E, Ho I, Rokhsar D. Meraculous-2D: haplotype-sensitive assembly of highly heterozygous genomes. arXiv preprint arXiv:1703.09852. 2017 Mar 29.

Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*. 2020;36:2896–2898.

Guignon V, Hueber Y, Rouard M, Bocs S, Couvin D, Lamotte F, Droc G, Dufayard J-F, El Hassouni N, Farcy C, et al. The South Green portal: a comprehensive resource for tropical and Mediterranean crop genomics. *Curr Plant Biol*. 2016;7–8:6–9.

Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013;29:1072–1075.

Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK Jr, Hannick LI, Maiti R, Ronning CM, Rusch DB, Town CD, et al. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res*. 2003;31:5654–5666.

Hancock JM. Genome size and the accumulation of simple sequence repeats: implications of new data from genome sequencing projects. *Genetica*. 2002;115:93–103.

Harenčár JG, Ávila-Lovera E, Goldsmith GR, Chen GF, Kay KM. Flexible drought deciduousness in a neotropical understory herb. *Am J Bot*. 2022;109:1262–1272.

Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, et al. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res*. 2004;32:D258–D261.

Hart AJ, Ginzburg S, Xu M (Sam), Fisher CR, Rahmatpour N, Mitton JB, Paul R, Wegrzyn J. EnTAP: bringing faster and smarter functional annotation to non-model eukaryotic transcriptomes. *Mol Ecol Res*. 2020;20:591–604.

Heslop-Harrison JS. Comparative genome organization in plants: from sequence and markers to chromatin and chromosomes. *Plant Cell*. 2000;12:617–636.

Heslop-Harrison J (Pat), Schmidt T. Plant nuclear genome composition. John Wiley & Sons, Ltd; 2012.

Jurka J, Kapitonov VV, Kohany O, Jurka MV. Repetitive sequences in complex genomes: structure and evolution. *Annu Rev Genom Hum Genet*. 2007;8:241–259.

Kay KM. Reproductive isolation between two closely related hummingbird pollinated neotropical gingers. *Evolution*. 2006;60:538–552.

Kay KM, Grossenbacher DL. Evolutionary convergence on hummingbird pollination in Neotropical *Costus* provides insight into the causes of pollinator shifts. *New Phytol*. 2022;236:1572–1583.

Kay KM, Reeves PA, Olmstead RG, Schemske DW. Rapid speciation and the evolution of hummingbird pollination in neotropical *Costus* subgenus *Costus* (Costaceae): evidence from nrDNA ITS and ETS sequences. *Am J Bot*. 2005;92:1899–1910.

Kay KM, Schemske DW. Pollinator assemblages and visitation rates for 11 species of Neotropical *Costus* (Costaceae). *Biotropica*. 2003;35:198–207.

Kay KM, Schemske DW. Natural selection reinforces speciation in a radiation of neotropical rainforest plants. *Evolution*. 2008;62:2628–2642.

Kay KM, Surget-Groba Y. The genetic basis of floral mechanical isolation between two hummingbird-pollinated Neotropical understorey herbs. *Mol Ecol*. 2022;31:4351–4363.

Kim Y-J, Lee J, Han K. Transposable elements: no more "Junk DNA". *Genom Inform*. 2012;10:226–233.
Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol Biol Evol*. 2017;34:1812–1819.
Lamb J, Yu W, Han F, Birchler J. Plant chromosomes from end to end: telomeres, heterochromatin and centromeres. *Curr Opin Plant Biol*. 2007;10:116–122.
Lee S-I, Kim N-S. Transposable elements and genome size variations in plants. *Genom Inform*. 2014;12:87–97.
Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34:3094–3100.
Maas PJM. Costoideae (Zingiberaceae). *Flora Neotrop*. 1972;8:1–139.
Maas PJM. Renealmia (Zingiberaceae-Zingiberoideae) Costoideae (Additions) (Zingiberaceae). *Flora Neotrop*. 1977;18:1–218.
Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO Update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol*. 2021;38:4647–4654.
Ming R, Yu Q, Moore PH, Paull RE, Chen NJ, Wang M-L, Zhu YJ, Schuler MA, Jiang J, Paterson AH. Genome of papaya, a fast growing tropical fruit tree. *Tree Genet Genom*. 2012;8:445–462.
Mutke J, Barthlott W. Patterns of vascular plant diversity at continental to global scale. *Biologiske Skrifter*. 2005;55:521–537.
Oppenheimer J, Rosen BD, Heaton MP, Vander Ley BL, Shafer WR, Schuetze FT, Stroud B, Kuehn LA, McClure JC, Barfield JP, et al. A reference genome assembly of American Bison, *Bison bison bison*. *J Hered*. 2021;112:174–183.
Prochnik S, Marri PR, Desany B, Rabinowicz PD, Kodira C, Mohiuddin M, Rodriguez F, Fauquet C, Tohme J, Harkins T, et al. The Cassava genome: current progress, future directions. *Trop Plant Biol*. 2012;5:88–94.
Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, Troll CJ, Fields A, Hartley PD, Sugnet CW, et al. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res*. 2016;26:342–350.
Ranz JM, Maurin D, Chan YS, Grotthuss M von, Hillier LW, Roote J, Ashburner M, Bergman CM. Principles of genome evolution in the *Drosophila melanogaster* species group. *PLoS Biol*. 2007;5:e152.
Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol*. 2020;21:245.
Schemske DW. A coevolved triad: *Costus woodsonii* (Zingiberaceae) its dipteran seed predator and ant mutualists. *Bull Ecol Soc Am*. 1978;59:89.
Schemske DW. The evolutionary significance of extrafloral nectar production by *Costus woodsonii* (Zingiberaceae): an experimental analysis of ant protection. *J Ecol*. 1980;68:959–967.
Schemske DW. Floral convergence and pollinator sharing in two bee-pollinated tropical herbs. *Ecology*. 1981;62:946–954.
Schemske DW. Ecological correlates of a neotropical mutualism: ant assemblages at *Costus* extrafloral nectaries. *Ecology*. 1982;63:932–941.
Schemske DW. Breeding system and habitat effects on fitness components in three neotropical *Costus* (zingiberaceae). *Evolution*. 1983;37:523–539.
Schemske DW, Pautler LP. The effects of pollen composition on fitness components in a neotropical herb. *Oecologia*. 1984;62:31–36.
Surget-Groba Y, Kay KM. Restricted gene flow within and between rapidly diverging Neotropical plant species. *Mol Ecol*. 2013;22:4931–4942.
Vargas OM, Goldston B, Grossenbacher DL, Kay KM. Patterns of speciation are similar across mountainous and lowland regions for a Neotropical plant radiation (Costaceae: *Costus*). *Evolution*. 2020;74:2644–2661.
Wellenreuther M, Bernatchez L. Eco-evolutionary genomics of chromosomal inversions. *Trends Ecol Evol*. 2018;33:427–440.
Workman R. High molecular weight DNA extraction from recalcitrant plant species for third generation sequencing. 2018. https://protocolexchange.researchsquare.com/article/nprot-6785/v1
Yost JM, Kay KM. The evolution of postpollination reproductive isolation in *Costus*. *Sex Plant Reprod*. 2009;22:247–255.