

## RESEARCH ARTICLE

# The Easter Egg Weevil (*Pachyrhynchus*) genome reveals syntenic patterns in Coleoptera across 200 million years of evolution

Matthew H. Van Dam<sup>1,2\*</sup>, Anlyn Anzano Cabras<sup>3</sup>, James B. Henderson<sup>2</sup>, Andrew J. Rominger<sup>4</sup>, Cynthia Pérez Estrada<sup>5</sup>, Arina D. Omer<sup>5</sup>, Olga Dudchenko<sup>5</sup>, Erez Lieberman Aiden<sup>5</sup>, Athena W. Lam<sup>2</sup>

**1** Entomology Department, Institute for Biodiversity Science and Sustainability, California Academy of Sciences, San Francisco, California, United States of America, **2** Center for Comparative Genomics, Institute for Biodiversity Science and Sustainability, California Academy of Science, San Francisco, California, United States of America, **3** Coleoptera Research Center, Institute for Biodiversity and Environment, University of Mindanao, Matina, Davao City, Philippines, **4** School of Biology and Ecology, University of Maine, Orono, Maine, United States of America, **5** The Center for Genome Architecture, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, United States of America

\* [matthewhvandam@gmail.com](mailto:matthewhvandam@gmail.com)



## OPEN ACCESS

**Citation:** Van Dam MH, Cabras AA, Henderson JB, Rominger AJ, Pérez Estrada C, Omer AD, et al. (2021) The Easter Egg Weevil (*Pachyrhynchus*) genome reveals syntenic patterns in Coleoptera across 200 million years of evolution. *PLoS Genet* 17(8): e1009745. <https://doi.org/10.1371/journal.pgen.1009745>

**Editor:** Gregory P. Copenhaver, The University of North Carolina at Chapel Hill, UNITED STATES

**Received:** January 20, 2021

**Accepted:** July 27, 2021

**Published:** August 30, 2021

**Copyright:** © 2021 Van Dam et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Part of the data are within the manuscript and its [Supporting Information](#) files and the other part is deposited in <https://doi.org/10.5061/dryad.cz8w9gj48> with identical data also in [https://www.dnazoo.org/assemblies/Pachyrhynchus\\_sulphureomaculatus](https://www.dnazoo.org/assemblies/Pachyrhynchus_sulphureomaculatus) Genome assembly from this project are archived under NCBI accession GCA\_019049505.1, RNA reads via SRA, bioproject: PRJNA731935.

## Abstract

Patterns of genomic architecture across insects remain largely undocumented or decoupled from a broader phylogenetic context. For instance, it is unknown whether translocation rates differ between insect orders. We address broad scale patterns of genome architecture across Insecta by examining synteny in a phylogenetic framework from open-source insect genomes. To accomplish this, we add a chromosome level genome to a crucial lineage, Coleoptera. Our assembly of the *Pachyrhynchus sulphureomaculatus* genome is the first chromosome scale genome for the hyperdiverse Phytophaga lineage and currently the largest insect genome assembled to this scale. The genome is significantly larger than those of other weevils, and this increase in size is caused by repetitive elements. Our results also indicate that, among beetles, there are instances of long-lasting (>200 Ma) localization of genes to a particular chromosome with few translocation events. While some chromosomes have a paucity of translocations, intra-chromosomal synteny was almost absent, with gene order thoroughly shuffled along a chromosome. This large amount of reshuffling within chromosomes with few inter-chromosomal events contrasts with patterns seen in mammals in which the chromosomes tend to exchange larger blocks of material more readily. To place our findings in an evolutionary context, we compared syntenic patterns across Insecta in a phylogenetic framework. For the first time, we find that synteny decays at an exponential rate relative to phylogenetic distance. Additionally, there are significant differences in decay rates between insect orders, this pattern was not driven by Lepidoptera alone which has a substantially different rate.

**Funding:** MHVD received grant funding from the NSF:DEB:1856402 (<https://www.nsf.gov/>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Author summary

Patterns of genomic architecture across insects remain largely undocumented or decoupled from a broader evolutionary context. For instance, it is unknown whether rates of gene order decay differ between insect orders. We address broad scale patterns of genome architecture across Insecta by examining synteny (shared gene order) in a phylogenetic framework from open-source insect genomes (143 complete chromosome assemblies in total). To accomplish this, we add a chromosome level genome to a crucial lineage, Coleoptera (beetles). Our assembly of the Easter Egg Weevil *Pachyrhynchus sulphureomaculatus* genome is the first chromosome scale genome for the hyperdiverse Phytophaga lineage and currently the largest insect genome assembled to this scale. We are the first to identify in beetles that genes stay localized on chromosomes for hundreds of millions of years, while their order along chromosomes gets completely shuffled over time. We are also the first to empirically demonstrate that synteny decay rates differ significantly between insect orders and that this pattern is not driven solely by Lepidoptera (moths and butterflies), which has a substantially different rate.

## Introduction

Beetles represent one of the most diverse groups of metazoans, with ~400,000 described species [1] and estimates of total diversity up to 0.9–2.1 million species [2]. Among beetles, weevils (Coleoptera: Curculionidae) are one of the most diverse insect groups (>60,000 described species [3]), encompassing a huge range of life history strategies and occupying every conceivable niche in a terrestrial ecosystem. With morphological forms specialized to ecological habits, such as feeding on fungi, seeds, pollen, wood, roots, and even kangaroo dung, weevils make an excellent system in which to study the evolution of different ecomorphologies [3,4]. Weevils belong to the group Phytophaga whose members comprise lineages that specialize on and have co-diversified with many plant lineages [5,6]. Given their vast diversity and economic importance as pollinators and crop pests, knowing more about the genomic architecture of beetles should be of broad applicability. However, to date, there are few available genomes resolved to chromosome level for Coleoptera and none for weevils or the hyperdiverse beetle lineage Phytophaga [7–10]. Here we present the first genome resolved to chromosome level for the Phytophaga beetle lineage *Pachyrhynchus sulphureomaculatus* Schultze, 1922 [11].

Recent advances in genome assembly techniques, such as in situ high throughput conformation capture technology (Hi-C) [12], have substantially enhanced our knowledge of genome architecture [13–15]. Increasing the accuracy and contiguity of genome assemblies has also been aided by using long-read sequencing technology in combination with in situ Hi-C [16–20]. These innovations have allowed researchers to not only reconstruct genomes to chromosome scale but also to do so relatively quickly and cheaply [21]. In addition, in situ Hi-C technology has shown that the 3D conformation of genomes is not random and that this conformation can influence gene expression and linkage [22]. The result of these new sequencing techniques has increased the number of high quality genomes for non-model insect species, including beetles [9,10,17,23–26]. Because in situ Hi-C orders scaffolds and corrects misjoins, we can study synteny (between chromosomes, unless otherwise specified) between organisms with more confidence [14,27]. In situ Hi-C is particularly important for assembling insect genomes which often have high heterozygosity as well as being composed of many repetitive elements, allowing for their assembly into chromosomes where other technologies produce significantly less contiguous and less accurate chromosome assemblies [14].

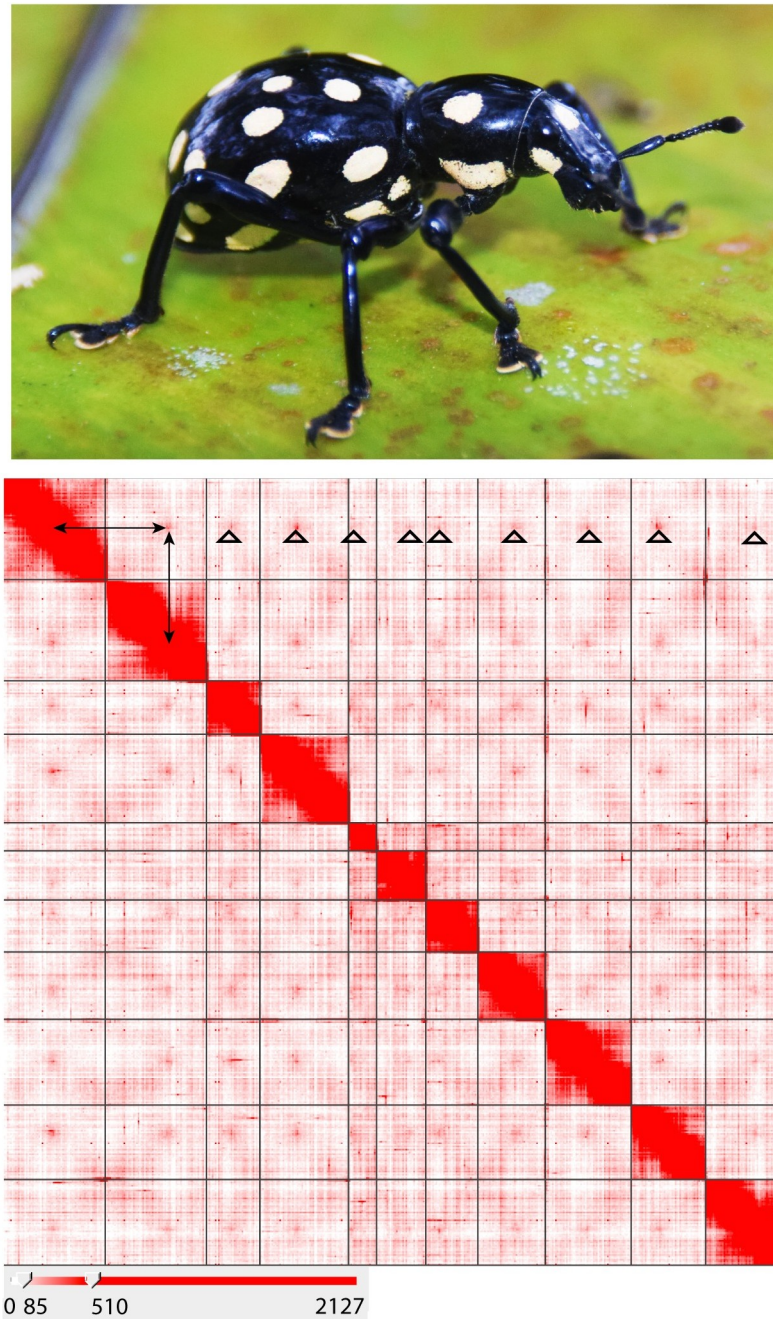
With the influx of new chromosome-level genomes, we can now begin to explore patterns of genome architecture within and between major insect lineages. For example, in Lepidoptera (butterflies and moths), genome architecture has been characterized as relatively stable with few (6%) orthologous loci being translocated [23,28–30]. Holocentric chromosomes observed throughout Lepidoptera are implicated in facilitating hybridization, [23,31–33] suggesting that genome architecture plays a significant role in their biology. In contrast to Lepidoptera, *Drosophila* species have many more translocations and rearrangements having monocentric centromeres [34]. The fungus *Cryptococcus neoformans* provides a clear example of how changing monocentric centromere position has a negative fitness costs [35]. In beetles, however, even a basic understanding of genomic architecture remains largely undocumented. The basic blueprints as revealed by in situ Hi-C maps of how a genome is organized (e.g.—with a Rab1-like conformation, i.e. grouping of telomeres and centromeres to the nuclear envelope, [36,37], holocentric chromosomes, chromosome domain territories, compartments, and topological associated domain loops) remain non-existent and therefore unplaced in a phylogenetic context. A general synthesis across insects linking these genomic architectural patterns to their function and potential influence on speciation remains incomplete. For example, do different insect orders have distinct rates of genomic rearrangements (the breakage of synteny between genes), or are the patterns we observe merely due their phylogenetic structure? The null expectation would be that there is no difference in synteny decay rate between insect orders. For the first time, we demonstrate that different insect orders do have distinct rates of synteny decay. To help accomplish this we also provide a new chromosome-level genome for Coleoptera.

## Results

### Sequencing and assembly results

Our goal was to obtain a genome with high contiguity and accuracy, we implemented a long-read sequencing strategy using PacBio long-reads in combination illumina short-reads that use in situ Hi-C library techniques to correct and reorder the scaffolds generated from the PacBio read assembly. From our PacBio library we sequenced a total of 87.5 Gbp with an N50 read length of 31,404 bp (see Table A in [S1 Raw Data Reports](#) for full report). From our in situ Hi-C library (we refer to the in situ Hi-C library or reads as “Hi-C” throughout), we sequenced a total of 228,169,567 paired reads after cleaning. Only 2.53% of our Hi-C reads were unmapped, and we had a total of 80,652,881 Hi-C contacts (intra/ inter-chromosomal interactions, i.e., chimeric read pairs). For a list of the intra-/inter-chromosomal contacts and long/ short range Hi-C contacts, see Table B in [S1 Raw Data Reports](#).

Next to correct read errors from our initial PacBio assembly we used iterations of RACON [38] followed by collapsing duplicate haplotigs not merged in the initial assembly. Our initial PacBio assembly after 3X polishing in RACON [38] consisted of 18,240 contigs and was 2,982,578,979 bp in total length. After removing duplicate haplotigs with *Purge Haplotigs* [39], 9,751 scaffolds and 2,052,097,903 bp remained. Next, we used our Hi-C reads to order our scaffolds into chromosomes and correct misjoins. Our initial Hi-C assembly resulted in 4,111 scaffolds and 2,057,226,403 bp total. The size increase is due to 500 bp insertions of Ns (the 3D-DNA default), between scaffolds merged into super-scaffolds. Running *Pilon* (v. 1.23) in “—fix bases” mode to remove homopolymer repeats and removal of mitochondrial and contaminant scaffolds (virus or bacteria) resulted in 4,093 scaffolds and 2,051,389,195 bp in the final assembly (see Figs 1 and S1 and Tables 1 and 2). The identity of a few other scaffolds not included in the main chromosomes are ambiguous (14 potential viruses and 31 potential bacteria). We retained these but did remove any with bacteria or virus as their best blast score. From the different versions of BUSCO [40] Insecta gene sets (1658 BUSCOs version 2, 1367



**Fig 1. *Pachyrhynchus sulphureomaculatus*, lateral habitus.** (photo by A. Cabras). Hi-C contact map heatmap of *Pachyrhynchus sulphureomaculatus* Schultze, 1922. Eleven chromosome boundaries are indicated by black lines. Heatmap scale lower left, range in counts of mapped Hi-C reads per megabase squared. Rabl-like pattern (grouping of telomeres and centromeres to the nuclear envelope) highlighted along chromosome 1, top row, top of open triangles point to contact between centromere regions, arrows indicate centromere to centromere contact between chromosomes 1 and 2. X-like pattern between adjacent off diagonal regions indicative of contact between distal portions of chromosomes.

<https://doi.org/10.1371/journal.pgen.1009745.g001>

version 4 and 5-beta), the percentage of complete genes varied (90.8% V2 (S2 Fig)), indicating a relatively complete assembly. Compared to other chromosome-level beetle genomes, we found a comparable number of complete BUSCO genes. However, the results vary somewhat

**Table 1. Summary statistics for final assembly.**

Number of scaffolds	4,093
Total size of scaffolds	2,051,389,195
Number of contigs	14,365
Number of contigs in scaffolds	10,283
Number of contigs not in scaffolds	4,082
Mean scaffold size	501,195
Median scaffold size	8,175
N50 scaffold length	215,921,627
scaffold %AT	33.11
scaffold %CG	16.87
scaffold %N	0.05
<b>% bp of assembly in chromosomes</b>	<b>97.52</b>

<https://doi.org/10.1371/journal.pgen.1009745.t001>

depending on which version of BUSCO and which genes were used (S2 Fig). We found a relatively low duplication rate compared to that found in two other beetle (*Photinus* firefly [8] and *Propylea* ladybeetle [9]) genomes that used primarily long-read and Hi-C sequencing in their assembly.

### Repeat content analyses

Weevils have a large distribution of genome sizes, to help investigate what is behind this pattern we analyzed each genome for its repetitive content, as repetitive content often accounts for large portions of a genome. At 2.05 Gbp, the *Pachyrhynchus sulphureomaculatus* genome is roughly 1.8 times as large as the next largest weevil (Curculionioidea) genome published to date, the 1.11 Gbp *Listronotus bonariensis*, the Argentine Stem Weevil [41], and 2.6 times the next largest, the 782 Mbp Red Palm Weevil, *Rhynchophorus ferrugineus* [42] genome. To help explain the size difference, we categorized the repeat content of *P. sulphureomaculatus*. The repeat content analyses from *RepeatMasker* shows that the genome of *P. sulphureomaculatus* consists of more than three quarters (76.36%) repetitive DNA, similar to the repeat percentage of *Listronotus*, which is the closest relative to *Pachyrhynchus*. Compared to other weevil genomes (Fig 2), *P. sulphureomaculatus* has roughly the same percentage of non-repetitive DNA as *Listronotus* and *Sitophilus*. However, the genomes of the two bark beetles of the subfamily Scolytinae (*Dendroctonus* and *Hypothenemus*), are ~1/12 the size of *P. sulphureomaculatus* and consist of only ~17% repetitive content. The *P. sulphureomaculatus* genome consisted of 73.1% interspersed repeats, with SINEs being 0.1%, LINEs 20.8%, LTR elements 2.6%, DNA elements 33% and unclassified repeats 16.6%. A sliding window analysis suggests that repetitive content tends to be found in a higher percentage towards the ends of the chromosomes in *P. sulphureomaculatus*, except in chromosome 5 (Fig 3).

### Genome annotation

As this is the first publically available weevil genome resolved to chromosome scale we wanted to provide an annotation of its genic content as this may prove an informative resource to other researchers. After removing low quality reads from our transcriptome library, a total of 20,551,938 paired reads remained. Our initial 3 transcriptome assemblies, *Trinity* de novo, *Trinity* genome guided assembly and *rnaSPAdes*, resulted in fairly similar assemblies, with each having a high number (~90%) of the BUSCO v.2 Arthropoda genes (see Table A in S1 BUSCO Analyses Results for details).

Table 2. Summary statistics for final assembly by chromosome.

Chromosome	length bp	# of contigs	number of N's (runs of 100)	percent N's in chromo.	N50	N50 reached in # of contigs
Chr_1	263,832,947	1388	138,700	0.05%	287319	280
Chr_2	253,284,860	1222	122,100	0.04%	326447	246
Chr_3	137,890,936	683	68,200	0.04%	315991	127
Chr_4	223,502,247	1131	113,000	0.05%	297265	217
Chr_5	69,931,891	236	23,500	0.03%	452330	50
Chr_6	125,299,487	655	65,400	0.05%	304080	131
Chr_7	132,078,125	624	62,300	0.04%	368684	112
Chr_8	173,282,956	836	83,500	0.04%	335626	165
Chr_9	215,921,627	1221	122,000	0.05%	280522	237
Chr_10	186,927,849	988	98,700	0.05%	290475	197
Chr_11	218,628,933	1074	107,300	0.04%	316887	219

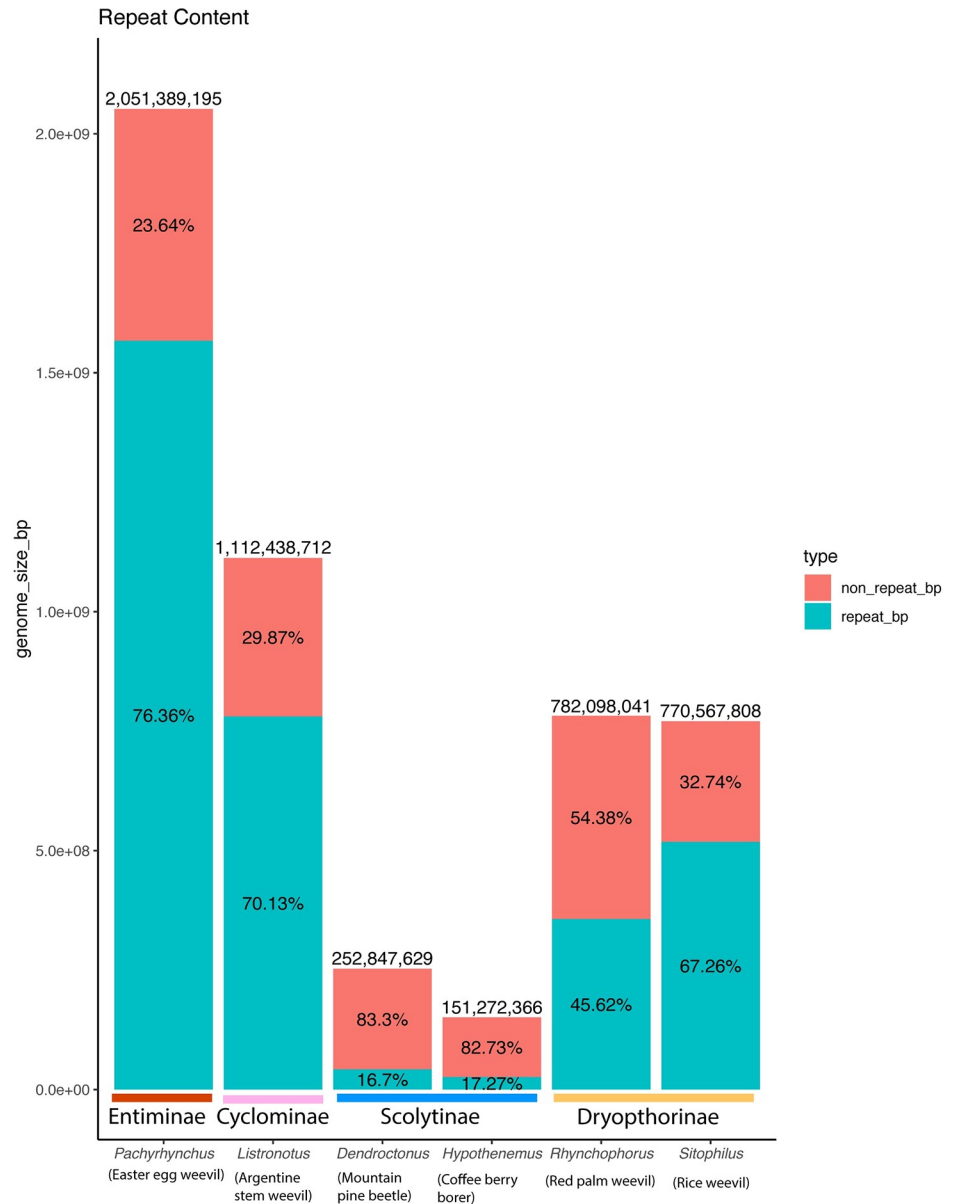
<https://doi.org/10.1371/journal.pgen.1009745.t002>

As the nuclei of cells between different species generally do not interact (except for viruses), and because Hi-C mapping will remove any non-*Pachyrhynchus* DNA from the chromosomes, we only annotated genes found within the 11 chromosomes comprising 2,000,581,858 bp. The *EvidenceModeler* analysis found that the *P. sulphureomaculatus* contained, 30,175 gene transcripts. After running an InterProScan (cross-referencing the results from *EvidenceModeler* with the protein databases) resulted in 18,741 gene models of which 19.01% are single exon genes. Of note are the large intron sizes on average 23,640 bp in length. For the details of results see [Table 3](#).

The gff, faa, gene model scores and tRNA annotations can be found [Table A](#), [Table B](#), [Table C](#), and [Table D](#) in [S1 Anno Results](#). Chromosome gene distribution is relatively even, with only a few regions enriched with genes ([Fig 3](#)). Also of note, the number of genes is larger than those found in the pine beetle *Dendroctonus ponderosae* (GCF\_000355655.1: 14,342 genes) also a weevil. *P. sulphureomaculatus* is more similar in number to those found in other phytophagous beetles who feed on plant foliage, such as the Colorado potato beetle *Leptinotarsa decemlineata* (GCF\_000500325.1: 16,533 genes) [43]. A more thorough examination of close relatives and more phylogenetically distant but ecologically similar species would need to be conducted to fully tease out why there are more gene models predicted in the foliage feeding species.

### Synteny across coleopteran chromosome-level genomes

Here we wanted to describe the syntenic patterns found in Coleoptera as this has not been attempted before to the best of our knowledge. To accomplish this, we mapped a BUSCO single copy gene set across the different taxa available, looking for any emergent patterns. We found that the BUSCO v.2 loci (1658 Insecta gene set), had a low level of translocations between chromosomes ([Fig 4](#)). Results show that within a chromosome, the order of BUSCO genes is not conserved ([Figs 4, 5 and 6](#)), with few long segments of synteny within a chromosome. Synteny is greatest between *P. sulphureomaculatus* and the five other Polyphaga beetles, and least between Adephaga (*Pogonus*) and *P. sulphureomaculatus*, however some of this difference may be in part to the *Pogonus* assembly having less of its contigs being localized into the chromosomes. Interestingly, there is more synteny between *P. sulphureomaculatus* and *Photinus pyralis* (firefly)[8] than between *P. sulphureomaculatus* and *Propylea japonica* (ladybird beetle), the closer relative of *P. sulphureomaculatus*, indicating that the lineage leading to *Propylea* has undergone many more chromosomal translocation events ([Figs 4 and 5](#)). Synteny

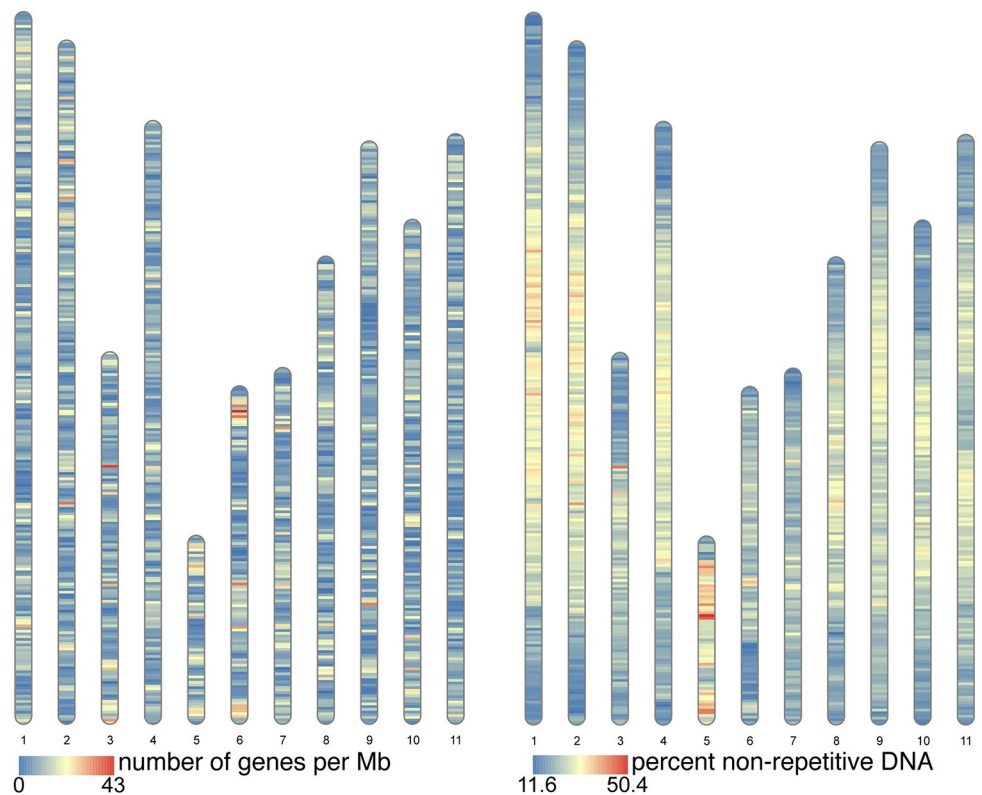


**Fig 2. Histogram of repeat content for weevil genomes examined.** Subfamily classification appears below the histograms. Latin names are in italic font with common names below in parentheses. Genome size largely corresponds to repeat content.

<https://doi.org/10.1371/journal.pgen.1009745.g002>

is greatest between *P. sulphureomaculatus* and the two Tenebrionoidea species (*Tribolium* and *Pyrochroa*), its closest relatives.

Given the divergence time between our taxa, when translocations do occur, their initial positions are lost due to a high level of reorganization producing a pattern of interwoven segments. For example, chromosomes 8 and 9 in *Pachyrhynchus* and the large chromosome 1 in *Tribolium* (Fig 5), have no large syntenic runs of genes or obvious places of translocation. In contrast, chromosome 9 of *Propylea* and chromosome 5 of *Pachyrhynchus* are still largely intact, with the homologous segment of chromosome 5 inserted into roughly the middle of *Propylea*'s chromosome 9. Lastly, we see another 2 fusion events in the *Rhagonycha* soldier



**Fig 3. Heat map of gene density and non-repetitive DNA per 1 Mb sliding window.** The 11 chromosomes are in the same order as in the Hi-C heat map (Fig 1) and fasta file of the genome. Repetitive content higher towards the distal portions of the chromosomes.

<https://doi.org/10.1371/journal.pgen.1009745.g003>

beetle. Here the chromosome number is reduced to 7 and we see 2 clear relatively recent fusion events on chromosome 2 and 4 (Fig 4). Given the relative amount of reshuffling along other parts of this chromosome, the ability to place the insertion indicates that this was a relatively recent event.

### Synteny across the insect tree of life

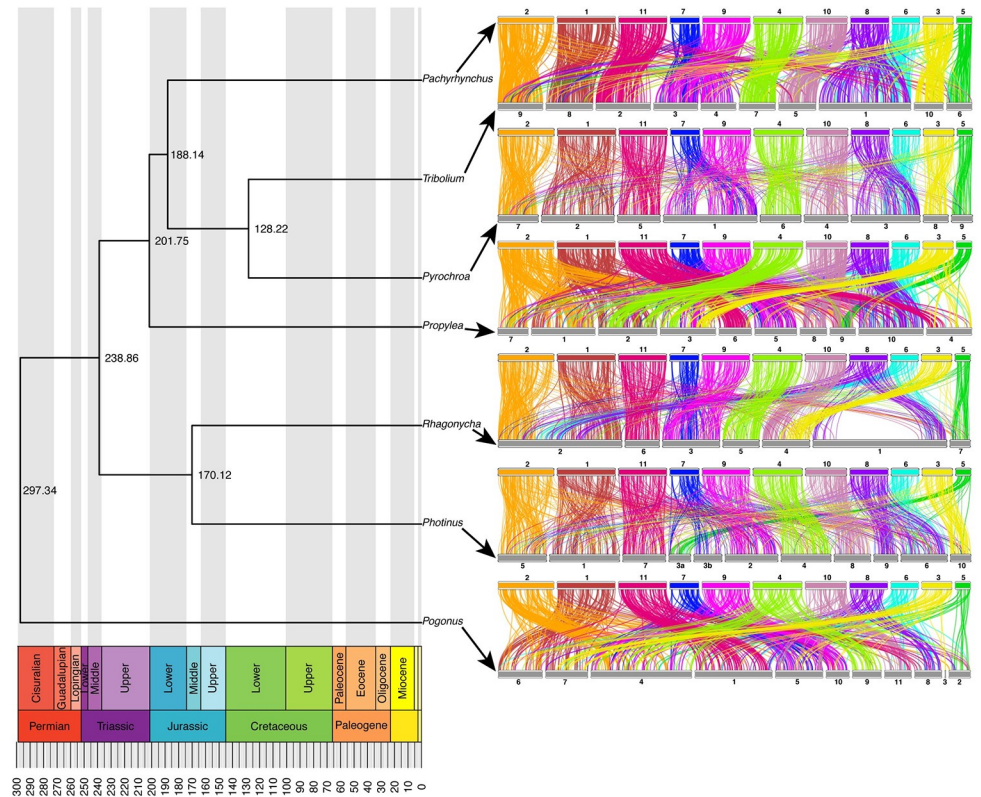
As we wanted to examine if insect orders have different synteny decay rates, we needed to have two pieces of information, a score for how syntenic two species are and their phylogenetic relatedness. For scoring synteny we computed the ENSEMBL Gene Order Conservations (GOC) scores [44] across all pairwise comparisons for our 143 taxa from the positions of their BUSCO version 5 genes. Species were chosen if their genome assemblies were recorded as chromosome level by NCBI or similar (using Hi-C for super scaffolding). The GOC pairwise matrix results can be found in Table A in *S1 Synteny Analyses*. To reconstruct the taxa's phylogenetic relationships, we recovered 1356 BUSCO Genes in a 50% complete matrix, totaling

**Table 3. Results from genome annotation, lengths in bp.**

Genomic class	Total	Total length	Mean length	Longest	Shortest	Mean # per gene	Mean length per gene
Genes	18,741	465,968,736	24,864	642,261	150	NA	NA
Exons	99,291–19% single exon genes	22,930,245	231	11,802	3	5	1,224
Introns	80,550	443,038,490	5,500	477,734	28	4.3	23,640

<https://doi.org/10.1371/journal.pgen.1009745.t003>





**Fig 4. Chromogram and ideograms of 7 beetle genomes which have chromosome level assemblies.** Chromosomes largely remain intact with few translocations relative to reshuffling within a chromosome. Colors correspond to the 11 chromosomes of *Pachyrhynchus sulphureomaculatus*, top row of ideogram plots. Each line represents a BUSCO gene connecting its position on the chromosome of *P. sulphureomaculatus* (top row, respectively) to its position on another species (lower row, respectively).

<https://doi.org/10.1371/journal.pgen.1009745.g004>

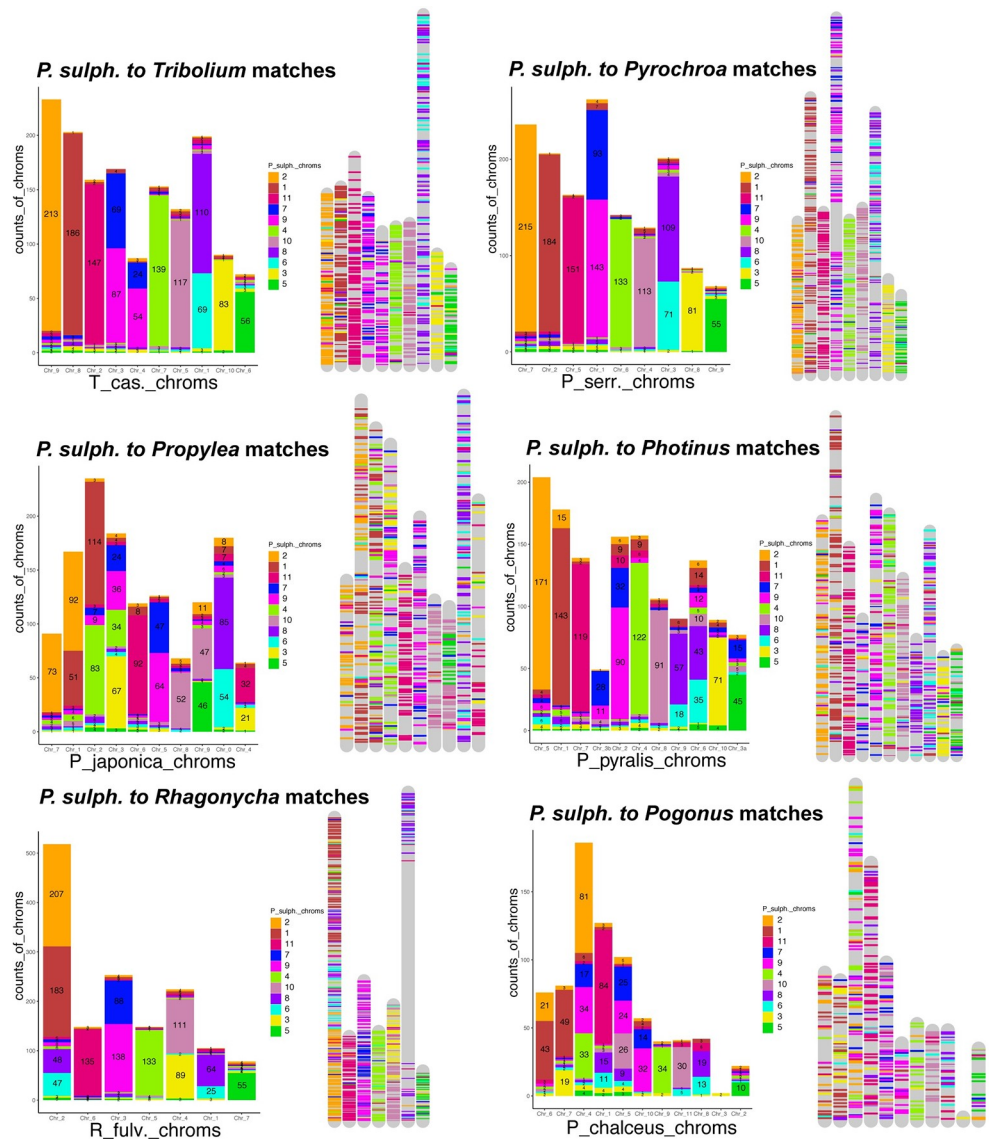
610,189 amino acids in length. The 50% complete matrix indicates the minimum number of taxa allowed in an alignment, loci below that percentage are removed from the analyses. The phylogenetic tree was calculated to get an estimate for the phylogenetic distance among taxa.

The phylogeny recovered many of the same clades as in [45] (Fig 7). While we primarily relied on chromosome scale assemblies that used Hi-C (or similar) to superscaffold into chromosomes some did not, such as the assembly of the carabid beetle *Pogonus*. Despite different assembly methods this assembly does not appear to be an outlier when we look at the synteny decay plot (Figs 7 and 8). We performed the regression analyses (below) with and without this taxon and it did not significantly alter the results, so we left it in all further analyses.

### Regression model results

As we wanted to calculate how synteny decays over phylogenetic distance and if insect orders have different rates, we first needed to avoid the lack of independence in pairwise distances (both along phylogenetic branches and in genomic position of genes) we used a permutational approach to evaluate the significance of the regression models we fit. This approach is consistent with widespread methods in ecology and evolutionary biology that preform regression analyses with distance matrices [46,47], for a full explanation see methods section.

The exponential decay model has the highest total model F-statistic and smallest p-value  $F_{9,3590} = 15,111, p = 2 \times 10^{-4}$  (compared to linear:  $F_{9,3590} = 3,493, p = 3 \times 10^{-4}$ ; power law:

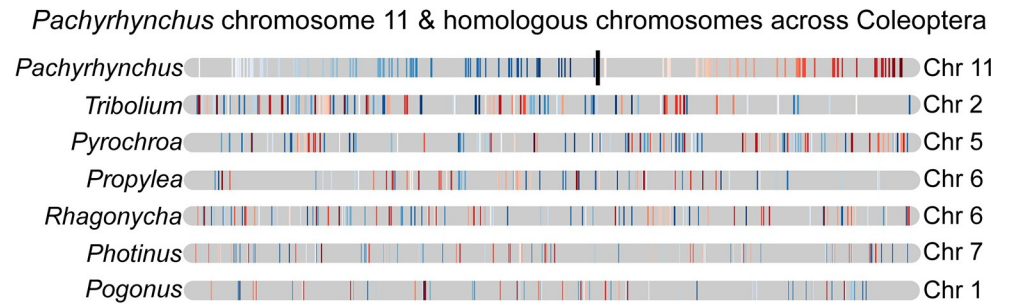


**Fig 5. Stacked bar plots and chromosome mappings of BUSCO genes' placements.** The Y-axis represents the counts of BUSCO genes from *Pachyrhynchus sulphureomaculatus* found on the corresponding chromosomes of another species. Colors correspond to *P. sulphureomaculatus* chromosomes. The numbering scheme (on X-axis) of chromosomes matches the names found in the genome's fasta file. While most chromosomes are primarily composed of one or two chromosomes, relative to *P. sulphureomaculatus*, the placement of the BUSCO genes are interleaved in many instances, indicating that while translocations are rare events reshuffling within a chromosome happens much more frequently.

<https://doi.org/10.1371/journal.pgen.1009745.g005>

$F_{9,3590} = 12,165$ ,  $p = 3 \times 10^{-4}$ ). This supports the exponential model as the best fitting model for the relationship between synteny and phylogenetic distance.

Using this best fitting exponential model, we then asked whether different insect orders show different rates of decay, again using permutational F-statistics. We find that the interaction between phylogenetic distance and order identity is statistically significant:  $F_{4,3590} = 1,344$ ,  $p = 4 \times 10^{-4}$ . We also find that this result is not driven solely by Lepidoptera; the analysis excluding Lepidoptera still finds a significant interaction between phylogenetic



**Fig 6. *Pachyrhynchus sulphureomaculatus* chromosome 11 and matching homologous chromosomes from taxa samples across the Coleoptera.** Top row, approximate position of *Pachyrhynchus* chromosome 11 centromere marked with black line, position derived from Hi-C contact map (see Fig 1). Colored lines correspond to the position of BUSCO genes found in *Pachyrhynchus* chromosome 11 are retained in the other species there is extensive reshuffling in their positions.

<https://doi.org/10.1371/journal.pgen.1009745.g006>

distance and order:  $F_{3,511} = 39$ ,  $p = 4 \times 10^{-4}$ . Results of the exponential decay model can be found in Fig 8.

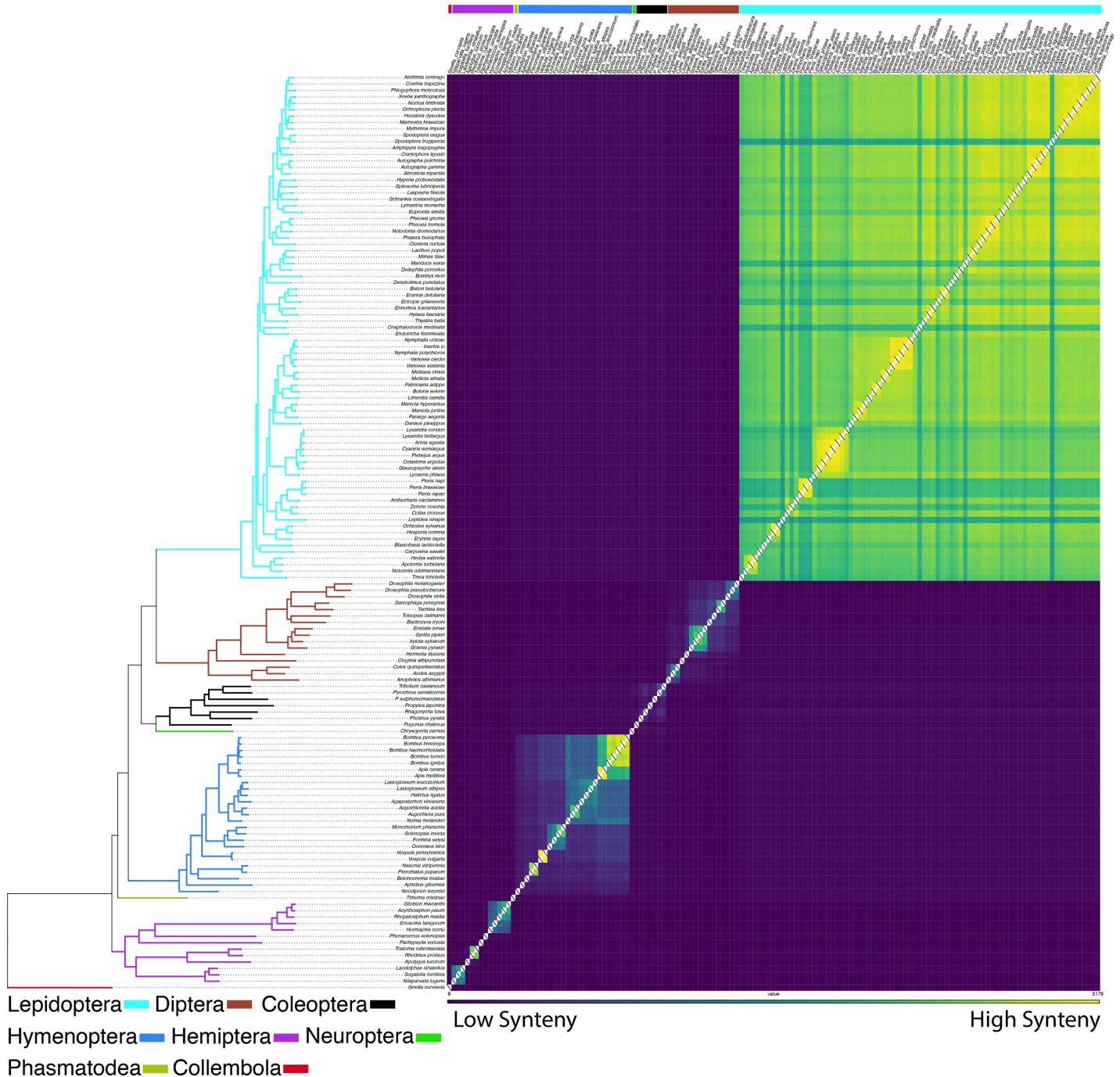
## Discussion

### Hi-C and long read sequencing resolve a large complex insect genome into chromosomes

The combination of long-read DNA and Hi-C sequencing was successful in resolving a large and highly repetitive insect genome. To date, this is the largest insect genome and one of the largest arthropod genomes assembled to chromosome scale, the horseshoe crab's (*Tachypleus tridentatus*) being only slightly larger (2.06 Gb vs 2.05 Gb) [48]. This is remarkable because the assembly of relatively large and highly repetitive insect genomes into highly contiguous ones such as this was previously unattainable [49]. Those efforts were hindered by repetitive contents breaking scaffolds or misjoining them [14,23,49]. The unusually large size of the *Pachyrhynchus* genome is mostly due to the inflated proportion of repetitive content, 76.4% of the genome (Fig 2). Again, highlighting the need for long sequencing reads to span the repetitive content. Here we used a single individual to create both our Hi-C and PacBio libraries. The main advantage over using multiple individuals is little loss of Hi-C reads mapped to the scaffolds; it also eliminates the need for isogenic lines to be established before sequencing. In our previous attempts to assemble a genome for *Pachyrhynchus*, we were greatly hindered by the loss of mappable reads when using multiple individuals. As long read sequencing improves in its capabilities of using a small amount (5–50 ng) of DNA, capitalizing on this combination of Hi-C and long-read sequencing will make it feasible to assemble chromosome scale genomes from single, very small insect specimens [19,50].

### Syntenic patterns in Coleoptera and divergent exponential decay rates of insect orders

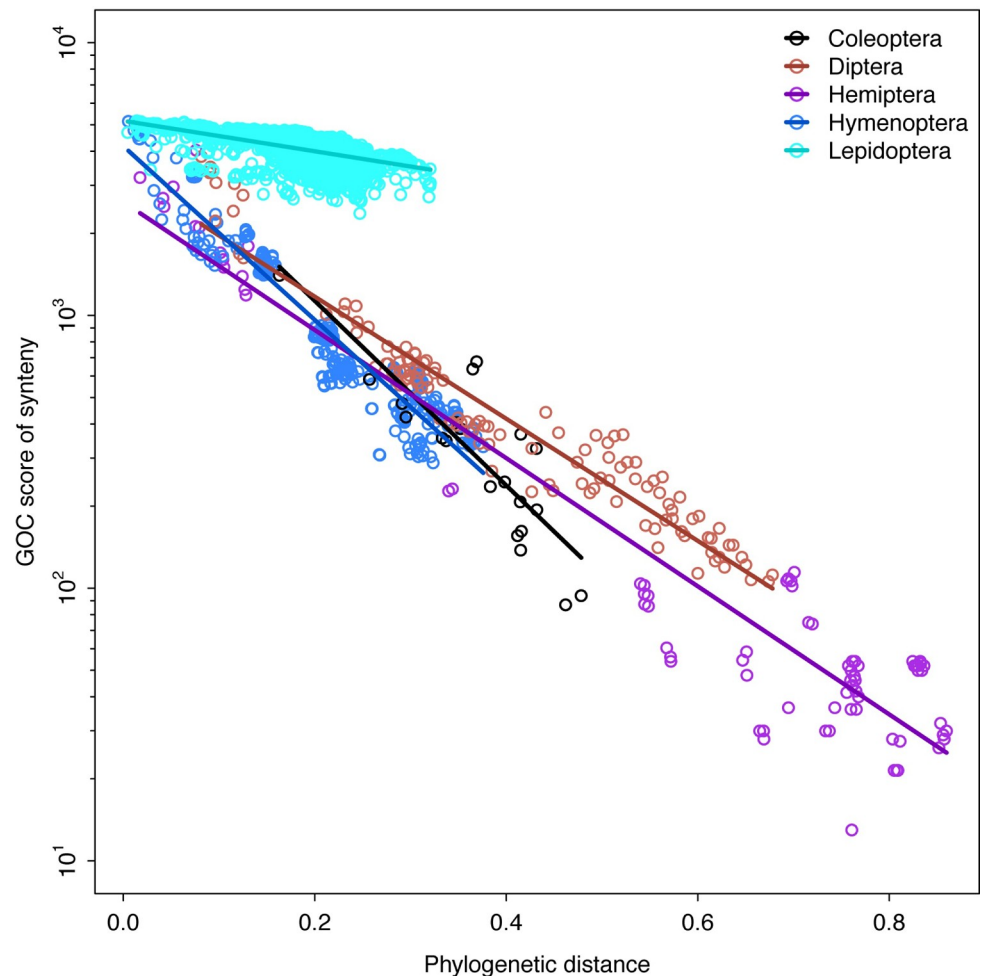
The conserved inter-chromosomal synteny (few chromosome translocations) between the beetle genomes is surprising given the divergence times of the different lineages. For example, we recovered chromosomes that have remained 80–92% intact for more than 200 Ma (Figs 4 and 5). By contrast, the order of the BUSCO genes inside of the chromosomes are highly rearranged, such as chromosomes 8 and 6 in *Pachyrhynchus* and chromosome 1 in *Tribolium* (Figs 4 and 5). This initial finding prompted us to examine whether similar patterns are observed



**Fig 7. Insect, gene order conservation score (GOC) of BUSCO genes.** Left, phylogeny of taxa in analyses, derived from BUSCO genes (610,189 AA sites), reconstructed via RAXML-ng, branches colored by insect order. Right, heat map from pairwise comparisons among insects with chromosome level genomes (only genes localized to chromosomes considers in analyses). Comparisons of gene order which are more syntenic (higher GOC scores) appear in yellow boxes, dark purple indicate less synteny between taxa pairs.

<https://doi.org/10.1371/journal.pgen.1009745.g007>

across other insect orders. A characteristic of Lepidoptera is having a high level of synteny across different families [23,30]. We find that relative to other insect orders sampled that Lepidoptera does have a lower rate of synteny decay. Here we performed the first formal test of this



**Fig 8. Relationship between synteny and phylogenetic distance across different insect orders.** Lines show the best-fitting exponential decay model. Note the log-transformed y-axis. Phylogenetic distance is calculated from a total tree height of 1. Higher values of the GOC score indicate more synteny, lower values less synteny. Synteny decay rate of Lepidoptera differs substantially, however other insect orders also have distinct rates.

<https://doi.org/10.1371/journal.pgen.1009745.g008>

untested (but often mentioned) observation [23,30]. Previous comparisons did not take into account phylogenetic relatedness. Closely related Lepidoptera have similar levels of synteny as other similarly closely related taxa (e.g. *Bombus* and *Apis* Fig 7). But as the phylogenetic distance increase between comparisons Lepidoptera tend to have higher levels of synteny than is found in other orders. In addition to the marked difference in synteny conservation, we also found that each order has a significantly different rate of decay (Fig 8). For example, in *Drosophila*, there is less synteny between members of this genus (~40 Ma) than across all of Lepidoptera, and Coleoptera and Hymenoptera tend to decay at an even faster rate than is seen in Diptera (Fig 8). These results of gene order conservation are consistent with research of *Drosophila* topological associated domains (TADs) that showed synteny break points at approximately every 6th gene between *D. melanogaster*, *D. virilis* and *D. busckii*, which have a similar level of divergence as the *Drosophila* taxa we examined, about 40 Ma of divergence [34]. In addition, the chromosomal rearrangement across *Drosophila* tends to occur at TAD boundaries, not inside the loops [34,51]. In *Anopheles* mosquitoes, the TAD structures seem to be associated with cytological structures as well [52]. In Diptera, despite having many

breakpoints, with relatively few chromosome translocations, their chromosomes largely remain intact [53]. However, in Coleoptera, unlike Mosquitos which show each chromosomal arm being conserved [14] we do not find this same level of conservation in Coleoptera sampled. This may be due to the larger phylogenetic distance between the beetle samples. However, despite this difference we find a somewhat similar syntenic pattern between the two orders, in that the chromosomes remain intact while also being highly shuffled (Figs 4 and 5). This large amount of reshuffling within chromosomes with few inter-chromosomal events contrasts with patterns seen in mammals in which the chromosomes tend to exchange larger blocks of material more readily [54–57].

Currently, chromosome-level genomes are not available for Trichoptera (caddisflies, the sister lineage to Lepidoptera) or early diverging lineages of Lepidoptera. With the addition of these lineages, we could determine whether the observed pattern of synteny conservation is found only in Lepidopteran crown groups or whether it is more widely dispersed across the entire Lepidopteran lineage. Additionally, there are many large orders of insect without a single genome resolved to the chromosome scale or just one, e.g. Psocoptera, Thysanoptera, Neuroptera and several others. A more complete and phylogenetically even sampling of Insecta would help to aid in understanding how changes in genomic architecture may affect other processes such as speciation.

### The genomic architecture of insects and its potential impacts on speciation

Another architectural feature of *Pachyrhynchus*' genome above the chromosome level includes the Rabl-like configuration of chromosomes, where centromeres and telomeres cluster at opposite/different regions of the nucleus. These features are important to note because they may serve an important evolutionary function, such as reducing chromosomal entanglements during interphase as well as regulating chromosomal compartmentalization [58,59]. Both major lineages of Diptera, the Nematocera (e.g. mosquitoes and Psychodidae) and Schizophora (e.g. *Drosophila*), have nucleus with a Rabl-like configuration [14,17,37]. These taxa span much of the phylogenetic distance across the dipteran lineage, and thus this pattern of chromosomal organization may be characteristic of Diptera. We also observe the Rabl-like configuration in *Pachyrhynchus* as well as in the Hi-C map of *Tribolium* (DNAZoo Consortium et al. 2020). Hi-C map observations published for the other taxa do not indicate any other obvious cases of the Rabl-like configuration within the Insecta. However, improving the quality of existing Hi-C maps would provide more evidence for this observation because a lack of valid Hi-C reads can obscure this type of chromosomal architecture.

The Hi-C maps from Tenebrionoidea and Phytophaga beetle lineages display chromosomes in the Rabl-like configuration, those of the other beetle genomes do not display this formation and are from similar tissue types to what we used [8]. It could be that this configuration is only restricted to the aforementioned lineages, more beetle genomes are required. The Rabl-like configuration is not just restricted to beetles and flies; it is also found in the yeast genome [58,60–62] as well as in wheat, barley and *Brassica* [30,63–65], and was originally described from salamander cells [36]. It is unclear how widespread the Rabl-like configuration is in Coleoptera. It is assumed that the Rabl-like configuration is found in all life stages, as appears to be the case in Diptera [14,17,52]. While the Rabl-like configuration is the predominant chromosomal arrangement observed thus far in Diptera and some Coleoptera, its evolutionary significance remains unclear. It has recently been demonstrated how changes in Condensin II impact chromosomes shape and territories which could possibly affect speciation rates by altering between few long chromosomes (with a Rabl-like configuration) and may smaller ones, as seen in Muntjac deer [66]. Our ability to detect genomic architecture's influence on

diversity, if any, is hindered by the sparse in cases, haphazard sampling of insect genomes. Rather than one to one comparison, it is more meaningful to describe patterns for a clade in a broader phylogenetic context. This will allow for the identification general patterns and potentially learning the mechanism as to why some taxa don't fit in.

## Conclusions

In summation, we have reconstructed one of the largest and most repetitive arthropod genomes. With the combination of Hi-C reads and PacBio long-read sequencing data, we were able to resolve a highly contiguous, chromosome-level genome. Across Coleoptera, we find a novel pattern where chromosomes remain relatively intact for hundreds of millions of years with few translocation events, yet their gene order within chromosomes is completely shuffled. Lastly, we find patterns of genomic architecture are clade specific across Insecta, with different insect orders having distinct rates of synteny decay.

## Methods

### Taxon selection and natural history

*Pachyrhynchus*, from the entirely flightless tribe Pachyrhynchini, is found from the Philippines to Papua New Guinea, Australia, Taiwan, Japan, and Indonesia [11,67]. They are known for their bright, iridescent and unique elytral markings, which they use as an aposematic signal to warn predators of their unpalatability [68]. Members of other weevil groups (e.g. *Polycatus*, *Eupyrigops*, *Neopyrigops*, *Alcidodes*) and long-horned beetles (e.g. *Doliops*, *Paradoliops*) mimic *Pachyrhynchus*' aposematic signals to ward off predators. Currently, the Pachyrhynchini has 17 known genera, with the majority found exclusively in the Philippines [11,69,70].

*Pachyrhynchus* Germar, 1824 has the widest geographic range among Pachyrhynchini. There are presently 145 species in the genus, of which 93% of are endemic to the Philippines [71], with the majority of species having a narrow geographic range, limited to a mountain range, island, or Pleistocene Aggregate Island Complex (PAIC) [72–74]. The general diagnostic characters of *Pachyrhynchus* Germar, 1824 include a head lacking a distinct transverse groove or distinct basal border, entire episternal suture, and antennal scape not reaching the hind eye [11]. *P. sulphureomaculatus* Schultze, 1922, is only recorded from Mindanao Island [11,71]. This species was described from material collected in South Cotabato but has recently been recorded (personal observations of A. Cabras) in other areas of Mindanao (e.g. Marilog, Davao City, Arakan, Cotabato, Mt. Kiamo, Bukidnon). This species belongs to the *P. venustus* group, conspicuous for their large size, prothorax with two dorsolateral spots in the middle a large, oblong spot at the lateral margins, and elytra with oval or oblong spots [11].

### Collection and extraction of DNA

Specimens were collected near the edge of the road in a secondary forest (HWY 81, Arakan, Cotabato, Philippines [N7.487059, E125.248795]). One individual was used for both in situ Hi-C and high molecular weight DNA libraries. A second individual was used for transcriptome sequencing. Individuals were collected live, then frozen and stored at -80°C until library preparation.

Beetle tissues were dissected carefully to avoid inclusion of contaminants from guts and impurities from chitinous cuticles. Half of the resulting tissues were used for Phenol Chloroform (PCI) based high molecular weight (HMW) DNA extraction for PacBio sequencing (the other half of the material was used as starting material for Hi-C library preparation, see below).

Tissues were homogenized on ice using a sterile razor blade. ATL buffer (140  $\mu$ l) and Proteinase K (60  $\mu$ l) were then added to the homogenized material and incubated at 65°C for 1 hr. The 200  $\mu$ l of resulting lysate was used as starting material for the PCI extraction following a PacBio recommended protocol [75]. Two additional rounds of PCI clean-up were performed to eliminate impurities such as chitin to meet the DNA requirement for PacBio sequencing. In particular, to achieve OD ratios of 1.8–2.0. DNA concentration was determined with the Qubit dsDNA HS Assay Kit (Invitrogen corp., Carlsbad, CA), and high molecular weight content was confirmed by running a Femto Pulse (Agilent, Santa Clara, USA).

### In situ Hi-C library preparation

Tissues from the same sample were homogenized using a sterile razor blade on ice. An in situ Hi-C library was prepared as described in [13] with a few modifications. Briefly, after the Streptavidin Pull-down step, the biotinylated Hi-C products underwent end repair, ligation and enrichment using the NEBNext UltraII DNA Library Preparation kit (New England Biolabs Inc, Ipswich, MA). Furthermore, titration of the number of PCR cycles was performed as described in [76].

### Transcriptome library preparation

RNA extraction was performed using tissues from a frozen sample. Tissue was extracted from the prothorax and abdomen with the digestive tract removed. The Monarch Total RNA Mini-prep kit (New England Biolabs Inc, Ipswich, MA) was used for extraction. The manufacturer's protocol for total RNA purification from tissue was followed [77]. RNA concentration was determined using the Qubit RNA HS Assay Kit (Invitrogen corp., Carlsbad, CA), and intact RNA content was confirmed by running a Bioanalyzer High Sensitivity RNA Analysis (Agilent, Santa Clara, USA). The resulting RNA was sent to Novogene Inc. for library preparation and sequencing, from which 12.5 Gbp of data were obtained.

### Genome sequencing and assembly

First, we performed an initial quality control of the in situ Hi-C library using the CPU version of *Juicer* v 1.5.7 [78] to determine if enough ligation motifs were present in the sample. To accomplish this, we first cleaned our reads with *fastp* [79] to remove sequencing adapters and low quality reads with default settings except for the more sensitive ‘—detect\_adapter\_for\_pe’ setting on. After passing the quality control of having >30% ligation motifs present, we proceeded to sequence the full library at higher coverage. We only considered ligation motifs as this was a de novo assembly without a closely related reference genome to align to the Hi-C reads. The full Hi-C library was sequenced on a paired-end (2x150 bp) lane on an Illumina HiSeq4000. High molecular weight DNA was sent to the QB3 Genomics facility at the University of California Berkeley for sequencing on a Pacific Biosciences Sequel II platform, sequencing one cell with CLR version 2 chemistry (PacBio, Menlo Park, CA, USA).

We used *PacBio Assembly Tool Suite pb-assembly* v 0.0.8 (which includes the FALCON assembly pipeline) to assemble the primary scaffolds. Next, we polished the primary assembly using 3 rounds of mapping the raw fastq reads using *minimap2* [80] followed by using RACON [38] to help error correct the initial assembly. This was followed by running the *Purge\_Haplotigs* [39] pipeline to eliminate haplotigs (alternative haplotype contigs) in the assembly. Next, using the CPU version of *Juicer* v 1.5.7, we created a site positions file for the restriction enzyme MboI using *Juicer's generate\_site\_positions.py* script, followed by running *Juicer* until it creates the mapping stats file and a “merged\_nodups” file. Then we used the 3D-DNA [14] pipeline with default settings to correct misjoins and place scaffolds into



chromosome groups. After generating a Hi-C heat map, we corrected any assembly errors manually via *Juicebox Assembly Tools* v 1.11.08 [21,78]. After, (Fig 1) we ran 3D-DNA's *run-asm-pipeline-post-review.sh* to produce a final assembly file and fasta. To polish our final assembly further, we aligned our Hi-C reads to our scaffolds using *bwa mem* followed by *SAMclip* and *SAMtools* 'view' [81] with options '-S -b -f 2 -q 1 -F 1536'. After grouping scaffolds into chromosomes, we divided each into a separate fasta (due to memory constraints) and used *Pilon* (v. 1.23) [82] in "—fix bases" mode as to not break our scaffolds and to fix any homopolymer repeat errors. The resulting assembly was used in all subsequent analyses.

### Removal of mitochondrial/contaminant DNA

To identify scaffolds that contained mitochondrial cytochrome oxidase subunit 1 (COI) DNA, we used BLAT v. 35 [83] using a reference sequence from *Pachyrhynchus smaragdinus* (S1 P79\_coI.fasta) to query our scaffolds. Once identified, these scaffolds were removed. We also used *blast* [84] with the nt database and default settings to identify contaminant (non-arthropod or undetermined) sequences and then removed these from the final assembly. These represented only a handful of sequences.

### Repeat content analyses

To address what is making the genome of *Pachyrhynchus sulphureomaculatus* so large relative to other complete weevil genomes (>85% Benchmarking Universal Single-Copy Orthologs BUSCO Insecta genes), we compared the repeat content of *P. sulphureomaculatus* to 5 other weevil genomes from NCBI (see Tables A and B in S1 RepeatMasker Results). We used the de novo *RepeatModeler* v. open-1.0.11 [85] repeat set combined with all rebase recs to first model for repeat content. Next, we used *RepeatMasker* v. 4.1.0 [85] to annotate and soft mask repeat content. For *Listronotus*, we downloaded the results from [41], who used comparable methodologies. We also calculated the percentage of repetitive content (bases soft masked) in a 1 Mb sliding window across the chromosomes in *R* using a custom script.

### Genome annotation

We first cleaned our reads with *fastp* and concatenated the unpaired cleaned reads. We performed 3 different initial reconstructions of the transcriptome: 1) *Trinity* v. 2.11.0 [86,87], de novo assembly using default settings, 2) *Trinity* genome guided assembly, where we first aligned our reads with *tophat* v. 2.1.1 [88], 3) *rnaSPAdes* [89] de novo assembly. Selecting the *rnaSPAdes* assembly, because it had the most single copy BUSCO V2 Arthropoda genes [40], we mapped our reads to this soft masked assembly using *HISAT2* v. 2.2.0 [90], and formatted a bam file using *SAMtools* 'view -b -f 3 -F 256 -q 10'. Next, we used *BRAKER* v. 2.1.5 [91] to create an annotated gff. This process used the bam file from *HISAT2* and results from a BUSCO search as 'seeding' genes to make the resulting gff. In addition, we used the *PASA* pipeline [92,93] which used our *rnaSPAdes* transcripts aligned to the genome assembly with *BLAT* [83] and *gmap* [94]. Lastly, we used *EVidenceModeler* [93] to evaluate our different annotations using the developers' recommended weights for each assembly type. To produce the final gene model gff, we used the potential gene models from *EVidenceModeler* and cross referenced these with several protein databases to validate and provide some curation of our gene models using *InterProScan* v 5.52–86.0 [95]. We used the following protein data bases: PFam, Panther, Prodom, Prosite, Tigrfams, Smart, Pirsf, Prints, Superfamily and CDD. Then searched the *EVidenceModeler* results using *blastn/blastp* against the *blast-nt* database, SwissProt, TrEMBL, orthodb10\_arthropoda, the results of which we only keep if one or more has a hit with e-

val > 1e-6 and then also match a protein domain from *InterProScan*. The best alignments from each database were used to create the final gene annotation result.

### Synteny across coleopteran and Insecta chromosome-level genomes

To examine the gene synteny between other Coleoptera genomes, we downloaded chromosome-level genomes from NCBI or supplied from the journal or authors website (see Table A in [S1 Insecta Trees and Calibrations](#)) [7–10,96]. We also used the unpublished genome assemblies (*Tribolium castaneum* [GCF\_000002335.3], *Bombyx mori* [GCA\_000151625.1], *Clogmia albipunctata* [clogmia.6], *Culex quinquefasciatus* [Cpip3], and *Rhodnius prolixus* [Rhodnius\_prolixus-3.0.3] as well as several others see Table A in [S1 Insecta Trees and Calibrations](#), generated by the DNA Zoo Consortium ([dnazoo.org](#)). The assemblies were based on the whole genome sequencing data from [10,97–100] as well as Hi-C data generated by the DNA Zoo Consortium and assembled using 3D-DNA [14] and Juicebox Assembly Tools [21]. Next, we identified the BUSCO v.2 loci, (1658 Insecta gene set) and extracted their coordinates for the single and fragmented loci. We then compared the coordinates of *Pachyrhynchus sulphureomaculatus* to the other Coleoptera genomes. Following, we calculated the number of loci found in *P. sulphureomaculatus* chromosomes and those in the other Coleoptera and calculated the percent conserved within a chromosome. To visualize the shared synteny, we plotted the different pairs using the R package *RIdeogram* [101].

Next, we investigated whether the observed synteny was distinctive within Coleoptera relative to other orders of insects, such as Lepidoptera, in which high levels of synteny between taxa have been recorded [23,28]. We used all insect genomes (with some exceptions) available from NCBI that were marked as “chromosome” level. (See Table A in [S1 Insecta Trees and Calibration](#) for a complete list.) We tried to sample evenly across insect orders. For example, we excluded the many *Drosophila* genomes as they are all phylogenetically close relatives, and this would cause over-representation (i.e., we want patterns of chromosomal evolution across Diptera, not just *Drosophila*). Instead, we sampled individual species across the phylogenetic breadth of the genus. In addition, we also gathered genomes from the literature. (See Table A in [S1 Insecta Trees and Calibration](#)) Next, we identified all BUSCO version 5-beta loci that were single copy and calculated the gene order conservation (GOC) score (see <https://m.ensembl.org/>) using a custom script (see Scripts A and B in [S1 Scripts](#)). We then only consider BUSCO genes localized in chromosomes. First, we ordered the BUSCO v5-beta genes by scaffold and position and then identified two genes upstream and downstream from a particular gene. Next, to determine if a set of 4 genes are in the same order in our target genome, they receive a score of 1, 0.75, 0.5, 0.25 or 0 based on whether 4, 3, 2, 1 or 0 genes are in the same order, respectively. Missing genes between the two genomes are discarded from comparisons. This process is repeated along the length of the two genomes. We then summed the scores for the four categories 0–100% and added these categories together (e.g., if 8 matched sets were found at 25% and 1 at 100%, the total score would be 5). We computed the total GOC scores for all pairwise comparisons among the 143 taxa. Next, to consider the effect of the phylogenetic relationships, we reconstructed the relationship among our taxa using the BUSCO gene sets' amino acids. We used custom scripts to identify a 50% complete matrix and used *mafft* with 1000 iterations and the “localpair” settings to align the sequences. Next, we used *trimAI* [102] with “automated1” settings to remove ambiguously aligned positions. RAxML-ng [103] with the LG+G8+F site rate substitution model was used to reconstruct the phylogeny for our exemplar taxa across Insecta. We dated the tree using dates (95% highest posterior density interval HPD) from previous studies [5,45,104] using the R package *ape* v.5.4 ‘makeChronos-Calib’ function [105] (see Tree A and Table B in [S1 Insecta Trees and Calibrations](#) for dates).

Calibration points can be found in Table B in [S1 Insecta Trees and Calibrations](#), from [5,45,104]. This calibration was done for visualization purposes only for the Coleoptera clade, as subsequent analyses do not use an ultrametric tree.

## Syntenic decay rate analysis

**Regression model methods.** We would like to know how synteny decays with phylogenetic distance and if different orders show different patterns of decay. To accomplish this, we will evaluate whether the decay in synteny is best fit by a linear, exponential, or power law relationship with phylogenetic distance using least squares regression models. However, because the pairwise distances (both along phylogenetic branches and in genomic position of genes) violate the independence assumptions of ordinary least squares regression models, we will use a permutational approach to evaluate the significance of the regression models we fit. This approach is consistent with widespread methods in ecology and evolutionary biology that perform regression analyses with distance matrices [46,47].

**Permutational algorithm.** We implement this permutational approach using a custom algorithm in the R programming language [106]. We use a custom algorithm because our analytical set-up is slightly different from other approaches, e.g., [46,47,107]. Unlike existing approaches, we are not making all pairwise comparisons, but rather only comparisons within orders (not across orders); we are also interested in the effect of one distance matrix (phylogeny) on another distance matrix (synteny) in combination with a categorical factor (taxonomic order).

We are forced to take a permutational approach because synteny can only be quantified in a pairwise fashion, obviating other methods such as independent contrasts (Harmon & Glor 2010). We use a simple permutation algorithm that does not take into account phylogenetic branch lengths [108] (unlike e.g., Harmon & Glor 2010 [107]) because phylogenetic distance is a key explanatory variable and constraining it in the permutations would lead to nonsensical null distributions. Our permutational algorithm leaves the structure of the phylogeny taxonomic classifications unaltered while permuting levels of divergence in synteny across the tips.

We evaluate which model (linear, exponential, or power law) best fits the data using a permutational estimate of the F statistic (i.e. the ratio of variance explained by the model versus residual variance) and its deviation from the null. We use the F statistic instead of AIC or BIC because these information theoretic and Bayesian model comparison criteria have been shown to perform poorly in distance matrix regression settings [109]. Similarly, to evaluate whether insect orders have different rates of decay in synteny we again use permutational tests based on F statistics (full R code and data found in Doc A in [S1 Synteny Analyses](#)). Code was first validated by comparing our calculations to standard R functions using simulated data. After validating code (F -statistics in agreement), we then analyzed how synteny decays with phylogenetic distance, and whether different orders behave differently. Because Lepidoptera represent the majority of data ( $n = 3,081$  out of 3,600 total data points), we also analyze the relationship between synteny and phylogenetic distance in the subset of data excluding Lepidoptera. We proceeded only with the exponential model as this proved to be the best fitting model.

## Supporting information

**S1 Anno Results. Contains the results of the genome annotation.** The faa, gff, and model scores results files as well as trna sequences of *P. sulphureomaculatus* assembly. A: Table\_A.gff: the gff file. B: Table\_B.faa: the faa file. C: Table\_C.tsv: the gene model scores file. D: Table\_D.

trna: the trna seqs.  
(ZIP)

**S1 BUSCO Analyses Results.** Contains: Table\_A.xlsx, Table\_B.csv. A: Table\_A.xlsx: lists the BUSCO results from the different transcriptome assemblies by method used. B: Table\_B.csv: lists the BUSCO results for the different versions of BUSCO insect e.g. 2, V4 and the associate percentages for single copy complete, complete and duplicated, fragmented and missing genes.  
(ZIP)

**S1 Fig. *P. sulphureomaculatus* scaffold bubble plot of coverage versus GC content.** Scaffolds included are from the unfiltered assembly. Taxonomic annotation provided via blastn alignment to the NCBI nt database.  
(PDF)

**S2 Fig. Stacked bar plot of Insecta BUSCO gene sets by category for chromosome-level beetle genomes.** Y-axis is the percent of BUSCO genes, X-axis labels are the genus names. The abbreviations in the legend are: D = duplicated, F = fragmented, M = missing and S = single.  
(PDF)

**S1 Insecta Trees and Calibrations.** Contains: A: Tree\_A.newick: chronogram used in Fig 4. B: Table\_A.xlsx: list of taxa used in synteny analyses by order, genus and species. With associate NCBI reference or similar. C: Table\_B.xlsx: calibration points used to create “Tree\_A.newick”, tree. D: Tree\_B.tre: all Insecta tree used in Fig 7 and synteny analyses.  
(ZIP)

**S1 P sulph HiC heatmap all chroms & scaffolds. Hi-C contact heat map of full assembly for *P. sulphureomaculatus*.**  
(PDF)

**S1 P79 colfasta. Sequenced used to as seed DNA to extract mtDNA from assembly.**  
(FASTA)

**S1 Raw Data Reports.** Contains: Table\_A.xlsx, Table\_B.docx. A: Table\_A.xlsx: Raw data report for PacBio sequences. B: Table\_B.docx: Summary of Hi-C reads mapped.  
(ZIP)

**S1 RepeatMasker Results.** Contains the RepeatMasker result tables: Table\_A.xlsx, Table\_B.docx. A: Table\_A.xlsx: The NCBI accession numbers used in repeatmasker analyses. B: Table\_B.docx: Table of results from RepeatMasker for *P. sulphureomaculatus*.  
(ZIP)

**S1 Scripts.** Contains: Script\_A.sh, Script\_B.sh. A: Script\_A.sh: script to create scaffold ordered BUSCOs. B: Script\_B.sh: uses results from Script\_B.sh to compute synteny scores.  
(ZIP)

**S1 Synteny Analyses.** Contains: A: Table\_A.txt: the GOC pairwise distances matrix. B: Doc\_A.pdf: instruction on how to preform synteny analyses. C: “synteny analyses/synteny/data/Insecta\_matrix\_matched\_to\_phylo\_mod3.txt”: GOC pairwise distances matrix. D: “synteny analyses/synteny/data/rescaled\_tree\_insecta6.csv”: pairwise phylogenetic distance matrix. E: “synteny analyses/synteny/R/syntPermAOV”: R function to perform correlation of GOC distance and phylogenetic distance by insect order. F: “Read\_me\_Example\_by A. Rominger synteny\_perm.pdf” step by step instruction on how synteny correlations were performed.  
(ZIP)

## Acknowledgments

We would like to thank the Ruth Tawan-tawan, Ceso II of the Philippines' Department of Environment and Natural Resources Region XI for help with the Gratuitous and export permits. We would also like to thank the University of Mindanao for the mobility support, and Milton N. Medina and Chrestine Torrejos of U.M. for help collecting specimens. We would like to thank Zane Colaric of B.C.M., for the help loading the QC library runs. We would also like to thank Sarah Crews of C.A.S. for help with the manuscript text. We would like to thank Chris Jiggins for his thoughtful comments that greatly improved the manuscript.

## Author Contributions

**Conceptualization:** Matthew H. Van Dam, Analyn Anzano Cabras, James B. Henderson, Cynthia Pérez Estrada, Erez Lieberman Aiden, Athena W. Lam.

**Data curation:** Matthew H. Van Dam, James B. Henderson, Andrew J. Rominger, Cynthia Pérez Estrada, Olga Dudchenko, Erez Lieberman Aiden.

**Formal analysis:** Matthew H. Van Dam, James B. Henderson, Andrew J. Rominger, Cynthia Pérez Estrada.

**Funding acquisition:** Matthew H. Van Dam, Analyn Anzano Cabras.

**Investigation:** Matthew H. Van Dam, Analyn Anzano Cabras, James B. Henderson, Cynthia Pérez Estrada.

**Methodology:** Matthew H. Van Dam, James B. Henderson, Andrew J. Rominger, Cynthia Pérez Estrada, Arina D. Omer, Olga Dudchenko, Athena W. Lam.

**Project administration:** Matthew H. Van Dam, Analyn Anzano Cabras.

**Resources:** Matthew H. Van Dam, Analyn Anzano Cabras, Erez Lieberman Aiden, Athena W. Lam.

**Software:** Matthew H. Van Dam, James B. Henderson, Andrew J. Rominger, Olga Dudchenko.

**Supervision:** Matthew H. Van Dam, Erez Lieberman Aiden.

**Validation:** Matthew H. Van Dam, James B. Henderson, Andrew J. Rominger.

**Visualization:** Matthew H. Van Dam, James B. Henderson, Andrew J. Rominger.

**Writing – original draft:** Matthew H. Van Dam, Analyn Anzano Cabras, Athena W. Lam.

**Writing – review & editing:** Matthew H. Van Dam, Analyn Anzano Cabras, James B. Henderson, Andrew J. Rominger, Cynthia Pérez Estrada, Olga Dudchenko, Erez Lieberman Aiden.

## References

1. Hammond P. Species inventory. Global Biodiversity. Status of the Earth's Living Resources. A Report Compiled by the World Conservation Monitoring Centre. Groombridge B, editor. Chapman and Hall, London; 1992.
2. Stork NE, McBroom J, Gely C, Hamilton AJ. New approaches narrow global species estimates for beetles, insects, and terrestrial arthropods. *Proc Natl Acad Sci U S A*. 2015; 112: 7519–7523. <https://doi.org/10.1073/pnas.1502408112> PMID: 26034274
3. Oberprieler RG, Marvaldi AE, Anderson RS. Weevils, weevils, weevils everywhere. *Zootaxa*. Magnolia Press; 2007. pp. 491–520. <https://doi.org/10.11646/zootaxa.1668.1.24>

4. Zimmerman EC. Australian weevils (Coleoptera: Curculionoidea), vol. I: Orthoceri: Anthribidae to Attelabidae: the primitive weevils. undefined. East Melbourne: CSIRO; 1994.
5. McKenna DD, Shin S, Ahrens D, Balke M, Beza-Beza C, Clarke DJ, et al. The evolution and genomic basis of beetle diversity. *Proc Natl Acad Sci U S A*. 2019; 116: 24729–24737. <https://doi.org/10.1073/pnas.1909655116> PMID: 31740605
6. Seppely M, Ioannidis P, Emerson BC, Pitteloud C, Robinson-Rechavi M, Roux J, et al. Genomic signatures accompanying the dietary shift to phytophagy in polyphagan beetles. *Genome Biol*. 2019; 20: 98. <https://doi.org/10.1186/s13059-019-1704-5> PMID: 31101123
7. Van Belleghem SM, Vangestel C, De Wolf K, De Corte Z, Möst M, Rastas P, et al. Evolution at two time frames: Polymorphisms from an ancient singular divergence event fuel contemporary parallel evolution. Schierup MH, editor. *PLOS Genet*. 2018; 14: e1007796. <https://doi.org/10.1371/journal.pgen.1007796> PMID: 30422983
8. Fallon TR, Lower SE, Chang CH, Bessho-Uehara M, Martin GJ, Bewick AJ, et al. Firefly genomes illuminate parallel origins of bioluminescence in beetles. *Elife*. 2018; 7. <https://doi.org/10.7554/eLife.36495> PMID: 30324905
9. Zhang L, Li S, Luo J, Du P, Wu L, Li Y, et al. Chromosome-level genome assembly of the predator *Pro-pylea japonica* to understand its tolerance to insecticides and high temperatures. *Mol Ecol Resour*. 2020; 20: 292–307. <https://doi.org/10.1111/1755-0998.13100> PMID: 31599108
10. Herndon N, Shelton J, Gerischer L, Ioannidis P, Ninova M, Dönitz J, et al. Enhanced genome assembly and a new official gene set for *Tribolium castaneum*. *BMC Genomics*. 2020;21. <https://doi.org/10.1186/s12864-019-6438-y> PMID: 31906847
11. Schultze W. Neunter Beitrag zur Coleopteren-Fauna der Philippinen. *Berliner Entomol Zeitschrift*. 1922; 1922: 36–45. <https://doi.org/10.1002/mmnd.192219220104>
12. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* (80-). 2009; 326. <https://doi.org/10.1126/science.1181369> PMID: 19815776
13. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014; 159: 1665–1680. <https://doi.org/10.1016/j.cell.2014.11.021> PMID: 25497547
14. O D, SS B, AD O, SK N, M H, NC D, et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. 2017; 356: 92–95. Available: <http://science.sciencemag.org/> <https://doi.org/10.1126/science.aal3327> PMID: 28336562
15. Eagen KP, Aiden EL, Kornberg RD. Polycomb-mediated chromatin loops revealed by a subkilobase-resolution chromatin interaction map. *Proc Natl Acad Sci U S A*. 2017; 114: 8764–8769. <https://doi.org/10.1073/pnas.1701291114> PMID: 28765367
16. Ghurye J, Pop M, Koren S, Bickhart D, Chin CS. Scaffolding of long read assemblies using long range contact information. *BMC Genomics*. 2017; 18: 527. <https://doi.org/10.1186/s12864-017-3879-z> PMID: 28701198
17. Matthews BJ, Dudchenko O, Kingan SB, Koren S, Antoshechkin I, Crawford JE, et al. Improved reference genome of *Aedes aegypti* informs arbovirus vector control. *Nature*. 2018; 563: 501–507. <https://doi.org/10.1038/s41586-018-0692-z> PMID: 30429615
18. Song C, Liu Y, Song A, Dong G, Zhao H, Sun W, et al. The *Chrysanthemum nankingense* Genome Provides Insights into the Evolution and Diversification of *Chrysanthemum* Flowers and Medicinal Traits. *Mol Plant*. 2018; 11: 1482–1491. <https://doi.org/10.1016/j.molp.2018.10.003> PMID: 30342096
19. Kingan SB, Urban J, Lambert CC, Baybayan P, Childers AK, Coates B, et al. A high-quality genome assembly from a single, field-collected spotted lanternfly (*Lycorma delicatula*) using the PacBio Sequel II system. *Gigascience*. 2019; 8: 1–10. <https://doi.org/10.1093/gigascience/giz122> PMID: 31609423
20. Sheffer M, Hoppe A, Krehenwinkel H, Uhl G, Kuss A, Jensen L, et al. Chromosome-level reference genome of the European wasp spider *Argiope bruennichi*: a resource for studies on range expansion and evolutionary adaptation. *bioRxiv*. 2020; 2020.05.21.103564. <https://doi.org/10.1101/2020.05.21.103564>
21. Dudchenko O, Shamim MS, Batra SS, Durand NC, Musial NT, Mostofa R, et al. The juicebox assembly tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under \$1000. *bioRxiv*. bioRxiv; 2018. p. 254797. <https://doi.org/10.1101/254797>
22. Sanborn AL, Rao SSP, Huang SC, Durand NC, Huntley MH, Jewett AI, et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl Acad Sci U S A*. 2015; 112: E6456–E6465. <https://doi.org/10.1073/pnas.1518552112> PMID: 26499245

23. Hill J, Rastas P, Hornett EA, Neethiraj R, Clark N, Morehouse N, et al. Unprecedented reorganization of holocentric chromosomes provides insights into the enigma of lepidopteran chromosome evolution. *Sci Adv*. 2019; 5. <https://doi.org/10.1126/sciadv.aau3648> PMID: 31206013
24. Lu S, Yang J, Dai X, Xie F, He J, Dong Z, et al. Chromosomal-level reference genome of Chinese peacock butterfly (*Papilio bianor*) based on third-generation DNA sequencing and Hi-C analysis. *Giga-science*. 2019; 8: 1–10. <https://doi.org/10.1093/gigascience/giz128> PMID: 31682256
25. Liu Q, Guo Y, Zhang Y, Hu W, Li Y, Zhu D, et al. A chromosomal-level genome assembly for the insect vector for Chagas disease, *Triatoma rubrofasciata*. *Gigascience*. 2019; 8. <https://doi.org/10.1093/gigascience/giz089> PMID: 31425588
26. Biello R, Singh A, Godfrey CJ, Fernández FF, Mugford ST, Powell G, et al. A chromosome-level genome assembly of the woolly apple aphid, *Eriosoma lanigerum* Hausmann (Hemiptera: Aphididae). *Mol Ecol Resour*. 2021; 21: 316–326. <https://doi.org/10.1111/1755-0998.13258> PMID: 32985768
27. Ghurye J, Rhie A, Walenz BP, Schmitt A, Selvaraj S, Pop M, et al. Integrating Hi-C links with assembly graphs for chromosome-scale assembly. Ioshikhes I, editor. *PLOS Comput Biol*. 2019; 15: e1007273. <https://doi.org/10.1371/journal.pcbi.1007273> PMID: 31433799
28. Ahola V, Lehtonen R, Somervuo P, Salmela L, Koskinen P, Rastas P, et al. The Glanville fritillary genome retains an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera. *Nat Commun*. 2014; 5: 1–9. <https://doi.org/10.1038/ncomms5737> PMID: 25189940
29. Davey JW, Chouteau M, Barker SL, Maroja L, Baxter SW, Simpson F, et al. Major improvements to the *Heliconius melpomene* genome assembly used to confirm 10 chromosome fusion events in 6 million years of butterfly evolution. *G3 Genes, Genomes, Genet*. 2016; 6: 695–708. <https://doi.org/10.1534/g3.115.023655> PMID: 26772750
30. Wang W, Guan R, Liu X, Zhang H, Song B, Xu Q, et al. Chromosome level comparative analysis of Brassica genomes. *Plant Mol Biol*. 2019; 99: 237–249. <https://doi.org/10.1007/s11103-018-0814-x> PMID: 30632049
31. Lukhtanov VA, Dinca V, Friberg M, Síchová J, Olofsson M, Vila R, et al. Versatility of multivalent orientation, inverted meiosis, and rescued fitness in holocentric chromosomal hybrids. *Proc Natl Acad Sci U S A*. 2018; 115: E9610–E9619. <https://doi.org/10.1073/pnas.1802610115> PMID: 30266792
32. Edelman NB, Frandsen PB, Miyagi M, Clavijo B, Davey J, Dikow RB, et al. Genomic architecture and introgression shape a butterfly radiation. *Science (80-)*. 2019; 366: 594–599. <https://doi.org/10.1126/science.aaw2090> PMID: 31672890
33. Marec F, Sahara K, Traut W. Meiotic pairing of sex chromosome fragments and its relation to atypical transmission of a sex-linked marker in *Ephesia kuehniella* (Insecta: Lepidoptera). *Heredity (Edinb)*. 2001; 87: 659–671. <https://doi.org/10.1046/j.1365-2540.2001.00958.x> PMID: 11903561
34. Renschler G, Richard G, Valsecchi CIK, Toscano S, Arrigoni L, Ramirez F, et al. Hi-C guided assemblies reveal conserved regulatory topologies on X and autosomes despite extensive genome shuffling. *Genes Dev*. 2019; 33: 1591–1612. <https://doi.org/10.1101/gad.328971.119> PMID: 31601616
35. Yadav V, Sun S, Coelho MA, Heitman J. Centromere scission drives chromosome shuffling and reproductive isolation. *Proc Natl Acad Sci*. 2020; 117: 7917–7928. <https://doi.org/10.1073/pnas.1918659117> PMID: 32193338
36. Rabl c. Über Zellteilung. *Morphologisches*. 1885 [cited 20 Dec 2020]. Available: <https://ci.nii.ac.jp/naid/10005431100/>
37. Csink AK, Henikoff S. Large-scale chromosomal movements during interphase progression in *Drosophila*. *J Cell Biol*. 1998; 143: 13–22. <https://doi.org/10.1083/jcb.143.1.13> PMID: 9763417
38. Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res*. 2017; 27: 737–746. <https://doi.org/10.1101/gr.214270.116> PMID: 28100585
39. Roach MJ, Schmidt SA, Borneman AR. Purge Haplotigs: Allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics*. 2018; 19: 460. <https://doi.org/10.1186/s12859-018-2485-7> PMID: 30497373
40. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015; 31: 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351> PMID: 26059717
41. Harrop TWR, Le Lec MF, Jauregui R, Taylor SE, Inwood SN, van Stijn T, et al. Genetic diversity in invasive populations of argentine stem weevil associated with adaptation to biocontrol. *Insects*. 2020; 11: 1–14. <https://doi.org/10.3390/insects11070441> PMID: 32674400
42. Hazzouri KM, Sudalaimuthasari N, Kundu B, Nelson D, Al-Deeb MA, Le Mansour A, et al. The genome of pest *Rhynchophorus ferrugineus* reveals gene families important at the plant-beetle interface. *Commun Biol*. 2020; 3. <https://doi.org/10.1038/s42003-020-1060-8> PMID: 32581279

43. Schoville SD, Chen YH, Andersson MN, Benoit JB, Bhandari A, Bowsher JH, et al. A model species for agricultural pest genomics: the genome of the Colorado potato beetle, *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae). *Sci Reports* 2018 81. 2018; 8: 1–18. <https://doi.org/10.1038/s41598-018-20154-1> PMID: 29386578
44. Aken BL, Achuthan P, Akanni W, Amode MR, Bernsdorff F, Bhai J, et al. Ensembl 2017. *Nucleic Acids Res.* 2017; 45: D635–D642. <https://doi.org/10.1093/nar/gkw1104> PMID: 27899575
45. Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, et al. Phylogenomics resolves the timing and pattern of insect evolution. *Science* (80-). 2014; 346: 763–767. <https://doi.org/10.1126/science.1257570> PMID: 25378627
46. Lichstein JW. Multiple regression on distance matrices: a multivariate spatial analysis tool. *Plant Ecol* 2006 1882. 2006; 188: 117–131. <https://doi.org/10.1007/S11258-006-9126-3>
47. McArtor DB, Lubke GH, Bergeman CS. Extending multivariate distance matrix regression with an effect size measure and the asymptotic null distribution of the test statistic. *Psychometrika*. 2017; 82: 1052–1077. <https://doi.org/10.1007/s11336-016-9527-8> PMID: 27738957
48. Zhou Y, Liang Y, Yan Q, Zhang L, Chen D, Ruan L, et al. The draft genome of horseshoe crab *Tachyplesus tridentatus* reveals its evolutionary scenario and well-developed innate immunity. *BMC Genomics*. 2020; 21: 137. <https://doi.org/10.1186/s12864-020-6488-1> PMID: 32041526
49. Li F, Zhao X, Li M, He K, Huang C, Zhou Y, et al. Insect genomes: progress and challenges. *Insect Molecular Biology*. Blackwell Publishing Ltd; 2019. pp. 739–758. <https://doi.org/10.1111/imb.12599> PMID: 31120160
50. Schneider C, Woehle C, Greve C, D'haese CA, Wolf M, Janke A, et al. Biodiversity genomics of small metazoans: high quality de novo genomes from single specimens of field-collected and ethanol-preserved springtails Running title: High quality genomes from single springtails. *bioRxiv*. 2020; 2020.08.10.244541. <https://doi.org/10.1101/2020.08.10.244541>
51. Liao Y, Zhang X, Chakraborty M, Emerson JJ. Topologically associating domains and their role in the evolution of genome structure and function in *Drosophila*. *bioRxiv*. 2020; 2020.05.13.094516. <https://doi.org/10.1101/2020.05.13.094516>
52. Lukyanchikova V, Nuriddinov M, Belokopytova P, Liang J, Reijnders M, Ruzzante L, et al. Anopheles mosquitoes revealed new principles of 3D genome organization in insects. *bioRxiv*. 2020; 2020.05.26.114017. <https://doi.org/10.1101/2020.05.26.114017>
53. Bracewell R, Chatla K, Nalley MJ, Bachtrog D. Dynamic turnover of centromeres drives karyotype evolution in *drosophila*. *Elife*. 2019; 8. <https://doi.org/10.7554/eLife.49002> PMID: 31524597
54. Chowdhary BP, Raudsepp T, Fröncke L, Scherthan H. Emerging patterns of comparative genome organization in some mammalian species as revealed by Zoo-FISH. *Genome Research*. Cold Spring Harbor Laboratory Press; 1998. pp. 577–589. <https://doi.org/10.1101/gr.8.6.577> PMID: 9647633
55. Kemkemer C, Kohn M, Cooper DN, Froenicke L, Högel J, Hameister H, et al. Gene synteny comparisons between different vertebrates provide new insights into breakage and fusion events during mammalian karyotype evolution. *BMC Evol Biol*. 2009; 9: 84. <https://doi.org/10.1186/1471-2148-9-84> PMID: 19393055
56. Brian Simison W, Parham JF, Papenfuss TJ, Lam AW, Henderson JB. An Annotated Chromosome-Level Reference Genome of the Red-Eared Slider Turtle (*Trachemys scripta elegans*). *Eyre-Walker A, editor*. 2020; 12: 456–462. <https://doi.org/10.1093/gbe/evaa063> PMID: 32227195
57. Deakin JE. Chromosome evolution in marsupials. *Genes*. MDPI AG; 2018. <https://doi.org/10.3390/genes9020072> PMID: 29415454
58. Mizuguchi T, Barrowman J, Grewal SIS. Chromosome domain architecture and dynamic organization of the fission yeast genome. *FEBS Letters*. Elsevier B.V.; 2015. pp. 2975–2986. <https://doi.org/10.1016/j.febslet.2015.06.008> PMID: 26096785
59. Pouokam M, Cruz B, Burgess S, Segal MR, Vazquez M, Arsuaga J. The Rab1 configuration limits topological entanglement of chromosomes in budding yeast. *Sci Rep*. 2019; 9. <https://doi.org/10.1038/s41598-019-42967-4> PMID: 31043625
60. Jin QW, Trelles-Sticken E, Scherthan H, Loidl J. Yeast nuclei display prominent centromere clustering that is reduced in nondividing cells and in meiotic prophase. *J Cell Biol*. 1998; 141: 21–29. <https://doi.org/10.1083/jcb.141.1.21> PMID: 9531545
61. Goto B, Okazaki K, science ON-J of cell, 2001 undefined. Cytoplasmic microtubular system implicated in de novo formation of a Rab1-like orientation of chromosomes in fission yeast. *jcs.biologists.org*. Available: <https://jcs.biologists.org/content/114/13/2427.short> PMID: 11559751
62. Kim S, Liachko I, Brickner DG, Cook K, Noble WS, Brickner JH, et al. The dynamic three-dimensional organization of the diploid yeast genome. *Elife*. 2017; 6. <https://doi.org/10.7554/eLife.23623> PMID: 28537556



63. Mascher M, Gundlach H, Himmelbach A, Beier S, Twardziok SO, Wicker T, et al. A chromosome conformation capture ordered sequence of the barley genome. *Nature*. 2017; 544: 427–433. <https://doi.org/10.1038/nature22043> PMID: 28447635
64. Concia L, Veluchamy A, Ramirez-Prado JS, Martin-Ramirez A, Huang Y, Perez M, et al. Wheat chromatin architecture is organized in genome territories and transcription factories. *Genome Biol*. 2020; 21: 104. <https://doi.org/10.1186/s13059-020-01998-1> PMID: 32349780
65. Santos AP, Shaw P. Interphase chromosomes and the Rab1 configuration: does genome size matter? *J Microsc*. 2004; 214: 201–206. <https://doi.org/10.1111/j.0022-2720.2004.01324.x> PMID: 15102067
66. Bauer CR, Hartl TA, Bosco G. Condensin II Promotes the Formation of Chromosome Territories by Inducing Axial Compaction of Polyploid Interphase Chromosomes. *PLoS Genet*. 2012; 8: e1002873. <https://doi.org/10.1371/journal.pgen.1002873> PMID: 22956908
67. Alonso-Zarazaga M. A world catalogue of families and genera of curculionoidea, insecta, coleptera, excepting scolytidae and platypodidae. 1999 [cited 20 Dec 2020]. Available: <http://www.sidalc.net/cgi-bin/wxis.exe/?IsisScript=COLPOS.xis&method=post&formato=2&cantidad=1&expresion=mfn=001035>
68. Tseng H-Y, Lin C-P, Hsu J-Y, Pike DA, Huang W-S. The Functional Significance of Aposematic Signals: Geographic Variation in the Responses of Widespread Lizard Predators to Colourful Invertebrate Prey. Osorio D, editor. *PLoS One*. 2014; 9: e91777. <https://doi.org/10.1371/journal.pone.0091777> PMID: 24614681
69. Yap S, Gapud V. Taxonomic review of the Genus *Metapocyrtus* Heller (Coleoptera: Curculionidae: Entiminae). *Philipp Entomol*. 2007 [cited 20 Dec 2020]. Available: <https://www.researchgate.net/publication/266260665>
70. Shi FM, Bian X, Chang YL. A new genus and two new species of the tribe Meconematini (Orthoptera: Tettigoniidae) from China. *Zootaxa*. 2013; 3681: 163–168. <https://doi.org/10.11646/zootaxa.3681.2.5> PMID: 25232596
71. Rukmane A. An annotated checklist of genus *Pachyrhynchus* (Coleoptera: Curculionidae: Pachyrhynchini). *Acta Biol. Univ. Daugavp*; 2018. Available: <http://sciences.lv/wp-content/uploads/2018/11/Rukmane.pdf>
72. Inger. Systematics and zoogeography of Philippine Amphibia. *Fieldiana*. 1954; 33: 182–531. Available: <https://ci.nii.ac.jp/naid/10018878211/>
73. Heaney L. Zoogeographic evidence for middle and late Pleistocene land bridges to the Philippine Islands. *Mod Quatern Res SE Asia*. 1985; 9: 127–144.
74. Brown RM, Siler CD. Spotted stream frog diversification at the Australasian faunal zone interface, mainland versus island comparisons, and a test of the Philippine ‘dual-umbilicus’ hypothesis. Ebach M, editor. *J Biogeogr*. 2014; 41: 182–195. <https://doi.org/10.1111/jbi.12192>
75. Pacbio. Extracting DNA Using Phenol-Chloroform. 2012 [cited 20 Dec 2020]. Available: <https://www.pacb.com/wp-content/uploads/2015/09/SharedProtocol-Extracting-DNA-using-Phenol-Chloroform.pdf>
76. Belton JM, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. Hi-C: A comprehensive technique to capture the conformation of genomes. *Methods*. 2012; 58: 268–276. <https://doi.org/10.1016/j.ymeth.2012.05.001> PMID: 22652625
77. NEB. Total RNA Purification from Tissues and Leukocytes using the Monarch Total RNA Miniprep Kit (NEB #T2010) | NEB. [cited 20 Dec 2020]. Available: <https://www.neb.com/protocols/2017/11/08/total-rna-purification-from-tissues-and-leukocytes-using-the-monarch-total-rna-miniprep-kit-neb-t2010>
78. Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, et al. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst*. 2016; 3: 95–98. <https://doi.org/10.1016/j.cels.2016.07.002> PMID: 27467249
79. Chen S, Zhou Y, Chen Y, Gu J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. Oxford University Press; 2018. pp. i884–i890. <https://doi.org/10.1093/bioinformatics/bty560> PMID: 30423086
80. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Biol I*, editor. *Bioinformatics*. 2018; 34: 3094–3100. <https://doi.org/10.1093/bioinformatics/bty191> PMID: 29750242
81. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25: 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352> PMID: 19505943
82. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. Wang J, editor. *PLoS One*. 2014; 9: e112963. <https://doi.org/10.1371/journal.pone.0112963> PMID: 25409509

83. Kent WJ. BLAT—The BLAST-Like Alignment Tool. *Genome Res.* 2002; 12: 656–664. <https://doi.org/10.1101/gr.229202> PMID: 11932250
84. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: Architecture and applications. *BMC Bioinformatics.* 2009; 10: 421. <https://doi.org/10.1186/1471-2105-10-421> PMID: 20003500
85. Smit A, Hubley R, Green P. RepeatMasker Open-4.0. *bioRxiv.* <http://www.repeatmasker.org>; 2015. <https://doi.org/10.1101/856591>
86. Haas BJ, Papanicolaou A, Yassour M GM, Philip D BJ, Couger MB ED, Li B, et al. Reference Generation and Analysis with Trinity. *nature.com.* 2014. Available: <https://www.nature.com/nprot/journal/v8/n8/full/nprot.2013.084.html>
87. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 2011; 29: 644–652. <https://doi.org/10.1038/nbt.1883> PMID: 21572440
88. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 2013; 14: R36. <https://doi.org/10.1186/gb-2013-14-4-r36> PMID: 23618408
89. Bushmanova E, Antipov D, Lapidus A, Pribelski AD. RnaSPAdes: A de novo transcriptome assembler and its application to RNA-Seq data. *Gigascience.* 2019; 8: 1–13. <https://doi.org/10.1093/gigascience/giz100> PMID: 31494669
90. Kim G, . . . RA-I, 2019 undefined. Foundational studies of Caribbean crustose coralline algae. DEPT, 2001 EVANS RD, CARY, NC . . .
91. Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. Whole-genome annotation with BRAKER. *Methods in Molecular Biology.* Humana Press Inc.; 2019. pp. 65–95. [https://doi.org/10.1007/978-1-4939-9173-0\\_5](https://doi.org/10.1007/978-1-4939-9173-0_5) PMID: 31020555
92. Campbell MA, Haas BJ, Hamilton JP, Mount SM, Robin CR. Comprehensive analysis of alternative splicing in rice and comparative analyses with Arabidopsis. *BMC Genomics.* 2006; 7: 327. <https://doi.org/10.1186/1471-2164-7-327> PMID: 17194304
93. Haas BJ. Analysis of alternative splicing in plants with bioinformatics tools. *Current Topics in Microbiology and Immunology.* Springer, Berlin, Heidelberg; 2008. pp. 17–37. [https://doi.org/10.1007/978-3-540-76776-3\\_2](https://doi.org/10.1007/978-3-540-76776-3_2) PMID: 18630745
94. Wu TD, Watanabe CK. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics.* 2005; 21: 1859–1875. <https://doi.org/10.1093/bioinformatics/bti310> PMID: 15728110
95. P J, D B, HY C, M F, W L, C M, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics.* 2014; 30: 1236–1240. <https://doi.org/10.1093/bioinformatics/btu031> PMID: 24451626
96. Generalovic TN, McCarthy SA, Warren IA, Wood JMD, Torrance J, Sims Y, et al. A high-quality, chromosome-level genome assembly of the Black Soldier Fly (*Hermetia illucens* L.). *G3 Genes|Genomes|Genetics.* 2021; 11. <https://doi.org/10.1093/G3JOURNAL/JKAB085> PMID: 33734373
97. Xia Q, Wang J, Zhou Z, Li R, Fan W, Cheng D, et al. The genome of a lepidopteran model insect, the silkworm *Bombyx mori*. *Insect Biochem Mol Biol.* 2008; 38: 1036–1045. <https://doi.org/10.1016/j.ibmb.2008.11.004> PMID: 19121390
98. Richards S, Gibbs RA, Weinstock GM, Brown S, Denell R, Beeman RW, et al. The genome of the model beetle and pest *Tribolium castaneum*. *Nature.* 2008; 452: 949–955. <https://doi.org/10.1038/nature06784> PMID: 18362917
99. Arensburger P, Megy K, Waterhouse RM, Abrudan J, Amedeo P, Antelo B, et al. Sequencing of *Culex quinquefasciatus* establishes a platform for mosquito comparative genomics. *Science (80-).* 2010; 330: 86–88. <https://doi.org/10.1126/science.1191864> PMID: 20929810
100. Mesquita RD, Vionette-Amaral RJ, Lowenberger C, Rivera-Pomar R, Monteiro FA, Minx P, et al. Genome of *Rhodnius prolixus*, an insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite infection. *Proc Natl Acad Sci U S A.* 2015; 112: 14936–14941. <https://doi.org/10.1073/pnas.1506226112> PMID: 26627243
101. Hao Z, Lv D, Ge Y, Shi J, Weijers D, Yu G, et al. Rideogram: Drawing SVG graphics to visualize and map genome-wide data on the idiograms. *PeerJ Comput Sci.* 2020; 6: 1–11. <https://doi.org/10.7717/peerj-cs.251> PMID: 33816903
102. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* 2009; 25: 1972–1973. <https://doi.org/10.1093/bioinformatics/btp348> PMID: 19505945

103. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014; 30: 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033> PMID: 24451623
104. Obbard DJ, MacLennan J, Kim KW, Rambaut A, O'Grady PM, Jiggins FM. Estimating divergence dates and substitution rates in the drosophila phylogeny. *Mol Biol Evol*. 2012; 29: 3459–3473. <https://doi.org/10.1093/molbev/mss150> PMID: 22683811
105. Paradis E, Schliep K. Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*. 2019; 35: 526–528. <https://doi.org/10.1093/bioinformatics/bty633> PMID: 30016406
106. R Core Team. R: A language and environment for statistical computing. <http://www.R-project.org>. Vienna, Austria.: R Foundation for Statistical Computing; 2020. Available: <https://ci.nii.ac.jp/naid/20001689445>
107. Harmon LJ, Glor RE. Poor statistical performance of the mantel test in phylogenetic comparative analyses. *Evolution (N Y)*. 2010; 64: 2173–2178. <https://doi.org/10.1111/j.1558-5646.2010.00973.x> PMID: 20163450
108. Lapointe Theodore Garland ois-J, Theodore Garland umonbealca, Lapointe F, Garland T. A Generalized Permutation Model for the Analysis of Cross-Species Data. *J Classif*. 18: 109–127. <https://doi.org/10.1007/s00357-001-0007-0>
109. Franckowiak RP, Panasci M, Jarvis KJ, Acuña-Rodríguez IS, Landguth EL, Fortin M-J, et al. Model selection with multiple regression on distance matrices leads to incorrect inferences. *PLoS One*. 2017; 12: e0175194. <https://doi.org/10.1371/journal.pone.0175194> PMID: 28406923