BMC Research Notes

DATA NOTE

Open Access

De novo transcriptome assembly and annotation of the third stage larvae of the zoonotic parasite *Anisakis pegreffii*

Marialetizia Palomba[1†], Pietro Libro[1†], Jessica Di Martino[1], Aurelia Rughetti[2], Mario Santoro[3], Simonetta Mattiucci[4,5*†] and Tiziana Castrignanò[1†]

Abstract

Objectives: *Anisakis pegreffii* is a zoonotic parasite requiring marine organisms to complete its life-history. Human infection (anisakiasis) occurs when the third stage larvae (L3) are accidentally ingested with raw or undercooked infected fish or squids. A new de novo transcriptome of *A. pegreffii* was here generated aiming to provide a robust bulk of data to be used for a comprehensive "ready-to-use" resource for detecting functional studies on genes and gene products of *A. pegreffii* involved in the molecular mechanisms of parasite-host interaction.

Data description: A RNA-seq library of *A. pegreffii* L3 was here newly generated by using Illumina TruSeq platform. It was combined with other five RNA-seq datasets previously gathered from L3 of the same species stored in SRA of NCBI. The final dataset was analyzed by launching three assembler programs and two validation tools. The use of a robust pipeline produced a high-confidence protein-coding transcriptome of *A. pegreffii*. These data represent a more robust and complete transcriptome of this species with respect to the actually existing resources. This is of importance for understanding the involved adaptive and immunomodulatory genes implicated in the "cross talk" between the parasite and its hosts, including the accidental one (humans).

Keywords: *Anisakis pegreffii*, Zoonotic parasite, Transcriptome, De novo assembly, Gene annotation

Objective

Anisakis pegreffii is a parasitic nematode belonging to the *A. simplex* (s.l.) species complex [1, 2]. It has a heteroxenous life cycle involving mainly cetaceans as definitive hosts, crustaceans as first intermediate hosts, fish, and squids as intermediate/paratenic ones. Its geographical distribution includes the Mediterranean Sea, the Iberian Atlantic coast waters, and the Austral region waters, between 30°S and 60°S. In humans, the accidental ingestion of third-stage larvae (L3) through the consumption of infected raw, undercooked, or improperly processed fish, causes a zoonosis, known as anisakiasis. Among the currently recognized nine biological species of the genus, so far only *A. pegreffii* and *A. simplex* (s.s.) cause anisakiasis [1, 3, 4].

The investigation of genes and proteins of *A. pegreffii* is crucial for understanding the parasite biological functions and its adaptation to abiotic and biotic conditions. It also represents a fundamental aspect to add knowledge about the molecular mechanisms involved in the evolutionary host-parasite interaction. Additionally, the molecules involved in the interaction between *A. pegreffii* and humans have not yet been elucidated. Finally, the absence of a suitable reference genome of this parasite species

†Marialetizia Palomba, Pietro Libro, Simonetta Mattiucci and Tiziana Castrignanò contributed equally to this work

*Correspondence: simonetta.mattiucci@uniroma1.it

[4] Department of Public Health And Infectious Diseases, Section of Parasitology, Sapienza University of Rome, P.le Aldo Moro, 5, 00185 Rome, Italy

Full list of author information is available at the end of the article

Table 1 Overview of data files/data sets

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data file 1	RNA-seq datasets from NCBI	Table file (.doc)	Figshare https://doi.org/10.6084/m9.figshare.19174214 [24]
Data file 2	RNA-seq dataset obtained in this study	Fastq file (.fastq)	NCBI https://identifiers.org/ncbi/bioproject:PRJNA752284 [25]
Data file 3	MultiQC reads quality results	Image file (.jpg)	Figshare https://doi.org/10.6084/m9.figshare.18480635 [26]
Data file 4	Trinity RNA-seq de novo transcriptome assembly	Fasta file (.fasta)	Figshare https://doi.org/10.6084/m9.figshare.18300896 [27]
Data file 5	rnaSPAdes RNA-seq de novo transcriptome assembly	Fasta file (.fasta)	Figshare https://doi.org/10.6084/m9.figshare.18301337 [28]
Data file 6	Oases RNA-seq de novo transcriptome assembly	Fasta file (.fasta)	Figshare https://doi.org/10.6084/m9.figshare.18480689 [29]
Data file 7	*Anisakis pegreffii* RNA-seq de novo transcriptome assembly	Fastq file (.fastq)	ENA https://identifiers.org/ena.embl:ERZ5400090 [30]
Data file 8	Unigenes	Fasta file (.fasta)	Figshare https://doi.org/10.6084/m9.figshare.18301772 [31]
Data file 9	Open reading frames (ORFs) prediction	Fasta file (.fasta)	Figshare https://doi.org/10.6084/m9.figshare.18302102 [32]
Data file 10	Functional annotation from non-redundant (nr) NCBI	Text file (.txt)	Figshare https://doi.org/10.6084/m9.figshare.18295190 [33]
Data file 11	Functional annotation from Swiss-Prot	Text file (.txt)	Figshare https://doi.org/10.6084/m9.figshare.18295970 [34]
Data file 12	Functional annotation from TrEMBL UniProt	Text file (.txt)	Figshare https://doi.org/10.6084/m9.figshare.18296603 [35]
Data file 13	Functional annotation from non-redundant (nr) protein NCBI	Text file (.txt)	Figshare https://doi.org/10.6084/m9.figshare.18296933 [36]
Data file 14	Functional annotation from Swiss-Prot Protein	Text file (.txt)	Figshare https://doi.org/10.6084/m9.figshare.18297410 [37]
Data file 15	Functional annotation from TrEMBL UniProt Protein	Text file (.txt)	Figshare https://doi.org/10.6084/m9.figshare.18297938 [38]
Data file 16	InterProScan results	Text file (.txt)	Figshare https://doi.org/10.6084/m9.figshare.18298319 [39]

could make it difficult to achieve those goals. Although several RNA-seq analyses of L3 *A. pegreffii* at different experimental conditions and from different larvae tissues were carried out [5–9], a complete "ready to use" transcriptome is missing.

Objective of this research was to provide a robust high-confidence protein-coding transcriptome of the L3 stage of *A. pegreffii* acquired from the assembly of data newly generated in the present study with those previously stored. The findings were to provide a more accurate de novo reference transcriptome of *A. pegreffii* that will allow to shed light on genes implicated in the "cross-talk" between the parasite and its natural and accidental hosts.

Data description

The input dataset for de novo assembly of *A. pegreffii* L3 was composed by six RNA-seq datasets (Table 1, Data file 1, 2): one obtained in the present study (PRJNA752284) (Table 1, Data file 2) and five retrieved from SRA of NCBI (PRJNA589243, PRJNA602791, PRJNA374530, PRJNA316941, PRJNA312925). In order to obtain the RNA-seq dataset in this study, A. *pegreffii* L3, collected from the viscera of fish from the Mediterranean Sea, were maintained in vitro culture for 24 h. RNA and DNA were extracted from nine L3 using TRIzol reagent, as previously described [10, 11]. The extracted RNA from each three L3 was pooled, and the quantity check was performed by using Agilent 2100 Bioanalyzer. The cDNA library was prepared using the TruSeq Stranded mRNA kit (Illumina). Ligated products of 200 bp were excised from agarose gels and PCR amplified. Products were single end sequenced on an Illumina TruSeq platform. Genetic/molecular identification of L3 *A. pegreffii* was performed by sequences analysis of mitochondrial (mtDNA *cox*2), and nuclear (EF1 $\alpha-1$ nDNA, *nas* 10 nDNA) gene loci, as previously described [12].

Bioinformatic analysis was performed using a High-Performance-Computing platform [13]. For each bioproject, the quality control of reads was performed running FastQC

v.0.11.2, before and after trimming step (Trimmomatic v.0.39 [14]). The quality assessment metrics for all trimmed data were aggregated with MultiQC v.1.9 [15]. Data file 3 (Table 1) shows both the mean read counts per quality scores and the mean quality scores in each base position higher than 35, for all the samples in the six analyzed bioprojects. A total of 393,512,048 cleaned reads (97% of whole raw reads) were obtained after the removal of the low-quality reads.

In order to construct a robust de novo transcriptome, three assembly tools with a multi-kmer approach were adopted: Trinity v.2.11.0 [16] (Table 1, Data file 4), rnaSPAdes v.3.14.1 [17] (Table 1, Data file 5) and Oases v.0.2.09 [18] (Table 1, Data file 6). Results for each assembler were merged with Transabyss v.2.0.1 [19] (Table 1, Data file 7). The merged assembly of *A. pegreffii* showed an average length of 939 bp and an N50 of 2859 bp. The assembly was validated with two algorithms: Busco v.4.1.4 [20] and Transrate v.1.0.3 [21]. A CD-HIT-est run v.4.8.1 was applied to the merged assembly to remove any redundant transcripts. A total of 394,635 unique genes were provided (Table 1, Data file 8) and a quality check was re-applied. A total of 260,872 ORFs were predicted by using Transdecoder v.5.5.0 [22] (Table 1, Data file 9).

The functional annotation of contigs was performed by using DIAMOND v.2.0.11 [23], calling both blastp and blastx functions against three databases (Nr, SwissProt and TremBL). The obtained results for blastx consisted in 86,982 (88.93%), 56,997 (58.47%) and 87,134 (89.39%) sequences against Nr (Table 1, Data file 10), SwissProt (Table 1, Data file 11) and TremBl (Table 1, Data file 12), respectively. Mapped transcripts listed in the Data file 10, yielded 38,972 matches (hits) with *A. simplex*. Blastp results also are available for Nr (Table 1, Data file 13), SwissProt (Table 1, Data file 14) and TremBl (Table 1, Data file 15). Output from InterProScan used to annotate protein signatures is available in Data file 16 (Table 1). In detail, 18,976 contigs were annotated: 5099 GO-annotated and 2800 KEGG-annotated.

Limitations

The *A. pegreffii* transcriptome here obtained was assembled with those RNA-seq data sets from the third larval stage of the parasite species. The single transcriptome available from the fourth stage larva of *A. pegreffii* [8] was not included in this analysis because the main aim of this analysis was to provide a robust and "ready to use" transcriptome of the infective stage (third larval stage) of the parasite also provoking the zoonotic disease (anisakiasis) to humans.

Abbreviations

L3: Third stage larvae; SRA: Sequence Read Archive; NCBI: National Center for Biotechnology Information; BUSCO: Benchmarking Universal Single-Copy Orthologs; Gb: Giga base-pair; bp: Base pair.

Acknowledgements

We acknowledge the CINECA for the availability of high-performance computing resources and, in particular, the ELIXIR-ITA HPC@CINECA initiative for providing HPC resources to our project (P.I. Simonetta Mattiucci, name of the project "Call ELIXIR-ITA CINECA (2020–2021)").

Author contributions

SM, MP, TC, AR, Conceptualization; MP, PL, JDM, SM, TC, Methodology; MP, AR, MS, SM, Material sampling; MP, SM, TC, PL, writing-original draft preparation; all authors writing-review and editing; SM, TC, supervision. All authors read and approved the final manuscript.

Funding

This study was supported by the Italian Ministry of Health, Ricerca Finalizzata (RF) 2018 – 12367986, title "Innovative approaches and parameters in the diagnosis and epidemiological surveillance of the *Anisakis*-related human diseases in Italy" (PI: Simonetta Mattiucci).

Availability of data and materials

The data described in this Data note can be freely and openly accessed on NCBI, ENA and figshare databases. Data from the following sex Bioprojects were used: PRJNA752284, PRJNA589243, PRJNA602791, PRJNA374530, PRJNA316941, PRJNA312925. The RNA-seq raw data here obtained have been deposited at ENA (ERZ54000). The other data files generated in the current study are available in the figshare database. Please see Table 1 and references [24–39] for details and links to the data.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Department of Ecological and Biological Sciences, Tuscia University, Viale dell'Università s/n, 01100 Viterbo, Italy. [2]Department of Experimental Medicine, "Sapienza" University of Rome, P.le Aldo Moro, 5, 00185 Rome, Italy. [3]Department of Integrative Marine Ecology, Stazione Zoologica Anton Dohrn, Villa Comunale, 1, 80121 Naples, Italy. [4]Department of Public Health And Infectious Diseases, Section of Parasitology, Sapienza University of Rome, P.le Aldo Moro, 5, 00185 Rome, Italy. [5]Laboratory affiliated to Istituto Pasteur Italia—Fondazione Cenci Bolognetti, Rome, Italy.

Received: 2 February 2022 Accepted: 7 June 2022
Published online: 25 June 2022

References

1. Mattiucci S, Cipriani P, Levsen A, Paoletti M, Nascetti G. Molecular epidemiology of *Anisakis* and Anisakiasis: an ecological and evolutionary road map. Adv Parasitol. 2018;99:93–263.
2. Mattiucci S, Palomba M, Nascetti G. Anisakis. Reference Module in Biomedical Sciences. 2021https://doi.org/10.1016/B978-0-12-818731-9.00075-6
3. Mattiucci S, Fazii P, De Rosa A, Paoletti M, Megna AS, Glielmo A, et al. Anisakiasis and gastroallergic reactions associated with *Anisakis pegreffii* infection. Italy Emerg Infect Dis. 2013;19:496–9.
4. Mattiucci S, Colantoni A, Crisafi B, Mori-Ubaldini F, Caponi L, Fazii P, et al. IgE sensitization to *Anisakis pegreffii* in Italy: comparison of two methods for the diagnosis of allergic anisakiasis. Parasite Immunol. 2017;39:12440.
5. Baird FJ, Su X, Aibinu I, Nolan MJ, Sugiyama H, Otranto D, et al. The *Anisakis* transcriptome provides a resource for fundamental and applied

studies on allergy-causing parasites. PLoS Negl Trop Dis. 2016;10: e0004845.
6. Cavallero S, Lombardo F, Su X, Salvemini M, Cantacessi C, D'Amelio S. Tissue-specific transcriptomes of *Anisakis simplex* (sensu stricto) and *Anisakis pegreffii* reveal potential molecular mechanisms involved in pathogenicity. Parasites Vectors. 2018;11:31.
7. Llorens C, Arcos SC, Robertson L, Ramos R, Futami R, Soriano B, Ciordia S, Careche M, González-Muñoz M, Jiménez-Ruiz Y, Carballeda-Sangiao N, Moneo I, Albar JP, Blaxter M, Navas A. Functional insights into the infective larval stage of *Anisakis simplex* s.s., *Anisakis pegreffii* and their hybrids based on gene expression patterns. BMC Genom. 2018;19:59.
8. Nam UH, Kim JO, Kim JO. De novo transcriptome sequencing and analysis of *Anisakis pegreffii* (Nematoda: Anisakidae) third-stage and fourth-stage larvae. J Nematol. 2020;52:e2020–41.
9. Wang X, Jia H, Gong H, Zhang Y, Mi R, Zhang Y, et al. Expression and functionality of allergenic genes regulated by simulated gastric juice in *Anisakis pegreffii*. Parasitol Int. 2021;80: 102223.
10. Palomba M, Paoletti M, Colantoni A, Rughetti A, Nascetti G, Mattiucci S. Gene expression profiles of antigenic proteins of third stage larvae of the zoonotic nematode *Anisakis pegreffii* in response to temperature conditions. Parasite. 2019;26:52.
11. Palomba M, Cipriani P, Giulietti L, Levsen A, Nascetti G, Mattiucci S. Differences in gene expression profiles of seven target proteins in third-stage larvae of *Anisakis simplex* (sensu stricto) by sites of infection in blue whiting (*Micromesistius poutassou*). Genes. 2020;11:559.
12. Palomba M, Paoletti M, Webb S, Nascetti G, Mattiucci S. A novel nuclear marker and development of an ARMS-PCR assay targeting the metallopeptidase 10 (nas 10) locus to identify the species of the *Anisakis simplex* (s. l.) complex (Nematoda, Anisakidae). Parasite. 2020;27:39.
13. Castrignanò T, Gioiosa S, Flati T, Cestari M, Picardi E, Chiara M, Fratelli M, Amente S, Cirilli M, Tangaro MA, Chillemi G, Pesole G, Zambelli F. ELIXIR-IT HPC@CINECA: high performance computing resources for the bioinformatics community. BMC Bioinform. 2020;21:352.
14. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.
15. Ewels P, Magnusson M, Lundin S, Kaller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32:3047–8.
16. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq: reference generation and analysis with trinity. Nat Protoc. 2013;8:1494–512.
17. Bushmanova E, Antipov D, Lapidus A, Prjibelski AD. rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. Gigascience. 2019;8:9.
18. Cédric C, Escudié F, Djari A, Yann G, Julien B, Klopp C. Compacting and correcting Trinity and Oases RNA-Seq de novo assemblies. PeerJ. 2017;5: e2988.
19. Robertson G, Schein J, Chiu R, Corbett R, Field M, Jackman SD, et al. De novo assembly and analysis of RNA-seq data. Nat Methods. 2010;7:909–12.
20. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31:3210–2.
21. Smith-Unna R, Boursnell C, Patro R, Hibberd JM, Kelly S. TransRate: reference-free quality assessment of *de novo* transcriptome assemblies. Genome Res. 2016;26:1134–44.
22. Tang S, Lomsadze A, Borodovsky M. Identification of protein-coding regions in RNA transcripts. Nucleic Acids Res. 2015;43: e78.
23. Buchfink B, Xie C, Huson D. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12:59–60.
24. Libro P, Di Martino J, Rughetti A, Santoro M, Mattiucci S, Castrignanò T, Palomba M. AP - Bioproject collection included in input dataset (Online resource). figshare. 2022. https://doi.org/10.6084/m9.figshare.19174214.v1
25. Palomba M, Libro P, Di Martino J, Rughetti A, Santoro M, Mattiucci S, Castrignanò T. Transcriptional changes in the *Anisakis pegreffii* third larval stage during human dendritic cells host-parasite interactions. https://identifiers.org/ncbi/bioproject:PRJNA752284
26. Libro P, Di Martino J, Rughetti A, Santoro M, Mattiucci S, Castrignanò T, Palomba M. AP - MultiQC reads quality results (Figure). figshare. 2022. https://doi.org/10.6084/m9.figshare.18480635.v1
27. Libro P, Di Martino J, Rughetti A, Santoro M, Mattiucci S, Castrignanò T, Palomba M. AP - Trinity RNA-Seq *de novo* transcriptome assembly (Online resource). figshare. 2022. https://doi.org/10.6084/m9.figshare.18300896.v1
28. Libro P, Di Martino J, Rughetti A, Santoro M, Mattiucci S, Castrignanò T, Palomba M. AP - rnaSPADES RNA-Seq *de novo* transcriptome assembly (Online resource). figshare. 2022. https://doi.org/10.6084/m9.figshare.18301337.v1
29. Libro P, Di Martino J, Rughetti A, Santoro M, Mattiucci S, Castrignanò T, Palomba M. AP - Oases RNA-Seq de novo transcriptome assembly (Online resource). figshare. 2022. https://doi.org/10.6084/m9.figshare.18480689.v1
30. Palomba M, Libro P, Di Martino J, Rughetti A, Santoro M, Mattiucci S, Castrignanò T. Trascriptome assembly of *Anisakis pegreffii*. Online resource. 2022. https://identifiers.org/ena.embl:ERZ5400090
31. Libro P, Di Martino J, Rughetti A, Santoro M, Mattiucci S, Castrignanò T, Palomba M. AP - Unigenes (Online resource). figshare. 2022. https://doi.org/10.6084/m9.figshare.18301772.v1.
32. Libro P, Di Martino J, Rughetti A, Santoro M, Mattiucci S, Castrignanò T, Palomba M. AP - Open reading frames (ORFs) prediction (Online resource). figshare. 2022. https://doi.org/10.6084/m9.figshare.18302102.v1
33. Libro P, Di Martino J, Rughetti A, Santoro M, Mattiucci S, Castrignanò T, Palomba M. AP - Functional annotation from non-redundant (nr) NCBI (Online resource). figshare. 2022. https://doi.org/10.6084/m9.figshare.18295190.v1
34. Libro P, Di Martino J, Rughetti A, Santoro M, Mattiucci S, Castrignanò T, Palomba M. AP - Functional annotation from Swiss-Prot (Online resource). figshare. 2022. https://doi.org/10.6084/m9.figshare.18295970.v1
35. Libro P, Di Martino J, Rughetti A, Santoro M, Mattiucci S, Castrignanò T, Palomba M. AP - Functional annotation from TrEMBL UniProt (Online resource). figshare. 2022. https://doi.org/10.6084/m9.figshare.18296603.v1
36. Libro P, Di Martino J, Rughetti A, Santoro M, Mattiucci S, Castrignanò T, Palomba M. AP - Functional annotation from non-redundant (nr) protein NCBI (Online resource). figshare. 2022. https://doi.org/10.6084/m9.figshare.18296933.v1
37. Libro P, Di Martino J, Rughetti A, Santoro M, Mattiucci S, Castrignanò T, Palomba M. AP - Functional annotation from Swiss-Prot Protein (Online resource). figshare. https://doi.org/10.6084/m9.figshare.18297410.v1
38. Libro P, Di Martino J, Rughetti A, Santoro M, Mattiucci S, Castrignanò T, Palomba M. AP - Functional annotation from TrEMBL UniProt Protein (Online resource). figshare. 2022. https://doi.org/10.6084/m9.figshare.18297938.v1
39. Libro P, Di Martino J, Rughetti A, Santoro M, Mattiucci S, Castrignanò T, Palomba M. AP - Interproscan results (Online resource). figshare. 2022. https://doi.org/10.6084/m9.figshare.18298319.v1

Publisher's Note