BMC Research Notes

DATA NOTE

Open Access

Genome sequence of the Mediterranean red coral *Corallium rubrum*

Philippe Ganot[1,2*], Tobias Rausch[4], Markus Hsi-Yang Fritz[4], Didier Zoccola[2], Xin Wang[5], Manuel Aranda[5], Vladimir Benes[4], Denis Allemand[1,3] and Sylvie Tambutté[1,2]

Abstract

Objectives *Corallium rubrum,* the precious red coral, is an octocoral endemic to the western Mediterranean Sea. Like most octocorals, it produces tiny, calcified structures called sclerites. Uniquely, it also produces a completely calcified axial skeleton that is a bright red color. This combination of color and hardness has made the red coral prized for centuries, leading to extensive fishing and trade for use in jewelry. Understanding how it produces this red skeleton is thus a central question in economics, culture, and biology. To gain insights into this process, we sequenced the *C. rubrum* genome.

Data description Our *C. rubrum* genome assembly is 655 megabases (Mb) in size, distributed across 2910 scaffolds with a very low level of unknown nucleotides (0.95%). We used a pipeline based on the MaSuRCA hybrid assembler, combining long PacBio reads and short Illumina reads, followed by several steps to improve the assembly, including scaffolding, merging, and polishing. This represents the third published genome of an octocoral and the first within the order Scleralcyonacea.

Keywords Precious corals, Octocoral, Scleralcyonacea, MaSuRCA, Hybrid Illumina PacBio, Biomineralization, Calcification

Objective

The precious red coral, *Corallium rubrum*, holds a special place in history. It was the first "coral" to be identified, lending its name to all other corals. Its use dates back to Neolithic times, valued not only for its beauty in jewelry but also for its symbolic, cultural, and even medicinal significance [1, 2].

C. rubrum plays a key role in the well-known Mediterranean coralligenous ecosystem, a biodiversity hotspot. This slow-growing species (millimeters per year) thrives at various depths, ranging from 5 to 800 m, with populations structured according to depth zone [3, 4]. However, overfishing and global warming have led to the decrease of its population, prompting the IUCN to classify it as Endangered [5].

Red corals (family Coralliidae) possess a unique trait: they produce sclerites (tiny calcified elements) like other octocorals, but also a fully calcified axial skeleton [6]. This skeleton's color can range from white (e.g., *Corallium konojoi*) to bright red, with *C. rubrum* being particularly prized for its intense hue [7, 8]. However, the genetic mechanisms controlling both calcification and color remain to be elucidated.

*Correspondence:
Philippe Ganot
pganot@centrescientifique.mc
[1] Research Unit on the Biology of Precious Corals CSM—CHANEL, 8 Quai Antoine 1er, Monaco, Principality of Monaco
[2] Department of Marine Biology, CSM, Centre Scientifique de Monaco, 8 Quai Antoine 1er, Monaco, Principality of Monaco
[3] CSM, Centre Scientifique de Monaco, 8 Quai Antoine 1er, Monaco, Principality of Monaco
[4] EMBL, Genomics Core Facility, Meyerhofstraße 1, 69117 Heidelberg, Germany
[5] Marine Science Program, BESE, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Kingdom of Saudi Arabia

Phylogenetically, the subphylum Anthozoa (phylum Cnidaria) encompasses two classes: Hexacorallia (sea anemones, reef-building corals) and Octocorallia (over 3500 species including soft corals, gorgonians, sea pens) [9]. The Hexacorallia phylogeny is now well resolved thanks to extensive genetic sequencing [10–15]. Conversely, despite recent revisionary systematics of Octocorallia using targeted capture of specific genetic sequences [16], only two octocorallian genomes [17, 18] (and a few transcriptomes) have been published to date.

Understanding the *C. rubrum* genome is crucial for further investigations into its biology (e.g., skeleton formation and color, genetic diversity), ecology, and evolutionary history. This knowledge is essential for developing effective conservation strategies to safeguard these vital marine organisms in the face of ongoing environmental challenges [19].

Data description

We report here the complete genome sequence of the red coral, *Corallium rubrum*, consisting of 655.3 megabases (Mb) assembled into 2910 scaffolds [20] (Table 1). The genomic DNA and total RNA were isolated from a single colony collected near Banyuls-sur-Mer, France [21].

Building the genome assembly

The genomic DNA was sequenced with 2 technologies: Illumina short reads (Paired-End and Mate-Pair) and PacBio long reads [27]. To construct the genome assembly, we initially attempted separate assemblies using only PacBio long reads (Canu-1.6 assembler [29]) or only Illumina short reads (MaSuRCA-3.2.2 assembler [30]) [22]. These initial assemblies had low N50 values (approximatively 36 kb and 21 kb, respectively), indicating a high degree of fragmentation.

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_file1	Figure 1: The sequenced colony of *Corallium rubrum*	Image (png)	Figshare: https://doi.org/10.6084/m9.figshare.25592979 [21]
Data_file2	Figure 2: Genome Assembly pipeline	Image (png)	Figshare: https://doi.org/10.6084/m9.figshare.25593036 [22]
Data_file3	Figure 3: Scaffolds length distribution	Image (png)	Figshare: https://doi.org/10.6084/m9.figshare.25592976 [23]
Data_file4	Supplementary methods for the assembly	Word document	Figshare: https://doi.org/10.6084/m9.figshare.25592973 [24]
Data_file5	MsGenomeScripts.txt	Text file (.txt)	Figshare: https://doi.org/10.6084/m9.figshare.25922611 [25]
Data_file6	Libraries_and_Metrics	Excel file (.xlsx)	Figshare: https://doi.org/10.6084/m9.figshare.25922689 [26]
Data_set1	Sequence Read Archive (SRA) Study Accession: SRP510411	fastq files (.fq.gz)	NCBI Sequence Read Archive (https://urldefense.proofpoint.com/v2/url?u=https-3A__identifiers.org_ncbi_insdc.sra-3ASRP510411&d=DwIFaQ&c=euGZstcaTDllvimEN8b7jXrwqOf-v5A_CdpgnVfiiMM&r=9y3wEglwtaFH048SAFCckQTUOmI7C5fjKXHTV_-EItc&m=NnuKU-1LNa7SgtjBH3sgc1rCIkiL7vn5hxmKZo4mgV5y-bEQrJUHF41qGiaaTEG3&s=84E2kBSXCVQluOtbyuxQxPilyTOXg8DAeIugOhaNGXA&e=) [27]
Data_set2	DNA-RNA STAR alignment	BAM file (.bam)	Dryad (https://urldefense.proofpoint.com/v2/url?u=https-3A__doi.org_10.5061_dryad.ksn02v7d6&d=DwIFaQ&c=euGZstcaTDllvimEN8b7jXrwqOf-v5A_CdpgnVfiiMM&r=9y3wEglwtaFH048SAFCckQTUOmI7C5fjKXHTV_-EItc&m=NnuKU-1LNa7SgtjBH3sgc1rCIkiL7vn5hxmKZo4mgV5y-bEQrJUHF41qGiaaTEG3&s=nkyRKBvM8VffFtIXa-U4EdRZlpr1gvpDUzchLcC1LOU&e=) [28]
Data_set3	The *Corallium rubrum* genome sequence (V152) CSM Bioproject: PRJCA026883 Biosample: SAMC3811372 Accession: GWHESWX00000000.1	Fasta file	Genome Sequence Archive (https://urldefense.proofpoint.com/v2/url?u=https-3A__ngdc.cncb.ac.cn_gwh_Assembly_84901_show&d=DwIFaQ&c=euGZstcaTDllvimEN8b7jXrwqOf-v5A_CdpgnVfiiMM&r=9y3wEglwtaFH048SAFCckQTUOmI7C5fjKXHTV_-EItc&m=NnuKU-1LNa7SgtjBH3sgc1rCIkiL7vn5hxmKZo4mgV5y-bEQrJUHF41qGiaaTEG3&s=zdeFrqGcfNgOEFLlkL2XVBjUQuTsVs82wQ6-zsYdeHo&e=) [20]

To overcome this limitation, we employed a hybrid assembly approach using both PacBio long reads and Illumina short reads with the MaSuRCA hybrid assembler (genome version 1.5.2). This approach yielded a significantly improved assembly with a larger size (686.6 Mb) distributed across 9,364 scaffolds. Importantly, the percentage of unknown nucleotides ("Ns") significantly decreased from 37.6% to 1.56% between the Illumina-only and the hybrid mode of MaSuRCA [22–25].

Scaffolding, merging, and polishing

Following the initial assembly, we employed a three-phase process to further refine and improve the genome sequence. The first phase involved scaffolding, which utilized information from mRNA sequences, RNA libraries, and a genome mate-pair library (using L_RNA scaffolder [31], Rascaf [32], and Sspace [33], respectively). This process successfully reduced the number of scaffolds by nearly half [23, 26].

In the second phase, we merged the V1.5.2 scaffolded assembly with the previous Illumina-only and PacBio-only assemblies using Metassembler, an algorithm that merges multiple genome assemblies into a single superior sequence [34].

The final phase focused on polishing the assembly using Pilon [35] and GapCloser [36]. These tools filled remaining gaps and polished the sequence, resulting in the final high-quality genome sequence (less than 1% Ns) of *Corallium rubrum*.

Data associated with each step of the pipeline is available in Data_Files2–5 [22–26].

Conclusions

A 2022 study revised the octocorallian phylogeny, establishing two orders: Scleralcyonacea and Malacalcyonacea [16]. However, only two complete octocorallian genomes have been published to date, both belonging to the order Malacalcyonacea: *Dendronephthya gigantea* (286 Mb) [18] and *Xenia sp.* (222 Mb) [17]. We present the first published genome sequence of a member of the order Scleralcyonacea, *Corallium rubrum*. The *C. rubrum* genome is estimated to be 2–3 times larger than previously sequenced octocorallian genomes (but similar in size to other scleractinian corals). Remarkably, transcriptome data suggests that it encodes for around 30,000 protein-coding genes [37, 38], a similar number for all three species. This indicates that the larger genome size of *C. rubrum* is likely attributed to an abundance of non-coding DNA, especially repetitive elements. These repetitive elements can significantly hinder De Bruijn graph-based genome assemblers reliant on short Illumina reads for assembly [39]. In support of this, we employed various assemblers and only achieved satisfactory metrics when we used MaSuRCA and a combination of short and long reads.

Limitations

The non-annotated *Corallium* rubrum genome

Our current *Corallium rubrum* genome sequence data has a key limitation: lack of predicted and annotated genes. Transcribed genes need to be identified downstream using either prediction algorithm and/or transcriptomic data. The first *C. rubrum* transcriptome was published in 2015 with another proposed in 2016 [37, 38]. In addition, to facilitate gene identification, we provide a link for the alignment (STAR [40]) of RNA reads against the *C. rubrum* genome (dataset 2, [28]), from which expressed genes can be extrapolated.

Haploid genome

Coral genomes are diploid and often highly heterozygous, in addition to sometimes display high level of tandemly duplicated genes and repeated sequences [13, 15, 41]. While we present a haploid representation of the genome, some scaffold portions might contain biased sequences resulting from uncaptured allelic duplications or misassembled repeats. Unfortunately, our current sequencing and assembling technology could not resolve a fully diploid assembly.

Forthcoming improvements

While advancements in sequencing technology, particularly long-read sequencing and assembling, promise improved accuracy and capability to handle repetitive regions [42], we are confident that our current *C. rubrum* genome assembly remains an accurate and valuable resource for various research areas, including genomics, metagenomics, and phylogenetic studies.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13104-024-07006-0.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Acknowledgements

We thank the foundation Paul HAMEL, CHANEL and the Government of the Principality of Monaco for financial support. We are very grateful to Pascal Romans (Observatoire Océanologique de Banyuls—Sorbonne Université / CNRS), for sample collection of the Corallium rubrum colony we sequenced.

Author contributions

Conceptualization and Methodology: PG, TR, MHF, VB, DA, ST; Biological Handling: PG, DZ; Bioinformatic Training of PG: TR, MHF, VB; Data Curation, sequence assembly and analysis: PG, TR, MHF, XW; Writing/Editing of the

manuscript: PG, TR, MA, DA, ST. All authors confirmed the final version of the manuscript.

Funding
We thank the foundation Paul HAMEL, CHANEL and the Government of the Principality of Monaco for financial support.

Availability of data and materials
The data described in this Data note can be freely and openly accessed: for Sequence Reads Archives (SRAs) on GeneBank under BioProject accession PRJNA1117673 (Study Accession: SRP510411; SRAs: SRR29212214-SRR29212226) [27]; for genome sequence on Genome Warehouse under BioProject accession PRJCA026883 (Biosample: SAMC3811372; Accession: GWHESWX00000000.1) [20]; for Genome/RNA_reads alignment on Dryad [28]; and for supplementary text, scripts and figures on Figshare [21–26]. Please see Table 1 for details and links to the data.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 14 June 2024 Accepted: 15 November 2024
Published online: 21 December 2024

References

1. Price LL, Narchi NE. Ethnobiology of *Corallium rubrum*: protection, healing, medicine, and magic. In: Narchi N, Price LL, editors. Ethnobiology of corals and coral reefs. Cham: Springer International Publishing; 2015. p. 73–86.
2. Moradi Z. The role of coral in art and architecture: an overview. Int J Aquat Biol. 2016;4(2):125–42.
3. Costantini F, Taviani M, Remia A, Pintus E, Schembri PJ, Abbiati M. Deep-water *Corallium rubrum* (L. 1758) from the Mediterranean Sea: preliminary genetic characterisation. Mar Ecol. 2010;31(2):261–9.
4. Pratlong M, Haguenauer A, Brener K, Mitta G, Toulza E, Garrabou J, et al. Separate the wheat from the chaff: genomic scan for local adaptation in the red coral *Corallium rubrum*. Peer Commun J. 2021;26(1): e31.
5. IUCN red_list *Corallium rubrum*. Available from: https://www.iucnredlist.org/species/50013405/110609252.
6. Bayer FM, Grasshoff M, Verseveldt J, editors. Illustrated trilingual glossary of morphological and anatomical terms applied to octocorallia [Internet]. Leiden: Brill; 1983. p. 75. Available from: https://repository.si.edu/handle/10088/6237.
7. Cvejic J, Tambutté S, Lotto S, Mikov M, Slacanin I, Allemand D. Determination of canthaxanthin in the red coral (*Corallium rubrum*) from Marseille by HPLC combined with UV and MS detection. Mar Biol. 2007;152(4):855–62.
8. Karampelas S, Fritsch E, Rondeau B, Andouche A, Métivier B. Identification of the endangered pink-to-red *Stylaster* corals by raman spectroscopy. Gems Gemol. 2009;45(1):48–52.
9. McFadden CS, Quattrini AM, Brugler MR, Cowman PF, Dueñas LF, Kitahara MV, et al. Phylogenomics, origin, and diversification of anthozoans (*Phylum cnidaria*). Syst Biol. 2021;70(4):635–47.
10. Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, et al. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science. 2007;317(5834):86–94.
11. Wang X, Liew YJ, Li Y, Zoccola D, Tambutte S, Aranda M. Draft genomes of the corallimorpharians *Amplexidiscus fenestrafer* and Discosoma sp. Mol Ecol Resour. 2017;17(6):e187–95.
12. Voolstra CR, Li Y, Liew YJ, Baumgarten S, Zoccola D, Flot JF, et al. Comparative analysis of the genomes of *Stylophora pistillata* and *Acropora digitifera* provides evidence for extensive differences between species of corals. Sci Rep. 2017;7(1):17583.
13. Buitrago-López C, Mariappan KG, Cárdenas A, Gegner HM, Voolstra CR. The genome of the cauliflower coral *Pocillopora verrucosa*. Genome Biol Evol. 2020;12(10):1911–7.
14. Shinzato C, Shoguchi E, Kawashima T, Hamada M, Hisata K, Tanaka M, et al. Using the *Acropora digitifera* genome to understand coral responses to environmental change. Nature. 2011;476(7360):320–3.
15. Noel B, Denoeud F, Rouan A, Buitrago-López C, Capasso L, Poulain J, et al. Pervasive tandem duplications and convergent evolution shape coral genomes. Genome Biol. 2023;24(1):123.
16. McFadden CS, Van Ofwegen LP, Quattrini AM. Revisionary systematics of Octocorallia (Cnidaria: Anthozoa) guided by phylogenomics. BSSB. 2022;1(3). Available from: https://ssbbulletin.org/index.php/bssb/article/view/8735.
17. Hu M, Zheng X, Fan CM, Zheng Y. Lineage dynamics of the endosymbiotic cell type in the soft coral Xenia. Nature. 2020;582(7813):534–8.
18. Jeon Y, Park SG, Lee N, Weber JA, Kim HS, Hwang SJ, et al. The draft genome of an octocoral, *Dendronephthya gigantea*. Genome Biol Evol. 2019;11(3):949–53.
19. Iwasaki N, Bavestrello G, Bo M, Hasegawa H, Tamenori Y. Editorial: Scientific approaches for the conservation and sustainable use of precious coral resources. Front Mar Sci. 2023;13(10):1285833.
20. Data_set3: *Corallium rubrum* genome sequence CSM. Available from: https://urldefense.proofpoint.com/v2/url?u=https-3A__ngdc.cncb.ac.cn_gwh_Assembly_84901_show&d=DwIFaQ&c=euGZstcaTDllvimEN8b7jXrwqOf-v5A_CdpgnVfiiMM&r=9y3wEgIwtaFH048SAFCckQTUOmI7C5fjKXHTV_-EItc&m=NnuKU-1LNa7SgtjBH3sgc1rClkiL7vn5hxmKZo4mgV5y-bEQrJUHF41qGiaaTEG3&s=zdeFrqGcfNgOEFLlkL2XVBjUQuTsVs82wQ6-zsYdeHo&e=.
21. Data_File1: Figure 1, The sequenced colony of *Corallium rubrum*. Available from: https://doi.org/10.6084/m9.figshare.25592979.
22. Data_File2: Figure 2, Genome Assembly pipeline. Available from: https://doi.org/10.6084/m9.figshare.25593036.
23. Data_File3: Figure 3, Scaffolds length distribution. Available from: https://doi.org/10.6084/m9.figshare.25592976.
24. Data_File4: Supplementary methods for the assembly. Available from: https://doi.org/10.6084/m9.figshare.25592973.
25. Data_File5: Scripts for the Genome assembly. Available from: https://doi.org/10.6084/m9.figshare.25922611.
26. Data_File6: Table, Libraries_and_Metrics. Available from: https://doi.org/10.6084/m9.figshare.25922689.
27. Data_set1: NCBI Sequence Read Archive [Internet]. Available from: https://urldefense.proofpoint.com/v2/url?u=https-3A__identifiers.org_ncbi_insdc.sra-3ASRP510411&d=DwIFaQ&c=euGZstcaTDllvimEN8b7jXrwqOf-v5A_CdpgnVfiiMM&r=9y3wEgIwtaFH048SAFCckQTUOmI7C5fjKXHTV_-EItc&m=NnuKU-1LNa7SgtjBH3sgc1rClkiL7vn5hxmKZo4mgV5y-bEQrJUHF41qGiaaTEG3&s=84E2kBSXCVQluOtbyuxQxPilyTOXg8DAelugOhaNGXA&e=.
28. Data_set2: DNA-RNA STAR alignment. Available from: https://urldefense.proofpoint.com/v2/url?u=https-3A__doi.org_10.5061_dryad.ksn02v7d6&d=DwIFaQ&c=euGZstcaTDllvimEN8b7jXrwqOf-v5A_CdpgnVfiiMM&r=9y3wEgIwtaFH048SAFCckQTUOmI7C5fjKXHTV_-EItc&m=NnuKU-1LNa7SgtjBH3sgc1rClkiL7vn5hxmKZo4mgV5y-bEQrJUHF41qGiaaTEG3&s=nkyRKBvM8VffFtIXa-U4EdRZlpr1gvpDUzchLcC1LOU&e=.
29. 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;27(5):722–36.
30. Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. The MaSuRCA genome assembler. Bioinformatics. 2013;29(21):2669–77.
31. Xue W, Li JT, Zhu YP, Hou GY, Kong XF, Kuang YY, et al. L_RNA_scaffolder: scaffolding genomes with transcripts. BMC Genomics. 2013;14:604.
32. Song L, Shankar DS, Florea L. Rascaf: improving genome assembly with rna sequencing data. Plant Genome. 2016;9(3):plantgenome2016.03.0027.
33. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding preassembled contigs using SSPACE. Bioinformatics. 2011;27(4):578–9.
34. Wences AH, Schatz MC. Metassembler: merging and optimizing de novo genome assemblies. Genome Biol. 2015;16(1):207.

35. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE. 2014;9(11):e112963.
36. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaSci. 2012;1(1):18.
37. Le Goff C, Ganot P, Zoccola D, Caminiti-Segonds N, Allemand D, Tambutté S. Carbonic anhydrases in cnidarians: novel perspectives from the Octocorallian *Corallium rubrum*. PLoS ONE. 2016;11(8):e0160368.
38. Pratlong M, Haguenauer A, Chabrol O, Klopp C, Pontarotti P, Aurelle D. The red coral (*Corallium rubrum*) transcriptome: a new resource for population genetics and local adaptation studies. Mol Ecol Resour. 2015;15(5):1205–15.
39. Tørresen OK, Star B, Mier P, Andrade-Navarro MA, Bateman A, Jarnot P, et al. Tandem repeats lead to sequence assembly errors and impose multi-level challenges for genome and protein databases. Nucleic Acids Res. 2019;47(21):10994–1006.
40. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21.
41. Shinzato C, Khalturin K, Inoue J, Zayasu Y, Kanda M, Kawamitsu M, et al. Eighteen coral genomes reveal the evolutionary origin of *Acropora* strategies to accommodate environmental changes. Mol Biol Evol. 2021;38(1):16–30.
42. De Coster W, Weissensteiner MH, Sedlazeck FJ. Towards population-scale long-read sequencing. Nat Rev Genet. 2021;22(9):572–87.

Publisher's Note