



Chromosome-level genomic assembly and allergome inference reveal novel allergens in the storage mite, *Tyrophagus putrescentiae**

YING ZHOU¹, PAVEL B. KLIMOV², XIAOHONG GU³, ZHIWEI YU³, XIAOCHUAN CUI⁴, QINGQING LI⁵, RUILIN PAN⁵, CUNYIN YUAN⁵, FANGFANG CAI⁵ & YUBAO CUI^{5,6}

¹Department of Pediatrics Laboratory, The Affiliated Wuxi Children's Hospital of Nanjing Medical University, Wuxi 214023, China

²Department of Biological Sciences, Purdue University, IN 47907, USA

³Department of Respiratory, The Affiliated Wuxi Children's Hospital of Nanjing Medical University, Wuxi 214023, China

⁴Department of General Medicine, The Affiliated Wuxi People's Hospital of Nanjing Medical University, Wuxi 214023, China

⁵Clinical Research Center, The Affiliated Wuxi People's Hospital of Nanjing Medical University, Wuxi 214023, China

⁶Corresponding author: ybcui1975@hotmail.com; <https://orcid.org/0000-0001-7736-8012>

*In: Zhang, Z.-Q., Fan, Q.-H., Heath, A.C.G. & Minor, M.A. (Eds) (2022) *Acarological Frontiers: Proceedings of the XVI International Congress of Acarology (1–5 Dec. 2022, Auckland, New Zealand)*. Magnolia Press, Auckland, 328 pp.

Tyrophagus putrescentiae is a common storage mite in house dust, stored grains and other food. It has been demonstrated as an important indoor allergen source resulting in IgE-mediated responses and allergic diseases in genetically predisposed patients. Identification and characterization of its allergens using whole genome sequencing is needed for improving allergy diagnostics and developing better immunotherapeutic vaccines. This study aimed to uncover previously unrecognized allergen and obtain a more complete *T. putrescentiae* genome by using a combination of three sequencing technologies (Illumina HiSeq 2000, PacBio Sequel and Hi-C). Short reads from the Illumina HiSeq X platform were obtained from a single library with an insert size of 400 bp, generating 15.560 Gb of data, whereas PacBio long reads were obtained from a 20-kb library, generating 19.969 Gb of data. After quality processing, PacBio reads were assembled into 169 contigs, with a contig N50 of 5,581,502 bp; 93.48% of Illumina reads were mapped onto this assembly. Among the 1066 Benchmarking Universal Single-Copy Orthologs (BUSCO), we identified 986 complete and 12 fragmented BUSCO core genes, accounting for 92.50% BUSCO genes in the genome. Prior to assembly, we removed non-mite DNA contamination after a metagenomic analysis. Our assembly had 145 scaffolds. After considering genome-wide Hi-C mapping information, we recovered eight pseudo-chromosomes with lengths between 9,159,145 and 16,055,500 bp, or 97.16% of the total genome sequence. We applied the 3D Hi-C approach to assemble the reference genome at the chromosome level, resulting in a contiguous 102.28-Mb final genome. All genomes of acariform mites sequenced to date are smaller than 200 Mb, while those of parasitiform mites usually have larger genomes up to 2.1 Gb in ticks. Among genomes of most other mites, particularly *Dermatophagoides farinae* and *D. pteronyssinus*, this genome also has one of the largest N50 values, indicating its high contiguity. Hence, this *T. putrescentiae* genome is one of the most complete ones of domestic mites sequenced to date. A total of 15.86 Mb of repeat sequences were identified, accounting for 15.51% of the genome. The top five repeat elements were simple repeats, unclassified repeats, long interspersed nuclear elements, DNA transposons, and long terminal repeats. Our transcriptome-based gene prediction pipeline recovered 18,763 protein-encoding genes, with an average coding DNA length of 1,329.46 bp, and 82,193 mRNA sequences, with the total mRNA length of 68,807,924 bp and average mRNA length of 837.15 bp. Combining our Illumina, PacBio, Hi-C, and transcriptome data and the published *T. putrescentiae* genome assembled from a single female using short Illumina reads at 1025x coverage (PRJNA598686), we predicted 39 allergen groups. These allergen-encoding genes were mapped on the mite chromosomes. The WHO/IUIS Allergen Nomenclature Sub-Committee lists 12 allergen groups in *T. putrescentiae* (Tyr p 1–3, 7, 8, 10, 13, 20, 28, and 34–36); therefore, 27 other predicted allergens from this study represent novel putative allergens plus a new isoform of Tyr p 20.0201. When tested against sera from sensitized patients, novel allergens group 11, 20.0201, and 32 elicited IgE responses in 72.7%, 45.5%, and 72.7% of cases, respectively. These novel allergens were accepted by the World Health Organization/IUIS Allergen Nomenclature Sub-committee (<http://www.allergen.org>). Our chromosome-scale assembly of *T. putrescentiae* genome and bioinformatics analyses revealed an almost complete allergome and also identified three novel allergen-encoding genes of this medically and agriculturally important mite. The results of this study provide a solid foundation for the development of component-resolved allergy diagnostics and improved immunotherapeutic vaccines for *T. putrescentiae*-induced allergies in the future.

Keywords: *Tyrophagus putrescentiae*, storage mite, allergen, genome, allergome