



OPEN

DATA DESCRIPTOR

Chromosome-level genome assembly of the spangled emperor, *Lethrinus nebulosus* (Forsskål 1775)

Lara Parata¹✉, Liam Anstiss¹, Emma de Jong¹, Adrienne Doran¹, Richard J. Edwards¹, Stephen J. Newman², Samuel D. Payet², Craig L. Skepper², Corey B. Wakefield², OceanOmics Centre*, OceanOmics Division* & Shannon Corrigan^{1,3}

Spangled emperor, *Lethrinus nebulosus* (Forsskål 1775), is a tropical marine fish of economic and cultural importance throughout the Indo-West Pacific. It is one of the most targeted recreational fishes in the Gascoyne Coast Bioregion of Western Australia where it serves as an indicator species for recreational fishing. Here, we present a highly accurate, near-gapless, chromosome-level, haplotype-phased reference genome assembly of *L. nebulosus* (*Lethrinus nebulosus* (Spangled Emperor) genome, fLetNeb1.1; PRJNA1074345), the first for the species and the first high-quality genome representative of the family Lethrinidae. The 1.09 Gb genome was assembled from PacBio HiFi and Dovetail Omni-C proximity ligation sequencing data. The contig N50 is 21–24 Mbp and BUSCO completeness greater than 99%. A preliminary gene annotation identified 24,583 genes with the predicted transcriptome achieving a BUSCO completeness score of 99.1%. This resource will facilitate genomic studies to inform the sustainable management of *L. nebulosus* and other Lethrinids.

Background & Summary

Spangled emperor, *Lethrinus nebulosus* (Forsskål 1775) is a tropical marine fish that is distributed throughout the Indo-West Pacific¹. *Lethrinus nebulosus* is common on reef habitat down to 150 m (Fig. 1) and is sought after by recreational, commercial, charter and customary fishers¹. In Australia, there are eight *L. nebulosus* stocks (at a management or jurisdictional level) from Perth on the southwest coast, across northern Australia to Sydney on the east coast². In the Ningaloo Marine Park of Western Australia, *L. nebulosus* is one of the most popular angling fish³. This species also plays important role in the management of the Gascoyne Demersal Scalefish Resource as it is an indicator species, where their stock status is used to infer that of other similar species in the multi-species fishery^{2,4}.

Dispersal of *L. nebulosus* individuals primarily occurs during the planktonic egg and larval phase, with adults generally considered site-attached⁵. Previous work indicates that there is a single genetic stock of *L. nebulosus* in northwestern Australia (Kimberley to Shark Bay), with evidence of genetic connectivity between the east and west coasts⁵. While large, connected populations typically harbour high genetic diversity and resilience, they may still experience localised population pressures and even declines, particularly in the context of a changing climate^{6,7}. For example, changes in environmental conditions and habitat could affect larval dispersal or mortality, leading to reduced juvenile recruitment or altered distributions. Furthermore, the relatively site-attached nature of adult fish^{2,5} suggests that populations could be susceptible to localised disturbances, with limited opportunities to move to more suitable environments. Together, these factors could influence the stock sustainability of *L. nebulosus*.

Here, we present a highly accurate, near-gapless, chromosome-level, haplotype-phased reference genome assembly of *L. nebulosus*. High-quality reference genome assemblies are increasingly considered foundational scientific resources that can facilitate a suite of molecular techniques to support sustainable management and biodiversity conservation. This high-quality genome assembly was generated using PacBio high fidelity (HiFi)

¹Minderoo OceanOmics Centre at UWA, Oceans Institute, University of Western Australia, Crawley, WA, Australia.

²Western Australian Fisheries and Marine Research Laboratories, Department of Primary Industries and Regional Development, Government of Western Australia, 39 Northside Drive, Hillarys, 6025, WA, Australia. ³Minderoo Foundation, Perth, WA, Australia. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: lara.parata@uwa.edu.au; oceanogenomes@uwa.edu.au

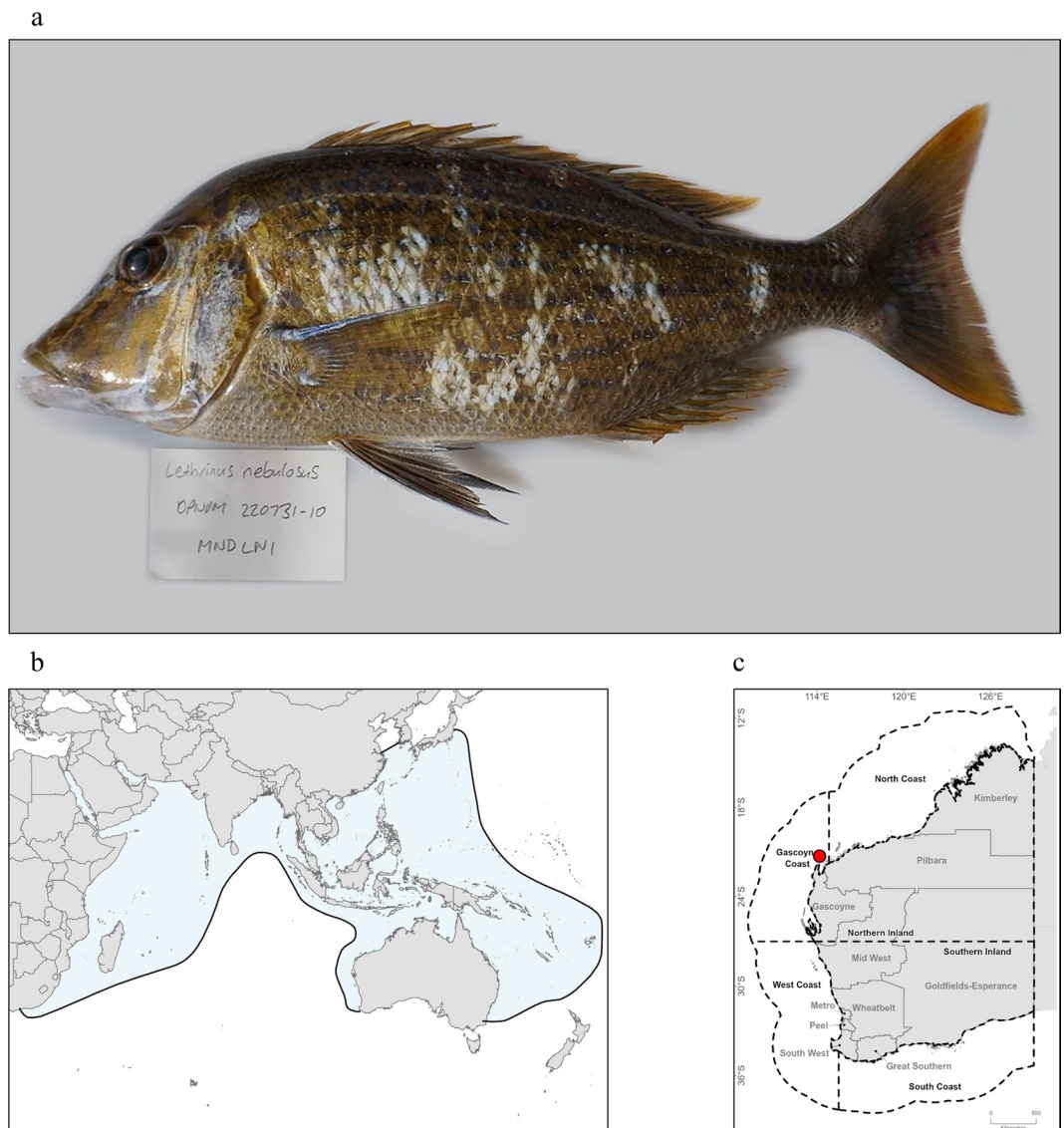


Fig. 1 Voucher image, species distribution and sampling location (a) voucher image of *Lethrinus nebulosus* (Forsskal 1775) specimen (fLetNeb1) sequenced in this study (b) a polygon (blue) of *Lethrinus nebulosus* (Forsskal 1775) distribution range inferred from the Global Biodiversity Information Facility (GBIF)⁴ records and Ocean Biodiversity Information System (OBIS)³⁵ sightings. (c) Map of the Gascoyne bioregion and voucher sampling location of fLetNeb1 marked with a red dot.

long-read sequences (PacBio, CA, USA) scaffolded with Dovetail Genomics Omni-C (Cantata Bio, CA, USA) data and annotated with Illumina short-read (Illumina, CA, USA) transcriptomic data. It represents the first high-quality reference genome for the species and the family Lethrinidae (emperors).

Methods

Sample collection. An adult *L. nebulosus* (356 mm fork length) was caught via fish trap, southwest of Barrow Island, Western Australia (within the Gascoyne Coast Bioregion) in July 2022. Gill and muscle tissue were rapidly aseptically dissected from the specimen to allow for immediate preservation by flash-freezing in liquid nitrogen. Sample collection from this non-model organism was conducted under marine field work conditions and this approach was determined to be the most expedient route to collect fresh flash frozen tissues that were likely to yield ultra-high molecular weight molecules amenable to long-read DNA and RNA sequencing (Department of Primary Industries and Regional Development exemption No. 251003922, see Ethics Declaration). The samples were stored at -80°C until further processing in the laboratory. Initial extraction and QC checks determined that the muscle sample yields insufficient to proceed with downstream workflows, so DNA and RNA extractions and sequencing were performed exclusively on gill tissue. In all cases, a margin was excised and the tissue rinsed thoroughly in deionised water prior to processing to minimise the chance of carryover contamination from parasites or other organisms.

Technology	Data (Gb)	Coverage (×)
PacBio HiFi	35	32
Omni-C	59	54
RNA-Seq	16	

Table 1. Raw sequencing data of *Lethrinus nebulosus* (fLetNeb1).

Genomic DNA extraction, library construction, and sequencing. High molecular weight (HMW) genomic DNA was extracted from approximately 25 mg of gill tissue. Tissue was homogenised and pelleted as per the PacBio Nanobind tissue kit (PacBio, CA, USA) protocol using the TissueRuptor II (QIAGEN, Hilden, Germany). Cell lysis, and DNA isolation was performed following the PacBio “Extracting HMW DNA from skeletal muscle using Nanobind” procedure (102-579-200, Dec 2022). The quantity, and fragment length distribution of extracted gDNA were determined using a Qubit 3 Fluorometer with the Qubit dsDNA Broad-Range Assay Kit (Thermo Fisher Scientific, MA, USA), a NanoDrop One (Thermo Fisher Scientific, MA, USA) and a Femto Pulse with the Genomic DNA 165 kb kit (Agilent, CA, USA). A 12,500 bp PacBio HiFi SMRTbell® library was prepared using the PacBio SMRTbell® prep kit 3.0 according to manufacturer’s instructions. The library was sequenced across one and a half SMRT Cells on a PacBio Sequel IIe, producing 35 Gb of HiFi data with an average read length of 9,842 bp (Table 1).

Chromatin conformation capture proximity ligation library construction and sequencing. Cryogenic grinding of flash-frozen *L. nebulosus* gill tissue in liquid nitrogen was used to facilitate the construction of a chromatin conformation capture proximity ligation library (Hi-C)⁸ using the Dovetail Omni-C proximity Ligation Assay kit, with the Dovetail Omni-C Module and Dovetail Library Module for Illumina kits (Cantata Bio, CA, USA) as per the manufacturer’s protocols. This method of acquiring Hi-C genome conformation data was chosen as it utilises a sequence-independent endonuclease, rather than restriction enzymes, to digest chromatin, providing more uniform sequencing coverage across the genome^{9,10}. Shallow sequencing of the Omni-C libraries was performed on an Illumina iSeq. 100 platform using a 2 × 150 bp paired-end run, generating 725,338 paired reads to assess library complexity. Deep sequencing was then carried out on an Illumina NextSeq. 2000 platform with a 2 × 150 bp paired-end run configuration (Table 1) to generate chromosome conformation data.

RNA extraction, library construction and sequencing. Flash-frozen gill tissue from the same *L. nebulosus* specimen was used for an RNA extraction using the Monarch® Total RNA Miniprep Kit (New England Biolabs, MA, USA) following the manufacturers recommendations. The quantity and quality of extracted RNA was assessed using a NanoDrop One, Qubit HS RNA Kit and TapeStation 4150 system High Sensitivity RNA ScreenTape (Agilent, CA, USA). RNA was then concentrated and purified using the Monarch® RNA Cleanup Kit. Subsequently, an RNA-seq library was constructed using the Illumina Stranded mRNA Prep before sequencing on an Illumina NovaSeq. 6000 using 2 × 150 bp reads on a SP flow cell (Table 1).

Genome assembly and quality assessment. A chromosome-level reference genome of *L. nebulosus* was assembled from the data¹¹ following Vertebrate Genomes Project workflows^{12,13}.

Raw PacBio HiFi reads were assessed for quality and sequencing adapter contamination using HiFiAdapterFilt v2.0¹⁴. Unassembled reads were passed through Meryl v1.3 (k = 31)¹⁵ and GenomeScope2 v2.0¹⁶ to generate a genome profile and genome size estimate of 1.09 Gb (Supplementary Figure S1). HiFiasm v0.19.8¹⁷ was used to produce phased haplotype contig-level assemblies based on filtered HiFi and Omni-C reads.

Omni-C reads were aligned to contig-level assemblies using BWA v0.7.17¹⁸, followed by detection of ligation junctions and removal of PCR duplicates with pairtools v1.0.2¹⁹. A final sorted bam file was generated for each haplotype using samtools v1.16.1²⁰. Sorted bam files containing alignment results of Omni-C reads to contig-level assemblies for each haplotype were produced following the Omni-C mapping pipeline²¹ and used to scaffold the assemblies with YAHS v1.2a.22²². The scaffolded assemblies underwent decontamination using FCS-GX v0.5.4²³ and Tiara v1.0.3²⁴ to eliminate sequences originating from foreign organisms or mitochondria.

PretextMap v0.1.9 (<https://github.com/sanger-tol/PretextMap>) generated Omni-C contact maps using Omni-C read alignments and decontaminated scaffold-level assemblies for each haplotype. Haplotype assemblies were then concatenated for dual manual curation using PretextView v0.2.5 (<https://github.com/sanger-tol/PretextView>). The quality of the final curated genome assemblies was assessed using gfastats v1.3.6²⁵, analysis of Benchmarking Universal Single-Copy Orthologs (BUSCO) v5.4.7²⁶ Actinopterygii genes (actinopterygii_odb10) and Merquy v1.3¹⁵.

The curated *L. nebulosus* genome assembly was 1,092,800,854 bp, with more than 98.9% of the Haplotype 1 and Haplotype 2 scaffolds assigned to chromosomes (Fig. 2, Table 2). BUSCO analysis indicated that an average of 99.35% complete genes were present in the final assemblies, of which 99.1% were single copy (Table 2). Merquy assessment of the base-level accuracy and completeness resulted in QV scores of 61.26 and 61.17 for Haplotype 1 and Haplotype 2, respectively, and a diploid completeness of 99.61% (Table 2). A chromSyn²⁷ synteny plot based on BUSCO collinearity blocks (minregion = 0 minbusco = 1) showed a high degree of synteny between haplotypes (Supplementary Figure S2).

Repeat annotation. A repeat library was generated with RepeatModeler v2.0.5²⁸ using the diploid *L. nebulosus* genome assembly, with the resulting library used to repeat mask the *L. nebulosus* haplotype 1 genome

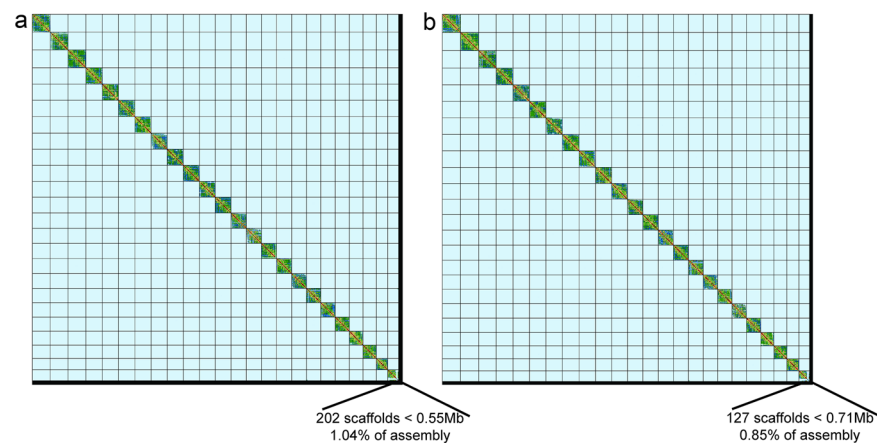


Fig. 2 Omni-C linked contact maps of the final curated genome assembly of *Lethrinus nebulosus* reference genome (fLetNeb1.1) (a) haplotype 1, and (b) haplotype 2. All chromosomes are ordered by length and the statistics detail how many scaffolds were not assigned to chromosomes, the maximum length of these scaffolds (Mb), followed by the percentage of the overall assembly they represent.

	VGP* standards (7.c.P6.Q50.C95)	<i>Lethrinus nebulosus</i>	
		Haplotype 1	Haplotype 2
Total length (Mb)	—	1092.8	
Number of contigs	—	422	395
Number of scaffolds	—	226	151
Contig NG50 (Mb)	>10	23.96	20.84
Scaffold NG50 (Mb)	= Chr. NG50	48.45	47.49
BUSCO complete (%)	>95	99.3% [S:99.1%, D:0.2%]	99.4% [S:99.1%, D:0.3%]
Gaps	<200 per Gb	196	244
Sequences assigned to chromosomes (%)	>95	98.96	99.15
QV	>50	61.26	61.17
		61.21	
Merquy completeness (%)	>95	80.26	80.06
		99.61	

Table 2. Genome assembly statistics of *Lethrinus nebulosus* reference genome (fLetNeb1.1). Notes: BUSCO complete ‘S’ denotes single copy and ‘D’ denotes duplicate. Merged columns reflect combined (i.e., haplotype 1 and haplotype 2) assembly statistics. *VGP; Vertebrate Genomes Project¹³

using RepeatMasker v4.1.5²⁸. The repeat annotation table was generated using the RepeatMasker buildSummary.pl script. Telomeres were predicted with Telociraptor (<https://github.com/slimsuite/telociraptor>) and TIDK (-s AACCT)²⁹ (Supplementary Figure S2). Overall, RepeatMasker identified 452,320,117 bp of repetitive elements, covering 39.91% of the genome (Table 3), these repetitive regions were enriched at the ends of the chromosomes (Fig. 3), as were the assembly gaps (Fig. 2, Supplementary Figure S2).

Gene annotation. The assembled *Lethrinus nebulosus* reference genome (fLetNeb1.1) was preliminarily annotated using the Eukaryotic Genome Annotation Pipeline - External (EGAPx) (<https://github.com/ncbi/egapx>). In total, 24,583 genes (29,167 isoforms) were annotated. The average coding sequence length was 1,858.5 bp with 10.25 exons per gene. The predicted transcriptome was 99.1% complete according to BUSCO (actinopterygii_odb10 n = 3640), with only 23 genes missing. Ribosomal rRNA genes were predicted with Barrnap v0.9 (-kingdom euk) (<https://github.com/tseemann/barrnap>). Complete ribosomal gene clusters were predicted at the ends of four chromosomes: 17, 18, 20 and 24 (Supplementary Figure S2) in addition to 26 unplaced hap1 scaffolds.

A final annotation of the *Lethrinus nebulosus* reference genome (fLetNeb1.1) will become available based on the NCBI RefSeq pipeline, accessible via umbrella BioProject PRJNA1074345 (<https://www.ncbi.nlm.nih.gov/bioproject/1046164>).

Data Records

All data (sequencing and assemblies) for the *Lethrinus nebulosus* reference genome (fLetNeb1) are deposited in NCBI under the umbrella BioProject PRJNA1074345. The PacBio HiFi, Dovetail Omni-C and Illumina RNA-seq data is accessioned under BioProject PRJNA1107547 and can be found at <https://identifiers.org/ncbi/insdc.sra:SRP505806> (2024)¹¹. The assembled genome and assembly information is deposited in NCBI GenBank

	Number of elements	Length (bp)	Percentage in genome (%)
Short interspersed nuclear elements (SINEs)	38,508	5,487,653	0.48
Long interspersed nuclear elements (LINEs)	134,447	37,170,753	3.28
Long terminal repeats (LTRs)	44,005	9,666,075	0.85
DNA transposon	770,686	144,634,861	12.76
Rolling circles	20,005	10,498,883	0.93
Unclassified	1,469,066	216,065,075	19.07
Small RNA	35,041	6,665,110	0.59
Satellites	15,617	2,746,411	0.24
Simple repeats	438,101	20,092,114	1.77
Low complexity	52,272	3,102,515	0.27

Table 3. Repetitive elements of the *Lethrinus nebulosus* reference genome (fLetNeb1.1) haplotype 1 assembly.

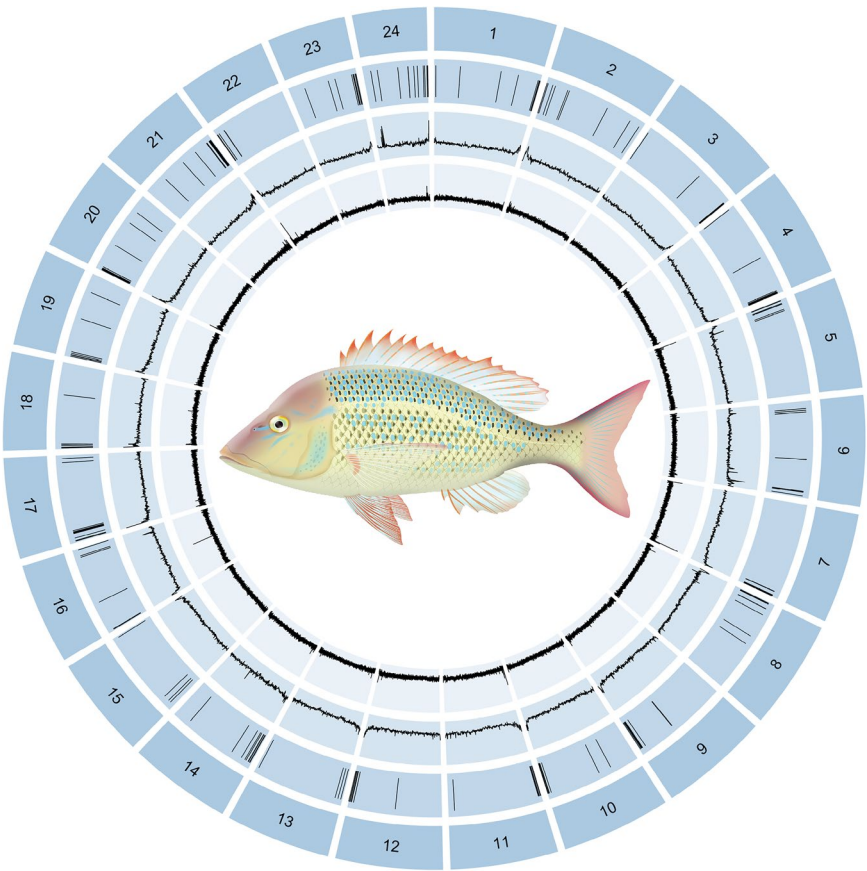


Fig. 3 Features of the phased haplotype 1 chromosomes of *Lethrinus nebulosus* reference genome (fLetNeb1.1). Concentric tracks from the outside inward represent chromosomes (numbered by length), gaps (gaps of unknown length appear as 100 bp in the assembly), GC content calculated using BEDTools v2.31.1³⁶ using a sliding window of 10,000 bp, and repeat density calculated using RepeatMasker v4.1.5 with a 10,000 bp sliding window. Visualisation created using the R package circlize v0.4.16^{37,38}. Fish graphic © Marinewise.com.au.

and is accessible at https://identifiers.org/ncbi/insdc.gca:GCA_045362495.1³⁰ (2024) and https://identifiers.org/ncbi/insdc.gca:GCA_045362485.1 (2024)³¹.

Technical Validation

Molecular validation of nominal specimen identity. Available genetic data indicates that *L. nebulosus* constitutes a single biological stock in Western Australia³². Nevertheless, cryptic diversity is recognised in other regions³³ across the species distribution range (Fig. 1). With this in mind, we performed a molecular validation of the nominal identity of our samples and voucher specimen, to confirm they represent *L. nebulosus* as opposed to cryptic diversity in the lineage. This step also served to confirm that the datasets contributing to the assembly (Omni-C and PacBio HiFi) matched the nominal species identity (i.e. internal quality control check to confirm

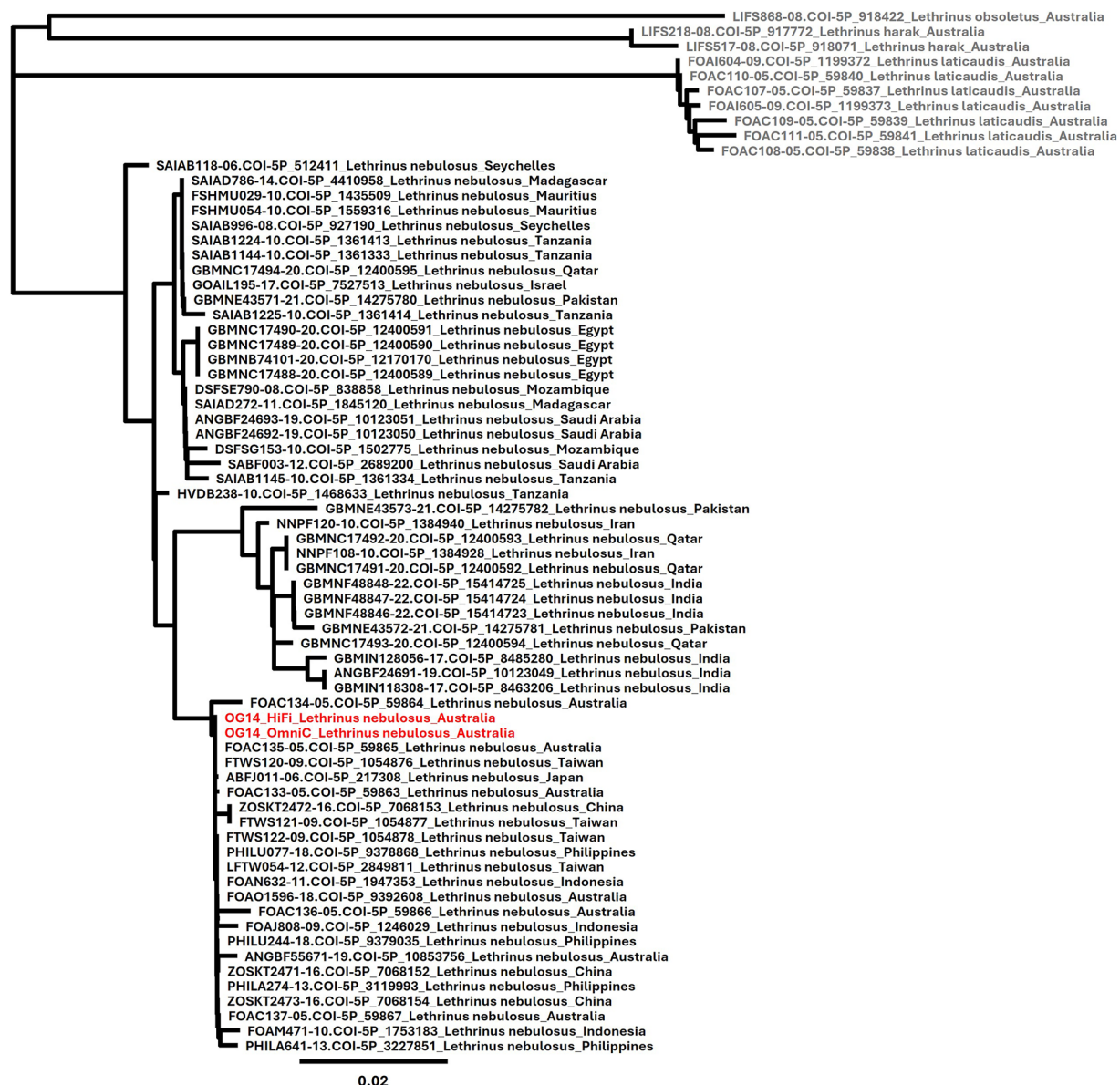


Fig. 4 Neighbour-joining tree (TrN + p) showing clustering of *Lethrinus nebulosus* COI haplotypes sampled throughout the species distribution range. Major nodes were supported with >50% bootstrap replicates. The barcodes recovered from data generated for the reference genome and specimen reported herein are shown in red.

that no tube or data swaps had occurred during sample processing). Complete mitochondrial genomes were assembled from both the PacBio HiFi and Omni-C data. The Cytochrome Oxidase I (COI) barcoding gene was mined from the mitochondrial genomes and placed in a phylogeny including all representatives of *L. nebulosus* that were available on The Barcode of Life Data Systems (BOLD; [v4.boldsystems.org](https://www.boldsystems.org)) database that included sampling location information. Sequences ($n = 70$) were trimmed to a consistent 472 bp region and aligned across all taxa, plus representatives of *Lethrinus obsoletus*, *Lethrinus harak* and *Lethrinus laticaudis* as outgroups, using Geneious Prime® 2024.0.7. Phylogenetic analyses were performed in PAUP* v4.0c. Internal model test options identified the Tamura–Nei (TrN) + γ gamma distribution (0.311) as the best model of sequence evolution. A neighbour-joining tree was estimated and bootstrapped with 1000 replicates (Fig. 4).

The analysis clustered *L. nebulosus* diversity into three major clades; the western Indian Ocean and Red Sea, the northern Indian Ocean, and the central Indo-Pacific (Fig. 4). Identical COI haplotypes were recovered from both the PacBio HiFi and Omni-C datasets, these clustered with other representatives nominally identified as *L. nebulosus* from Australia and the Indo-Pacific. We therefore feel comfortable that this reference genome represents *L. nebulosus* from Western Australia.

Quality control of nucleic acids and extracts. The quality and quantity of extracted DNA was assessed using a Qubit 3 Fluorometer, a NanoDrop One Spectrophotometer and an Agilent Femto Pulse. The DNA extract

was determined to be of high molecular weight, with a concentration 326 ng/μl, a Qubit:NanoDrop ratio of 0.84 and an average length of 21,681 bp. The quality of the Omni-C lysate was assessed using a Qubit 3 Fluorometer and 4150 TapeStation system. The Omni-C lysate had a yield of 938 ng and a 91% Chromatin Digestion Efficiency (CDE) metric between the recommended range of 100–2500 bp. RNA quality and quantity was determined using a Qubit 3 Fluorometer and a 4150 TapeStation system. The RNA extract had a concentration of 162 ng/μl and an RNA integrity number (RIN) value of 7.4.

Evaluation of the genome assembly. The quality of the haplotype phased *Lethrinus nebulosus* reference genome (fLetNeb1.1) assembly was calculated using gfastats v1.3.6²⁵, Merqury v1.3¹⁵ and BUSCO v 5.4.7²⁶ using the Actinopterygii lineage database (actinopterygii_odb10). Gfastats v1.3.6 indicated that the N50 (a measure of contiguity) of each haplotype was greater than 20 Mb, with more than 98% of the sequences assigned to chromosomes (Table 2). In addition, both Merqury and BUSCO determined that the *Lethrinus nebulosus* reference genome (fLetNeb1.1) is more than 99% complete (Table 2). DepthKopy v1.5.1³⁴ was used to establish predicted copy numbers for genes, scaffolds, and 100 kb windows based on the single-copy read depth (33.0×) calculated from Complete BUSCO genes. DepthKopy profiles for Complete and Duplicated BUSCO genes are very similar (Supplementary Figure S3) with 27/29 (93.1%) Duplicated BUSCO genes predicted to be genuine biological gene duplications. Profiles for annotated genes and sliding 100 kb windows both confirm that the majority of the genome has been assembled without duplication or collapse (Supplementary Figure S3). Together these metrics indicate that the haplotype phased assembly obtained for *Lethrinus nebulosus* reference genome (fLetNeb1) is high-quality and meets standards, and quality metrics of the Earth BioGenome Project³⁵ and the Vertebrate Genomes Project¹³.

Ethics declaration. In Western Australia, the Animal Welfare Act 2002 does not require the Department of Primary Industries and Regional Development (DPIRD) to obtain a permit to use animals for scientific purposes unless the species are outside the provisions of the Fish Resources Management Act 1994 and Fish Resources Management Regulations 1995. Nonetheless, sampling was undertaken in strict adherence to the DPIRD Policy for the handling, use and care of marine fauna for research purposes.

Code availability

No custom scripts or code was used in the validation of the dataset. The data analyses used standard bioinformatic tools specified in the methods. The data and code used to generate the figures is available on GitHub: https://github.com/MinderooFoundation/Data_desc_fLetNeb1.

Received: 15 October 2024; Accepted: 20 February 2025;

Published online: 13 March 2025

References

1. Froese, R. & Pauly, D. *Fishbase. World Wide Web electronic publication. Version (02/2024)*, www.fishbase.org (2024).
2. Wakefield, C. *et al.* Spangled Emperor (2023). (Fisheries Research and Development Corporation, 2023).
3. Cresswell, A. K. *et al.* Disentangling the response of fishes to recreational fishing over 30 years within a fringing coral reef reserve network. *Biol Conserv* **237**, 514–524, <https://doi.org/10.1016/j.biocon.2019.06.023> (2019).
4. Secretariat, G. GBIF: GBIF Backbone Taxonomy. <https://doi.org/10.15468/39omei>. Accessed via <https://www.gbif.org/species/2374885> (2024).
5. Berry, O., England, P., Marriott, R. J., Burrige, C. P. & Newman, S. J. Understanding age-specific dispersal in fishes through hydrodynamic modelling, genetic simulations and microsatellite DNA analysis. *Molecular Ecology* **21**, 2145–2159, <https://doi.org/10.1111/j.1365-294X.2012.05520.x> (2012).
6. Taylor, R. S. *et al.* High genetic load without purging in caribou, a diverse species at risk. *Curr Biol* **34**, <https://doi.org/10.1016/j.cub.2024.02.002> (2024).
7. Vandewoestijne, S., Schtickzelle, N. & Baguette, M. Positive correlation between genetic diversity and fitness in a large, well-connected metapopulation. *Bmc Biol* **6**, <https://doi.org/10.1186/1741-7007-6-46> (2008).
8. Belton, J. M. *et al.* Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* **58**, 268–276, <https://doi.org/10.1016/j.ymeth.2012.05.001> (2012).
9. Liu, N. *et al.* Seeing the forest through the trees: prioritising potentially functional interactions from Hi-C. *Epigenet Chromatin* **14**, <https://doi.org/10.1186/s13072-021-00417-4> (2021).
10. Yamaguchi, K. *et al.* Technical considerations in Hi-C scaffolding and evaluation of chromosome-scale genome assemblies. *Molecular Ecology* **30**, 5923–5934, <https://doi.org/10.1111/mec.16146> (2021).
11. Parata, L. *et al.* *Lethrinus nebulosus* (Spangled Emperor), fLetNeb1, sequence data, <https://identifiers.org/ncbi/insdc.sra:SRP505806> (2024).
12. Larivière, D. *et al.* Scalable, accessible and reproducible reference genome assembly and evaluation in Galaxy. *Nature Biotechnology* **42**, 367–370, <https://doi.org/10.1038/s41587-023-02100-3> (2024).
13. Rhie, A. *et al.* Towards complete and error-free genome assemblies of all vertebrate species. *Nature* **592**, 737–746 (2021).
14. Sim, S. B., Corpuz, R. L., Simmonds, T. J. & Geib, S. M. HiFiAdapterFilt, a memory efficient read processing pipeline, prevents occurrence of adapter sequence in PacBio HiFi reads and their negative impacts on genome assembly. *BMC Genomics* **23**, 157, <https://doi.org/10.1186/s12864-022-08375-1> (2022).
15. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome biology* **21**, 1–27 (2020).
16. Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications* **11**, 1432, <https://doi.org/10.1038/s41467-020-14998-3> (2020).
17. Cheng, H. *et al.* Haplotype-resolved assembly of diploid genomes without parental data. *Nature Biotechnology* **40**, 1332–1335, <https://doi.org/10.1038/s41587-022-01261-x> (2022).
18. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595, <https://doi.org/10.1093/bioinformatics/btp698> (2010).
19. Abdennur, N. *et al.* Pairtools: From sequencing data to chromosome contacts. *Plos Comput Biol* **20**, <https://doi.org/10.1371/journal.pcbi.1012164> (2024).

20. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10** <https://doi.org/10.1093/gigascience/giab008> (2021).
21. D Genomics. *Omni-C: From fastq to final valid pairs bam file*, (2021).
22. Zhou, C., McCarthy, S. A. & Durbin, R. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics* **39**, btac808, <https://doi.org/10.1093/bioinformatics/btac808> (2023).
23. Astashyn, A. *et al.* Rapid and sensitive detection of genome contamination at scale with FCS-GX. *Genome Biology* **25** <https://doi.org/10.1186/s13059-024-03198-7> (2024).
24. Karlicki, M., Antonowicz, S. & Karnkowska, A. Tiara: deep learning-based classification system for eukaryotic sequences. *Bioinformatics* **38**, 344–350, <https://doi.org/10.1093/bioinformatics/btab672> (2022).
25. Formenti, G. *et al.* Gfastats: conversion, evaluation and manipulation of genome sequences using assembly graphs. *Bioinformatics* **38**, 4214–4216, <https://doi.org/10.1093/bioinformatics/btac460> (2022).
26. Manni, M., Berkeley, M. R., Seppey, M., Simao, F. A. & Zdobnov, E. M. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular Biology and Evolution* **38**, 4647–4654, <https://doi.org/10.1093/molbev/msab199> (2021).
27. Edwards, R. J., Dong, C., Park, R. F. & Tobias, P. A. A phased chromosome-level genome and full mitochondrial sequence for the dikaryotic myrtle rust pathogen, *Austropuccinia psidii*. *BioRxiv*, 2022.2004. 2022.489119 (2022).
28. Flynn, J. M. *et al.* RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences* **117**, 9451–9457 (2020).
29. A Telomere Identification Toolkit. *Zenodo*. <https://doi.org/10.5281/zenodo.10091385> (2023).
30. Parata, L. *et al.* *Lethrinus nebulosus* (Spangled Emperor). *Genbank* https://identifiers.org/ncbi/insdc.gca:GCA_045362495.1 (2024).
31. Parata, L. *et al.* *Lethrinus nebulosus* (Spangled Emperor). *Genbank* https://identifiers.org/ncbi/insdc.gca:GCA_045362485.1 (2024).
32. Johnson, M., Hebbert, D. & Moran, M. Genetic analysis of populations of north-western Australian fish species. *Marine and Freshwater Research* **44**, 673–685, <https://doi.org/10.1071/MF9930673> (1993).
33. Healey, A. J. E. *et al.* Genetic analysis reveals harvested *Lethrinus nebulosus* in the Southwest Indian Ocean comprise two cryptic species. *ICES Journal of Marine Science* **75**, 1465–1472, <https://doi.org/10.1093/icesjms/fsx245> (2018).
34. Chen, S. H. *et al.* Chromosome-level de novo genome assembly of (New South Wales waratah) using long-reads, linked-reads and Hi-C. *Molecular Ecology Resources* **22**, 1836–1854, <https://doi.org/10.1111/1755-0998.13574> (2022).
35. Lewin, H. A. *et al.* Earth BioGenome Project: Sequencing life for the future of life. *P Natl Acad Sci USA* **115**, 4325–4333, <https://doi.org/10.1073/pnas.1720115115> (2018).
36. OBIS. *Ocean Biodiversity Information System*. Intergovernmental Oceanographic Commission of UNESCO, <https://obis.org/> (2024).
37. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842, <https://doi.org/10.1093/bioinformatics/btq033> (2010).
38. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize Implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812, <https://doi.org/10.1093/bioinformatics/btu393> (2014).

Acknowledgements

We acknowledge traditional owners throughout Western Australia, and their continuing connection to the land, sea, and community, including those with connections to Barrow Island from where this specimen was collected. We pay our respects to all Aboriginal people and communities and their cultures, and to Elders both past and present. Logistical support for this project was provided by the Department of Primary Industries and Regional Development (DPIRD), Government of Western Australia. We gratefully acknowledge the crew of RV *Naturaliste* (DPIRD) for their contributions to sample acquisition. Data described herein was generated via generous support of Minderoo Foundation and resources provided by the Pawsey Supercomputing Research Centre with funding from the Australian Government and the Government of Western Australia.

Author contributions

L.P. contributed to data acquisition, genome assembly and curation, data interpretation, drafting and critical review of the manuscript. E.D.J. contributed to genome assembly and curation, and critical review of the manuscript. A.D. and L.A. contributed to data acquisition and critical review of the manuscript. R.J.E. contributed to genome assembly and curation, data interpretation and critical review of the manuscript. S.C. contributed to project administration, sample acquisition, data interpretation and critical review of the manuscript. C.L.S., C.B.W. and S.J.N. contributed sample acquisition, specimen identification and preservation, and critical review of the manuscript. S.D.P. contributed data interpretation and critical review of the manuscript. All authors read and approved the final manuscript. OceanOmics Division and OceanOmics Centre are a consortium of authors contributing to infrastructure, program administration, and delivery that facilitated the production of this dataset.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-025-04690-w>.

Correspondence and requests for materials should be addressed to L.P.

Reprints and permissions information is available at www.nature.com/reprints.

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



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

© The Author(s) 2025

OceanOmics Centre

Anna Depiazzi¹, Ibrahim Faseeh¹, Lauren Huet¹, Sang Huynh¹, Laura Missen¹ and Tyler Peirce¹

OceanOmics Division

Marcelle E. Ayad³, Philipp E. Bayer^{1,3}, Adam J. Bennett^{1,3}, Stephen J. Burnell^{1,3}, Madalyn K. Cooper³, Matthew W. Fraser^{1,3}, Priscila Goncalves^{1,3}, Anya Kardailsky^{1,3}, Georgia M. Nester^{1,3}, Jessica Pearce^{1,3}, Eric J. Raes^{1,3}, Sebastian Rauschert^{1,3}, Julie C. Robidart^{1,3} & Ebony M. Thorpe^{1,3}