



Macrosynteny analysis shows the absence of ancient whole-genome duplication in lepidopteran insects

Yoichiro Nakatani^{a,1} and Aoife McLysaght^{a,1}

Whole-genome duplication (WGD) is considered a key evolutionary event for genetic innovation and has occurred in diverse eukaryotic lineages (1). In PNAS, Li et al. (2) report multiple WGD events in hexapods, which is surprising because WGD is thought to be disruptive in animals with chromosomal sex determination and therefore expected to be rare (1). The argument by Li et al. (2) is based on gene-tree and molecular-clock analyses (except for a scaffold-level synteny analysis of the silkworm genome). However, gene-tree-based (or molecular-clock-based) approaches were previously shown to be potentially misleading (3, 4), whereas chromosome-scale macrosynteny-based approaches are more reliable and conclusive (5, 6). Here we examine the hypothesized Lepidoptera-specific genome duplication by macrosynteny analysis of the chromosome-level silkworm genome assembly (7). Contrary to the scaffold-scale synteny analysis performed by Li et al. (2), our chromosome-scale analysis shows the absence of WGD in silkworm and suggests that several large-scale genome duplications proposed by Li et al. may be artificial.

To investigate the proposed Lepidoptera-specific genome duplication, we assessed the analysis method used by Li et al. (2). They performed a BLASTP search with a nonstringent *E*-value cutoff and chained the BLASTP hits to obtain colinear paralogous regions. We visualized this analysis on the silkworm chromosomes (Fig. 1 *A* and *B*) and found that BLASTP hits were distributed randomly over the entire genome, which suggested that the resulting chains were unreliable. Indeed, several chromosomes had chains on multiple chromosomes (Fig. 1*B*), which cannot be explained by a single WGD. Furthermore, these chains were not especially enriched with Lepidoptera-specific paralogs in the Ensembl database (8) (see Fig. 1 legend for details).

Unlike the argument in ref. 2, our synteny analysis indicated absence of WGD. First, the distributions of

Lepidoptera-specific and silkworm-specific paralogs were similarly scattered (Fig. 1 *C* and *D*), suggesting that Lepidoptera-specific paralogs were created by the same process, namely small-scale duplications. Alternatively, extensive rearrangements might have shuffled the paralog distribution in the silkworm genome; however, this scenario is unlikely because macrosynteny conservation is strong between silkworm and beetle (Fig. 2). Second, we observed lack of doubly conserved synteny (5, 6, 9) in outgroup insect genomes (Fig. 2). We expect duplicated silkworm chromosomes (*i*) to have similar ortholog distributions in outgroup genomes and (*ii*) to share a large number of Lepidoptera-specific paralogs. However, silkworm chromosomes with similar ortholog distributions in beetle tended to have distinct distributions in honey bee (Fig. 2) and have only a small number of paralogs (Fig. 1*C*). These observations indicate that genome restructuring among these insect lineages occurred by rearrangements, and not by large-scale genome duplication.

In conclusion, the absence of WGD in silkworm suggests that the gene-tree-based detection of WGD is unreliable and casts doubt on the six ancient large-scale genome duplications reported in ref. 2. In addition, a recent paper extensively studied the accuracy of K_S analyses and warned of the risk of spurious WGD detection (10). For these reasons, we suggest that proposed large-scale genome duplications should be verified by chromosome-scale genome assembly, macrosynteny analysis, and reconstruction of the pre-WGD genome structure (5, 6, 9).

Acknowledgments

We thank all members of the A.McL. research group for valuable discussions and Karsten Hokamp for technical assistance and support with computational resources. This work is supported by funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013)/European Research Council Grant Agreement 309834.

^aDepartment of Genetics, Smurfit Institute of Genetics, Trinity College Dublin, University of Dublin, Dublin 2, Ireland

Author contributions: Y.N. performed research; and Y.N. and A.McL. wrote the paper.

The authors declare no conflict of interest.

Published under the PNAS license.

¹To whom correspondence may be addressed. Email: yoichiro.nakatani@tcd.ie or aoife.mclysaght@tcd.ie.

Published online January 23, 2019.

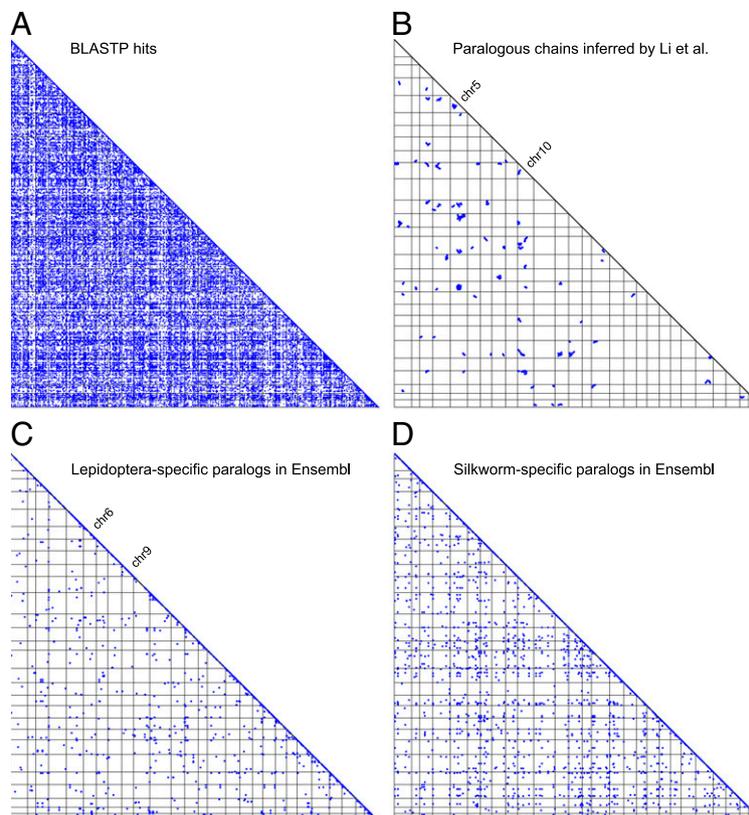


Fig. 1. Chaining of unreliable BLASTP hits results in spurious detection of large-scale genome duplications. (A) The triangular plot shows BLASTP hits with an E -value cutoff of 10^{-5} . Both axes represent the silkworm chromosomes (1 to 28 from top to bottom and from left to right). (B) Li et al. (2) chained the BLASTP hits and presented 83 high-confidence chains as dataset S6 in ref. 2. The plot shows paralog pairs in the 83 chains. Several chromosomes (e.g., chr5 and chr10) have paralogous chains on multiple chromosomes, which contradicts the model that suggests these paralogs were created by a single WGD. Furthermore, these chains do not consist of Lepidoptera-specific paralogs exclusively: They include 7 out of 4,206 Lepidoptera-specific paralog pairs, 14 out of 4,105 silkworm-specific paralog pairs, and 68 out of 7,174 Metazoa-specific paralog pairs in Ensembl Metazoa (metazoa.ensembl.org). (C) The plot shows Lepidoptera-specific paralogs (annotated as *Obtectomera* in Ensembl Metazoa). The genome-wide distribution of Lepidoptera-specific paralogs disagrees with the chains inferred by Li et al. (2) and suggests the absence of chromosome-scale duplications in the silkworm lineage. For example, lack of Lepidoptera-specific paralogs between chr6 and chr9 indicates that the two chromosomes were not created by Lepidoptera-specific genome duplication, although they have similar ortholog distributions in outgroup insect genomes (Fig. 2). (D) Silkworm-specific paralogs (annotated as *Bombyx mori* in Ensembl Metazoa) were plotted. The similarity of the Lepidoptera-specific and silkworm-specific paralog distributions suggests that these paralogs were created by the same mechanism (namely, small-scale duplication).

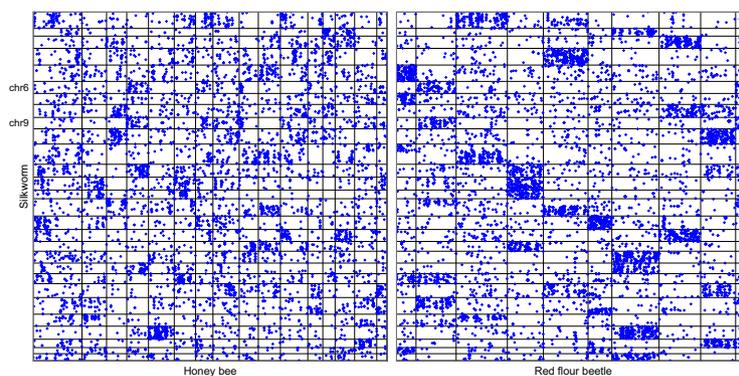


Fig. 2. The silkworm genome exhibits a strong macrosynteny conservation with the honey bee (*Apis mellifera*) and red flour beetle (*Tribolium castaneum*) genomes. The y axis represents the silkworm chromosomes and the x axes represent the bee chromosomes (1 to 16 from left to right) and the beetle linkage groups (X, 2 to 10 from left to right). Orthologs in Ensembl Metazoa were plotted. This pattern of macrosynteny conservation in the silkworm genome suggests that the genome structure evolution in these insects was driven mainly by interchromosomal rearrangements and not by large-scale structural duplication. For example, silkworm chromosomes chr6 and chr9 have similar ortholog distributions in the two x-axis species, but the two chromosomes share no Lepidoptera-specific paralogs (Fig. 1C), suggesting that chr6 and chr9 were created by fission in the lepidopteran lineage and not by WGD.

- 1 Gregory TR (2011) *The Evolution of the Genome* (Elsevier, New York).
- 2 Li Z, et al. (2018) Multiple large-scale gene and genome duplications during the evolution of hexapods. *Proc Natl Acad Sci USA* 115:4713–4718.
- 3 Wolfe KH (2001) Yesterday's polyploids and the mystery of diploidization. *Nat Rev Genet* 2:333–341.
- 4 Durand D, Hoberman R (2006) Diagnosing duplications—Can it be done? *Trends Genet* 22:156–164.
- 5 Kellis M, Birren BW, Lander ES (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* 428:617–624.
- 6 Jaillon O, et al. (2004) Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* 431:946–957.
- 7 The International Silkworm Genome Consortium (2008) The genome of a lepidopteran model insect, the silkworm *Bombyx mori*. *Insect Biochem Mol Biol* 38:1036–1045.
- 8 Kersey PJ, et al. (2018) Ensembl genomes 2018: An integrated omics infrastructure for non-vertebrate species. *Nucleic Acids Res* 46:D802–D808.
- 9 Nakatani Y, McLysaght A (2017) Genomes as documents of evolutionary history: A probabilistic macrosynteny model for the reconstruction of ancestral genomes. *Bioinformatics* 33:i369–i378.
- 10 Tiley GP, Barker MS, Burleigh JG (2018) Assessing the performance of *Ks* plots for detecting ancient whole genome duplications. *Genome Biol Evol* 10:2882–2898.