Title

Chromosome level assembly of the comma butterfly (*Polygonia c-album*)

Authors

Maria de la Paz Celorio-Mancera¹d (0000-0003-0296-0577), Pasi Rastas²d (0000-0003-2768-1339), Rachel A. Steward¹ (0000-0001-8610-334X), Soren Nylin¹ (0000-0003-4195-8920), Christopher W. Wheat¹* (0000-0003-1863-2340)

¹Stockholm University, Faculty of Science, Department of Zoology
²Institute of Biotechnology, University of Helsinki, Finland
dBoth authors contributed equally.
*Author for Correspondence: Christopher W. Wheat, Department of Zoology, Population Genetics, Faculty of Science, Stockholm University, Stockholm, Sweden, Tel. +46 721 958586, E-mail: chris.wheat@zoologi.su.se

Abstract

The comma butterfly (*Polygonia c-album*, Nymphalidae, Lepidoptera) is a model insect species, most notably in the study of phenotypic plasticity and plant-insect coevolutionary interactions. In order to facilitate the integration of genomic tools with a diverse body of ecological and evolutionary research, we assembled the genome of a Swedish comma using 10X sequencing, scaffolding with matepair data, genome polishing, and assignment to linkage groups using a high-density linkage map. The resulting genome is 373 Mb in size, with a scaffold N50 of 11.7Mb and contig N50 of 11.2Mb. The genome contained 90.1% of single-copy Lepidopteran orthologs in a BUSCO analysis of 5286 genes. A total of 21,004 gene-models were annotated on the genome using RNAseq data from larval and adult tissue in combination with proteins from the Arthropoda database, resulting in a high-quality annotation for which functional annotations were generated. We further documented the quality of the

© The Author(s) 2021. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
chromosomal assembly via synteny assessment with *Melitaea cinxia*. The resulting annotated, chromosome-level genome will provide an important resource for investigating coevolutionary dynamics and comparative analyses in Lepidoptera.

**Keywords**

linkage map, butterfly genome, quantitative annotation assessment, comparative genomics, Polygonia c-album

**Significance statement**

The *Polygonia c-album* butterfly is considered a model species for the study of evolutionary interactions between insects and their host plants. However, it is conspicuously absent in genomic and genetics literature. We provide a chromosome-level genome for this species in order to facilitate the integration of functional and population genomic research with ecology, physiology and evolutionary findings. Assessment of our annotated genes suggest a high quality de novo assembly. Chromosome level assembly accuracy was validated via alignment with the genome of another nymphalid species, *Melitaea cinxia*.

**Introduction**

Butterflies have long served as model species for a wide range of research, from ecology to studies of developmental evolution (Boggs, Watt and Erhlich, 2003). Within this diverse field of species and questions, research using the comma butterfly, *Polygonia c-album* (Nymphalidae, Lepidoptera), has made extensive contributions, in particular to the study of plant-insect coevolution. Most butterflies are specialists at the level of individual plant families, making the host plant repertoire of *P. c-album* notable as it includes several families in four different plant orders. Two additional observations make the dramatically higher diversity of
the host plant repertoire of *P. c-album* even more interesting. First, related species from the same tribe (Nymphalini) have a host repertoire that is mostly a subset of the *P. c-album* hosts (Nylin 1988). Second, the larvae of these other species often can feed on the diverse hosts of *P. c-album*, even when females of these species no longer use those plants for oviposition (Janz et al. 2001). These patterns suggest that the diverse host plant repertoire of *P. c-album* reflects the suite of host plants used during the evolution of the tribe, an observation that inspired the “oscillation hypothesis” of host range and speciation (Janz and Nylin 2008). The oscillation hypothesis is an important alternative to the classical coevolution hypothesis, for explaining the striking diversification of phytophagous insects as well the ecological and evolutionary patterns seen in other coevolutionary interactions, including pollination and emerging infectious diseases (Braga et al. 2018; Hamm and Fordyce 2015; Hardy and Otto 2014; Hoberg and Brooks 2015; Sedivy et al. 2011).

Research specifically using *P. c-album* itself as the model has also generated many other insights into insect-plant systems, concerning host repertoires of adults vs. larvae (Nylin and Janz 1996), preference-performance correlations (Janz et al. 1994), female host search strategies, and neural constraint and plasticity (Carlsson et al. 2011; Gamberale-Stille et al. 2019; Schäpers et al. 2015; van Dijk et al. 2017;) and genetics of host use within and among populations (Nygren et al. 2006). Other research areas that are making considerable use of *P. c-album* and close relatives as model species include effects of temperature and climate change (Audusseau et al. 2013; Braschler and Hill 2007; Hodgson et al. 2011) as well as seasonal plasticity, life history regulation and seasonal polyphenism (Eriksson et al. 2020; Hiroyoshi et al. 2018; Inoue et al. 2005).

Finally, two studies have investigated transcriptome plasticity in larvae depending on host plants, using RNASeq and GeneFishing, respectively (Celorio-Mancera et al. 2013; Heidel-
Fischer et al. 2009) but the analysis and interpretation of results were constrained by the lack of a published genome. 

In order to facilitate insights at the genomic level into these extensively studied coevolutionary dynamics and plastic phenotypes, here we present a chromosomal assembly of the *P. c-album* genome, the result of combining Illumina sequencing data from 10 X and matepair data, with a high-density linkage map. Together with our validated functional annotation, this genomic resource will greatly facilitate future studies using the species as a model, as well as provide an important genome for comparative evolutionary analyses of the Lepidoptera.

**Results and Discussion**

**Genome assembly**

Using 197 Million 10X reads, 11 genomes were assembled using Supernova with a range of data input (15 to 100%; Supplementary F1), an optimal assembly using 70% data was identified, based contiguity and lowest percentage of missing BUSCOs (scaffold N50 of 76.5 Kb and 4.2% missing BUSCOs; Figure 1). Scaffolding with a 3kb mate-pair library increased the N50 to 519.1 kbp, with subsequent haplotype merging further increasing N50 to 572.5 kbp. Genome polishing with Pilon using three different mapping programs found that bam files generated by NGM outperformed the rest by displaying the longest N50, high genome completeness and lowest recovery of duplicates which may indicate erroneous assembly of haplotypes (Supplementary F2), and was thus used for downstream steps.

**Linkage map**

To generate a chromosome-level assembly, we used a linkage mapping dataset, which also provide information on recombination rate, providing insights into the relationship between
physical and genetic distance. Using RAD-seq data from 287 sexed individuals, composed of two families with full-sibs and corresponding parents, we identified 84422 candidate SNPs, which allowed us to identify 12541 markers in 31 linkage groups. This is consistent with the reported *P. c-album* karyotype of 30 autosomes and one sex chromosome (Robinson 1971). Comparisons between physical and genetic distance revealed variable recombination landscapes across chromosomes (Figure 1), with an overall high level of recombination across chromosomes typical of butterfly species (Martin, et al. 2016). Using this we were able to anchor 1,366 scaffolds, of which we could orient 550, totaling 86% and 69% of the assembly length, respectively. The resulting chromosome-level assembly consisted of 31 scaffolds with an N50 of 11.7Mb, with 13,625 unplaced scaffolds (ranging from 502 to 390,522 bp in length, an N50 = 4,741 bp, and total length of 51.7 Mbp). We then validated the chromosome structure of our assembly via alignment to the chromosome-level assembly of *M. cinxia*, which last shared a common ancestor with *P. c-album* ca. 42 million years ago (Chazot, et al. 2019), finding a high concordance across chromosomes (Figure 2).

*Genome annotation and validation*

The chromosomal genome completeness was assessed using BUSCO, which identified 90.1% of the Lepidoptera ortholog dataset (N=5286) as complete and single copy, 0.3% duplicated, 4.7% fragmented and 4.9% missing (Supplementary F2). We next compared genome annotations generated either using a protein sequence dataset for Arthropoda, RNA-Seq data from our focal species, or both, using Braker2 v.2.1.5, expecting similar ability to predict genes when training the algorithm with the different datasets (Brůna et al., 2021). Quantitatively, the RNA-Seq dataset allowed the software to predict 358 more unique genes than the protein dataset regardless of isoform number per gene. When the number of genes was limited to only those consisting of one isoform, the protein dataset predicted 1011 more. For a qualitative
assessment, the longest ortholog hit ratio (OHR) (O’Neil, et al. 2010; Hornett and Wheat 2012) between the predicted gene sequences by either dataset (protein- or transcript-based) and the B. mori gene set (Figure 2) was calculated, finding no differences in gene prediction success between the two algorithm-training strategies. When using both the Arthropod protein database and the P. c-album RNAseq data, and considering only one isoform per gene, the joined training set predicted 199 more unique genes than when using the RNA-seq set and 812 less unique genes when using the protein set only. However, there was an improvement in training capacity of the algorithm when using both datasets together, as more complete homologs in the P. c-album genome were identified (Figure 2); using both datasets generated more accurate annotations.

Here we report a chromosome-level genome assembly for the Nymphalid butterfly P. c-album. Using a linkage map we were able to place 86% our assembly into a chromosomal context, with the number of chromosomes and their genic content highly syntenic in comparison to a related butterfly ca. 42 million years divergent. Quantitative assessment of alternative genome polishing methods, as well as genome annotation methods, supports our chosen pipeline for a high-quality assembly. Together with our validated functional annotation, this genomic resource will greatly facilitate future studies using the species as a model, as well as provide an important genome for comparative evolutionary analyses of the Lepidoptera.

**Material and Methods**

**Biological samples**

Material for the genome was generated from P. c-album butterflies, collected in Stockholm area (years: 2013-2015). The laboratory population was inbred for five generations, with a last generation, female pupa used for DNA extraction. The offspring (F1 pupae) of two additional
mating pairs (wild female x inbred male) was used for the mapping analysis (Family B = 140 F1, and Family H=141 F1).

**DNA and RNA sampling and sequencing**

Two DNA extraction protocols were used. A phenol-chloroform procedure combining salting out extraction (Woronik et al., 2019) was used for inbred female pupae, and a robot-based protocol for the samples used for linkage mapping, following manufacturer’s instructions (KingFisher Cell and Tissue DNA Kit and the KingFisher Duo Prime System, Thermo Scientific, MA, USA). RNA was extracted from adult antenna, tarsi and larval gut tissue (34 samples total, approx. 30 M reads) using a phenol-guanidinium thiocyanate protocol (TRIzol, Thermo Fisher Scientific, MA, USA) and followed by either BCP (1-bromo, 3- chloropropane; Merck KGaA, Darmstadt, Germany) or column-based chemistry (Direct-zol RNA MiniPrep, Zymo, CA, USA). Samples were quantified using fluorometry (Qubit, Thermo Scientific, MA, USA) and the corresponding library preparation, sequencing and Falcon assemblies were performed at the National Genomics Infrastructure Sweden (NGI, Stockholm, Sweden). For genome assembly, three Illumina DNA libraries were produced: short insert (180 bp), long insert (3 Kb) mate-pair libraries (Illumina TruSeq PCR-free and Nextera libraries, respectively), and 10X Chromium Genome library, all sequenced on Illumina HiSEqX. Read trimming and quality filtering used bbmap, following previous work (Woronik et al., 2019).

**Genome Assembly**

De novo genome assemblies were made using Supernova v. 1.21 (Weisenfeld, et al. 2017) by following the authors’ recommendation to barcode subsample genomes smaller than 1.6 Gb, i.e. using the parameters “--bcfrac” and “--maxreads”. Assemblies were generated for barcode fractions 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.5, 0.6, 0.7, 0.9 and 1.0 and subsequently evaluated
using QUAST version 4.5.4 (Gurevich, et al. 2013), MultiQC v. 1.3 (Ewels, et al. 2016) and BUSCO v. 2.0.1 (Simao, et al. 2015) with its “eukaryota_odb9” dataset. Scaffolding of this genome using the mate-pair data started with read data filtering using NEXTCLIP (Leggett, et al. 2014) followed by scaffolding using BESST v2.0 (Sahlin, et al. 2016). Alternative haplotypes were then merged, after soft masking repeats using RED v. 05/22/2015 (Girgis 2015), using Haplomerger2 v. 20180603 (Huang, et al. 2017), with detected tandem duplicates removed settings. Polishing of the merged genome using the short insert library was performed using Pilon v. 1.23 (Walker, et al. 2014), with reads mapped using NextGenMap-0.5.0 (Sedlazeck, et al. 2013), bwa v. 0.7.17 (Li and Durbin 2009) and SNAP (Zaharia, et al. 2017).

**Linkage map construction and scaffold anchoring**

We obtained RAD-seq data from 287 sexed individuals composed of two families with full-sibs and corresponding parents. The data was generated with nextRAD methology implemented by SNPsaurus (Oregon, USA). Reads were mapped to the genome using bwa mem and together with samtools (Li et al.,2009) sorted individual bam files were created. The samtools mpileup and Lep-MAP3 (Rastas, 2017) pipeline was used to get genotype likelihoods for the map construction. The map construction pipeline used default parameters, except 1) ZLimit=2 in ParentCall2 to call Z/W markers, 2) dataTolerance=0.0001 in Filtering2, 3) informativeMask=2 in SeparateChromosomes2 in order to find linkage groups robustly using only non-recombining female information and lodLimit=30 and lodDifference=5 in JoinSingles2All to add male informative markers to the map, recombination2=0 and informativeMask=13 and calculate Intervals in OrderMarkers2 to ignore the non-recombining female information in the final maps and to output information on the map uncertainty. The scaffold anchoring was obtained using a preliminary version of Lep-Anchor (Rastas, 2020) using the linkage map. This linkage map

http://mc.manuscriptcentral.com/gbe
was re-evaluated in the found scaffold order (parameter evaluateOrder in OrderMarkers2), and based on these maps some minor manual fixes on 5 chromosomes were performed.

**Genome annotation, validation and functional annotation**

After soft-masking the final genome version, annotations were performed using Braker2 v. 2.1.5 (Brůna, et al. 2021 and references herein), in the genome mode, training Augustus using either the RNA-Seq, the protein mode or both, with reference proteins from the Arthropoda section of OrthoDB v. 10 (Kriventseva et al., 2019) and our RNA-Seq data mapped against the genome using HiSat2 2.1.0 (Kim, et al. 2019).

The qualities of the RNA, protein, and RNA + protein annotations were assessed using the longest ortholog hit ratio (OHR) (O’Neil, et al. 2010; Hornett and Wheat 2012). Protein sequences in each annotation were collapsed (CD-Hit; 90% identity) and converted to protein databases (NCBI BLAST v. 2.5.0). Protein sequences from a published Bombyx mori annotation (accessed from NCBI; GCF_000151625.1_ASM15162v1) were then blasted against the databases. The longest hit for each B. mori protein was identified and OHR was calculated as the ratio of the hit length to that of the B. mori protein.

**Comparative analysis of chromosomal structure**

We used nucmer (MUMmer4 v. 4.0.0beta2; Marçais, et al. 2018) to align the polished genome to that of the best chromosome level genome assembly from the same subfamily (Nymphalinae) as P. c-album, *Melitaea cinxia* (Blande, et al. 2020). The alignment file was filtered to retain only those aligned sequences that were longer than 200 bp and had less than 90% identity between the two genomes in contigs that were at least 1Mbp long. Alignments were visualized using the R-package circlize (Gu, et al. 2014).
Data Availability

The chromosome level assembly fasta sequence file of *P. c-album* is available on ENA (Accession number ERZ1744298). The scripts in the bioinformatic pipeline are available at https://github.com/bioinfowheat/Polygonia_calbum_genomics.

Literature cited

Audusseau H, Nylin S, Janz N 2013. Implications of a temperature increase for host plant range: predictions for a butterfly. Ecology and Evolution 3: 3021-3029. doi: 10.1002/ece3.696

Barnett DW, Garrison EK, Quinlan AR, Stromberg MP, Marth GT 2011. BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics 27: 1691-1692. doi: 10.1093/bioinformatics/btr174

Blande D, et al. 2020. Improved chromosome level genome assembly of the Glanville fritillary butterfly (*Melitaea cinxia*) based on SMRT Sequencing and linkage map. bioRxiv: 2020.11.03.364950. doi: 10.1101/2020.11.03.364950

Boggs CL, Watt WB and PR Ehrlich, editors.2003. Butterflies: ecology and evolution taking flight. Chicago: University of Chicago Press.

Braga MP, Guimaraes PR, Wheat CW, Nylin S, Janz N 2018. Unifying host-associated diversification processes using butterfly-plant networks. Nature Communications 9. doi: 10.1038/s41467-018-07677-x

Braschler B, Hill JK 2007. Role of larval host plants in the climate-driven range expansion of the butterfly *Polygonia c-album*. Journal of Animal Ecology 76: 415-423. doi: 10.1111/j.1365-2656.2007.01217.x

Bruna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR Genomics and Bioinformatics 3. doi: 10.1093/nargab/lqaa108

Carlsson MA, et al. 2011. Odour Maps in the Brain of Butterflies with Divergent Host-Plant Preferences. Plos One 6. doi: 10.1371/journal.pone.0024025

Celorio-Mancera MP, et al. 2013. Mechanisms of macroevolution: polyphagous plasticity in butterfly larvae revealed by RNA-Seq. Molecular Ecology 22: 4884-4895. doi: 10.1111/mec.12440

Chazot N, et al. 2019. Priors and Posteriors in Bayesian Timing of Divergence Analyses: The Age of Butterflies Revisited. Systematic Biology 68: 797-813. doi: 10.1093/sysbio/syz002

Eriksson M, Janz N, Nylin S, Carlsson MA 2020. Structural plasticity of olfactory neuropils in relation to insect diapause. Ecology and Evolution 10: 14423-14434. doi: 10.1002/ece3.7046

Ewels P, Magnusson M, Lundin S, Kaller M 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32: 3047-3048. doi: 10.1093/bioinformatics/btw354
Gamberale-Stille G, Schapers A, Janz N, Nylin S 2019. Selective attention by priming in host search behavior of 2 generalist butterflies. Behavioral Ecology 30: 142-149. doi: 10.1093/beheco/ary146

Girgis HZ 2015. Red: an intelligent, rapid, accurate tool for detecting repeats de-novo on the genomic scale. BMC Bioinformatics 16. doi: 10.1186/s12859-015-0654-5

Gu Z, Gu L, Eils R, Schlesner M, Brors B 2014. circlize implements and enhances circular visualization in R. Bioinformatics 30: 2811-2812. doi: 10.1093/bioinformatics/btu393

Gurevich A, Saveliev V, Vyahhi N, Tesler G 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29: 1072-1075. doi: 10.1093/bioinformatics/btt086

Hamm CA, Fordyce JA 2015. Patterns of host plant utilization and diversification in the brush-footed butterflies. Evolution 69: 589-601. doi: 10.1111/evo.12593

Hardy NB, Otto SP 2014. Specialization and generalization in the diversification of phytophagous insects: tests of the musical chairs and oscillation hypotheses. Proceedings of the Royal Society B-Biological Sciences 281. doi: 10.1098/rspb.2013.2960

Heidel-Fischer HM, et al. 2009. Phylogenetic relatedness and host plant growth form influence gene expression of the polyphagous comma butterfly (Polygonia c-album). BMC Genomics 10. doi: 50610.1186/1471-2164-10-506

Hiroyoshi S, Reddy GVP, Mitsuhashi J 2018. Effects of photoperiod, temperature and aging on adult diapause termination and post-diapause development in female Asian comma butterflies, Polygonia c-aureum Linnaeus (Lepidoptera: Nymphalidae). Journal of Comparative Physiology a-Neuroethology Sensory Neural and Behavioral Physiology 204: 849-858. doi: 10.1007/s00359-018-1284-y

Hoberg EP, Brooks DR 2015. Evolution in action: climate change, biodiversity dynamics and emerging infectious disease. Philosophical Transactions of the Royal Society B-Biological Sciences 370. doi: 10.1098/rstb.2013.0553

Hodgson JA, et al. 2011. Predicting insect phenology across space and time. Global Change Biology 17: 1289-1300. doi: 10.1111/j.1365-2486.2010.02308.x

Hornett E, Wheat C 2012. Quantitative RNA-Seq analysis in non-model species: assessing transcriptome assemblies as a scaffold and the utility of evolutionary divergent genomic reference species. BMC Genomics 13: 361.

Huang SF, Kang MJ, Xu AL 2017. HaploMerger2: rebuilding both haploid sub-assemblies from high-heterozygosity diploid genome assembly. Bioinformatics 33: 2577-2579. doi: 10.1093/bioinformatics/btx220

Inoue M, et al. 2005. Purification of Bombyx neuropeptide showing summer-morph-producing-hormone (SMPH) activity in the Asian comma butterfly, Polygonia c-aureum. Journal of Insect Science 5: 8-8.

Janz N, Nylin S. 2008. The oscillation hypothesis of host-plant range and speciation. In. Specialization, speciation and radiation: the evolutionary biology of herbivorous insects.

Janz N, Nylin S, Wedell N 1994. Host-plant utilization in the comma butterfly - sources of variation and evolutionary implications. Oecologia 99: 132-140. doi: 10.1007/bf00317093
Kim D, Paggi JM, Park C, Bennett C, Salzberg SL 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 37: 907-915. doi: 10.1038/s41587-019-0201-4

Kriventseva EV, et al. 2019. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic Acids Research 47: D807-D811. doi: 10.1093/nar/gky1053

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. doi: 10.1093/bioinformatics/btt702

Li H, et al. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078-2079. doi: 10.1093/bioinformatics/btp352

Marcais G, et al. 2018. MUMmer4: A fast and versatile genome alignment system. Plos Computational Biology 14. doi: 10.1371/journal.pcbi.1005944

Martin SH, et al. 2016. Natural Selection and Genetic Diversity in the Butterfly Heliconius melpomene. Genetics 203: 525-+. doi: 10.1534/genetics.115.183285

Nygren GH, Nylin S, Stefanescu C 2006. Genetics of host plant use and life history in the comma butterfly across Europe: varying modes of inheritance as a potential reproductive barrier. Journal of Evolutionary Biology 19: 1882-1893. doi: 10.1111/j.1420-9101.2006.01174.x

Nylin S 1988. Host plant specialization and seasonality in a polyphagous butterfly, Polygonia c-album (Nymphalidae). Oikos 53: 381-386. doi: 10.2307/3565539

Nylin S, Janz N 1996. Host plant preferences in the comma butterfly (Polygonia c-album): Do parents and offspring agree? Ecoscience 3: 285-289.

O'Neil ST, et al. 2010. Population-level transcriptome sequencing of nonmodel organisms Erynnis propertius and Papilio zelicaon. BMC genomics 11: 310-310. doi: 10.1186/1471-2164-11-310

Rastas P 2017. Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data. Bioinformatics 33: 3726-3732. doi: 10.1093/bioinformatics/btx494

Rastas P 2020. Lep-Anchor: automated construction of linkage map anchored haploid genomes. Bioinformatics 36: 2359-2364. doi: 10.1093/bioinformatics/btz978

Robinson R. 1971. Lepidoptera genetics. Oxford: Pergamon.p. 575

Sahlin K, Chikhi R, Arvestad L 2016. Assembly scaffolding with PE-contaminated mate-pair libraries. Bioinformatics 32: 1925-1932. doi: 10.1093/bioinformatics/btw064

Schäpers A, Carlsson MA, Gamberale-Stille G, Janz N 2015. The Role of Olfactory Cues for the Search Behavior of a Specialist and Generalist Butterfly. Journal of Insect Behavior 28: 77-87. doi: 10.1007/s10905-014-9482-0

Sedivy C, Muller A, Dorn S 2011. Closely related pollen generalist bees differ in their ability to develop on the same pollen diet: evidence for physiological adaptations to digest pollen. Functional Ecology 25: 718-725. doi: 10.1111/j.1365-2435.2010.01828.x
Sedlazeck FJ, Rescheneder P, von Haeseler A 2013. NextGenMap: fast and accurate read mapping in highly polymorphic genomes. Bioinformatics 29: 2790-2791. doi: 10.1093/bioinformatics/btt468

Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31: 3210-3212. doi: 10.1093/bioinformatics/btv351

van Dijk LJA, Janz N, Schapers A, Gamberale-Stille G, Carlsson MA 2017. Experience-dependent mushroom body plasticity in butterflies: consequences of search complexity and host range. Proceedings of the Royal Society B-Biological Sciences 284: 8. doi: 10.1098/rspb.2017.1594

Walker BJ, et al. 2014. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. Plos One 9. doi: 10.1371/journal.pone.0112963

Weisenfeld NI, Kumar V, Shah P, Church DM, Jaffe DB 2017. Direct determination of diploid genome sequences. Genome Research 27: 757-767. doi: 10.1101/gr.214874.116

Woronik A, et al. 2019. A transposable element insertion is associated with an alternative life history strategy. Nat Commun 10: 5757. doi: 10.1038/s41467-019-13596-2

Zaharia M, et al. 2011. Faster and More Accurate Sequence Alignment with SNAP. arXiv preprint arXiv 1111.
Figure 1. Pipeline for genome assembly and linkage map construction of the P. c-album genome. Details of the results from each step are indicated within each box.

178x104mm (300 x 300 DPI)
Figure 2. Genome validation and genetic diversity. (a) Ortholog homology ratio improved with the combination of RNA-Seq and protein data. There was greater homology between B. mori proteins and proteins predicted using both the RNA-Seq and protein trained annotation (blue) than using either RNA-Seq trained annotation (red) or protein-trained annotation (yellow) only.  (b) Synteny between the M. cinxia genome (colored chromosomes) and the P. c-album genome (non-colored linkage groups).