# Correction

## EVOLUTION

Correction for "Structural complexity and molecular heterogeneity of a butterfly ejaculate reflect a complex history of selection," by Camille Meslin, Tamara S. Cherwin, Melissa S. Plakke, Brandon S. Small, Breanna J. Goetz, Nathan I. Morehouse, and Nathan L. Clark, which was first published June 19, 2017; 10.1073/pnas.1707680114 (*Proc Natl Acad Sci USA* 114:E5406–E5413).

The authors note that Jason Hill and Christopher W. Wheat should be added to the author list. Jason Hill should be added between Melissa S. Plakke and Brandon S. Small, and Christopher W. Wheat should be added between Breanna J. Goetz and Nathan I. Morehouse. Both Jason Hill and Christopher W. Wheat should be credited with performing research and analyzing data. The author line, affiliation line, and author contributions have been corrected online and appear below.

The authors also note that the following paragraphs should have appeared in the *Methods* section:

**Sample Collection And Genome Sequencing of *Pieris napi*.** Pupal DNA was isolated from a fourth-generation inbred cohort that originated from a wild-caught female collected in Skåna, Sweden, using a standard salt extraction (1). Illumina sequencing was used for all data generation used in genome construction. A 180-bp paired end (PE) and the two mate-pair (MP) libraries were constructed at Science for Life Laboratory, the National Genomics Infrastructure, Sweden (SciLifeLab), using one PCR-free PE DNA library (180 bp) and two Nextera MP libraries (3 kb and 7 kb) all from a single individual. All sequencing was performed on Illumina HiSeq 2500 High Output mode, PE 2 × 100 bp by SciLifeLab. An additional two 40-kb MP fosmid-jumping libraries were constructed from a sibling used in the previous library construction. Genomic DNA, isolated as above, was shipped to Lucigen Co. for the fosmid-jumping library construction and sequencing was performed on an Illumina MiSeq using 2 × 250-bp reads (2).

***P. napi* Genome Assembly.** Nearly 500 M read pairs of data were generated, providing ∼285× genomic coverage. The 3-kb and 7-kb MP pair libraries were filtered for high-confidence true MPs using Nextclip v0.8 (3). All read sets were then quality filtered, the ends trimmed of adapters and low-quality bases, and screened of common contaminants using BBDuk v37.51 (BBTools, Brian Bushnell). Insert size distributions were plotted to assure high library quality. The 180-bp-, 3-kb-, and 7-kb-read datasets were used with AllPaths-LG r50960 for initial contig generation and scaffolding (4). AllPaths-LG was run with *haploidify = true* to compensate for the high degree of heterozygosity. The conserved single-copy ortholog content of the initial contig assembly was assessed at 78% for *P. napi* by CEGMA v2.5 (5). A further round of superscaffolding using the 40-kb libraries alongside the 3-kb and 7-kb libraries was performed using SSPACE v2, resulting in an N50 value of 300 kbp (6).

**Alignment of Orthologs Between *P. napi* and *P. rapae*.** Protein-coding exons from *P. rapae* genes were BLASTed against the *P. napi* scaffolds. The best hit below an E-value of $1 \times 10^{-5}$ was used if it reciprocally BLASTed back to the original query sequence from *P. rapae*. These exons were concatenated into protein-coding sequences, and a pairwise alignment was made for $d_N/d_S$ estimation using CODEML (7).

**Camille Meslin[a], Tamara S. Cherwin[b,c], Melissa S. Plakke[c], Jason Hill[d], Brandon S. Small[b], Breanna J. Goetz[c], Christopher W. Wheat[d], Nathan I. Morehouse[c,e,1], and Nathan L. Clark[b,c,1]**

[a]Institut National de la Recherche Agronomique (INRA), Institute of Ecology and Environmental Sciences of Paris (IEES-Paris), 78026 Versailles Cedex, France; [b]Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA 15260; [c]Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260; [d]Department of Zoology, Stockholm University, S-10691 Stockholm, Sweden; and [e]Department of Biological Sciences, University of Cincinnati, Cincinnati, OH 45221

Author contributions: C.M., N.I.M., and N.L.C. designed research; C.M., T.S.C., M.S.P., J.H., B.S.S., B.J.G., C.W.W., N.I.M., and N.L.C. performed research; C.M., N.I.M., and N.L.C. contributed new reagents/analytic tools; C.M., T.S.C., M.S.P., J.H., B.S.S., B.J.G., C.W.W., N.I.M., and N.L.C. analyzed data; and C.M., M.S.P., N.I.M., and N.L.C. wrote the paper.

1. Aljanabi SM, Martinez I (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Res* 25:4692–4693.
2. Bentley DR, et al. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456:53–59.
3. Leggett RM, Clavijo BJ, Clissold L, Clark MD, Caccamo M (2014) NextClip: An analysis and read preparation tool for Nextera Long Mate Pair libraries. *Bioinformatics* 30:566–568.
4. Gnerre S, et al. (2011) High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci USA* 108:1513–1518.
5. Parra G, Bradnam K, Korf I (2007) CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23:1061–1067.
6. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W (2011) Scaffolding preassembled contigs using SSPACE. *Bioinformatics* 27:578–579.
7. Bielawski JP, Yang Z (2005) *Maximum likelihood methods for detecting adaptive protein evolution. Statistical Methods in Molecular Evolution*, ed Nielsen R (Springer, Berlin), pp 103–124.



Published online February 26, 2018.

www.pnas.org/cgi/doi/10.1073/pnas.1801459115