Check for updates

DATA NOTE

The complete genome sequences of two species of seventeen-year cicadas: *Magicicada septendecim* and *Magicicada septendecula* [version 1; peer review: 2 approved with reservations]

Harold B. White¹, Stacy Pirro ²¹Department of Chemistry and Biochemistry, University of Delaware, Delaware, USA²Department Biodiversity, Iridian Genomes, Bethesda, MD, 20817, USA

V1 First published: 16 Mar 2021, 10:215
<https://doi.org/10.12688/f1000research.27309.1>
Latest published: 16 Mar 2021, 10:215
<https://doi.org/10.12688/f1000research.27309.1>

Abstract

The genus *Magicicada* (Hemiptera: Cicadidae) includes the periodical cicadas of Eastern North America. Spending the majority of their long lives underground, the adult cicadas emerge every 13 or 17 years to spend 4-6 weeks as adult to mate. We present the whole genome sequences of two species of 17-year cicadas, *Magicicada septendecim* and *Magicicada septendecula*. The reads were assembled by a *de novo* method followed by alignments to related species. Annotation was performed by GeneMark-ES. The raw and assembled data is available via NCBI Short Read Archive and Assembly databases.

Keywords

Genome, assembly, arthropoda, insecta, hemiptera

This article is included in the **Genomics and Genetics** gateway.

Open Peer Review

Approval Status ? ?

	1	2
version 1		
16 Mar 2021	view	view
<hr/>		
1. Hu Li , China Agricultural University, Beijing, China		
2. Shuai Zhan , Chinese Academy of Sciences, Shanghai, China		
Any reports and responses or comments on the article can be found at the end of the article.		

Corresponding author: Stacy Pirro (info@iridiangenomes.org)

Author roles: White HB: Resources; Pirro S: Data Curation, Investigation, Resources

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Iridian Genomes (IRGEN-38085).

Copyright: © 2021 White HB and Pirro S. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: White HB and Pirro S. **The complete genome sequences of two species of seventeen-year cicadas: *Magicicada septendecim* and *Magicicada septendecula* [version 1; peer review: 2 approved with reservations]** F1000Research 2021, 10:215 <https://doi.org/10.12688/f1000research.27309.1>

First published: 16 Mar 2021, 10:215 <https://doi.org/10.12688/f1000research.27309.1>

Introduction

Periodical cicadas in North America spend 13 or 17 years in the larval stage underground, and emerge in very large numbers for 4–6 weeks to mate and lay eggs. This strategy, known as “predator satiation” is intended to ensure that after all predators have eaten as much as possible, most cicadas will survive (Williams & Simon, 1995). The emergence occurring in prime-numbered years is thought to be a mechanism to avoid competition between species for egg-laying sites and accidental cross-species mating as the emergence of the 13- and 17-year cicadas would only coincide once every 221 years (Tanaka *et al.*, 2009).

The length of time spent in the larval stage is thought to be dependent on a single gene, although this has not yet been demonstrated at the genomic level (Cox & Carlton, 1991).

Complete genome sequences for these two species will assist with studies on taxonomy, longevity, and the timing of long-term larval development.

Methods

Wild caught specimens of *Magicicada septendecim* and *Magicicada septendecula* from a small premature emergence of Brood X (2017) collected in Newark, Delaware, USA were used in this study. DNA extraction was performed using the Qiagen DNAeasy genomic extraction kit for tissue, using the standard process. A paired-end sequencing library was constructed using the Illumina TruSeq kit, according to the manufacturer’s instructions. The library was sequenced on an Illumina Hi-Seq platform in paired-end, 2 × 150bp format.

The resulting fastq files were trimmed of adapter/primer sequence and low-quality regions with Trimmomatic v0.33 (Bolger *et al.*, 2014). The trimmed sequence was assembled by SPAdes v2.5 (Bankevich *et al.*, 2012) followed by a finishing step using RagTag v1.0.0 (Alonge, 2020) to make additional contig joins based on conserved regions in related insect species: *Rhopalosiphum maidis* (GCA_003676215), *Euschistus*

heros (GCA_003667255), and *Aphis glycines* (GCA_0099285-15). Default parameters were used for all assembly steps.

Annotation was performed using GeneMark-ES v2.0 (Lomsadze *et al.*, 2005). Annotation was performed fully *de novo* without a curated training set and using default parameters.

Results

The genome assembly for *Magicicada septendecim* yielded a total sequence length of 1,579,033,894 with an N50 value of 983 kb and 27,124 gene models.

The genome assembly of *Magicicada septendecula* yielded 1,585,977,997 with an N50 value of 281 kb and 28,651 gene models.

Data availability

Genome data available from NCBI’s Short Read Archive (SRA):

Magicicada septendecim, Accession number SRR6782667: <https://www.ncbi.nlm.nih.gov/sra/SRR6782667>

Magicicada septendecula, Accession number SRR6792649: <https://www.ncbi.nlm.nih.gov/sra/SRR6792649>

Assembled genomes available from NCBI’s Assembly database:

Magicicada septendecim, Accession number GCA_011326945: https://www.ncbi.nlm.nih.gov/assembly/GCA_011326945.1/

Magicicada septendecula, Accession number GCA_011763675: https://www.ncbi.nlm.nih.gov/assembly/GCA_011763675.1/

Author information

Harold B. White is now retired from Department of Chemistry and Biochemistry, University of Delaware.

References

- Alonge M: **Ragtag: Reference-guided genome assembly correction and scaffolding**. GitHub archive. 2020.
- Bankevich A, Nurk S, Antipov D, *et al.*: **SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing**. *J Comput Biol*. 2012; **19**(5): 455–477.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bolger AM, Lohse M, Usadel B: **Trimmomatic: A flexible trimmer for Illumina Sequence Data**. *Bioinformatics*. 2014; **30**(15): 2114–20.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cox RT, Carlton CE: **Evidence of genetic dominance of the 13-year life cycle in periodical cicadas (Homoptera: Cicadidae: Magicicada spp.)**. *Am Midl Nat*. 1991; **125**(1): 63–74.
[Publisher Full Text](#)

- Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, *et al.*: **Gene identification in novel eukaryotic genomes by self-training algorithm**. *Nucleic Acids Res*. 2005; **33**(20): 6494–6506.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Tanaka Y, Yoshimura J, Simon C, *et al.*: **Allee effect in the selection for prime-numbered cycles in periodical cicadas**. *Proc Natl Acad Sci U S A*. 2009; **106**(22): 8975–8979.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Williams KS, Simon C: **The ecology, behavior, and evolution of periodical cicadas**. *Annu Rev Entomol*. 1995; **40**(1): 269–295.
[Publisher Full Text](#)

Open Peer Review

Current Peer Review Status: ? ?

Version 1

Reviewer Report 12 May 2021

<https://doi.org/10.5256/f1000research.30176.r83461>

© 2021 Zhan S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Shuai Zhan

Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China

The fascinating life cycle of *Periodical cicadas* attracts public and scientific interests from around the globe. In this Data note, the authors report the genome references of two such cicadas species, *Magicicada septendecim* and *M. septendecula*. The available genomes would no doubt benefit the community in various areas. However, it is difficult to judge the actual quality of the yielded genomes, due to the incomplete method description and lacking of necessary quality evaluation. I think this is important for the community to determine whether the resource is reliable and to what extent.

Here are some specific concerns:

- It is unclear to me how the genome is generated, which is critical to assess whether the assembling approach is reasonable. The authors should provide key information such as how many libraries being generated for a single species and the insert size information for the library(s). It seems that only one paired-end library was constructed, thus I cannot imagine that how a single library would yield genome assemblies for two species and how the assembly could reach a scaffold level with the N50 size as long as hundreds of Kb.
- The authors claimed that the assembling approach includes the alignment step to related species, such as some aphids. This is uncommon for *de novo* assembling a genome reference, particularly for an insect species, given the high level of divergence between families and species.
- The current note only presents the genome size and N50 size, but lacks of most other key properties to assess the quality of a genome assembly, such as the expected genome size, the completeness, the redundancy ratio, the GC ratio, the potential contamination ratio, etc.
- The current gene prediction method is unsatisfying and could be substantially improved by applying an integrated approach. Using GeneMark alone is uncommon for annotating a large eukaryotic genome. Instead, this gene predictor is more common to be used as partial *ab initio* evidence of a comprehensive gene annotation, which is usually achieved by

combining transcriptome, homology of related species, and several lines of *ab initio* signatures (e.g. GeneMark and other predictors).

- I'm not sure whether '**data note**' has a word limit or some special requirements, but the current form of the manuscript is obviously too dense. To allow replication by other researchers, the current method section should be provided with necessary details, such as the parameters for each software.

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

Partly

Are the protocols appropriate and is the work technically sound?

Partly

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

No

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

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Insect genomics

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

Reviewer Report 12 April 2021

<https://doi.org/10.5256/f1000research.30176.r81859>

© 2021 Li H. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Hu Li

Department of Entomology and MOA Key Lab of Pest Monitoring and Green Management, College of Plant Protection, China Agricultural University, Beijing, China

The authors presented the genome assembly of *Magicicada septendecim* and *Magicicada septendecula* using Illumina platform, which is of great help to the *Magicicada* evolution study. Using the RagTag method to correct, orient and scaffold the assembled sequence is a novel idea for the short reads assembly. However, there are still some limitations which might be improved.

In this manuscript, detailed information of sequencing samples counts, male or female, were not

given. It also lacks the amount of data sequenced by Illumina platform, genome assembly and annotation completeness assessed by BUSCO, and repetitive sequence annotation.

Additionally, one of my major concerns is that, we usually use a long reads platform such as PacBio or ONT to sequence the genome; could short sequence be assembled completely? Although *de novo* method by using SPAdes and RagTag software in this study works well to assemble short sequences, when aligned to related species, contigs lower than 1 kb were filtered and some genome sequences might be lost.

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

Yes

Are the protocols appropriate and is the work technically sound?

Yes

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

Partly

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

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: comparative genomics, phylogenomics, population genomics

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

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research