

SCIENTIFIC DATA

OPEN

DATA DESCRIPTOR

The survey and reference assisted assembly of the *Octopus vulgaris* genome

Ilaria Zarrella¹, Koen Herten^{2,3}, Gregory E. Maes^{3,4,5}, Shuaishuai Tai⁶, Ming Yang⁶, Eve Seuntjens⁷, Elena A. Ritschard⁸, Michael Zach⁸, Ruth Styfhals^{7,9}, Remo Sanges⁹, Oleg Simakov⁸, Giovanna Ponte^{1,9} & Graziano Fiorito⁹

Received: 18 October 2018

Accepted: 4 February 2019

Published online: 01 April 2019

The common octopus, *Octopus vulgaris*, is an active marine predator known for the richness and plasticity of its behavioral repertoire, and remarkable learning and memory capabilities. Octopus and other coleoid cephalopods, cuttlefish and squid, possess the largest nervous system among invertebrates, both for cell counts and body to brain size. *O. vulgaris* has been at the center of a long-tradition of research into diverse aspects of its biology. To leverage research in this iconic species, we generated 270 Gb of genomic sequencing data, complementing those available for the only other sequenced congeneric octopus, *Octopus bimaculoides*. We show that both genomes are similar in size, but display different levels of heterozygosity and repeats. Our data give a first quantitative glimpse into the rate of coding and non-coding regions and support the view that hundreds of novel genes may have arisen independently despite the close phylogenetic distance. We furthermore describe a reference-guided assembly and an open genomic resource (CephRes-gdatabase), opening new avenues in the study of genomic novelties in cephalopods and their biology.

Background & Summary

Octopus vulgaris is a benthic, neritic species belonging to the class Cephalopoda. It occurs from the coastline to the outer edge of the continental shelf, inhabiting various marine habitats at depths spanning from 0 to 200 m. *O. vulgaris* is one of the most widely distributed species belonging to the genus, and is an important commercially harvested resource for human consumption. It is found worldwide in temperate and tropical waters¹⁻³. Throughout its distribution range, the animal undertakes limited seasonal migrations: mostly found in deep waters in winter and shallow waters in summer.

O. vulgaris is perhaps the most famous and best studied of all octopus species, largely due to the initiative of Professor John Z. Young^{4,5}. Since the late 1940s, Young carried out at the Stazione Zoologica Anton Dohrn of Napoli (Italy) a systematic analysis of the neural structures underlying behavioural plasticity in this animal^{6,7}. Based on this contribution, the anatomy of *O. vulgaris* nervous system⁸ and its physiology and life history^{4,9-11} have been well characterized. It is the phenomenological proximity of behavioral traits and phylogenetic distance in respect to higher vertebrates that guaranteed the short, but wide success of cephalopods^{5,10}. *O. vulgaris* in particular became a “model of the brain”^{12,13}, and more recently a case for studying the evolution of cognition

¹Association for Cephalopod Research ‘CephRes’, Napoli, Italy. ²Genomics Core, KU Leuven, Leuven, Belgium.

³Laboratory for Cytogenetics and Genome Research, Center for Human Genetics, KU Leuven, Leuven, Belgium.

⁴Laboratory of Biodiversity and Evolutionary Genomics, KU Leuven, Leuven, Belgium. ⁵Centre for Sustainable

Tropical Fisheries and Aquaculture, Comparative Genomics Centre, College of Science and Engineering, James Cook

University, Townsville, 4811, QLD, Australia. ⁶BGI Genomics, BGI-Shenzhen, Shenzhen, 518083, China. ⁷Laboratory of

Developmental Neurobiology, Department of Biology, KU Leuven, Leuven, Belgium. ⁸Department of Molecular

Evolution and Development, University of Vienna, Vienna, Austria. ⁹Department of Biology and Evolution of Marine

Organisms, Stazione Zoologica Anton Dohrn Napoli, Napoli, Italy. These authors contributed equally: Ilaria Zarrella,

Koen Herten, Shuaishuai Tai and Ming Yang. These authors jointly supervised this work: Gregory E. Maes, Oleg

Simakov, Giovanna Ponte and Graziano Fiorito. Correspondence and requests for materials should be addressed to

O.S. (email: oleg.simakov@univie.ac.at) or G.F. (email: graziano.fiorito@szn.it)

in invertebrates^{7,14–19}. Researchers still use *O. vulgaris* as an organism to study behavioural and neural plasticity including learning and memory recall^{5,12,20}, regeneration^{21–24} and sophisticated cognition^{7,14–17,25}.

Currently available genomic resources for molluscs are scarce, considering the species abundance and the commercial value of the phylum Mollusca. Publicly available molluscan genomes include a dozen representatives from bivalves, gastropods^{26–42} and to-date only three cephalopods, namely the California two-spot octopus *Octopus bimaculoides*⁴³ and, more recently, for *Callistoctopus minor*⁴⁴ and *Euprymna scolopes*⁴⁵.

Although the first step towards cephalopod genetics was made over 30 years ago⁴⁶, cephalopod research is only slowly entering the genomics era^{10,47}. Obtaining high quality cephalopod genomes has been impeded due to their large size (e.g., *O. bimaculoides*: 2.7 Gb; Gregory, 2018 - Animal Genome Size Database, <http://www.genomesize.com>), heterozygosity and high abundance of repeat regions^{43,47,48}. However, several collaborative genome projects are currently underway for a variety of cephalopod species such as the nautilus, *Sepia officinalis*, *Idiosepius paradoxus* and *Doryteuthis pealeii*.

Cephalopods arose more than 500 Mya and diverged into over 800 current living species with highly diversified life styles and body plans⁴⁸. Translocations, duplications, exon shuffling and gene conversions occurred within the cephalopod genome during evolution, which might explain the development of different morphological novelties, such as the prehensile arms, the unique jet propulsion system, the ink sac and sophisticated sensory and neural systems⁴⁹. The analysis of *O. bimaculoides* genome revealed an extensive expansion of particular gene families, including protocadherins and the C2H2 superfamily of zinc-finger transcription factors⁴³, as well as novel octopus-specific genes expressed in specialized structures such as suckers, skin and brain (for review see also Shigeno *et al.*¹⁸). These genome-level novelties are accompanied by other sophisticated innovations such as extensive RNA editing, particularly in the nervous system cells^{50–52}. Furthermore, partial genome sequencing of several cephalopods showed that repeat elements, in particular transposable elements, are abundant^{53,54}. Indeed, the genome of *O. bimaculoides* revealed that over 45% of the genome is comprised of repetitive elements⁴³.

The study of cephalopod biological innovations^{10,18,43,55} is driven by the unique scientific value of these animals for evolutionary genomics, neuroscience and cognition^{7,10,18,25,43,55–58} which continues the heritage of the discovery of the action potential in the squid giant axon, a seminal contribution to neuroscience⁵⁹. Furthermore, the phylogenetic relationships within the cephalopods have not yet been fully elucidated and biological research would benefit from more cephalopod genomes^{60,61}.

In line with those previous and current efforts, and to promote data sharing among cephalopod researchers^{10,47}, we present the sequence and draft assembly of the common octopus, *Octopus vulgaris*, genome. It is noteworthy to report that the two species (*i.e.*, *O. vulgaris* and *O. bimaculoides*), although both belonging to the same genus, go through a substantially different life cycle since the paralarval stage is absent in *O. bimaculoides*⁶². Therefore, the two species represent different biological and physiological adaptations among closely related species. The genomic sequencing of both octopus species and our online platform to browse these data will allow for future comparative genomics studies, revealing key genomic innovations and facilitating the discovery of the molecular basis of intricate processes such as learning, regeneration and the evolution of complex brains.

Methods

Genomic DNA preparation. An adult male belonging to the species *O. vulgaris* Cuvier, 1797 (450 g body weight) was caught by fishermen from the Bay of Naples in 2011^{1,2} and immediately humanely-killed^{63,64}. Given the high rate of heterozygosity in marine organisms^{65,66}, tissue from a single individual was used to extract the genomic DNA (to avoid contamination, spermatophores were used). Spermatophores in octopus are stored within the Needham's sac, structure that was dissected following Chapko and coworkers⁶⁷. Tissue (124 mg) was used to extract the genomic DNA following the recommended phenol-chloroform extraction protocol by the Beijing Genomics Institute (BGI)-Shenzhen. Briefly, tissue lysis occurred overnight at 56 °C after adding 3.0 ml of lysis buffer containing proteinase K (300 µg; Sigma-Aldrich, Saint Louis, Missouri, United States) and RNase A (100 µg; Sigma-Aldrich, Saint Louis, Missouri, United States). DNA was then extracted with phenol (2X), phenol:chloroform, chloroform and was subsequently precipitated. Genomic DNA was dissolved in TE buffer to reach a final concentration of 1 µg/µl.

Genome sequencing and quality control. A total of four genomic DNA libraries (with different insert sizes: 170, 250, 500 and 800 bp) were constructed following the Illumina library preparation protocols. Briefly, to construct the paired-end libraries DNA was fragmented by Adaptive Focused Acoustics technology (Covaris) and tested via gel-electrophotometry, the fragmented DNA combined with End Repair Mix (20 °C for 30 min). After purification, DNA ends were blunted and an A base was added to the 3' ends. DNA adaptors with a single T-base 3'-end overhang were ligated to the above products. Ligation products were purified on 2% agarose gels to recover the target fragments and were purified from the gels (Qiagen Gel Extraction kit, 28704). Several rounds of PCR amplification with PCR Primer Cocktail and PCR Master Mix were performed to enrich the Adapter-ligated DNA fragments. Then the PCR products selected by running another 2% agarose gel to recover the target fragments and the gel purified (QIAquick Gel Extraction kit, QUIAGEN). The final library was quantified by assessing the average molecule length (Agilent 2100 Bioanalyzer), and by Real-Time qRT-PCR. A total of 277 Gb of raw data were generated by Illumina HiSeq 2000 at BGI.

All libraries were sequenced in a paired-end mode with read lengths of 100 bp or 150 bp. Reads were filtered and trimmed (100 bp to 95 bp, 150 bp to 145 bp) using SOAPnuke software (<https://github.com/BGI-flexlab/SOAPnuke>)⁶⁸ which yielded 250 Gb of data. Low-quality reads, reads with adaptor sequences and duplicated reads were filtered, and if the quality of bases at the head or tail of the reads was low, we directly trimmed them from 100 bp to 95 bp (PE100) or from 150 bp to 145 bp (PE150). The remaining high-quality data were used in the further analysis. *SGA PreQC v0.10.14*⁶⁹ modules were run per library and on the combined libraries to estimate various genome parameters (Table 1 and Table 2).

Library ID	Insert Size(bp)	Read Length (bp)	Data (Gb)	Sequence Depth (X)
SZAXPI006102-158	170	100	82.15	29.34
SZAXPI006612-13	250	150	52.25	18.66
SZAXPI005989-166	500	100	62.05	22.16
SZAXPI005988-169	800	100	53.59	19.14
Total	—	—	250.04	89.30

Table 1. Main statistics from *O. vulgaris* sequencing data.

K-mer_num	Peak_depth	Genome Size	Used Bases	Used Reads
212,679,899,304	76	2,798,419,727	249,873,643,000	2,324,608,981

Table 2. k-mer = 17 raw read statistics for *Octopus vulgaris* genome data.

	# scaffolds	genome size	N50/L50	N75/L75	Ns/100 kbp	Complete BUSCOs	Fragmented BUSCOs
ABYSS k41 scaffolds	26,350,077	3.30 Gb	1,488 bp 199,442	767 bp 503,977	979.41	112	50
ABYSS k81 scaffolds	8,918,381	3.31 Gb	2,627 bp 195,104	980 bp 496,991	706.92	275	286
Redundans k81	1,157,969	2.10 Gb	3,958 bp 149,577	2,126 bp 330,514	3,961.18	390	319
Chromosomer k81	77,683	1.78 Gb	263,097 bp 1,607	56,379 bp 5,018	19,504.19	505	88
<i>O. bimaculoides</i>	151,674	2.34 Gb	485,615 bp 1,300	215,581 bp 3,077	15,346.35	773	28

Table 3. Assembly statistics for *Octopus vulgaris*. Statistics were generated with QUAST and a default threshold of 500 bp. See text for details.

Draft genome assembly. We applied Assembly By Short Sequencing 2.0.2 (ABYSS^{70,71}) for both k-mer sizes that were suggested by SGA PreQC. The quality of assemblies (ABYSS kmer41 and ABYSS kmer81) was evaluated by QUAST 4.3⁷². A summary of various statistics is shown in Table 3. Based on the QUAST analysis the optimal kmer size for the ABYSS assembly was estimated to be 81. Since a higher heterozygosity rate of the genome was predicted based on these initial results, the Redundans 0.13 c⁷³ tool was used to reduce the number of ABYSS contigs from the initial assemblies. Redundans reduces contigs by removing highly similar contigs. These highly similar contigs are originally the different alleles of the same genomic position, but are too different for the De Bruijn graph method to be assembled into the same contig (too much variation inside one kmer). Redundans collapses and scaffolds these reduced contigs into single genomic locations. Redundans reduced the number of scaffolds of the draft genome over seven (7) times, while improving assembly statistics (see Table 3).

Reference Assisted Scaffolding. Given the availability of a relatively good reference genome of a related species (*O. bimaculoides*)⁴³, a reference assisted scaffolding tool was used to optimize the genome. The reduced scaffolds were aligned to the *O. bimaculoides* genome using blastn⁷⁴ of the blast+ toolkit 2.8.0-alpha. These alignments were used by chromosomer 0.1.3 (<https://github.com/gtamazian/Chromosomer>) to scaffold the reduced scaffolds according to the given genome.

Assessment of draft genomes. An assessment of the draft genomes (ABYSS, Redundans and chromosomer) was performed by looking for the highly conserved genes using BUSCO 3.0.2⁷⁵. The Metazoa odb9 database was used, supplying 978 orthologs. The number of complete orthologs increased with each improvement of the assembly (Table 3), confirming the gain in assembly quality of the final chromosomer version. The final genome build has over 50% complete BUSCOs, and 10% fragmented BUSCOs (orthologs found, but scattered over multiple scaffolds).

Data Records

The draft genome(s) of *O. vulgaris* as shown in Table 3 has been made publicly available on the genome browser and data repository of the Association for Cephalopod Research that initiated this work (http://www.cephalopodresearch.org/ceph_gdatab/) in collaboration with the Department of Molecular Evolution and Development, University of Vienna. This web resource is based on the browser originally designed by University of California, Santa Cruz (UCSC)⁷⁶ and will be maintained and curated to keep track of all present and upcoming octopus genomes. It includes comparative genomics tracks such as read mapping and whole genome alignment between the two octopus species. Raw reads have also been deposited to the NCBI SRA⁷⁷. The reference-guided assembly has been deposited at GenBank⁷⁸ and its original version is also provided in the associated FigShare record (chromosomer.fa) together with its annotation (gene_models.chromosomer.gff), and other assemblies listed in Table 3 (*Octopus vulgaris* genome assemblies⁷⁹). Table 2 and Table 3 summarize statistics about *O. vulgaris* genome as deduced from our current sequencing data and Fig. 1 shows the kmer (17mer) distribution determining the overall sequencing depth (Table 1 and 2).

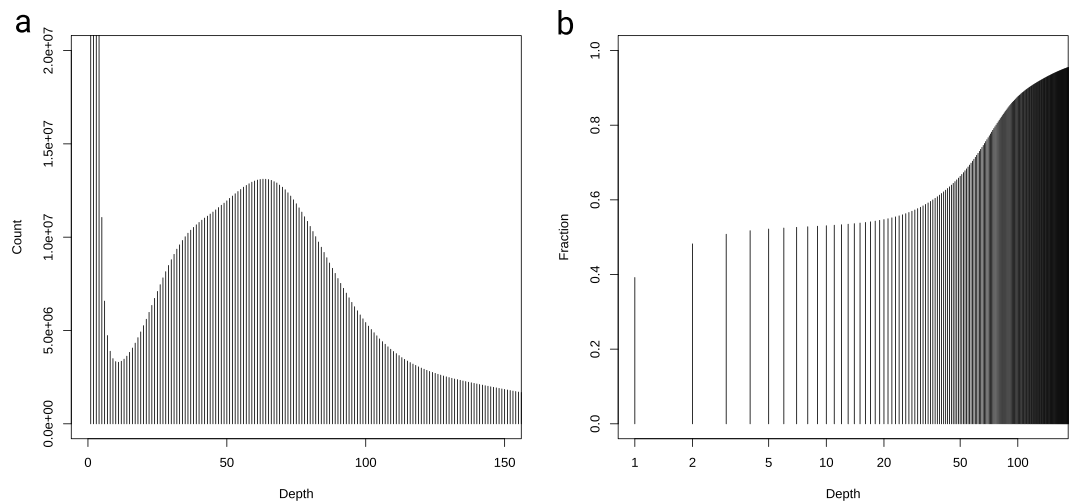


Fig. 1 Sequencing depth and genome repetitiveness estimation from 17mer counts in the raw read data. **(a)** 17mer depth analysis using raw data showing elevated levels of heterozygosity. **(b)** Cumulative proportion of 17mers as a function of their depth showing that at least half of the genome occurs at depth 10 or more.

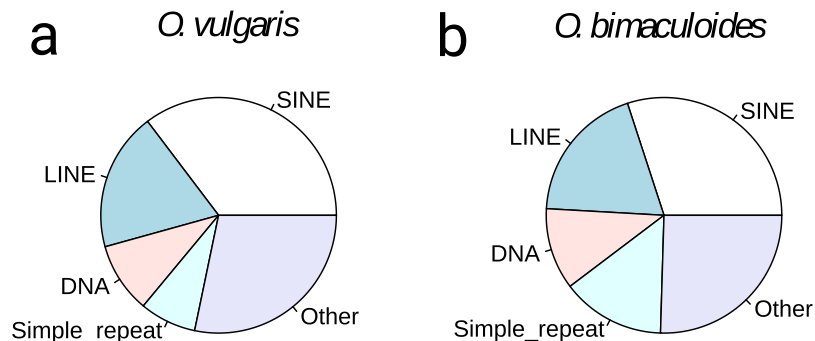


Fig. 2 Proportions of the most abundant repetitive element classes in *Octopus vulgaris* compared to *Octopus bimaculoides* based on the *ab initio* reconstruction of repetitive elements using the DNAPipeTE pipeline. **(a)** Repeat proportions in the *Octopus vulgaris* genome. **(b)** Repeat proportions in the *Octopus bimaculoides* genome. In both genomes, SINE elements are the most abundant repeat classes. While the total number of repeats is similar in both genomes, differences in the proportions can be attributed to individual expansions of repeat elements that occurred independently in both lineages.

Technical Validation

Quality control. The quantity and integrity of the genomic DNA was analysed via agarose gel electrophoresis and with a NanoDrop spectrophotometer (Thermo Fisher Scientific; concentration of 1 µg/µl, $A_{260}/A_{280} = 1.84$ and $A_{230}/A_{260} = 2.2$). DNA integrity was analysed with Agilent Bioanalyzer 2100.

Quality control DNA library. To assess the quality of Illumina reads FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc) was performed on all raw data. *Trimmomatic v0.36*⁸⁰ was not able to identify any significant adaptor sequence contamination within the raw data. The data were mapped to the PhiX control library (Illumina, Inc) using Bowtie2 v2.3.4⁸¹ and no matches were found.

Sequencing depth assessment. We used jellyfish 2.2. 10⁸² on the raw read data using kmer size of 17bp. This resulted in a depth of sequencing histogram (Fig. 1) showing sequencing depth peak of around 76x. Using the kmer depth curve and the cumulative read depth (Fig. 1), repetitiveness, and heterozygosity was conducted independent of the genome assemblies (see Tables 2 and 3). The genome was estimated to be around 2.4 Gb in length with a relatively high heterozygosity rate (>1.1%) and large repetitiveness (>50%).

Genome properties and future steps

To gain information on the genetic distance between the two closely related species *O. vulgaris* and *O. bimaculoides*, we mapped all the available raw sequence data from *O. vulgaris* against the genome of *O. bimaculoides*⁸³ and found that 74–84% of the data aligned, but that a high percentage (20–50%) was able to align multiple times. The significant proportion of multiple mapping reads suggests that, similar to the *O. bimaculoides* genome, *O. vulgaris* genome has a large number (at least 50%) of repetitive elements, confirmed by the cumulative read depth analysis

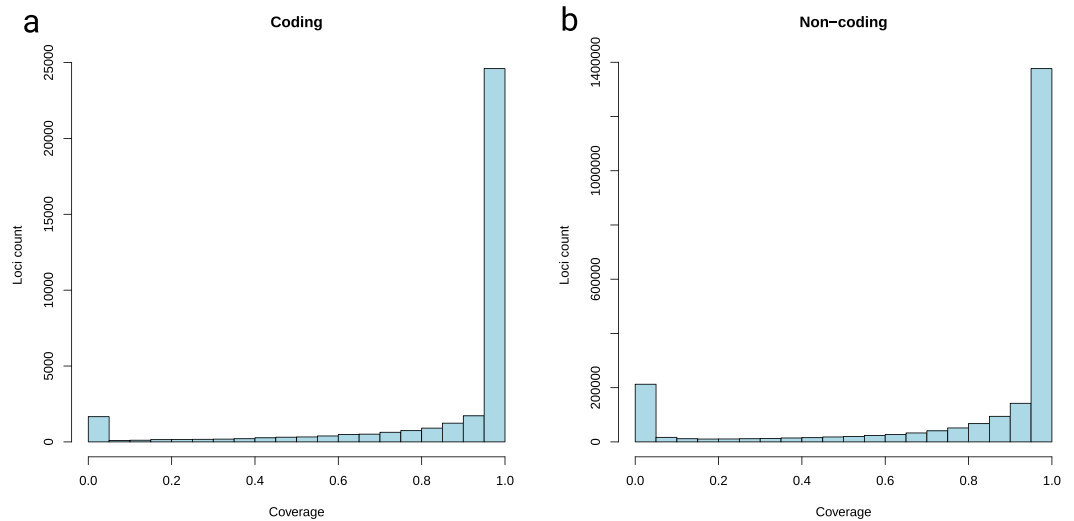


Fig. 3 Comparison of coding and non-coding region conservation between *Octopus bimaculoides* and *Octopus vulgaris*. **(a)** Alignment coverage in the coding genomic regions. **(b)** Alignment coverage in the non-coding, non-repetitive genomic regions. Coverage shows the proportion of nucleotides that are covered in *O. bimaculoides* assembly with *O. vulgaris* read mapping in both coding and non-coding non-repetitive regions of at least 100 bp. The main peak at 1 (100% coverage) indicates the presence of a complete region in *O. vulgaris* genome at very low sequence divergence, whereas the secondary peak at 0 indicates regions of *O. bimaculoides* genome that are not matching in *O. vulgaris* read data (see text for analysis).

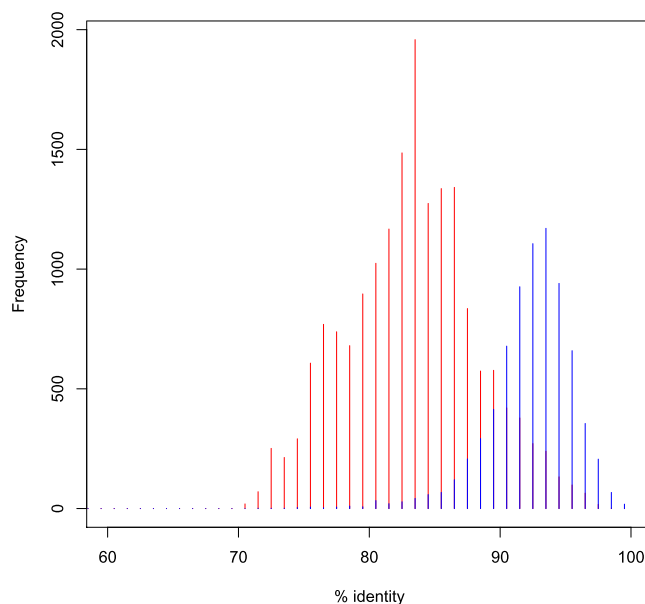


Fig. 4 Comparison of whole genome alignments using MEGABLAST among the available octopod genomes. Only the longest scoring alignment between any given pair of two scaffolds or contigs was considered. Red: percentage nucleotide identity between *Callistoctopus minor* to *Octopus bimaculoides*. Blue: percentage nucleotide identity between *Octopus vulgaris* to *O. bimaculoides*.

(Fig. 1). *Ab initio* repeat analysis using dnaPipeTE⁸⁴ revealed similar classes of octopus specific short interspersed nuclear elements (SINE) to be over-represented (Fig. 2), yet the proportions were strikingly different, despite the close phylogenetic distance. This indicates high activity of repetitive elements in the common octopus genome.

Profiling *O. bimaculoides* regions with read coverage from *O. vulgaris*, we found that 23,509 *O. bimaculoides* genes were covered at 90% or more of their coding sequence length by *O. vulgaris* reads (Fig. 3). Approximately 50% of those genes had a Pfam annotation, including gene families previously reported to have undergone major expansions in the *O. bimaculoides* genome, such as zinc fingers and protocadherins. This is in strong contrast to only 1,570 *O. bimaculoides* genes with no *O. vulgaris* read coverage, with just 14% of those having a Pfam annotation. Those candidates represent very recent novel or highly diverged genes and their number indicates

a relatively high rate of novel gene formation in octopus genomes. To investigate non-coding evolution among cephalopods, we furthermore compared the mapping rates to non-repetitive non-coding regions of 100 bp and longer. Again, we found the majority of those loci are covered at 90% length or higher. However, the relative proportion of *O. bimaculoides* regions not covered by any reads was higher than for the genes, indicating a higher turnover rate for the non-coding, potentially regulatory, sequences (Fig. 3).

To evaluate the completeness of our assemblies, raw reads were mapped using Bowtie2 v2.3.4 against both ABySS kmer81 and kmer41 assemblies. For ABySS kmer 41, at least 99.94% of all the reads were mapped while the percentage of uniquely mapped reads was only around 33–50%. For the ABySS kmer81 assembly, percentages were at least 98% and between 31 and 57%, respectively.

We used our assemblies to estimate whole-genome divergences between the available octopus genomes. Mapping of the scaffolds of 10 kb and longer against the *O. bimaculoides* genome using MEGABLAST resulted in the overall sequence similarity of 92.4% in the aligned regions of 1 kb and above (Fig. 4). This divergence of around 8% between the two species is higher than the estimated heterozygosity rate of 1.1% in *O. vulgaris* and lower than the divergence between *O. bimaculoides*⁸³ and the recently released data of *C. minor* (82.4% similarity) (Fig. 4, and ref.⁸⁵) from a different genus, providing for the first whole-genome divergence estimates within this clade.

Our assemblies confirm that abundant repeat regions make it difficult to improve the genome based on the currently available sequence data. Future steps will include long read sequencing technology such as proximity-ligation based assemblies (e.g., Dovetail, PhaseGenomics) or longer read technologies (e.g., PacBio) to optimize the current assemblies.

References

- De Luca, D., Catanese, G., Procaccini, G. & Fiorito, G. An integration of historical records and genetic data to the assessment of global distribution and population structure in *Octopus vulgaris*. *Front. Ecol. Evol.* **2**, 55 (2014).
- De Luca, D., Catanese, G., Procaccini, G. & Fiorito, G. *Octopus vulgaris* (Cuvier, 1797) in the Mediterranean Sea: Genetic diversity and population structure. *Plos One* **11**, e0149496 (2016).
- Amor, M. D. *et al.* Morphological assessment of the *Octopus vulgaris* species complex evaluated in light of molecular-based phylogenetic inferences. *Zool. Scr.* **46**, 275–288 (2017).
- Wells, M. J. *Octopus: physiology and behaviour of an advanced invertebrate*. (Springer Science & Business Media, 1978).
- Marini, G., De Sio, F., Ponte, G. & Fiorito, G. In *Learning and Memory: A Comprehensive Reference (Second Edition)* Vol. Volume 1 - Learning Theory and Behavior (Menzel, Randolph - volume Editor) (ed John H. Byrne) 441–462 (Academic Press, Elsevier, 2017).
- Borrelli, L. & Fiorito, G. In *Learning and Memory: A Comprehensive Reference* (ed. Byrne, J. J.) 605–627 (Academic Press, 2008).
- Edelman, D. B. & Seth, A. K. Animal consciousness: a synthetic approach. *Trends Neurosci.* **32**, 476–484 (2009).
- Young, J. Z. *The anatomy of the nervous system of Octopus vulgaris*. (Oxford University Press, 1971).
- Zarella, I., Ponte, G., Baldascino, E. & Fiorito, G. Learning and memory in *Octopus vulgaris*: a case of biological plasticity. *Curr. Opin. Neurobiol.* **35**, 74–79, <https://doi.org/10.1016/j.conb.2015.06.012> (2015).
- O'Brien, C. E., Roumbedakis, K. & Winkelmann, I. E. The Current State of Cephalopod Science and Perspectives on the Most Critical Challenges Ahead From Three Early-Career Researchers. *Front. Physiol.* **9**, 700, <https://doi.org/10.3389/fphys.2018.00700> (2018).
- Nixon, M. & Young, J. Z. *The brains and lives of Cephalopods*. 1–392 (Oxford University, 2003).
- Young, J. Z. Computation in the learning system of cephalopods. *Biol. Bull.* **180**, 200–208 (1991).
- Young, J. Z. *A Model of the Brain*. (Clarendon Press, 1964).
- Mather, J. A. Cognition in cephalopods. *Adv. Stud. Behav.* **24**, 317–353 (1995).
- Mather, J. A. Cephalopod consciousness: behavioural evidence. *Conscious. Cogn.* **17**, 37–48 (2008).
- Mather, J. A., Anderson, R. C. & Wood, J. B. *Octopus: The Ocean's Intelligent Invertebrate*. 1–208 (Timber Press, 2010).
- Mather, J. A. & Dickel, L. Cephalopod complex cognition. *Curr. Opin. Behav. Sci.* **16**, 131–137 (2017).
- Shigeno, S., Andrews, P. L. R., Ponte, G. & Fiorito, G. Cephalopod Brains: An Overview of Current Knowledge to Facilitate Comparison With Vertebrates. *Front. Physiol.* **9**, 952, <https://doi.org/10.3389/fphys.2018.00952> (2018).
- Edelman, D. B. How Octopuses See the World and Other Roads Less Traveled: Necessity Versus Sufficiency and Evolutionary Convergence in the Study of Animal Consciousness. *J. Shellfish Res.* **30**, 1001 (2011).
- Young, J. Z. In *Cephalopod Neurobiology* (eds Abbott, J. N., Williamson, R. & Maddock, L.) 431–443 (Oxford University Press, 1995).
- Imperadore, P. Nerve regeneration in the cephalopod mollusc *Octopus vulgaris*: a journey into morphological, cellular and molecular changes including epigenetic modifications. PhD Thesis (Università della Calabria, 2017).
- Imperadore, P. & Fiorito, G. Cephalopod Tissue Regeneration: consolidating over a century of knowledge. *Front. Physiol.* **9**, 593, <https://doi.org/10.3389/fphys.2018.00593> (2018).
- Imperadore, P. *et al.* Nerve regeneration in the cephalopod mollusc *Octopus vulgaris*: label-free multiphoton microscopy as a tool for investigation. *J. Royal Soc. Interface* **15**, 20170889, <https://doi.org/10.1098/rsif.2017.0889> (2018).
- Zullo, L., Fossati, S. M., Imperadore, P. & Nödl, M. T. Molecular determinants of Cephalopod muscles and their implication in muscle regeneration. *Front. Cell. Dev. Biol.* **5**, 53 (2017).
- Huffard, C. L. Cephalopod neurobiology: an introduction for biologists working in other model systems. *Invert. Neurosci.* **13**, 11–18 (2013).
- Barghi, N., Concepcion, G. P., Olivera, B. M. & Lluisma, A. O. Structural features of conopeptide genes inferred from partial sequences of the *Conus tribblei* genome. *Mol. Genet. Genomics* **291**, 411–422, <https://doi.org/10.1007/s00438-015-1119-2> (2016).
- Chen, H. *et al.* Sequencing and de novo assembly of the Asian clam (*Corbicula fluminea*) transcriptome using the Illumina GAIIx method. *Plos One* **8**, e79516 (2013).
- Du, X. *et al.* The pearl oyster *Pinctada fucata martensii* genome and multi-omic analyses provide insights into biomineralization. *Gigascience* **6**, gix059, <https://doi.org/10.1093/gigascience/gix059> (2017).
- Gómez-Chiarri, M., Warren, W. C., Guo, X. & Proestou, D. Developing tools for the study of molluscan immunity: the sequencing of the genome of the eastern oyster, *Crassostrea virginica*. *Fish Shellfish Immunol.* **46**, 2–4 (2015).
- Modica, M. V., Lombardo, F., Franchini, P. & Oliverio, M. The venomous cocktail of the vampire snail *Colubraria reticulata* (Mollusca, Gastropoda). *BMC Genomics* **16**, 441 (2015).
- Murgarella, M. *et al.* A First Insight into the Genome of the Filter-Feeder Mussel *Mytilus galloprovincialis*. *Plos One* **11**, e0151561, <https://doi.org/10.1371/journal.pone.0151561> (2016).
- Peñarrubia, L., Sanz, N., Pla, C., Vidal, O. & Viñas, J. Using Massive Parallel Sequencing for the Development, Validation, and Application of Population Genetics Markers in the Invasive Bivalve Zebra Mussel (*Dreissena polymorpha*). *Plos One* **10**, e0120732, <https://doi.org/10.1371/journal.pone.0120732> (2015).
- Raghavan, N. & Knight, M. The snail (*Biomphalaria glabrata*) genome project. *Trends Parasitol.* **22**, 148–151 (2006).

34. Sadamoto, H. *et al.* De novo sequencing and transcriptome analysis of the central nervous system of mollusc *Lymnaea stagnalis* by deep RNA sequencing. *Plos One* **7**, e42546 (2012).
35. Schell, T. *et al.* An Annotated Draft Genome for *Radix auricularia* (Gastropoda, Mollusca). *Genome Biol. Evol.* **9**, 585–592, <https://doi.org/10.1093/gbe/evx032> (2017).
36. Simakov, O. *et al.* Insights into bilaterian evolution from three spiralian genomes. *Nature* **493**, 526 (2013).
37. Sun, J. *et al.* Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes. *Nat. Ecol. Evol.* **1**, 0121 (2017).
38. Uliano-Silva, M. *et al.* A hybrid-hierarchical genome assembly strategy to sequence the invasive golden mussel, *Limnoperna fortunei*. *Gigascience* **7**, gix128, <https://doi.org/10.1093/gigascience/gix128> (2018).
39. Wang, S. *et al.* Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nat. Ecol. Evol.* **1**, 0120 (2017).
40. Xing, Q. *et al.* Genome-wide identification, characterization and expression analyses of two TNFRs in Yesso scallop (*Patinopecten yessoensis*) provide insight into the disparity of responses to bacterial infections and heat stress in bivalves. *Fish Shellfish Immunol.* **52**, 44–56, <https://doi.org/10.1016/j.fsi.2016.03.010> (2016).
41. Zhang, G. *et al.* The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **490**, 49 (2012).
42. Zhou, X. *et al.* The complete mitochondrial genome of *Pomacea canaliculata* (Gastropoda: Ampullariidae). *Mitochondrial DNA A* **27**, 884–885 (2016).
43. Albertin, C. B. *et al.* The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature* **524**, 220–224 (2015).
44. Kim, B.-M. *et al.* The genome of common long-arm octopus *Octopus minor*. *Gigascience* **7**, giy119, <https://doi.org/10.1093/gigascience/giy119> (2018).
45. Belcaid, M. *et al.* Symbiotic organs shaped by distinct modes of genome evolution in cephalopods. *Proc. Natl. Acad. Sci. USA.*, **201817322**, <https://doi.org/10.1073/pnas.1817322116> (2019).
46. Walker, W. F. & Ford Doolittle, W. 5S rRNA sequences from four marine invertebrates and implications for base pairing models of metazoan sequences. *Nucleic Acids Res* **11**, 5159–5164 (1983).
47. Albertin, C. B. *et al.* Cephalopod genomics: A plan of strategies and organization. *Stand. Genomic Sci.* **7**, 175–188 (2012).
48. Kröger, B., Vinther, J. & Fuchs, D. Cephalopod origin and evolution: A congruent picture emerging from fossils, development and molecules. *BioEssays* **33**, 602–613 (2011).
49. Packard, A. Cephalopods and fish: the limits of convergence. *Biol. Rev.* **47**, 241–307 (1972).
50. Garrett, S. & Rosenthal, J. J. C. RNA Editing Underlies Temperature Adaptation in K⁺ Channels from Polar Octopuses. *Science* **335**, 848 (2012).
51. Liscovitch-Brauer, N. *et al.* Trade-off between Transcriptome Plasticity and Genome Evolution in Cephalopods. *Cell* **169**, 191–202. e111, <https://doi.org/10.1016/j.cell.2017.03.025> (2017).
52. Garrett, S. C. & Rosenthal, J. J. C. A Role for A-to-I RNA Editing in Temperature Adaptation. *Physiology Bethesda* **27**, 362–369 (2012).
53. Yoshida, M.-A. *et al.* Genome structure analysis of molluscs revealed whole genome duplication and lineage specific repeat variation. *Gene* **483**, 63–71 (2011).
54. Petrosino, G. The transcriptional landscape of the nervous system of *Octopus vulgaris*. PhD Thesis (Università degli Studi di Napoli Federico II, 2015).
55. Lee, P. N., Callaerts, P., de Couet, H. G. & Martindale, M. Q. Cephalopod Hox genes and the origin of morphological novelties. *Nature* **424**, 1061–1065 (2003).
56. Nakajima, R., Shigeno, S., Zullo, L., De Sio, F. & Schmidt, M. R. Cephalopods between science, art, and engineering: a contemporary synthesis. *Front. Commun* **3**, 20 (2018).
57. Shigeno, S., Parnaik, R., Albertin, C. B. & Ragsdale, C. W. Evidence for a cordal, not ganglionic, pattern of cephalopod brain neurogenesis. *Zool. Lett* **1**, 26 (2015).
58. Shigeno, S., Takenori, S. & Boletzky, S. V. In *Cephalopods - Present and Past* (eds Tanabe, K., Shigeta, Y., Sasaki, T. & Hirano, H.) 23–34 (Tokai University Press, 2010).
59. Keynes, R. D. The role of giant axons in studies of the nerve impulse. *BioEssays* **10**, 90–93 (1989).
60. Allcock, A. L., Lindgren, A. & Strugnell, J. The contribution of molecular data to our understanding of cephalopod evolution and systematics: a review. *J. Nat. Hist.* **49**, 1373–1421 (2015).
61. Sanchez, G. *et al.* Genus-level phylogeny of cephalopods using molecular markers: current status and problematic areas. *PeerJ* **6**, e4331, <https://doi.org/10.7717/peerj.4331> (2018).
62. Sweeney, M. J., Roper, C. F., Mangold, K. M., Clark, M. & Boletzky, S. V. “Larval” and juvenile cephalopods: a manual for their identification. *Smithson. Contrib. Zool.* **513**, 1–282 (1992).
63. Andrews, P. L. R. *et al.* The identification and management of pain, suffering and distress in cephalopods, including anaesthesia, analgesia and humane killing. *J. Exp. Mar. Biol. Ecol.* **447**, 46–64 (2013).
64. Fiorito, G. *et al.* Guidelines for the Care and Welfare of Cephalopods in Research - A consensus based on an initiative by CephRes, FELASA and the Boyd Group. *Lab. Anim.* **49**, 1–90 (2015).
65. DeWoody, J. & Avise, J. Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. *J. Fish Biol.* **56**, 461–473 (2000).
66. Bazin, E., Glémin, S. & Galtier, N. Population size does not influence mitochondrial genetic diversity in animals. *Science* **312**, 570–572 (2006).
67. Chapko, M. K. *et al.* *Devilfish. A Practical Guide to the Dissection of Octopus*. 1–65 (Wayne Senior High School, 1962).
68. Chen, Y. *et al.* SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience* **7**, gix120, <https://doi.org/10.1093/gigascience/gix120> (2018).
69. Simpson, J. T. Exploring genome characteristics and sequence quality without a reference. *Bioinformatics* **30**, 1228–1235 (2014).
70. Jackman, S. D. *et al.* ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. *Genome Res.* **27**, 768–777 (2017).
71. Simpson, J. T. *et al.* ABySS: a parallel assembler for short read sequence data. *Genome Res.* **19**, 1117–1123 (2009).
72. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUASt: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).
73. Przytycz, L. P. & Gabaldón, T. Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Res.* **44**, e113, <https://doi.org/10.1093/nar/gkw294> (2016).
74. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
75. Waterhouse, R. M. *et al.* BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Mol. Biol. Evol.* **35**, 543–548 (2018).
76. Kent, W. J. *et al.* The Human Genome Browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
77. *NCBI Sequence Read Archive*, <http://identifiers.org/ncbi/insdc.sra:SRP162501> (2018).
78. Octopus vulgaris isolate Ov12053, whole genome shotgun sequencing project. *GenBank*, <http://identifiers.org/ncbi/insdc:RXHP00000000> (2018).
79. Zarrella, I. *et al.* The survey and reference assisted assembly of the Octopus vulgaris genome. *figshare* <https://doi.org/10.6084/m9.figshare.c.4266164> (2019).

80. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
81. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357 (2012).
82. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
83. *NCBI Assembly*, https://www.ncbi.nlm.nih.gov/assembly/GCF_001194135.1/ (2015).
84. Goubert, C. *et al.* De Novo Assembly and Annotation of the Asian Tiger Mosquito (*Aedes albopictus*) Repeatome with dnaPipeTE from Raw Genomic Reads and Comparative Analysis with the Yellow Fever Mosquito (*Aedes aegypti*). *Genome Biol. Evol.* **7**, 1192–1205, <https://doi.org/10.1093/gbe/evv050> (2015).
85. Kim, B.-M. *et al.* Supporting data for “The genome of common long-arm octopus *Octopus minor*”. *GigaScience Database*, <https://doi.org/10.5524/100503> (2018).

Acknowledgements

We are recognizing the contribution of Drs A. Cole (currently at University of Vienna), N. Tublitz (Oregon University), and C. Di Cristo (University of Sannio and CephRes) for their generous support at an earlier stage of this project. The contribution of the State Key Laboratory of Agricultural Genomics (grant No. 2011DQ782025) to GEM is also appreciated. EAR and OS are supported by a grant from the Austrian Science Fund (FWF): P30686-B29. We thank the COST Action (FA1301) that facilitated the networking initiative underlying this study. Genome browsers are integrated in the CephRes databases-website platform (http://www.cephalopodresearch.org/ceph_gdatab/) based on a bilateral agreement with University of Vienna. This study was supported by BGI-Shenzhen and by the Association for Cephalopod Research ‘CephRes’ a no-profit organization.

Author Contributions

I.Z. and G.P. provided material for DNA sequencing and drafted an early first version of the manuscript. S.T. and M.Y. generated the data and performed a preliminary analysis. K.H. and G.E.M. further analysed the data and run following steps, and developed a further version of the manuscript. O.S. contributed to the conception of the work and contributed to data analysis and to paper writing. M.Z. established the database resource. G.F. and G.P. conceived the work, performed the sampling and contributed to the paper writing. All authors contributed to writing and editing the final manuscript.

Additional Information

Competing Interests: The authors declare no competing interests.

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 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

© The Author(s) 2019