

# Genome Analyses of a New *Mycoplasma* Species from the Scorpion *Centruroides vittatus*

Tsunemi Yamashita,<sup>\*1</sup> Douglas D. Rhoads,<sup>†</sup> and Jeff Pummill<sup>\*</sup>

<sup>\*</sup>Department of Biological Sciences, Arkansas Tech University, Russellville, AR 72801, <sup>†</sup>Department of Biological Sciences, and <sup>‡</sup>High Performance Computing Center, University of Arkansas-Fayetteville, Fayetteville, AR 72701

**ABSTRACT** Arthropod *Mycoplasma* are little known endosymbionts in insects, primarily known as plant disease vectors. *Mycoplasma* in other arthropods such as arachnids are unknown. We report the first complete *Mycoplasma* genome sequenced, identified, and annotated from a scorpion, *Centruroides vittatus*, and designate it as *Mycoplasma vittatus*. We find the genome is at least a 683,827 bp single circular chromosome with a GC content of 42.7% and with 987 protein-coding genes. The putative virulence determinants include 11 genes associated with the virulence operon associated with protein synthesis or DNA transcription and ten genes with antibiotic and toxic compound resistance. Comparative analysis revealed that the *M. vittatus* genome is smaller than other *Mycoplasma* genomes and exhibits a higher GC content. Phylogenetic analysis shows *M. vittatus* as part of the Hominis group of *Mycoplasma*. As arthropod genomes accumulate, further novel *Mycoplasma* genomes may be identified and characterized.

## KEYWORDS

arachnid  
scorpions  
*Mycoplasma*  
genome

The investigation of the microbiome has shown important relationships exist between hosts and their symbionts. In eukaryotic genomics projects, symbiont genomes are often revealed via significant variation in GC/AT nucleotide content. In arthropod genomes, Mollicutes as a bacterial class have gained attention due to their reduced genome, symbiotic existence, and their occurrence as pathogens (Thompson *et al.* 2011, Browning and Citti 2014). Within the Mollicutes, the *Mycoplasma* are significant as animal pathogens and arthropod symbionts (Thompson *et al.* 2011, Leclercq *et al.* 2014). Our knowledge of arthropod *Mycoplasma* is under-represented, but the collection of additional arthropod genomes will improve the catalog of *Mycoplasma* diversity as contaminants in arthropod genome assemblies (Leclercq *et al.* 2014). As Mollicute genomes accumulate, taxonomic revisions with genomic data provide evolutionary insights into these prokaryotes (Hicks *et al.* 2014, Bolaños *et al.* 2015). Here, we report a new *Mycoplasma* genome from

the striped scorpion, *Centruroides vittatus*, and we designate this *Mycoplasma* as *M. vittatus*.

## MATERIALS AND METHODS

### Genome sequencing and assembly

Total genomic DNA was extracted from three scorpions collected in Pope County, AR with the Qiagen genomic-tip and genomic DNA buffer set (Qiagen, Inc.). The genomic DNA quality was analyzed through 0.9% agarose gel electrophoresis and UV spectroscopy. One genomic DNA sample was sent to the University of Arkansas for Medical Sciences DNA Sequencing Core Facility for library generation and 300bp paired-end sequencing (600 total bp) on a Illumina MiSeq. Two genomic DNA samples were sent to the National Center for Genome Resources (NCGR, NM) for PacBio 20K library generation and 10 SMRT cell sequencing for each individual genome. The *de novo* assembly was conducted at the Arkansas High Performance Computing Center at the University of Arkansas. Sequence read data quality control check was conducted with FastQC (0.11.5). The PacBio data were assembled with the Canu Pipeline Assembler (v1.3), while the MiSeq data were assembled using Spades (St. Petersburg genome assembler, ver. 3.6.1). Quality assessment of assemblies was conducted with Quast (Quality Assessment Tool, ver. 4.0). One PacBio assembly (Q1133) was polished with Quiver (PacBio, Inc.) and also with Pilon (Walker *et al.* 2014) as the data from this DNA appeared to be noisier based on fastqc analyses. We aligned our three assemblies

Copyright © 2019 Yamashita *et al.*

doi: <https://doi.org/10.1534/g3.118.200858>

Manuscript received November 1, 2018; accepted for publication February 5, 2019; published Early Online February 14, 2019.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supplemental material available at Figshare: <https://doi.org/10.25387/g3.6979928>.

<sup>1</sup>Corresponding Author: Dept. of Biological Sciences, Arkansas Tech University, 1701 N. Boulder Ave., Russellville, AR 72801, E-mail: [tyamashita@atu.edu](mailto:tyamashita@atu.edu)

against each other with Mauve (V2.4.0) to compare variation among them (Darling *et al.* 2004). We visualized and compared these draft genomes with CGViewer for blastn comparisons (Grant and Stothard 2008). Prokaryote genome annotation was conducted with the Rapid Annotation using Subsystem Technology (RAST) at <http://rasat.nmpdr.org> and BASys: <https://www.basys.ca/> (Van Domselaar *et al.* 2005, Aziz *et al.* 2008). The genes were categorized into subsystems with the SEED browser at <http://www.theseed.org/> (Overbeek *et al.* 2005). Related genomes were identified through RAST and distance matrices generated for selected genomes using the GGDC server for Genome-to-Genome Distance Calculator 2.1 at <http://ggdc.dsmz.de/ggdc.php> (Meier-Kolthoff *et al.* 2013). This in-silico reciprocal BLASTn method produces robust distances among bacterial taxa as it incorporates Genome Blast Distance Phylogeny and allows confidence interval estimation. We also included the *M. pulmonis* and *Candidatus Hepatoplasma crinochetorum* genomes in our RAST and SEED analyses to compare genome and protein features to the *M. vittatus* genome. Phylogenetic relationships were constructed from distance matrices using the Neighbor-Joining algorithm implemented at Trex-online

(<http://trex.uqam.ca>) with the subsequent trees visualized in FigTree v1.4.3 (<http://tree.bio.ed.ac.uk/>).

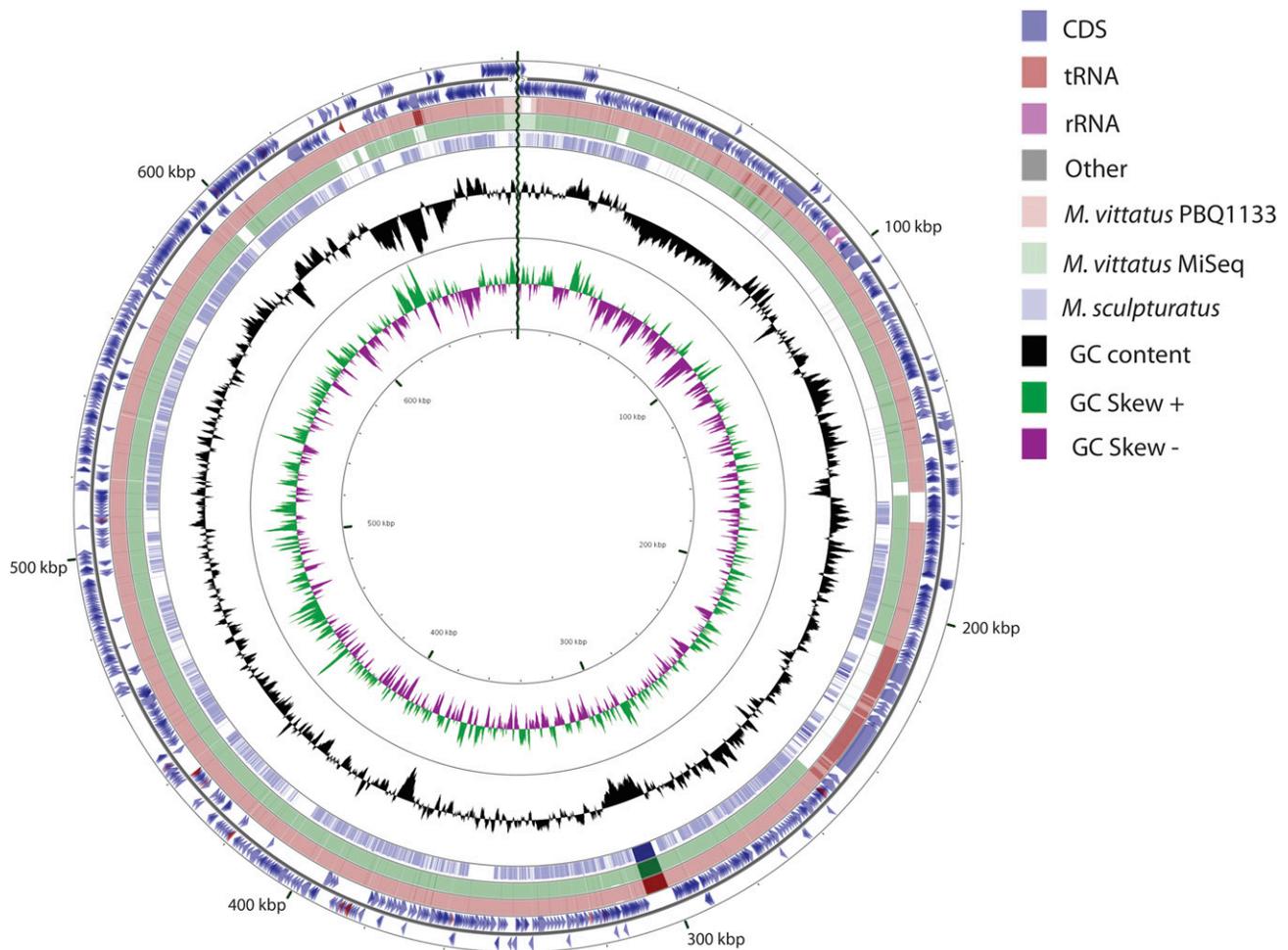
### Data availability

The genome sequences of the scorpion *Mycoplasma* are available under the Genbank Accession numbers RYZU01000001.1 and RYZV01000001.1. Two of the genomes were identical, thus only unique genomes were submitted. Supplemental Material, Figure S1 contains genome alignments. Table S1 contains predicted genome results. Table S2 contains functional category delineations. Table S3 contains gene predictions in virulence. Table S4 contains genes predicted in metabolism. Supplemental material available at Figshare: <https://doi.org/10.25387/g3.6979928>.

## RESULTS

### Identification of *M. vittatus* contigs

We assembled three draft *Mycoplasma* genomes from three different scorpions, *Centruroides vittatus*, via data produced as a part of the *C. vittatus* genome project (Yamashita T, Rhoads D, and Pummill J,



**Figure 1** Genome architecture of *M. vittatus* with comparison to *M. sculpturatus*. Moving from outside to inside, the first circle shows the position coordinates for the genome sequence. The second circle shows the predicted locations of the protein coding sequences (in blue) on the plus and minus strands, with tRNA in brown and rRNA in purple. The third pink circle shows the blast alignment of the *M. vittatus* PBQ1133, while the fourth green and fifth blue circles shows the blast alignment of the *M. vittatus* MiSeq and *M. sculpturatus*, respectively. The sixth circle shows the mean centered GC content, with the average GC as baseline and outward projections as higher than average and inwardly projections as lower than average. The seventh ring shows the GC Skew with above zero values in green and below zero values in purple.

unpublished data). We identified contigs in our MiSeq and PacBio assemblies from total scorpion DNA that differed markedly in GC content (43.7% GC content vs. 32% in the scorpion genome). There was a single contig in one of the PacBio assemblies (Q1171) of 683,827 bp. The other PacBio assembly (Q1133) contained two contigs of 511,437 and 163,546 bp (total 674,983). In the MiSeq assembly there were four contigs of 353,646, 177,906, 55,555 and 53,089 bp (total 640,196). The single contig in the Q1171 assembly did not appear to contain any terminal redundancy so it does not appear to represent a complete genome. However, based on Mauve alignments with the two contigs from the Q1133 assembly the Q1171 contig appears to represent a nearly complete assembly (Figure S1). Based on the PacBio assemblies the four contigs from the MiSeq assembly were ordered into a single contig with gaps. Figure 1 shows CGViewer for blastn comparisons and reveals that the two PacBio assemblies differ only for a small region at about 170 bp and reinforces that we have a nearly complete genome between the two assemblies (neither assembly contains additional data not in the other assembly). The MiSeq assembly primarily lacks a region from about 210 to 250 kbp, which is highly conserved between the two PacBio assemblies. This region consists entirely of “hypothetical proteins” according to the RAST annotation and is not recognized as a prophage by the PHASTER server (<http://phaster.ca>).

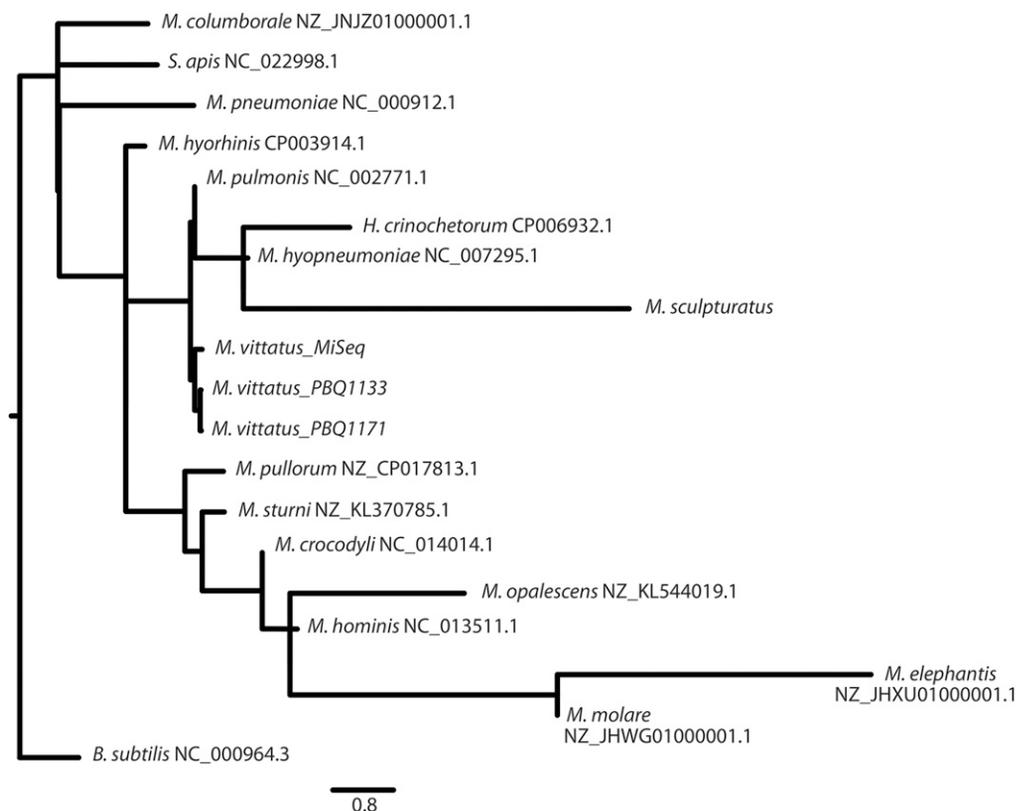
### Comparative and evolutionary analysis

We compared the three *M. vittatus* draft genomes to 13 representative mollicute genomes from NCBI. In addition, we extracted 40 contigs representing more than 16 Mbp of likely *Mycoplasma* contigs from a recently released draft genome (NW\_019384690.1)

for the related scorpion, *Centruroides sculpturatus* (referred to as *M. sculpturatus* and also included in the CGViewer comparison in Figure 1) (Schwager *et al.* 2017). The Neighbor Joining phylogeny based upon genome distances among 17 genomes based on pairwise distances from the GGDC analyses indicates that *M. vittatus* clusters with the Hominis group of *Mycoplasma* with *M. pulmonis*, the most similar taxon as revealed through our phylogenetic analyses (Figure 2). The comparative genome data are shown in Table S5. The *M. pulmonis* genome is 963,879 bp, 280,052 bp longer than *M. vittatus* and shows 15.9% sequence similarity (51,423 identical nucleotide sites of 322,418 nucleotides from 37 aligned locally colinear blocks) identified from a Mauve alignment conducted in Geneious 10.2.3 (<https://www.geneious.com>). The RAST server comparison to close strains (*i.e.*, most similar annotated genome) identifies similarities and differences in coding sequences between these two genomes, 350 genes were identified with a known function and 173 genes were identified as similar between the two genomes with 30 genes unique to *M. pulmonis* and 145 genes to *M. vittatus*.

### Genome features

The *M. vittatus* genome from one of the PacBio assemblies contains a 683,827 bp single contig with 42.7% GC content, which appears to deviate from the low GC content seen in other *Mycoplasma* (Figure 1). A total of 987 protein coding genes were identified that occupy 89.44% of the genome. Table S1 shows the summary of the genome content in the *M. vittatus*. The non-coding RNAs include 29 tRNA's, one large subunit RNA, and one small subunit RNA (Table S1). For the protein coding genes, 297 (30%) were classified into 18 subsystems and 673 genes (68%) were classified as unknown



**Figure 2** A phylogenetic trees created from GGDC output distances in a Neighbor-Joining tree. The scorpion *Mycoplasma* fall within the hominis group of *Mycoplasma*. The scale represents the number of substitutions per site.

function (Table S2). Figure S1 shows a comparison of genomic structure among the three *M. vittatus*. Table S5 shows the genome comparison of *M. vittatus* to *M. pulmonis* and *M. sculptratus*.

We identified 21 genes associated with virulence, disease, and defense (Table S2 & S3). Ten genes were those associated with the *Mycobacterium* virulence operon and six genes associated with fluoroquinolone resistance. Ten genes associated with the virulence operon were clustered into three genomic regions: one with three RNA polymerase genes, a second with two LSU ribosomal protein genes with a translation initiation factor, and a third with two SSU ribosomal proteins with a translation factor G gene. Four genes were identified for membrane transport as ABC transporters (Table S3). Interestingly, no genes associated with adhesion, toxins, nor anti-bacterial peptides were identified, which is also reflected in a SEED analysis of *M. pulmonis*. And, a single gene associated with CRISPR proteins was identified: Cas1.

*Mycoplasma* are known for reduced biosynthetic activities. We identified 297 genes for metabolic activities, but only two genes associated with the respiratory dehydrogenases were identified (Table S4). No genes associated with the ATP synthase complex, nor the electron accepting reactions were identified. The remainder were listed in the following categories: protein metabolism 93, carbohydrates metabolism 38, DNA metabolism 32, RNA metabolism 23, potassium metabolism 2, respiration 2, Lipid metabolism 1, and sulfur metabolism 1.

## DISCUSSION

The *M. vittatus* genome is substantially smaller than *M. pulmonis*, but slightly larger than the genome of *H. crinochetorum* (657,101bp). With a GC content of 42.70%, it also exhibits a higher GC content than other *Mycoplasma* with GC contents between 20–40% (Thompson *et al.* 2011).

As *Mycoplasma* exist primarily as intracellular symbionts (Chen *et al.* 2017), transporter systems are crucial in obtaining nutrients from their hosts. In the *M. vittatus*, four genes identified in the transporter system fall into the ABC transporter system (Table S3). In *H. crinochetorum* only five genes were identified through the SEED subsystem analysis in the transporter system with none associated with ABC transporters: three genes were identified as involved with protein translocation across cytoplasmic systems and two as cation transporters. *M. pulmonis* exhibited 22 genes associated with membrane transport with the bulk (11/22) in the ABC transporter category.

*M. vittatus* shows 21 genes in virulence, disease and defense (Table S3). Six genes produce proteins to resist fluoroquinolones, namely in DNA replication (Gyrase and Topoisomerase subunits). The rest are associated with invasion and intracellular resistance. In *H. crinochetorum* 14 genes are virulence, disease, and defense with four genes associated with fluoroquinolone resistance (Gyrase and Topoisomerase subunits), eight in invasion and intracellular resistance, one in Copper homeostasis, and one in multidrug resistance efflux pumps. *M. pulmonis* exhibited 14 genes associated with virulence, disease, and defense with the bulk (9/14) in the invasion and intracellular resistance subsystem.

Our phylogenetic analysis indicates *M. vittatus* is nested in the Hominis group of *Mycoplasma* along with a distinct *Mycoplasma* in the related scorpion *C. sculptratus* (Figure 2). This phylogeny mirrors other *Mycoplasma* phylogenies (Thompson *et al.* 2011, Liu *et al.* 2012, Hicks *et al.* 2014). In comparison with the nearest similar species, *M. pulmonis*, *M. vittatus* exhibits major variation in genome size, gene number, GC content, and average gene length. In fact, *M. vittatus* possesses a genome significantly smaller than those of other species within the Hominis group in our phylogenetic tree (683,827bp

vs. mean = 887,259), and suggests arthropod *Mycoplasma* may house reduced genomes when compared to those symbiotic with vertebrates. These differences also suggest complex evolutionary histories and selection pressures are needed to produce diverse genomes in the *Mycoplasma* phylogeny. Further genomic studies in arthropod genomics should reveal other intricate relationships between arthropods and their *Mycoplasma* symbionts.

## ACKNOWLEDGMENTS

This publication was made possible by the Arkansas INBRE program, supported by a grant from the National Institute of General Medical Sciences, (NIGMS), P20 GM103429 from the National Institutes of Health.

## AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: T.Y. and D.R. Performed the experiments: T.Y., D.R., and J.P. Analyzed the data: T.Y., D.R., and J.P. Contributed reagents/materials/analysis tools: T.Y., D.R., and J.P. Wrote the paper: T.Y., and D.R. All authors read and approved the final manuscript. The authors declare that they have no competing interests. The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the manuscript writing, and in the publication decision.

## LITERATURE CITED

- Aziz, R. K., D. Bartels, A. A. Best, M. DeJongh, T. Disz *et al.*, 2008 The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9: 75. <https://doi.org/10.1186/1471-2164-9-75>
- Bolaños, L. M., L. E. Servin-Garcidueñas, and E. Martínez-Romero, 2015 Arthropod-*Spiroplasma* relationship in the genomic era. *FEMS Microbiol. Ecol.* 91: 1–8. <https://doi.org/10.1093/femsec/fiu008>
- Browning, G. F., and C. Citti, 2014 Preface, *Mollicutes: Molecular biology and pathogenesis*, edited by Browning, G. F., and C. Citti. Caister Academic Press, Norfolk, UK.
- Chen, S., H. Hao, P. Zhao, F. Thiaucourt, Y. He *et al.*, 2017 Genome-Wide Analysis of the First Sequenced *Mycoplasma capricolum* subsp. *capripneumoniae* Strain M1601. *G3: Genes, Genomes. Genetics* 7: 2899–2906. <https://doi.org/10.1534/g3.117.300085>
- Darling, A. C. E., B. Mau, F. R. Blattner, and N. T. Perna, 2004 Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14: 1394–1403. <https://doi.org/10.1101/gr.2289704>
- Grant, J. R., and P. Stothard, 2008 The CGView Server: a comparative genomics tool for circular genomes. *Nucleic Acids Res.* 36: W181–W184. <https://doi.org/10.1093/nar/gkn179>
- Hicks, C. A. E., E. N. Barker, C. Brady, C. R. Stokes, C. R. Helps *et al.*, 2014 Non-ribosomal phylogenetic exploration of *Mollicute* species: New insights into *haemoplasma* taxonomy. *Infect. Genet. Evol.* 23: 99–105. <https://doi.org/10.1016/j.meegid.2014.02.001>
- Leclercq, S., J. Dittmer, D. Bouchon, and R. Cordaux, 2014 Phylogenomics of “*Candidatus Hepatoplasma crinochetorum*,” a lineage of *Mollicutes* associated with noninsect arthropods. *Genome Biol. Evol.* 6: 407–415. <https://doi.org/10.1093/gbe/evu020>
- Liu, W., L. Fang, M. Li, S. Li, S. Guo *et al.*, 2012 Comparative Genomics of *Mycoplasma*: Analysis of Conserved Essential Genes and Diversity of the Pan-Genome. *PLoS One* 7: e35698. <https://doi.org/10.1371/journal.pone.0035698>
- Overbeek, R., T. Begley, R. M. Butler, J. V. Choudhuri, H. Y. Chuang *et al.*, 2005 The Subsystems Approach to Genome Annotation and its Use in the Project to Annotate 1000 Genomes. *Nucleic Acids Res.* 33: 5691–5702. <https://doi.org/10.1093/nar/gki866>
- Schwager, E. E., P. P. Sharma, T. Clarke, D. J. Leite, T. Wierschin *et al.*, 2017 The house spider genome reveals an ancient whole-genome

- duplication during arachnid evolution. *BMC Biol.* 15: 62. <https://doi.org/10.1186/s12915-017-0399-x>
- Thompson, C. C., N. Y. Vieira, A. C. P. Vicente, and F. L. Thompson, 2011 Towards a genome based taxonomy of *Mycoplasmas*. *Infect. Genet. Evol.* 11: 1798–1804. <https://doi.org/10.1016/j.meegid.2011.07.020>
- Van Domselaar, G.H., P. Stothard, S. Shrivastava, J.A. Cruz, A. Guo *et al.*, 2005 BASys: a web server for automated bacterial genome annotation. *Nucleic Acids Res.* Jul 1;33(Web Server issue):W455–9. <https://doi.org/10.1093/nar/gki593>
- Walker, B. J., T. Abeel, T. Shea, M. Priest, A. Abouelliel *et al.*, 2014 Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS One* 9: e112963. <https://doi.org/10.1371/journal.pone.0112963>

*Communicating editor: B. Oliver*