

DATA NOTE

Open Access



Draft genome assemblies for two species of *Escallonia* (Escalloniales)

Andre S. Chanderbali^{1*}, Christopher Dervinis², Ioana G. Anghel³, Douglas E. Soltis¹, Pamela S. Soltis¹ and Felipe Zapata⁴

Abstract

Objectives *Escallonia* (Escalloniaceae) belongs to the Escalloniales, a diverse clade of flowering plants with unclear placement in the tree of life. *Escallonia* species show impressive morphological and ecological diversity and are widely distributed across three hotspots of biodiversity in the Neotropics. To shed light on the genomic substrate of this radiation and the phylogenetic placement of Escalloniales as well as to generate useful data for comparative evolutionary genomics across flowering plants, we produced and annotated draft genomes for two species of *Escallonia*.

Data description Genomic DNA from *E. rubra* and *E. herrerae* was sequenced with Oxford Nanopore sequencing chemistry, generating 3.4 and 12 million sequence reads with an average read length of 9.4 and 9.1 Kb (approximately 31 and 111 Gb of sequence data), respectively. In addition, we generated Illumina 100-bp paired-end short read data for *E. rubra* (approximately 75 Gb of sequence data). The *Escallonia rubra* genome was 566 Mb, with 3,233 contigs and an N50 of 285 Kb. The assembled genome for *E. herrerae* was 994 Mp, with 5,760 contigs and an N50 of 317 Kb. The genome sequences were annotated with 31,038 (*E. rubra*) and 47,905 (*E. herrerae*) protein-coding gene models supported by transcriptome/protein evidence and/or Pfam domain content. BUSCO assessments indicated completeness levels of approximately 98% for the genome assemblies and 88% for the genome annotations.

Keywords Andes, Campanulidae, *Escallonia rubra*, *Escallonia herrerae*, Genomics, Oxford Nanopore Technology

Objective

Escalloniales comprise approximately 130 species of herbs, shrubs, and trees that grow in diverse habitats ranging from desolate rocky outcrops to rain forests across South America, Australia, Southeast Asia, and the Indian Ocean islands [1]. It is not known how and when Escalloniales diversified so extensively and colonized the Southern Hemisphere because the phylogenetic relationships within Escalloniales and between Escalloniales and other flowering plant lineages remain elusive. Escalloniales are part of the more inclusive clade Campanulidae, a hyperdiverse group of flowering plants with approximately 35,000 species [2]. Yet, the precise phylogenetic relationships among the major lineages of Campanulidae

*Correspondence:

Andre S. Chanderbali
achander@ufl.edu

¹Florida Museum of Natural History, University of Florida, Gainesville, FL, USA

²School of Forest, Fisheries, and Geomatics Sciences, University of Florida, Gainesville, FL, USA

³Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA, USA

⁴Department of Ecology and Evolutionary Biology and Center for Tropical Research, Institute of the Environment and Sustainability, University of California, Los Angeles, CA, USA



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

have not been clearly resolved with strong support by current molecular data [3–7]. Clarifying these relationships is critical to elucidate the mechanisms of phenotypic evolution and geographic diversification for a large group of angiosperms [8, 9]. Within Escalloniales, the genus *Escallonia* represents a remarkable radiation across three hotspots of biodiversity in the mountains of South America [10, 11]. *Escallonia* species grow from sea level to snow line, and from temperate to tropical regions, showing distinct adaptations related to environmental stress such as extreme temperature and water availability.

Table 1 Overview of data files/data sets

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data File 1	Materials and Methods	Text (.docx)	Figshare: https://doi.org/10.6084/m9.figshare.24263800 [17]
Data set 1	Illumina reads of <i>E. rubra</i> genomic DNA	DNA sequence (.fastq)	NCBI: https://identifiers.org/ncbi/insdc.sra:SRX21711620 [18]
Data set 2	ONT reads of <i>E. rubra</i> genomic DNA	DNA sequence (.fastq)	NCBI: https://identifiers.org/ncbi/insdc.sra:SRX21711620 [19]
Data set 3	ONT reads of <i>E. herrerae</i> genomic DNA	DNA sequence (.fastq)	NCBI: https://identifiers.org/ncbi/insdc.sra:SRX21711620 [20]
Data set 4	<i>E. rubra</i> whole genome sequence	DNA sequence (.fasta)	NCBI: https://identifiers.org/ncbi/insdc.gca:GCA_033065855.1 [21]
Data set 5	Assembly metrics for <i>E. rubra</i> whole genome sequence	Spreadsheet (xlsx)	Figshare:(https://doi.org/10.6084/m9.figshare.24263788.v1) [22]
Data set 6	Annotation of <i>E. rubra</i> whole genome sequence	General Feature Format (.gff)	NCBI: https://identifiers.org/ncbi/insdc.gca:GCA_033065855.1 [21]
Data set 7	BUSCO summary statistics for <i>E. rubra</i> whole genome sequence and annotated proteins	Portable Network Graphic (.PNG)	Figshare:(https://doi.org/10.6084/m9.figshare.24265801.v1) [23]
Data set 8	<i>E. herrerae</i> whole genome sequence	DNA sequence (.fasta)	NCBI: https://identifiers.org/ncbi/insdc.gca:GCA_033070095.1 [24]
Data set 9	Assembly metrics for <i>E. herrerae</i> whole genome sequence	Spreadsheet (xlsx)	Figshare:(https://doi.org/10.6084/m9.figshare.24263785.v1) [25]
Data set 10	Annotation of <i>E. herrerae</i> whole genome sequence	General Feature Format (.gff)	NCBI: https://identifiers.org/ncbi/insdc.gca:GCA_033070095.1 [24]
Data set 11	BUSCO summary statistics for <i>E. herrerae</i> whole genome sequence and annotated proteins	Portable Network Graphic (.PNG)	Figshare:(https://doi.org/10.6084/m9.figshare.24265804.v1) [26]

Further, groups of closely related *Escallonia* species have diversified independently along elevational gradients in the tropical Andes, Southern Brazil, and the temperate Andes, suggesting that repeated ecological divergence may play an important role in *Escallonia* speciation [10]. Thus, *Escallonia* is emerging as a notable system to uncover the ecological and evolutionary processes underpinning tropical plant adaptation, speciation, and the nature of plant species [12]. To begin investigating the genomic substrate and biological processes underlying the radiations in *Escallonia* and Escalloniales, we hereby report the draft genomes of two *Escallonia* species. These data will also be relevant for broader comparative genomics studies across flowering plants.

Data description

Methodology - Leaf tissues from a single *Escallonia rubra* plant and an *Escallonia herrerae* plant cultivated at the University of California Botanical Garden at Berkeley (Voucher numbers: UCBG92.1500 *E. rubra*, UCBG64.0493 *E. herrerae*) were used for genomic DNA extraction and sequencing (Table 1; Data File 1). For *E. rubra*, isolated DNA was prepared following the Nextera XT DNA Library Prep Kit guideline and sequenced on an Illumina HiSeq 4000 system to generate 100-bp paired-end WGS reads (Table 1; Data Set 1; 376 million paired-end reads). In addition, we sequenced high-molecular-weight genomic DNA for both *E. rubra* and *E. herrerae* using the Oxford Nanopore Technology (ONT) PromethION 24 A series and the LSK114 ligation prep kit and R10.4.1 flow cells to generate approximately 140 Gb of sequence data (Table 1; Data Sets 2 and 3); 3.4 and 12 million sequence reads with an average read length of 9.4 and 9.1 Kb (approximately 31 and 111 Gb of sequence data), for *E. rubra* and *E. herrerae*, respectively. We used the Canu genome assembler [13] to generate contigs with ONT data. These were then polished (for *E. rubra*) using WGS sequences through NextPolish [14] and deduplicated using Purge Haplotigs [15].

Genome descriptions

Escallonia rubra - The *Escallonia rubra* genome assembly (Table 1, Data Set 4) consists of 3,233 contigs (N50=285 kb) with a total length of 566 Mb (Table 1, Data Set 5). The genome annotation includes 31,028 gene models supported by transcriptome and protein sequences and/or the presence of Pfam domains (Table 1; Data Set 6). BUSCO (Benchmarking Universal Single-Copy Orthologs) analyses based on conserved single-copy eudicot genes [16] indicate completeness levels of 97.8% for the genome sequence and 87.8% for the genome annotation (Table 1; Data Set 7).

Escallonia herrerae - The *Escallonia herrerae* genome assembly (Table 1, Data Set 8) consists of 5,760 contigs

(N50=317 kb) with a total length of 944 Mb (Table 1, Data Set 9). The genome annotation includes 47,905 gene models supported by transcriptome and protein sequences and/or the presence of Pfam domains (Table 1, Data Set 10). BUSCO analyses, relying on conserved single-copy eudicot genes [16], indicate completeness levels of 97.8% for the genome sequence and 87.8% for the genome annotation (Table 1, Data Set 11).

Limitations

The base chromosome number of *Escallonia* is $n=12$ [27], but our assemblies consist of 3,233 and 5,760 contigs for *E. rubra* and *E. herrerae*, respectively. As such, additional genome assembly and sequencing technologies, such as Hi-C, are needed to generate chromosome-level assemblies suitable for chromosome-scale comparative genomics.

Abbreviations

ONT	Oxford Nanopore Technology
BUSCO	Benchmarking Universal Single-Copy Orthologs
WGS	Whole Genome Shotgun

Acknowledgements

Holly Forbes, Curator of the University of California Botanical Garden at Berkeley, kindly provided access to plant tissue.

Author contributions

ASC, DES, PSS, and FZ conceived the project. IGA collected tissues. CD extracted DNA and prepared the library. ASC analyzed the data and produced the annotated genome assemblies. FZ wrote the initial draft of the manuscript. All authors read, revised, and approved the final manuscript.

Funding

This work was supported by startup funds from UCLA to FZ. The funding body played no role in the design of the study or the collection, analysis, and interpretation of data, or in writing the manuscript.

Data Availability

The data described in this Data Note can be freely and openly accessed at NCBI under accession number PRJNA1014744. Please see Table 1 and references [17–25] for details and links to the data. Detailed methodology is available on the Figshare repository [17].

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 4 November 2023 / Accepted: 12 December 2023

Published online: 02 January 2024

References

1. Stevens P. Angiosperm Phylogeny Website. Version 14, July 2017 [and more or less continuously updated since]. 2001. <http://www.mobot.org/MOBOT/research/APweb/>.

2. Beaulieu JM, O'Meara BC. Can we build it? Yes we can, but should we use it? Assessing the quality and value of a very large phylogeny of campanulid angiosperms. *Am J Bot*. 2018;105:417–32.
3. Tank DC, Donoghue MJ. Phylogeny and phylogenetic nomenclature of the Campanulidae based on an expanded sample of genes and Taxa. *Syst Bot*. 2010;35:425–41.
4. Li H-T, Yi T-S, Gao L-M, Ma P-F, Zhang T, Yang J-B, et al. Origin of angiosperms and the puzzle of the Jurassic gap. *Nat Plants*. 2019;5:461–70.
5. Stull GW, Soltis PS, Soltis DE, Gitzendanner MA, Smith SA. Nuclear phylogenomic analyses of asterids conflict with plastome trees and support novel relationships among major lineages. *Am J Bot*. 2020;107:790–805.
6. Zhang C, Zhang T, Luebert F, Xiang Y, Huang C-H, Hu Y, et al. Asterid Phylogenomics/Phylotranscriptomics Uncover Morphological Evolutionary Histories and support phylogenetic Placement for numerous whole-genome duplications. *Mol Biol Evol*. 2020;37:3188–210.
7. Baker WJ, Bailey P, Barber V, Barker A, Bellot S, Bishop D, et al. A Comprehensive Phylogenomic platform for exploring the Angiosperm Tree of Life. *Syst Biol*. 2022;71:301–19.
8. Beaulieu JM, Donoghue MJ. Fruit evolution and diversification in campanulid angiosperms. *Evolution*. 2013;67:3132–44.
9. Beaulieu JM, Tank DC, Donoghue MJ. A Southern Hemisphere origin for campanulid angiosperms, with traces of the break-up of Gondwana. *BMC Evol Biol*. 2013;13:80.
10. Zapata F. A multilocus phylogenetic analysis of Escallonia (Escalloniaceae): diversification in montane South America. *Am J Bot*. 2013;100:526–45.
11. Sede SM, Dürnhöfer SI, Morello S, Zapata F. Phylogenetics of Escallonia (Escalloniaceae) based on plastid DNA sequence data. *Bot J Linn Soc*. 2013;173:442–51.
12. Jacobs SJ, Grundler MC, Henriquez CL, Zapata F. An integrative genomic and phenomic analysis to investigate the nature of plant species in Escallonia (Escalloniaceae). *Sci Rep*. 2021;11:24013.
13. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res*. 2017;gr.215087.116.
14. Hu J, Fan J, Sun Z, Liu S. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics*. 2020;36:2253–5.
15. Roach MJ, Schmidt SA, Borneman AR. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics*. 2018;19:460.
16. Seppey M, Manni M, Zdobnov EM. BUSCO: Assessing Genome Assembly and Annotation Completeness. In: Kollmar M, editor. *Gene Prediction: methods and protocols*. New York, NY: Springer; 2019. pp. 227–45.
17. Chanderbali A, Dervinis C, Anghel I, Soltis DE, Soltis PS, Zapata F. Materials and Methods.docx. <https://doi.org/10.6084/m9.figshare.24263800.v2>. 2023.
18. Illumina sequencing of Escallonia rubra - SRA. - NCBI. <https://identifiers.org/ncbi/insdc.sra:SRX21711620>. 2023.
19. Nanopore sequencing of Escallonia rubra - SRA. - NCBI. <https://identifiers.org/ncbi/insdc.sra:SRX21711620>. 2023.
20. Nanopore sequencing of Escallonia herrerae - SRA. - NCBI. <https://identifiers.org/ncbi/insdc.sra:SRX21711620>. 2023.
21. Escallonia rubra genome assembly ASM3306585v1. NCBI. https://identifiers.org/ncbi/insdc.gca:GCA_03306585.1. 2023.
22. Chanderbali A, Zapata F, Dervinis C, Anghel I, Soltis DE, Soltis PS. Assembly metrics for E. Rubra whole genome sequence. 2023. <https://doi.org/10.6084/m9.figshare.24263788.v1>.
23. BUSCO summary statistics for E. rubra whole genome sequence and annotated proteins. figshare. 2023. <https://doi.org/10.6084/m9.figshare.24265801.v1>.
24. Escallonia herrerae genome assembly ASM3307009v1. NCBI. https://identifiers.org/ncbi/insdc.gca:GCA_03307009.1. 2023.
25. Chanderbali A, Dervinis C, Anghel I, Soltis DE, Soltis PS, Zapata F. Assembly metrics for E. Herrerae whole genome sequence. 2023. <https://doi.org/10.6084/m9.figshare.24263785.v1>.
26. BUSCO summary statistics for E. herrerae whole genome sequence and annotated proteins. figshare. 2023. <https://doi.org/10.6084/m9.figshare.24265804.v1>.

27. Hanson L, Brown RL, Boyd A, Johnson MAT, Bennett MD. First nuclear DNA C-values for 28 angiosperm genera. *Ann Bot.* 2003;91:31–8.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.