DATA NOTE

# The chromosomal genome sequence of the sponge *Crambe crambe* (Schmidt, 1862) and its associated microbial metagenome sequences

[version 1; peer review: 2 approved, 1 approved with reservations]

Manuel Maldonado[1], Lucia Pita[2,3], Ute Hentschel[4], Dirk Erpenbeck[5], Graeme Oatley[6], Elizabeth Sinclair[6], Eerik Aunin[6], Noah Gettle[6], Camilla Santos[6], Michael Paulini[6], Haoyu Niu[6], Victoria McKenna[6], Rebecca O'Brien[6],
Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory Team,
Wellcome Sanger Institute Scientific Operations: Sequencing Operations,
Wellcome Sanger Institute Tree of Life Core Informatics Team,
EBI Aquatic Symbiosis Genomics Data Portal Team,
Aquatic Symbiosis Genomics Project Leadership

[1]Center for Advanced Studies of Blanes (CEAB-CSIC), Girona, Spain
[2]Institute of Marine Sciences – CSIC, Barcelona, Spain
[3]Integrated Marine Ecology group, Institute of Marine Research IIM-CSIC, Vigo, Spain
[4]GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany
[5]Ludwig-Maximilians University of Munich, Munich, Germany
[6]Tree of Life, Wellcome Sanger Institute, Hinxton, England, UK

**v1** **First published:** 23 May 2025, **10**:275
https://doi.org/10.12688/wellcomeopenres.24154.1
**Latest published:** 23 May 2025, **10**:275
https://doi.org/10.12688/wellcomeopenres.24154.1

## Abstract

We present a genome assembly from an individual *Crambe crambe* (Porifera; Demospongiae; Poecilosclerida; Crambeidae). The host genome sequence is 143.20 megabases in span. Most of the assembly is scaffolded into 18 chromosomal pseudomolecules. The mitochondrial genome has also been assembled and is 19.53 kilobases in length. Several symbiotic prokaryotic genomes were assembled as MAGs, including two relevant sponge symbionts, the *Candidatus* Beroebacter blanensis/*AqS2* clade (Tethybacterales, Gammaproteobacteria) of LMA sponges, and the widely distributed archaeal *Nitrosopumilus* sp. clade.

**Open Peer Review**

**Approval Status** ✓ ? ✓

| | 1 | 2 | 3 |
|---|---|---|---|
| **version 1** 23 May 2025 | ✓ view | ? view | ✓ view |

1. **Shikai Liu**, Ocean University of China, Qingdao, China

2. **Weizhi Song**, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong
**Sihan Li**, The Hong Kong University of

**Keywords**
Crambe crambe, marine sponge, genome sequence, chromosomal, Poecilosclerida

This article is included in the Tree of Life gateway.

Science and Technology Department of Ocean Science, Hong Kong, Hong Kong

3. **Paco Cardenas**, Uppsala University, Uppsala, Sweden

Any reports and responses or comments on the article can be found at the end of the article.

**Corresponding author:** Aquatic Symbiosis Genomics Project Leadership (Mark.Blaxter@sanger.ac.uk)

**Author roles: Maldonado M**: Investigation, Resources, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Pita L**: Resources; **Hentschel U**: Funding Acquisition, Project Administration, Writing – Review & Editing; **Erpenbeck D**: Investigation; **Oatley G**: Investigation; **Sinclair E**: Investigation; **Aunin E**: Formal Analysis; **Gettle N**: Formal Analysis; **Santos C**: Formal Analysis; **Paulini M**: Formal Analysis; **Niu H**: Project Administration; **McKenna V**: Project Administration; **O'Brien R**: Project Administration;

**Competing interests:** No competing interests were disclosed.

**Grant information:** This work was funded by the Gordon and Betty Moore Foundation through a grant (GBMF8897) to the Wellcome Sanger Institute to support the Aquatic Symbiosis Genomics Project, and by Wellcome through core funding to the Wellcome Sanger Institute (220540). Collection and preservation were funded by a grant of the Spanish Government (PID2019-108627RB-I00) to MM.
*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*



**How to cite this article:** Maldonado M, Pita L, Hentschel U *et al.* **The chromosomal genome sequence of the sponge *Crambe crambe* (Schmidt, 1862) and its associated microbial metagenome sequences [version 1; peer review: 2 approved, 1 approved with reservations]** Wellcome Open Research 2025, **10**:275 https://doi.org/10.12688/wellcomeopenres.24154.1

**First published:** 23 May 2025, **10**:275 https://doi.org/10.12688/wellcomeopenres.24154.1

## Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Porifera; Demospongiae; Heteroscleromorpha; Poecilosclerida; Crambeidae; Crambe (in: sponges) *Crambe*; *Crambe crambe* Vosmaer, 1880 (NCBI: txid3722).

## Background

*Crambe crambe* (Schmidt, 1862) is probably the most abundant sponge species in the sublittoral rocky bottoms of the Atlantic-Mediterranean region. It is a bright red encrusting sponge that grows at both well-lit and poorly lit sites, forming patches of up to 0.5 $m^2$ (Pansini & Pronzato, 1990; Turon *et al.*, 1998). As additional macroscopic clues for species identification, oscula and their radially converging excurrent channels are often visible on the sponge surface, which is slippery to the touch. The sponge grows not only on rocks, but also on barnacles and on the shells of the red oyster *Spondylus gaederopus*.

Due to its abundance, the species is ecologically important in many ways. For instance, its skeletal growth represents a substantial silicon sink for the sublittoral system (Maldonado *et al.*, 2005). The sponge also provides food and habitat for a variety of marine organisms, including recruitment habitat for juvenile ophiuroids (Turon *et al.*, 2000) and small benthic fish. *C. crambe* produces various bioactive compounds that interact chemically with many community members (Becerro *et al.*, 1994; Becerro *et al.*, 1997), some of which have potential pharmaceutical applications derived from their antibacterial, antifungal, and anti-tumour properties, among others (El-Demerdash *et al.*, 2018). Given its biotechnological potential, attempts have been made to farm the species (Padiglia *et al.*, 2018). Despite its abundance and ecological versatility (or perhaps because of it), the species is thought to be a surviving relict of the Jurassic oceans. This hypothesis is supported by the observation that the formation of all four spicule types is only possible at a silicate concentration ≥100 μM – concentrations which are likely to have occurred in Jurassic seas before the ecological expansion of diatoms (Maldonado *et al.*, 1999). Secondly, the biogeographic distribution of the genus *Crambe* shows a clear Tethyan pattern (Maldonado *et al.*, 2001).

Regarding the microbiome, the sponge is a species with low microbial abundance. While most of the few microbes occur in low abundance extracellularly in the mesohyl and around the skeletal spongin fibres, some of the microbes have been documented by electron microscopy to be contained within vesicles in the cytoplasm of bacteriocytes that appear to contain a single microbial species per cell (Carrier *et al.*, 2022; Maldonado, 2007). Gammaproteobacteria, ammonia-oxidising *Nitrosopumilus* sp. (Archaea) and a single taxon, *Candidatus* Beroebacter blanensis, dominate the microbial community. This latter symbiont clade appears to be vertically transmitted (Turon *et al.*, 2024). It was originally classified as *Betaproteobacteria* (Croué *et al.*, 2013), but was later identified as *Ca.* Beroebacter blanensis, belonging to a novel bacterial order, *Candidatus* (*Ca.*) Tethybacterales within the Gammaproteobacteria and consisting mainly of sponge symbionts (Taylor *et al.*, 2021). The well characterized symbiont "AqS2" of *Amphimedon queenslandica* is the nearest phylogenetic relative of the *B. blanensis* clade, which displays genome reduction and limited metabolic capabilities, likely reflecting an adaptation to a symbiotic lifestyle within the sponge host (Gauthier *et al.*, 2016).

The sexual condition of the species is hermaphroditism. It is worth noting that its spermatozoa are highly atypical within the phylum. They are very elongated and V-shaped, with the flagellum inserted in an antero-lateral position next to a true acrosome (Riesgo & Maldonado, 2009; Tripepi *et al.*, 1984). This general organisation of the spermatozoon, which closely resembles that of Phoronida spermatozoa, appears to be common in the order Poecilosclerida but not in other sponges. Fertilisation is internal, and embryos are incubated for several months, until they develop into bright red, non-tufted parenchymella larvae (Maldonado & Bergquist, 2002; Uriz *et al.*, 2001). In western Mediterranean populations, larval release extends from mid-July to mid-August, and larval production can be as high as 76 embryos per $cm^2$ of sponge tissue (Uriz *et al.*, 1998), which would explain the abundance of adults.

The sequencing of the whole-chromosome genome of *C. crambe* will facilitate in-depth understanding of the genomic basis of this species biology, as well as its ecology and evolution. This genome will be particularly useful for investigating the evolution of sexual strategies in Demospongiae, as well as for clarifying between-family relationships within the order Poecilosclerida. Together with the genome sequences of *C. crambe* microbial symbionts presented here, the novel data will enable targeted examination of the molecular basis of sponge silicate metabolism and skeleton formation, alkaloid metabolism, and sponge-microbe interactions in the role of carbon cycling, among other key questions in sponge symbiosis.

## Genome sequence report

The genome was sequenced from an adult *Crambe crambe* (Figure 1) collected from Blanes, Girona, Spain. A total of

**Figure 1. Photograph of the *Crambe crambe* (odCraCram1) specimen used for genome sequencing.**

**Table 1. Genome data for *Crambe crambe*, odCraCram1.1.**

| | | | |
|---|---|---|---|
| **Project accession data** | | | |
| Assembly identifier | odCraCram1.1 | | |
| Species | *Crambe crambe* | | |
| Specimen | odCraCram1 | | |
| NCBI taxonomy ID | 3722 | | |
| BioProject | PRJEB65618 | | |
| BioSample ID | Genome sequencing: SAMEA9361910<br>Hi-C scaffolding: SAMEA9361908 | | |
| Isolate information | odCraCram1: (genome and Hi-C sequencing) | | |
| **Assembly metrics** | | | |
| Consensus quality (QV) | 58.1 | | |
| BUSCO* | C:78.8%[S:78.0%,D:0.8%],F:9.4%,M:11.8%,n:954 | | |
| Percentage of assembly mapped to chromosomes | 98.69% | | |
| Organelles | Mitochondrial genome: 19.53 kb | | |
| **Sequencing information** | | | |
| **Platform** | **Run accession** | **Read count** | **Base count (Gb)** |
| **Hi-C Illumina NovaSeq 6000** | ERR12512721 | 1.13e+09 | 170.77 |
| **PacBio Revio** | ERR12015695 | 9.82e+06 | 67.94 |
| **Genome assembly** | | | |
| Assembly accession | GCA_963924555.1 | | |
| *Accession of alternate haplotype* | GCA_963924525.1 | | |
| Span (Mb) | 143.20 | | |
| Number of contigs | 178 | | |
| Contig N50 length (Mb) | 3.5 | | |
| Number of scaffolds | 124 | | |
| Scaffold N50 length (Mb) | 7.7 | | |
| Longest scaffold (Mb) | 9.77 | | |

* BUSCO scores based on the metazoa_odb10 BUSCO set using version 5.4.3. C = complete [S = single copy, D = duplicated], F = fragmented, M = missing, n = number of orthologues in comparison. A full set of BUSCO scores is available at https://blobtoolkit.genomehubs.org/view/Crambe_crambe/dataset/GCA_963924555.1/busco.

459-fold coverage in Pacific Biosciences single-molecule HiFi long reads was generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 62 missing joins or mis-joins and removed 18 haplotypic duplications, reducing the assembly length by 2.19% and the scaffold number by 29.78%, also decreasing the scaffold N50 by 0.31%.

The final assembly has a total length of 143.20 Mb in 124 sequence scaffolds with a scaffold N50 of 7.7 Mb (Table 1). The snail plot in Figure 2 provides a summary of the assembly statistics, while the distribution of assembly scaffolds on GC proportion and coverage is shown in Figure 3. The cumulative assembly plot in Figure 4 shows curves for subsets of scaffolds assigned to different phyla. Most (98.69%) of the assembly sequence was assigned to 18 chromosomal-level scaffolds. Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 5; Table 2). While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been

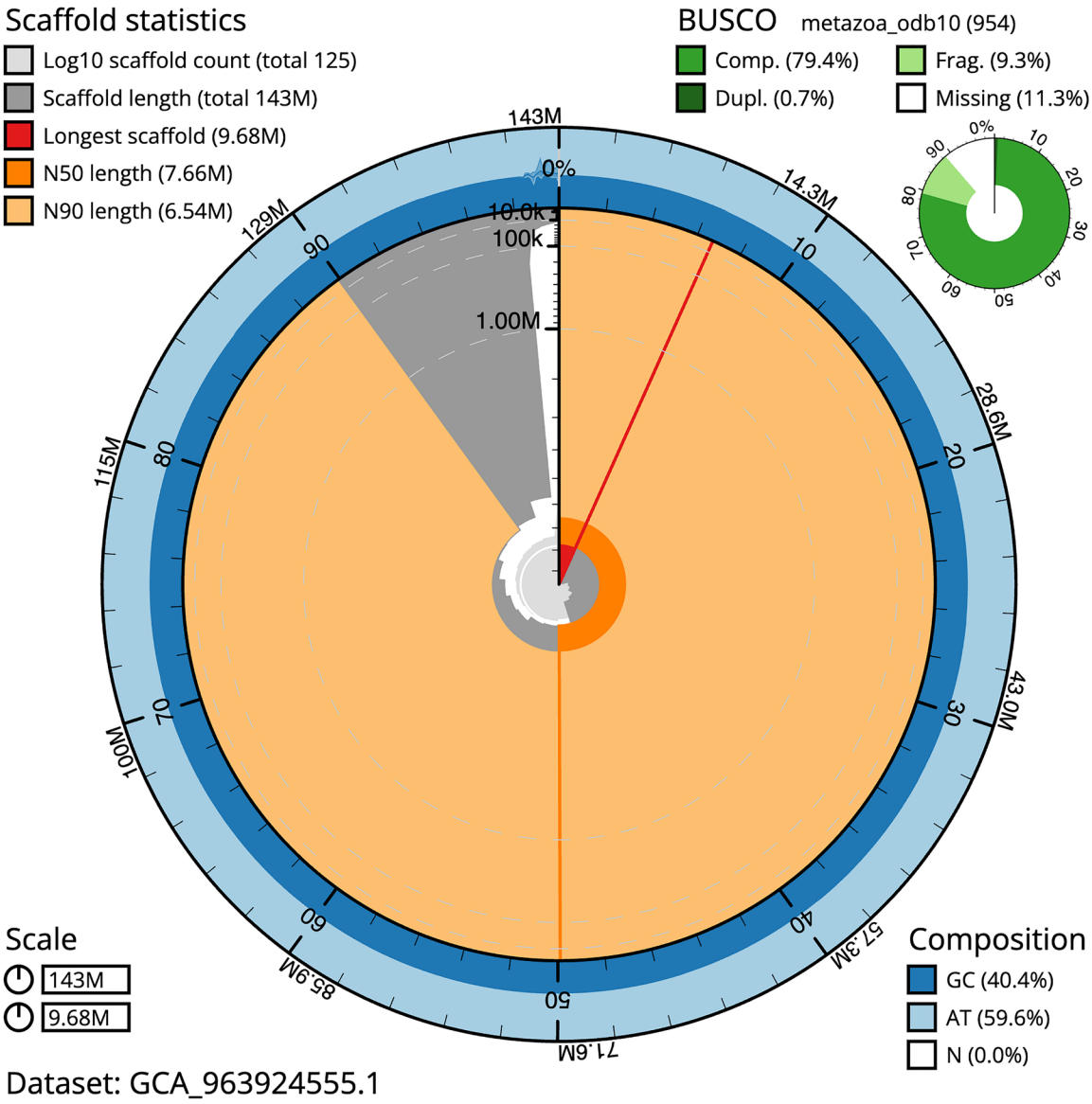


**Figure 2. Genome assembly of *Crambe crambe*, odCraCram1.1: metrics.** The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 143,197,480 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (9,683,886 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (7,656,483 and 6,535,638 bp), respectively. The pale grey spiral shows the cumulative scaffold count on a log scale with white scale lines showing successive orders of magnitude. The blue and pale-blue area around the outside of the plot shows the distribution of GC, AT and N percentages in the same bins as the inner plot. A summary of complete, fragmented, duplicated and missing BUSCO genes in the metazoa_odb10 set is shown in the top right. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Crambe_crambe/dataset/GCA_963924555.1/snail.

deposited. The mitochondrial genome was also assembled and can be found as a contig within the multifasta file of the genome submission.

The estimated Quality Value (QV) of the final assembly is 58.1. The assembly has a BUSCO v5.4.3 completeness of 78.8% (single = 78.0%, duplicated = 0.8%), using the metazoa_odb10 reference set ($n$ = 954).

## Metagenome report

Sixteen binned genomes were generated from the metagenome assembly (Figure 6), of which three were classified as high-quality metagenome assembled genomes (MAGs) (see methods). The completeness values for these assemblies range from approximately 20% to 100% with contamination below 7%. A cladogram of the binned metagenomes is shown in Figure 7. For details on binned genomes see Table 3.

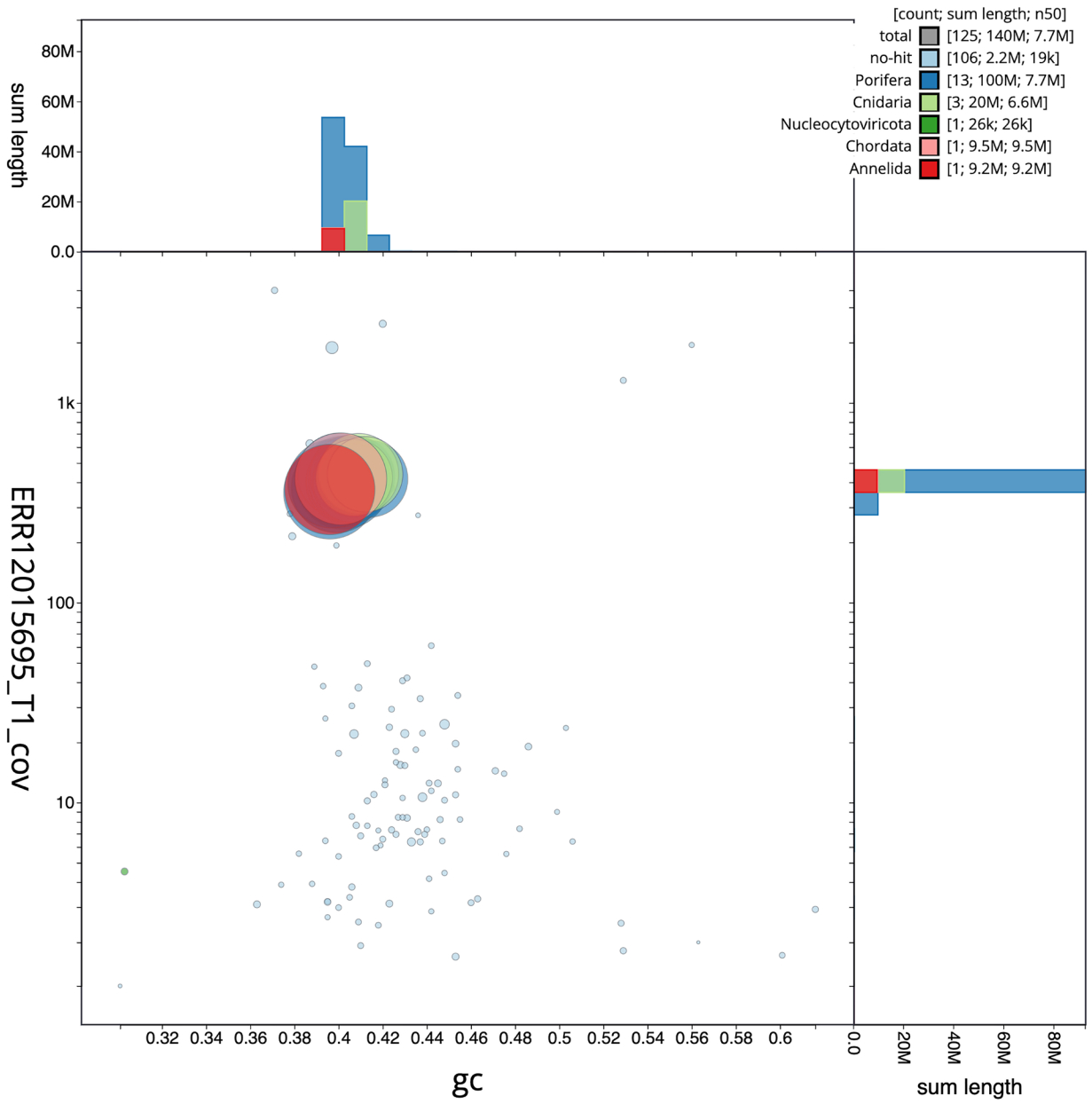


**Figure 3. Genome assembly of *Crambe crambe*, odCraCram1.1: BlobToolKit GC-coverage plot.** Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Crambe_crambe/dataset/GCA_963924555.1/blob.

## Methods

### Sample acquisition

A specimen of *Crambe crambe* (specimen ID GHC0000181, ToLID odCraCram1) was collected from Blanes, Girona, Spain (latitude 41.67, longitude 2.80) on 2021-02-01 by SCUBA diving. The specimen was collected and identified by Manuel Maldonado (CEAB-CSIC) and preserved by snap-freezing.

### Nucleic acid extraction

The workflow for high molecular weight (HMW) DNA extraction at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory includes a sequence of core procedures: sample preparation; sample homogenisation, DNA extraction, fragmentation, and clean-up. Protocols are available on protocols.io (Denton *et al.*, 2023). In sample preparation, the

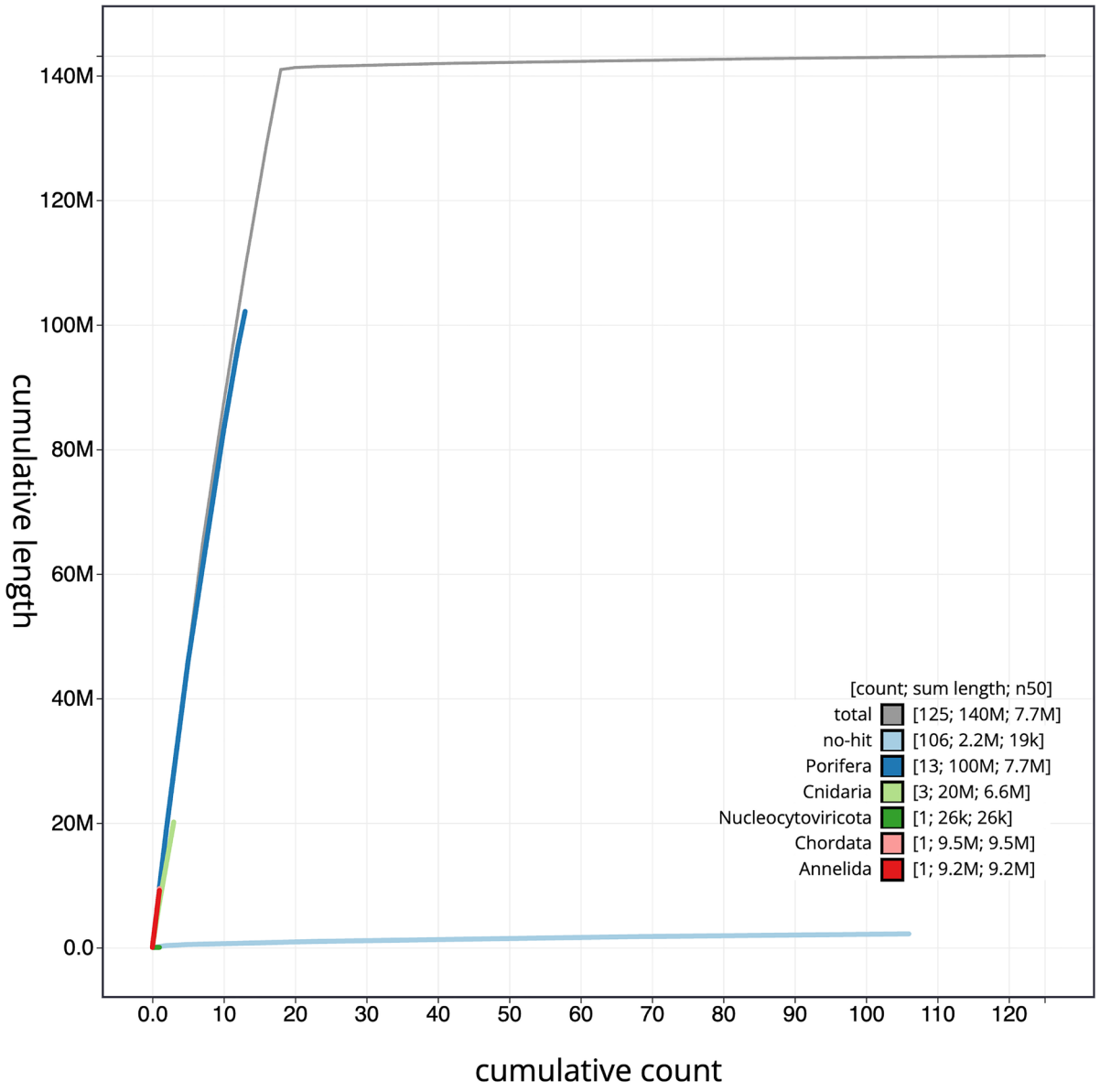


**Figure 4. Genome assembly of *Crambe crambe*, odCraCram1.1: BlobToolKit cumulative sequence plot.** The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Crambe_crambe/dataset/GCA_963924555.1/cumulative.

odCraCram1 sample was weighed and dissected on dry ice (Jay *et al.*, 2023). Prior to DNA extraction, the sponge sample was bathed in "L buffer" (10 mM Tris, pH 7.6, 100 mM EDTA, 20 mM NaCl), minced into small pieces using a scalpel and the cellular interior separated from the mesohyl using forceps (Lopez, 2022). HMW DNA was extracted using the Manual MagAttract v1 protocol (Strickland *et al.*, 2023b). DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system (Todorovic *et al.*, 2023). Sheared DNA was purified by solid-phase reversible immobilisation (Strickland *et al.*, 2023a), using AMPure PB beads to eliminate shorter fragments and concentrate the DNA. The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer, Qubit Fluorometer and Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

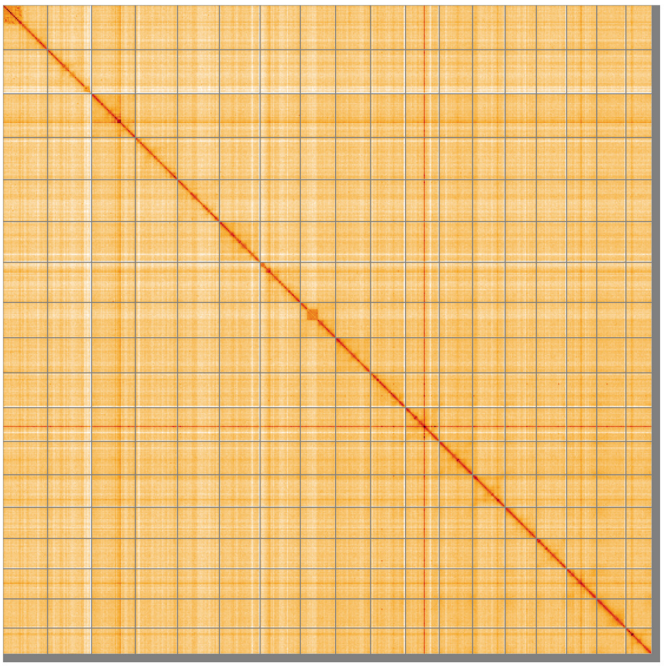

**Figure 5. Genome assembly of *Crambe crambe*, odCraCram1.1: Hi-C contact map of the odCraCram1.1 assembly, visualised using HiGlass.** Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=IeGb4iyXTOqWeiUVkpdMrA.

**Table 2. Chromosomal pseudomolecules in the genome assembly of *Crambe crambe*, odCraCram1.**

| INSDC accession | Name | Length (Mb) | GC% |
|---|---|---|---|
| OZ004581.1 | 1 | 9.57 | 39.5 |
| OZ004582.1 | 2 | 9.68 | 40.0 |
| OZ004583.1 | 3 | 9.49 | 40.0 |
| OZ004584.1 | 4 | 9.08 | 40.0 |
| OZ004585.1 | 5 | 9.18 | 39.5 |
| OZ004586.1 | 6 | 8.84 | 40.0 |
| OZ004587.1 | 7 | 8.73 | 39.5 |
| OZ004588.1 | 8 | 7.66 | 40.0 |
| OZ004589.1 | 9 | 7.63 | 40.5 |
| OZ004590.1 | 10 | 7.57 | 40.5 |
| OZ004591.1 | 11 | 7.33 | 40.5 |
| OZ004592.1 | 12 | 7.24 | 41.0 |
| OZ004593.1 | 13 | 7.06 | 40.5 |
| OZ004594.1 | 14 | 6.8 | 40.5 |
| OZ004595.1 | 15 | 6.57 | 40.5 |
| OZ004596.1 | 16 | 6.54 | 41.5 |
| OZ004597.1 | 17 | 6.34 | 41.0 |
| OZ004598.1 | 18 | 5.67 | 40.5 |
| OZ004599.1 | MT | 0.02 | 37.0 |

## Sequencing

Pacific Biosciences HiFi circular consensus DNA sequencing libraries were constructed according to the manufacturers' instructions. DNA sequencing was performed by the Scientific Operations core at the WSI on a Pacific Biosciences Revio instrument. Hi-C data were also generated from tissue of odCraCram1 using the Arima2 kit and sequenced on the Illumina NovaSeq 6000 instrument.

## Host genome assembly and curation

Assembly was carried out with Hifiasm (Cheng *et al.*, 2021) and haplotypic duplication was identified and removed with purge_dups (Guan *et al.*, 2020). The assembly was then scaffolded with Hi-C data (Rao *et al.*, 2014) using YaHS (Zhou *et al.*, 2023). The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023), which runs MitoFinder (Allio *et al.*, 2020) and uses these annotations to select the final mitochondrial contig and to ensure the general quality of the sequence. Table 4 contains a list of relevant software tool versions and sources.

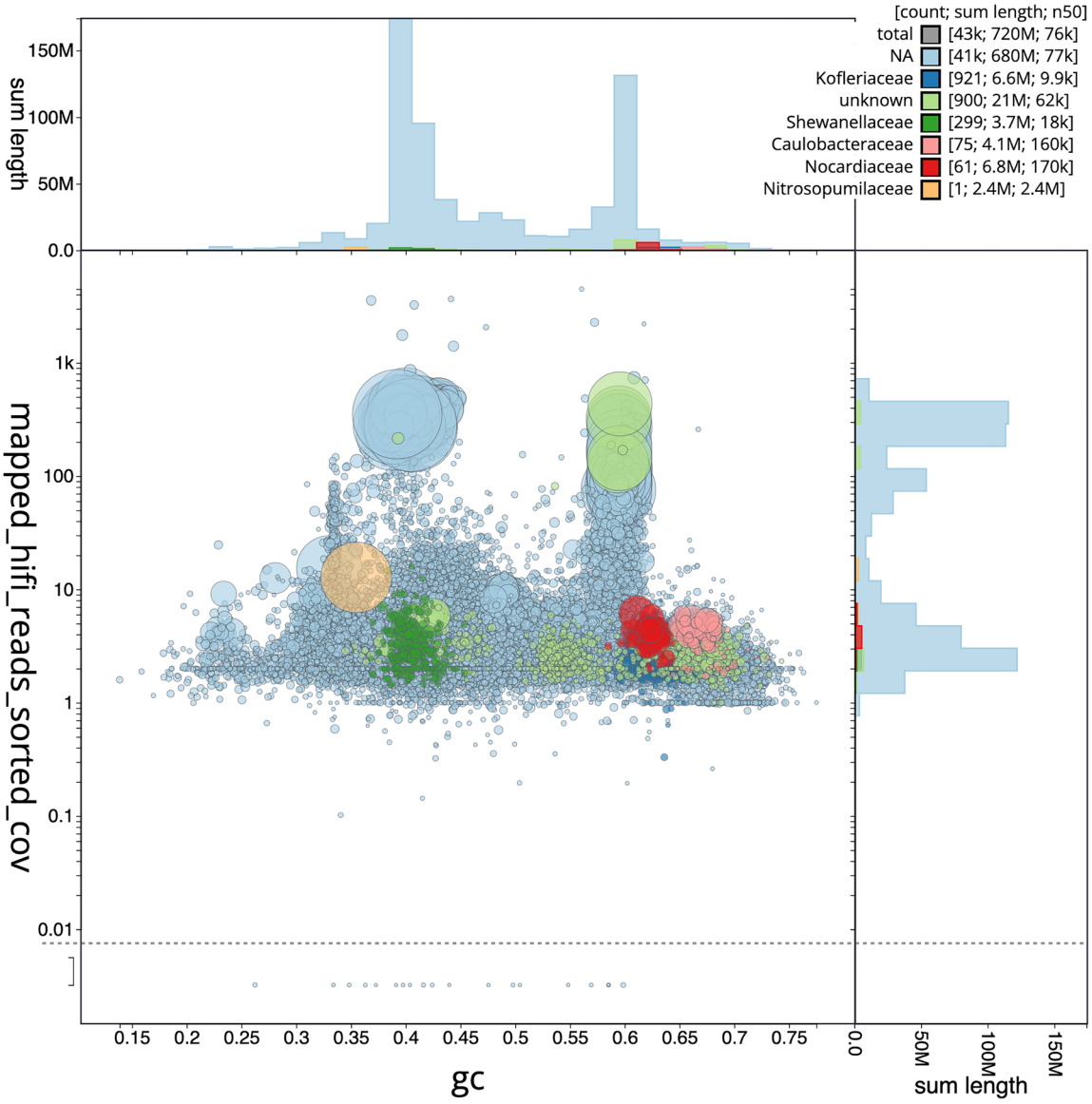


**Figure 6. Blob plot of base coverage in mapped against GC proportion for sequences in the metagenome of *Crambe crambe*.** Binned metagenomes are coloured by family. Circles are sized in proportion to sequence length on a square root scale, ranging from 501 to 4,126,685. Histograms show the distribution of sequence length sum along each axis An interactive version of this figure may be viewed here.

The assembly was checked for contamination and corrected using the TreeVal pipeline (Pointon *et al.*, 2023). Manual curation was primarily conducted using PretextView (Harry, 2022), with additional insights provided by JBrowse2 (Diesh *et al.*, 2023) and HiGlass (Kerpedjiev *et al.*, 2018). Any identified contamination, missed joins, and mis-joins were corrected, and duplicate sequences were tagged and removed. The curation process is documented at https://gitlab.com/wtsi-grit/rapid-curation.

## Taxonomic verification

Molecular markers obtained from the assembly were used to reconstruct the phylogenetic position of the sample. In an alignment using MAFFT v7.450 (Katoh & Standley, 2013), the COI barcoding fragment ("Folmer" fragment) of the sample was found to be identical to haplotype 1 from a dedicated study on *Crambe crambe* (Duran *et al.*, 2004, AF526297), besides samples from other studies on this species as published in NCBI Genbank.

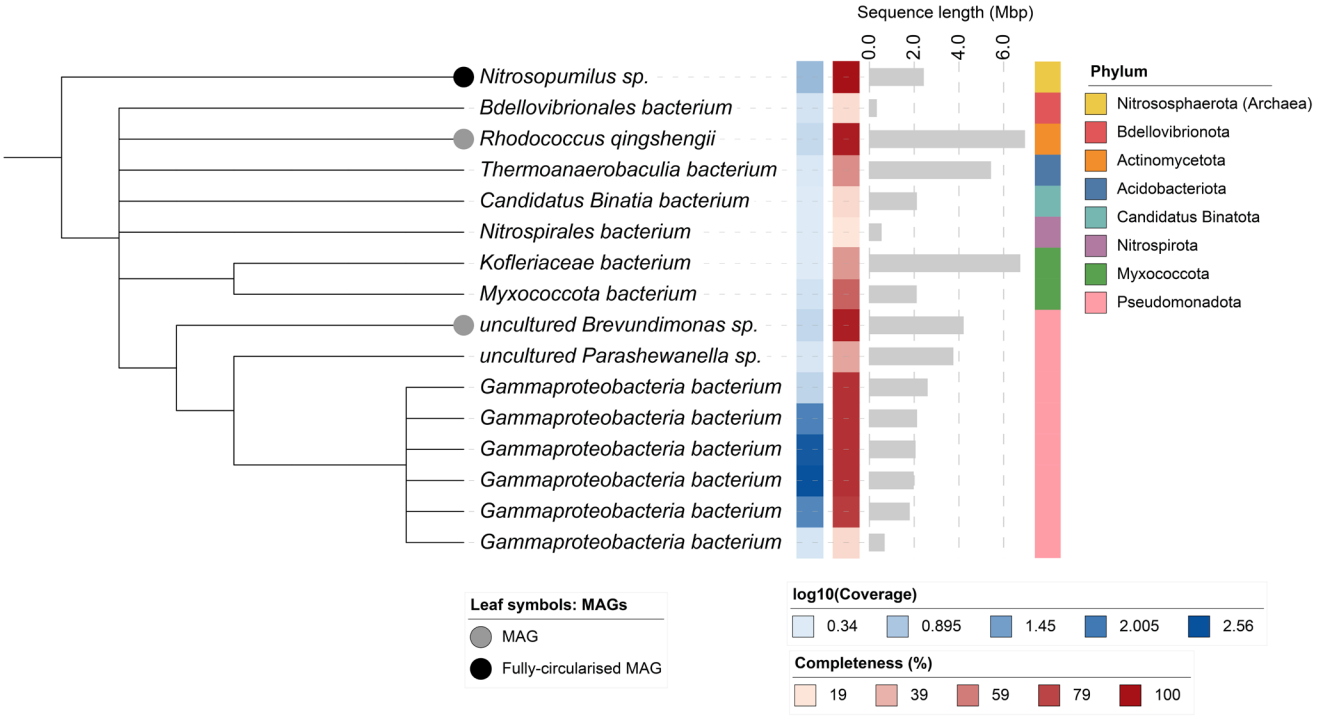


**Figure 7. Cladogram showing the taxonomic placement of metagenome bins, constructed using NCBI taxonomic identifiers with *taxonomizr* and annotated in iTOL.** Colours indicate phylum-level taxonomy. Additional tracks show sequencing coverage ($log_{10}$), estimated genome size (Mbp), and completeness. Bins that meet the criteria for MAGs are marked with a grey circle; the single fully circularised MAG is marked in black.

## Host assembly quality assessment

The Merqury.FK tool (Rhie *et al.*, 2020), run in a Singularity container (Kurtzer *et al.*, 2017), was used to evaluate *k*-mer completeness and assembly quality for the primary and alternate haplotypes using the *k*-mer databases ($k$ = 31) that were computed prior to genome assembly. The analysis outputs included assembly QV scores and completeness statistics.

A Hi-C contact map was produced for the final version of the assembly. The Hi-C reads were aligned using bwa-mem2 (Vasimuddin *et al.*, 2019) and the alignment files were combined using SAMtools (Danecek *et al.*, 2021). The Hi-C alignments were converted into a contact map using BEDTools (Quinlan & Hall, 2010) and the Cooler tool suite (Abdennur & Mirny, 2020). The contact map is visualised in HiGlass (Kerpedjiev *et al.*, 2018).

The blobtoolkit pipeline is a Nextflow port of the previous Snakemake Blobtoolkit pipeline (Challis *et al.*, 2020). It aligns the PacBio reads in SAMtools and minimap2 (Li, 2018) and generates coverage tracks for regions of fixed size. In parallel, it queries the GoaT database (Challis *et al.*, 2023) to identify all matching BUSCO lineages to run BUSCO (Manni *et al.*, 2021). For the three domain-level BUSCO lineages, the pipeline aligns the BUSCO genes to the UniProt Reference Proteomes database (Bateman *et al.*, 2023) with DIAMOND blastp (Buchfink *et al.*, 2021). The genome is also divided into chunks according to the density of the BUSCO genes from the closest taxonomic lineage, and each chunk is aligned to the UniProt Reference Proteomes database using DIAMOND blastx. Genome sequences without a hit are chunked using seqtk and aligned to the NT database with blastn (Altschul *et al.*, 1990). The blobtools suite combines all these outputs into a blobdir for visualisation.

The blobtoolkit pipeline was developed using nf-core tooling (Ewels *et al.*, 2020) and MultiQC (Ewels *et al.*, 2016), relying on the Conda package manager, the Bioconda initiative (Grüning *et al.*, 2018), the Biocontainers infrastructure (da Veiga Leprevost *et al.*, 2017), as well as the Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017) containerisation solutions.

## Metagenome assembly

The metagenome assembly was generated using metaMDBG (Benoit *et al.*, 2024) and binned using MetaBAT2 (Kang *et al.*, 2019), MaxBin (Wu *et al.*, 2014), bin3C (DeMaere & Darling, 2019), and MetaTOR. The resulting bin sets of each binning algorithm were optimised and refined using DAS Tool (Sieber *et al.*, 2018). PROKKA (Seemann, 2014) was used to identify tRNAs and rRNAs in each bin, CheckM (Parks *et al.*, 2015) (checkM_DB release 2015-01-16) was

**Table 3. Quality metrics and taxonomic assignments of the binned metagenomes.**

| NCBI taxon | Taxid | GTDB taxonomy | Quality | Size (bp) | Contigs | Circular | Mean coverage | Completeness (%) | Contamination (%) |
|---|---|---|---|---|---|---|---|---|---|
| Nitrosopumilus sp. | 2024843 | g__Nitrosopumilus | High | 2,406,465 | 1 | Yes | 11.72 | 100.00 | 0.00 |
| uncultured Brevundimonas sp. | 213418 | g__Brevundimonas | High | 4,185,465 | 75 | Partial | 4.39 | 95.06 | 4.27 |
| Rhodococcus qingshengii | 334542 | s__Rhodococcus qingshengii | High | 6,934,702 | 61 | No | 4.2 | 95.58 | 0.00 |
| Gammaproteobacteria bacterium | 1913989 | f__AqS2 | Medium | 1,777,560 | 2 | No | 61.05 | 83.22 | 1.22 |
| Gammaproteobacteria bacterium | 1913989 | f__AqS2 | Medium | 1,964,849 | 1 | Yes | 364.77 | 87.49 | 0.61 |
| Gammaproteobacteria bacterium | 1913989 | f__AqS2 | Medium | 2,039,060 | 1 | Yes | 278.56 | 87.49 | 0.61 |
| Myxococcota bacterium | 2818507 | f__UBA6930 | Medium | 2,090,974 | 133 | No | 3.11 | 69.12 | 6.72 |
| Gammaproteobacteria bacterium | 1913989 | f__AqS2 | Medium | 2,109,981 | 4 | No | 74.99 | 87.49 | 1.83 |
| Gammaproteobacteria bacterium | 1913989 | g__UBA1858 | Medium | 2,576,596 | 26 | No | 4.79 | 87.52 | 1.97 |
| Thermoanaerobaculia bacterium | 2651171 | f__UBA5704 | Medium | 5,408,624 | 319 | No | 2.54 | 52.97 | 1.44 |
| Bdellovibrionales bacterium | 2053517 | g__JACOND01 | Low | 314,367 | 16 | No | 2.89 | 23.10 | 0.00 |
| Nitrospirales bacterium | 2358460 | g__Bin75 | Low | 524,459 | 58 | No | 2.22 | 19.47 | 0.00 |
| Gammaproteobacteria bacterium | 1913989 | g__UBA1858 | Low | 655,358 | 63 | No | 2.67 | 24.99% | 1.77% |
| Candidatus Binatia bacterium | 2838779 | g__JAAXHF01 | Low | 2,103,514 | 277 | No | 2.23 | 24.60% | 0.00% |
| uncultured Parashewanella sp. | 2547967 | g__Parashewanella | Low | 3,728,522 | 299 | Partial | 2.61 | 43.78% | 5.25% |
| Kofleriaceae bacterium | 2212474 | f__Haliangiaceae | Low | 6,724,497 | 921 | Partial | 2.17 | 48.48% | 5.11% |



Wellcome Open Research 2025, 10:275 Last updated: 13 JUN 2025

**Table 4. Software tools: versions and sources.**

| Software tool | Version | Source |
|---|---|---|
| BEDTools | 2.30.0 | https://github.com/arq5x/bedtools2 |
| bin3C | 0.3.3 | https://github.com/cerebis/bin3C |
| Blast | 2.14.0 | ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/ |
| BlobToolKit | 4.3.7 | https://github.com/blobtoolkit/blobtoolkit |
| BUSCO | 5.4.3 and 5.5.0 | https://gitlab.com/ezlab/busco |
| bwa-mem2 | 2.2.1 | https://github.com/bwa-mem2/bwa-mem2 |
| CheckM | 1.2.1 | https://github.com/Ecogenomics/CheckM |
| Cooler | 0.8.11 | https://github.com/open2c/cooler |
| DAS Tool | - | https://github.com/cmks/DAS_Tool |
| DIAMOND | 2.1.8 | https://github.com/bbuchfink/diamond |
| dRep | 3.4.0 | https://github.com/MrOlm/drep |
| fasta_windows | 0.2.4 | https://github.com/tolkit/fasta_windows |
| FastK | 427104ea91c78c3b8b8b49f1a7d6bbeaa869ba1c | https://github.com/thegenemyers/FASTK |
| GoaT CLI | 0.2.5 | https://github.com/genomehubs/goat-cli |
| GTDB-TK | 2.3.2 | https://github.com/Ecogenomics/GTDBTk |
| Hifiasm | 0.19.5-r587 | https://github.com/chhylp123/hifiasm |
| HiGlass | 44086069ee7d4d3f6f3f0012569789ec138f42b84 aa44357826c0b6753eb28de | https://github.com/higlass/higlass |
| MaxBin | 2.7 | https://sourceforge.net/projects/maxbin/ |
| MerquryFK | d00d98157618f4e8d1a9190026b19b471055b22e | https://github.com/thegenemyers/MERQURY.FK |
| MetaBat2 | 2.15-15-gd6ea400 | https://bitbucket.org/berkeleylab/metabat/src/master/ |
| MetaTOR | - | https://github.com/koszullab/metaTOR |
| MitoHiFi | 2 | https://github.com/marcelauliano/MitoHiFi |
| MultiQC | 1.14, 1.17, and 1.18 | https://github.com/MultiQC/MultiQC |
| Nextflow | 23.04.0-5857 | https://github.com/nextflow-io/nextflow |
| PretextView | 0.2 | https://github.com/wtsi-hpag/PretextView |
| PROKKA | 1.14.5 | https://github.com/vdejager/prokka |
| purge_dups | 1.2.5 | https://github.com/dfguan/purge_dups |
| samtools | 1.16.1, 1.17, and 1.18 | https://github.com/samtools/samtools |
| Seqtk | 1.3 | https://github.com/lh3/seqtk |
| Singularity | 3.9.0 | https://github.com/sylabs/singularity |
| TreeVal | 1.0.0 | https://github.com/sanger-tol/treeval |
| YaHS | 1.1a.2 | https://github.com/c-zhou/yahs |

used to assess bin completeness/contamination, and GTDB-TK (Chaumeil *et al.*, 2022) (GTDB release 214) was used to taxonomically classify bins. Taxonomic replicate bins were identified using dRep (Olm *et al.*, 2017), with default settings (95% ANI threshold). The final bin set was filtered for bacteria and archaea. All bins were assessed for quality and categorised as metagenome-assembled genomes (MAGs) if they met the following criteria: contamination ≤ 5%, presence

of 5S, 16S, and 23S rRNA genes, at least 18 unique tRNAs, and either ≥ 90% completeness or ≥ 50% completeness with fully circularised chromosomes. Bins that did not meet these thresholds, or were identified as taxonomic replicates of MAGs, were retained as 'binned metagenomes' provided they had ≥ 50% completeness and ≤ 10% contamination. A cladogram based on NCBI taxonomic assignments was generated using the 'taxonomizr' package in R. The tree was visualised and annotated using iTOL (Letunic & Bork, 2024). Software tool versions and sources are given in Table 4.

### Wellcome Sanger Institute – Legal and Governance

The materials that have contributed to this genome note have been supplied by a Tree of Life collaborator. The Wellcome Sanger Institute employs a process whereby due diligence is carried out proportionate to the nature of the materials themselves, and the circumstances under which they have been/are to be collected and provided for use. The purpose of this is to address and mitigate any potential legal and/or ethical implications of receipt and use of the materials as part of the research project, and to ensure that in doing so we align with best practice wherever possible. The overarching areas of consideration are:

- Ethical review of provenance and sourcing of the material
- Legality of collection, transfer and use (national and international)

Each transfer of samples is undertaken according to a Research Collaboration Agreement or Material Transfer Agreement entered into by the Tree of Life collaborator, Genome Research Limited (operating as the Wellcome Sanger Institute) and in some circumstances other Tree of Life collaborators.

## Data availability

European Nucleotide Archive: *Crambe crambe*. Accession number PRJEB65618; https://identifiers.org/ena.embl/PRJEB65618. The genome sequence is released openly for reuse. The *Crambe crambe* genome sequencing initiative is part of the Aquatic Symbiosis Genomics (ASG) project (https://www.ebi.ac.uk/ena/browser/view/PRJEB43743). All raw sequence data and the assembly have been deposited in INSDC databases. The genome will be annotated using available RNA-Seq data and presented through the Ensembl pipeline at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in Table 1 and Table 2.

### Author information

Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory Team are listed here: https://doi.org/10.5281/zenodo.10066175.

Members of the Wellcome Sanger Institute Scientific Operations: Sequencing Operations are listed here: https://doi.org/10.5281/zenodo.10043364.

Members of the Wellcome Sanger Institute Tree of Life Core Informatics team are listed here: https://doi.org/10.5281/zenodo.10066637.

Members of the European Bioinformatics Institute ASG Data Portal team are listed here: https://doi.org//10.5281/zenodo.10076466.

Members of the Wellcome Sanger Institute/Aquatic Symbiosis Genomics Project Leadership are listed here: https://doi.org/10.5281/zenodo.10184833.

## References


Abdennur N, Mirny LA: **Cooler: scalable storage for Hi-C data and other genomically labeled arrays.** *Bioinformatics.* 2020; **36**(1): 311–316.
PubMed Abstract | Publisher Full Text | Free Full Text

Allio R, Schomaker-Bastos A, Romiguier J, *et al.*: **MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics.** *Mol Ecol Resour.* 2020; **20**(4): 892–905.
PubMed Abstract | Publisher Full Text | Free Full Text

Altschul SF, Gish W, Miller W, *et al.*: **Basic Local Alignment Search Tool.** *J Mol Biol.* 1990; **215**(3): 403–410.
PubMed Abstract | Publisher Full Text

Bateman A, Martin MJ, Orchard S, *et al.*: **UniProt: the universal protein knowledgebase in 2023.** *Nucleic Acids Res.* 2023; **51**(D1): D523–D531.
PubMed Abstract | Publisher Full Text | Free Full Text

Becerro MA, Uriz MJ, Turón X: **Trends in space occupation by the encrusting sponge *Crambe crambe*: variation in shape as a function of size and environment.** *Mar Biol.* 1994; **121**: 301–307.
Publisher Full Text

Becerro MA, Uriz MJ, Turon X: **Chemically-mediated interactions in benthic organisms: the chemical ecology of *Crambe crambe* (Porifera, Poecilosclerida).** *Hydrobiologia.* 1997; **355**(1–3): 77–89.
Publisher Full Text

Benoit G, Raguideau S, James R, *et al.*: **High-quality metagenome assembly from long accurate reads with metaMDBG.** *Nat Biotechnol.* 2024; **42**(9): 1378–1383.
PubMed Abstract | Publisher Full Text | Free Full Text

Buchfink B, Reuter K, Drost HG: **Sensitive protein alignments at Tree-of-Life scale using DIAMOND.** *Nat Methods.* 2021; **18**(4): 366–368.
PubMed Abstract | Publisher Full Text | Free Full Text

Carrier TJ, Maldonado M, Schmittmann L, *et al.*: **Symbiont transmission in marine sponges: reproduction, development, and metamorphosis.** *BMC Biol.* 2022; **20**(1): 100.
PubMed Abstract | Publisher Full Text | Free Full Text

Challis R, Kumar S, Sotero-Caio C, *et al.*: **Genomes on a Tree (GoaT): a versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic Tree of Life [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2023; **8**: 24.
PubMed Abstract | Publisher Full Text | Free Full Text

Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit – interactive quality assessment of genome assemblies.** *G3 (Bethesda).* 2020; **10**(4): 1361–1374.
PubMed Abstract | Publisher Full Text | Free Full Text

Chaumeil PA, Mussig AJ, Hugenholtz P, *et al.*: **GTDB-Tk v2: memory friendly classification with the genome taxonomy database.** *Bioinformatics.* 2022; **38**(23): 5315–5316.
PubMed Abstract | Publisher Full Text | Free Full Text

Cheng H, Concepcion GT, Feng X, *et al.*: **Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm.** *Nat Methods.* 2021; **18**(2): 170–175.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Croué J, West NJ, Escande ML, *et al.*: **A single betaproteobacterium dominates the microbial community of the crambescidine-containing sponge *Crambe crambe*.** *Sci Rep.* 2013; **3**: 2583.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

da Veiga Leprevost F, Grüning BA, Alves Aflitos S, *et al.*: **BioContainers: an open-source and community-driven framework for software standardization.** *Bioinformatics.* 2017; **33**(16): 2580–2582.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Danecek P, Bonfield JK, Liddle J, *et al.*: **Twelve years of SAMtools and BCFtools.** *GigaScience.* 2021; **10**(2): giab008.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

DeMaere MZ, Darling AE: **bin3C: exploiting Hi-C sequencing data to accurately resolve metagenome-assembled genomes.** *Genome Biol.* 2019; **20**(1): 46.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Denton A, Yatsenko H, Jay J, *et al.*: **Sanger Tree of Life wet laboratory protocol collection V.1.** *protocols.io.* 2023.
**Publisher Full Text**

Diesh C, Stevens GJ, Xie P, *et al.*: **JBrowse 2: a modular genome browser with views of synteny and structural variation.** *Genome Biol.* 2023; **24**(1): 74.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Duran S, Pascual M, Turon X: **Low levels of genetic variation in mtDNA sequences over the western Mediterranean and Atlantic range of the sponge *Crambe crambe* (Poecilosclerida).** *Mar Biol.* 2004; **144**(1): 31–35.
**Publisher Full Text**

El-Demerdash A, Atanasov AG, Bishayee A, *et al.*: ***Batzella*, *Crambe* and *Monanchora*: highly prolific marine sponge genera yielding compounds with potential applications for cancer and other therapeutic areas.** *Nutrients.* Multidisciplinary Digital Publishing Institute, 2018; **10**(1): 33.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Ewels P, Magnusson M, Lundin S, *et al.*: **MultiQC: summarize analysis results for multiple tools and samples in a single report.** *Bioinformatics.* 2016; **32**(19): 3047–3048.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Ewels PA, Peltzer A, Fillinger S, *et al.*: **The nf-core framework for community-curated bioinformatics pipelines.** *Nat Biotechnol.* 2020; **38**(3): 276–278.
**PubMed Abstract** | **Publisher Full Text**

Gauthier MEA, Watson JR, Degnan SM: **Draft genomes shed light on the dual bacterial symbiosis that dominates the microbiome of the coral reef sponge *Amphimedon queenslandica*.** *Front Mar Sci.* 2016; **3**: 196.
**Publisher Full Text**

Grüning B, Dale R, Sjödin A, *et al.*: **Bioconda: sustainable and comprehensive software distribution for the life sciences.** *Nat Methods.* 2018; **15**(7): 475–476.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Guan D, McCarthy SA, Wood J, *et al.*: **Identifying and removing haplotypic duplication in primary genome assemblies.** *Bioinformatics.* 2020; **36**(9): 2896–2898.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Harry E: **PretextView (Paired REad TEXTure Viewer): a desktop application for viewing pretext contact maps**. 2022.
**Reference Source**

Jay J, Yatsenko H, Narváez-Gómez JP, *et al.*: **Sanger Tree of Life sample preparation: triage and dissection.** *protocols.io.* 2023.
**Publisher Full Text**

Kang DD, Li F, Kirton E, *et al.*: **MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies.** *PeerJ.* 2019; **7**: e7359.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Katoh K, Standley DM: **MAFFT multiple sequence alignment software version 7: improvements in performance and usability.** *Mol Biol Evol.* 2013; **30**(4): 772–80.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Kerpedjiev P, Abdennur N, Lekschas F, *et al.*: **HiGlass: web-based visual exploration and analysis of genome interaction maps.** *Genome Biol.* 2018; **19**(1): 125.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Kurtzer GM, Sochat V, Bauer MW: **Singularity: scientific containers for mobility of compute.** *PLoS One.* 2017; **12**(5): e0177459.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Letunic I, Bork P: **Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool.** *Nucleic Acids Res.* 2024; **52**(W1): W78–W82.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Li H: **Minimap2: pairwise alignment for nucleotide sequences.** *Bioinformatics.* 2018; **34**(18): 3094–3100.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Lopez J: **Squeeze-enrichment of intact cells (eukaryotic and prokaryotic) from marine sponge tissues prior to routine DNA extraction.** *protocols.io.* 2022.
**Reference Source**

Maldonado M: **Intergenerational transmission of symbiotic bacteria in oviparous and viviparous demosponges, with emphasis on intracytoplasmically-compartmented bacterial types.** *J Mar Biol Assoc UK.* 2007; **87**(6): 1701–1713.
**Publisher Full Text**

Maldonado M, Bergquist PR: **Phylum porifera**. In: Young, C. M., Sewell, M. A., and Rice, M. E. (eds.) *Atlas of marine invertebrate larvae*. San Diego: Academic Press, 2002; 21–50.
**Reference Source**

Maldonado M, Carmona MC, Uriz MJ, *et al.*: **Decline in Mesozoic reef-building sponges explained by silicon limitation.** *Nature.* 1999; **401**(6755): 785–788.
**Publisher Full Text**

Maldonado M, Carmona MC, Van Soest RWM, *et al.*: **First record of the sponge genera *Crambe* and *Discorhabdella* for the eastern Pacific, with description of three new species.** *J Nat Hist.* 2001; **35**(9): 1261–1276.
**Publisher Full Text**

Maldonado M, Carmona MC, Velásquez Z, *et al.*: **Siliceous sponges as a silicon sink: an overlooked aspect of benthopelagic coupling in the marine silicon cycle.** *Limnol Oceanogr.* 2005; **50**(3): 799–809.
**Publisher Full Text**

Manni M, Berkeley MR, Seppey M, *et al.*: **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**(10): 4647–4654.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Merkel D: **Docker: lightweight Linux containers for consistent development and deployment.** *Linux J.* 2014; **2014**(239): 2, [Accessed 2 April 2024].
**Reference Source**

Olm MR, Brown CT, Brooks B, *et al.*: **dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication.** *ISME J.* 2017; **11**(12): 2864–2868.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Padiglia A, Ledda FD, Padedda BM, *et al.*: **Long-term experimental *in situ* farming of *Crambe crambe* (Demospongiae: Poecilosclerida).** *PeerJ.* 2018; **6**: e4964.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Pansini M, Pronzato R: **Observations on the dynamics of a Mediterranean sponge community.** In: Rützler, K. (ed.) *New perspectives in sponge biology*. Washington, DC. London: Smithsonian Institution Press, 1990; 404–415.

Parks DH, Imelfort M, Skennerton CT, *et al.*: **CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes.** *Genome Res.* 2015; **25**(7): 1043–55.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Pointon DL, Eagles W, Sims Y, *et al.*: **sanger-tol/treeval v1.0.0 – Ancient Atlantis**. 2023.
**Publisher Full Text**

Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features.** *Bioinformatics.* 2010; **26**(6): 841–842.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Rao SSP, Huntley MH, Durand NC, *et al.*: **A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.** *Cell.* 2014; **159**(7): 1665–1680.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Rhie A, Walenz BP, Koren S, *et al.*: **Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1): 245.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Riesgo A, Maldonado M: **An unexpectedly sophisticated, V-shaped spermatozoon in Demospongiae (Porifera): reproductive and evolutionary implications.** *Biol J Linn Soc.* 2009; **97**(2): 413–426.
**Publisher Full Text**

Schmidt O: **Die Spongien des Adriatischen Meeres**. Engelmann, W. (ed.) . Leipzig, 1862.

Seemann T: **Prokka: rapid prokaryotic genome annotation.** *Bioinformatics.* 2014; **30**(14): 2068–2069.
**PubMed Abstract** | **Publisher Full Text**

Sieber CMK, Probst AJ, Sharrar A, *et al.*: **Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy.** *Nat Microbiol.* 2018; **3**(7): 836–843.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Strickland M, Cornwell C, Howard C: **Sanger Tree of Life fragmented DNA clean up: manual SPRI.** *protocols.io.* 2023a.
**Publisher Full Text**

Strickland M, Moll R, Cornwell C, *et al.*: **Sanger Tree of Life HMW DNA extraction: manual MagAttract.** *protocols.io.* 2023b.
**Publisher Full Text**

Taylor JA, Palladino G, Wemheuer B, *et al.*: **Phylogeny resolved, metabolism**

**revealed: functional radiation within a widespread and divergent clade of sponge symbionts.** *ISME J.* 2021; **15**(2): 503–519.
PubMed Abstract | Publisher Full Text | Free Full Text

Todorovic M, Sampaio F, Howard C: **Sanger Tree of Life HMW DNA fragmentation: diagenode Megaruptor®3 for PacBio HiFi.** *protocols.io.* 2023.
Publisher Full Text

Tripepi S, Longo O, La Camera R: **A new pattern of spermiogenesis in the sponge Crambe crambe: preliminary observations.** In: Csanády, A., Röhlich, P., and Szabó, D. (eds.) *Eighth European congress on electron microscopy*. Budapest: Programme Committee, 1984; 2073–2074.

Turon M, Ford M, Maldonado M, *et al.*: **Microbiome changes through the ontogeny of the marine sponge *Crambe crambe*.** *Environ Microbiome.* 2024; **19**(1): 15.
PubMed Abstract | Publisher Full Text | Free Full Text

Turon X, Codina M, Tarjuelo I, *et al.*: **Mass recruitment of *Ophiothrix fragilis* (Ophiuroidea) on sponges: settlement patterns and post-settlement dynamics.** *Mar Ecol Prog Ser.* 2000; **200**: 201–212.
Publisher Full Text

Turon X, Tarjuelo I, Uriz MJ: **Growth dynamics and mortality of the encrusting sponge *Crambe crambe* (Poecilosclerida) in contrasting habitats: correlation with population structure and investment in defence.** *Funct Ecol.* 1998; **12**(4): 631–639.
Publisher Full Text

Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, *et al.*: **MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads.** *BMC Bioinformatics.* 2023; **24**(1): 288.
PubMed Abstract | Publisher Full Text | Free Full Text

Uriz MJ, Maldonado M, Turon X, *et al.*: **How do reproductive output, larval behaviour, and recruitment contribute to adult spatial patterns in Mediterranean encrusting sponges?** *Mar Ecol Prog Ser.* 1998; **167**: 137–148.
Publisher Full Text

Uriz MJ, Turon X, Becerro MA: **Morphology and ultrastructure of the swimming larvae of *Crambe crambe* (Demospongiae, Poecilosclerida).** *Invertebr Biol.* 2001; **120**(4): 295–307.
Publisher Full Text

Vasimuddin M, Misra S, Li H, *et al.*: **Efficient architecture-aware acceleration of BWA-MEM for multicore systems.** In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).* IEEE, 2019; 314–324.
Publisher Full Text

Wu YW, Tang YH, Tringe SG, *et al.*: **MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm.** *Microbiome.* 2014; **2**(1): 26.
PubMed Abstract | Publisher Full Text | Free Full Text

Zhou C, McCarthy SA, Durbin R: **YaHS: yet another Hi-C scaffolding tool.** *Bioinformatics.* 2023; **39**(1): btac808.
PubMed Abstract | Publisher Full Text | Free Full Text

# Open Peer Review

## Current Peer Review Status: ✔ ? ✔

**Version 1**

Reviewer Report 13 June 2025

https://doi.org/10.21956/wellcomeopenres.26645.r124356



✔ **Paco Cardenas**
Museum of Evolution, Uppsala University, Uppsala, Sweden

This article describes how the whole genome of the marine sponge *Crambe crambe* was sequenced and assembled into 18 chromosomes, along with several MAGs.
Being a sponge systematist and not a genomics bioinformatician, I could only judge the biological part of the article, not the genomics and metagenomics technical parts.
In the background, the authors mention that this species is a surviving relict. They give two arguments but miss a third important one, the one given by time-calibrated phylogenies suggesting the emergence of the order Poecilosclerida around the Triassic (Plese et al., 2021) (Ref 1).
In Figure 1, a scale is missing.

**References**

1. Plese B, Kenny N, Rossi M, Cárdenas P, et al.: Mitochondrial evolution in the Demospongiae (Porifera): Phylogeny, divergence time, and genome biology. *Molecular Phylogenetics and Evolution*. 2021; **155**. Publisher Full Text

**Is the rationale for creating the dataset(s) clearly described?**
Yes

**Are the protocols appropriate and is the work technically sound?**
Yes

**Are sufficient details of methods and materials provided to allow replication by others?**
Yes

**Are the datasets clearly presented in a useable and accessible format?**
Yes

***Competing Interests:*** No competing interests were disclosed.

***Reviewer Expertise:*** Porifera systematics and biology, Deep-sea habitats, Museum collections, Historical collections.

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Reviewer Report 12 June 2025

https://doi.org/10.21956/wellcomeopenres.26645.r124358



?

**Weizhi Song**
The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong
**Sihan Li**
The Hong Kong University of Science and Technology Department of Ocean Science, Hong Kong, Hong Kong

This article presents the chromosomal genome sequence of the sponge *Crambe crambe* (Schmidt, 1862) and its associated microbial metagenome sequences. *Crambe crambe* is abundant in the sublittoral rocky bottoms of the Atlantic-Mediterranean region, and its genome (including that of its symbionts) was sequenced as part of the Aquatic Symbiosis Genomics Project. Cutting-edge sequencing technologies, including PacBio HiFi and Hi-C sequencing, were employed to generate the genome assemblies. Technical details for generating the assemblies are provided, and the resulting data can be publicly accessed via the European Nucleotide Archive. This dataset serves as a valuable resource for sponge-related research, ranging from investigations into sponge host phylogeny to studies on sponge-microbe symbiosis. However, several aspects require improvement before it can be accepted for publication. Please see details below.

The units used for sequence length or assembly size are inconsistent throughout the article. For example, “Gb” and “Mb” are used in Tables 1 and 2; “M” and “k” appear in Figures 2, 3, 4, and 6; while “bp” and “Mbp” are used in Figure 7. Please ensure consistent format of length units in the article.

Most of the figures (Figures 2 to 7) in this article appear to have been automatically generated, with little refinement for publication quality. Some of them are not informative and should be removed from the article. Specifically:

Figure 2: It is not informative and should be removed from the article, the relevant statistical details are already clearly provided in figure legends. In addition, it is redundant to Table 1 in several aspects, including reports on BUSCO assessments, Scaffold N50, and Longest scaffold. Furthermore, there are inconsistencies between the statistics reported in Table 1 and those shown in Figure 1. For instance, the BUSCO completeness is listed as 78.8% in Table 1, but 79.4% in Figure 1. Please double check on these details.

Figure 3: The “sum length” on both the x- and y-axes doesn’t make much sense and better to be removed from the figure.

Figure 4: It is not informative, including a link to the plot within the article should be sufficient.

Figure 5: Please enlarge it for better visibility. It appears that there may be issues with chromosomes 3, 8, and 11 (from left to right) (By the way, I couldn’t find an option to attach figures in this online review system to illustrate the issues I’ve identified). Could this be further refined or discussed in the main text?

Figure 6: Similar to Figure 3, the “sum length” panels are meaningless, especially for the microbial metagenome, and should be removed from the figure. Also, the link to Figure 6 is not accessible.

Figure 7: “Sequence length (Mbp)” in should be revised to “Genome Size (Mbp)”.

Figure 7: I recommend building a phylogenetic tree for the generated MAGs, rather than a cladogram with lots of polytomies.

Figure 7: Many of the leaves on the tree share the same name (e.g., *Gammaproteobacteria bacterium*). Please assign a unique identifier to each, such as the MAG name or accession number.

Figure 7: Using a black symbol to indicate that a MAG is circularized is fine. However, the meaning of the two grey symbols is unclear to me, as all of them appear to be MAGs. If the intention is to distinguish between high-, medium-, or low-quality MAGs, please clarify that explicitly.

Table 3: Some values in the “Completeness” and “Contamination” columns include a “%” symbol, while others do not. Please ensure consistency in formatting throughout the table.

Table 4: there is no need to retain Table 4. Please ensure that all software tools used in data analysis, along with their versions, are properly cited in the “Methods” section. I noticed that software version is missing from the main text.

“Metagenome assembly” section:
GTDB release r214 came out in 2023 (two years ago). It may be worth considering reclassifying these MAGs using the most recent GTDB release (e.g., r226).

“Metagenome assembly” section:
“All bins were assessed for quality and categorised as metagenome-assembled genomes (MAGs) if they met the following criteria: …”. Is there a reference supporting the criteria used here?

**Is the rationale for creating the dataset(s) clearly described?**

Yes

**Are the protocols appropriate and is the work technically sound?**

Partly

**Are sufficient details of methods and materials provided to allow replication by others?**

Yes

**Are the datasets clearly presented in a useable and accessible format?**
Partly

***Competing Interests:*** No competing interests were disclosed.

***Reviewer Expertise:*** bioinformatics and sponge-microbe symbiosis

**We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however we have significant reservations, as outlined above.**

Reviewer Report 09 June 2025

https://doi.org/10.21956/wellcomeopenres.26645.r124362



**Shikai Liu**
Ocean University of China, Qingdao, China

In this work, Maldonado et al presented a genome assembly from an individual Crambe crambe, with genome sequence of 143.2 Mb in span. The assmbly is scaffolded into 18 pseudochromomes. A size of 19.53 Kb mitogenome was also assembled for this species. Additionally, several symbiotic prokaryotic genomes were also assembled as MAGs. The sequencing was performed with PacBio HiFi incombination with Hi-C data, which is a well-recognized whole genome sequencing strategy. The data was well presented and should be valuable to the research community.

**Is the rationale for creating the dataset(s) clearly described?**
Yes

**Are the protocols appropriate and is the work technically sound?**
Yes

**Are sufficient details of methods and materials provided to allow replication by others?**
Yes

**Are the datasets clearly presented in a useable and accessible format?**
Yes

***Competing Interests:*** No competing interests were disclosed.

***Reviewer Expertise:*** Aquatic animal genomics and genetic breeding

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**