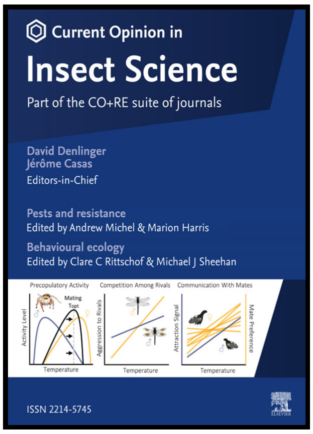

Exploring new genomic territories with emerging model insectsGenome-enabled model insect systems

Romain Feron, Robert M. Waterhouse

PII: S2214-5745(22)00037-2

DOI: https://doi.org/10.1016/j.cois.2022.100902

Reference: COIS100902

To appear in: *Current Opinion in Insect Science*

Received date: 28 January 2022
Revised date: 24 February 2022
Accepted date: 4 March 2022

Please cite this article as: Romain Feron and Robert M. Waterhouse, Exploring new genomic territories with emerging model insectsGenome-enabled model insect systems, *Current Opinion in Insect Science,* (2022) doi:https://doi.org/10.1016/j.cois.2022.100902

Special issue: "Genomics of Emerging Model Insects"

Exploring new genomic territories with emerging model insects

Romain Feron[1], Robert M. Waterhouse[1,*]

[1] Department of Ecology and Evolution, University of Lausanne, and the Swiss Institute of Bioinformatics,1015, Lausanne, Switzerland

[*] Correspondence: robert.waterhouse@unil.ch

Abstract

Improvements in reference genome generation for insects and across the tree of life are extending the concept and utility of model organisms beyond traditional laboratory-tractable supermodels. Species or groups of species with comprehensive genome resources can be developed into model systems for studying a large variety of biological phenomena. Advances in sequencing and assembly technologies are supporting these emerging genome-enabled model systems by producing resources that are increasingly accurate and complete. Nevertheless, quality controls including assessing gene content completeness are required to ensure that these data can be included in expanding catalogues of high-quality references that will greatly advance understanding of insect biology and evolution.

Keywords: insect genomics; reference genomes; genome assembly quality; BUSCO

Short Title: Genome-enabled model insect systems

Introduction

Model organisms can be described as non-human species that are studied to advance the understanding of biological phenomena, with traditional model species being easily bred in the laboratory and amenable to experimental manipulation [1]. The common ancestry of living organisms means that insights from such models also inform knowledge of molecular and genetic mechanisms underlying common biological functions across the tree of life. Representing insects is the renowned model, the fruit fly *Drosophila melanogaster*, with ground-breaking work on fields from genetics and heredity to behaviour, physiology, development, immunity, and countless others [2]. A major contributing factor to the success of *Drosophila* as a versatile model over the last two decades was the establishment of a reference genome assembly and its functional genomic element annotations [3]. Developing new models with reference genomes and experimental tools analogous to those available for *Drosophila* can be challenging, but is important for diversifying the systems we use to learn about organismal biology [4,5]. Currently, substantial advances in sequencing technologies mean that it can be more readily feasible to generate a high-quality genome for a new species than it is to rear in the laboratory. This genomics revolution is opening up a whole new set of possibilities considering a shift from the traditional model organism to the concept of species or groups of species that offer the ability to develop new model systems for studying a large variety of biological phenomena at many different levels [6,7].

Conserved orthologues help gauge gene content completeness of accumulating genome resources

Recent surveys of the current status of available genome resources for insects focus on taxonomic representation, assembly quality metrics, gene content completeness, and sequencing technology usage [8–10]. These highlight the continued rapid accumulation since previous surveys e.g. [11,12], and show current biases in species sampling with several insect orders still lacking publicly available resources. Notably, long-read data, e.g. from approaches developed by Pacific Biosciences (PacBio) or Oxford Nanopore Technologies (ONT), are helping to improve assembly contiguity and produce more complete and accurate representations of new and upgraded insect genomes. For these resources to support the development of emerging model systems, they need to be of the highest possible quality, not only in terms of assembly statistics but also with respect to gene content representation.

The need to assess quality in terms of expected gene content prompted the proposal of Benchmarking Universal Single-Copy Orthologues (BUSCOs) [13]. BUSCO relies on the expectation that single-copy orthologues present in most species within a taxonomic lineage should be identifiable in any new genome from a species in the same clade. The BUSCO lineage datasets are built by identifying near-universal single-copy orthologues from the OrthoDB orthology resource [13–15]. Using these to evaluate assemblies starts with BUSCO sequence searches to guide gene predictions, then orthology classifications identify complete, duplicated, or fragmented BUSCOs. The numbers of identifiable BUSCOs provide an indication of gene content completeness based on expected subsets of evolutionarily conserved genes for a given lineage. High completeness scores thereby imply that a genome assembly confidently represents the complete gene repertoire.

Development of BUSCO assessments resulted in an initial implementation [16] with three lineage datasets relevant for insects (Eukaryota, Metazoa, and Arthropoda) based on orthology data from OrthoDB v8 [14]. Subsequent updates in BUSCO v3 [17] provided four more lineage datasets within Arthropoda (Insecta n=1’658 BUSCOs, Endopterygota n=2’442, Diptera n=2’799, Hymenoptera n=4’415), using orthologues from OrthoDB v9 [15]. The latest BUSCO releases [18] now provide additional insect-relevant lineage datasets for Hemiptera (n=2’510) and Lepidoptera (n=5’286), from OrthoDB v10 [19]. The assessments provide measures of data quality, and protocols for applications to insect genomic data [20] and for the wider uses [21] help users to identify the best available genomic resources. BUSCO completeness is also recognised as an important quality check of resources for new model systems and for cataloguing eukaryotic genomic biodiversity, e.g. the Earth BioGenome Project (EBP) standards recommendations for genome generation include achieving recovery of more than 90% single-copy conserved genes [22].

Using results from the Arthropoda Assembly Assessment Catalogue (A^3Cat) [10,23] to survey BUSCO completeness of insect genome assemblies deposited at the United States National Centre for Biotechnology Information (NCBI) shows that while many do meet the EBP’s standards recommendations, quality in terms of gene content completeness still varies dramatically (Figure 1). Thus while the NCBI may currently offer more than 2’500 assemblies for insects, fewer than half of these achieve a complete and single-copy BUSCO score >90% and most do not yet reach the EBP’s standard of having the majority of sequences assigned to chromosomes. Notably however, accuracy-enhanced long-read technologies together with scaffolding approaches such as high-throughput chromatin conformation capture (Hi-C) are more consistently producing high-quality new genome resources, which are greatly expanding the possibilities for developing new insect model systems.

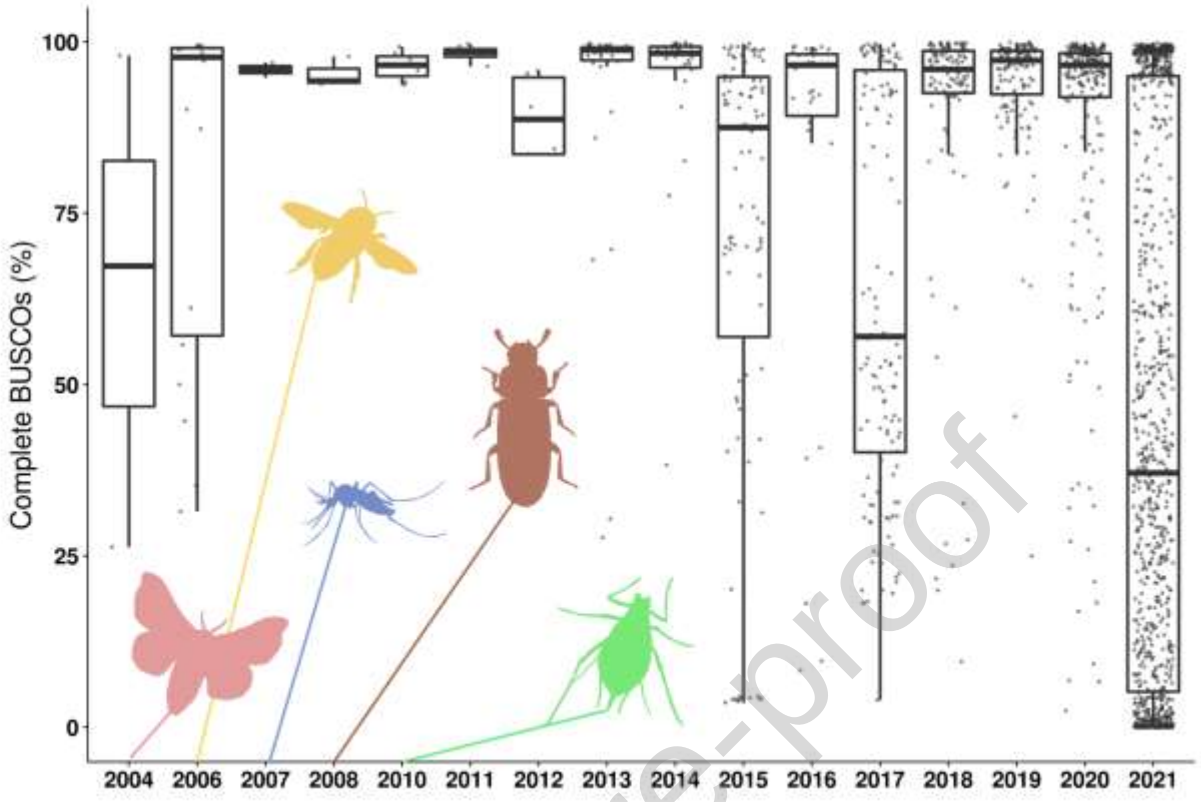

Figure 1. Benchmarking Universal Single-Copy Orthologue (BUSCO) completeness of insect genome assemblies deposited at the United States National Centre for Biotechnology Information (NCBI). The boxplots show distributions per year of the percentage of complete BUSCOs assessed using the Arthropoda lineage dataset for insect assemblies available from the NCBI Assembly database. The first decade is characterised by a slowly increasing number of genome assembly releases, usually for what are regarded as some of the most charismatic and well-studied model insect species, and mostly showing high BUSCO completeness. The subsequent years are characterised by a much faster rate of growth in the numbers of genome assembly releases, accompanied by large variations in quality in terms of gene content completeness. The large numbers of low-completeness assemblies deposited in 2017 and 2021 comprise mainly those for lepidopteran species. Insect silhouettes depict, from left to right: *Bombyx mori* silkmoth, *Apis mellifera* honey bee, *Aedes aegypti* mosquito, *Tribolium castaneum* beetle, and *Acyrthosiphon pisum* pea aphid, linked to the year their genome was first published. Boxplots show the median, first and third quartiles, and lower and upper extremes of the distribution (1.5 x interquartile range). Data are sourced from the Arthropoda Assembly Assessment Catalogue (A^{3}Cat) [10,23], data for 2021 are shown only for assemblies available up to June 11.

Emerging insect model systems are supported by high-quality genome resources

Advances in taxonomic sampling of insects for genome sequencing have been reviewed for ants and other Hymenoptera [24,25], hemipterans [26], beetles [27], flies and other Diptera [28,29], butterflies and other Lepidoptera [30], and many others [9,11,31]. Here we focus on a selection of recent examples of high-quality genomics resources (Table 1) that are supporting the use of new species or groups of species to develop and expand emerging model systems that help advance understanding of insect biology and evolution.

Mayflies have long been the focus of many ecological studies, and together with dragonflies and damselflies they form the sister group to all other winged insect lineages. Recent establishment of a continuous culture system of the *Cloeon dipterum* mayfly [32] allows for comprehensive life-stage and tissue sampling for detailed transcriptional profiling. Combining short reads with ONT sequencing data enabled the assembly of its relatively compact genome of 180 Megabasepairs (Mbp) in 1'395 scaffolds with 96%-97% complete BUSCOs (Table 1), and annotated with 16'357 protein-coding genes. These resources lay the foundations for investigating genomic adaptations to aquatic and aerial life and the origin of insect wings in this emerging model system [32].

Combining long-reads with Hi-C data is proving to be an effective approach for generating chromosome-level assemblies. This was recently demonstrated by Sun *et al.* [33] for five of 17 new high-quality bumblebee genomes (Table 1), where comparisons revealed how the 25-chromosome karyotype of social parasite species derived from the ancestral karyotype of 18 chromosomes. These resources are helping to set up the *Bombus* genus as a new model for quantifying genetic and genomic variation underlying important ecological and behavioural traits of key pollinators. Along with other new bumblebee reference genomes [34,35], they also offer opportunities to explore genetic factors influencing the plastic and adaptive responses impacting insect resilience to climate change [36].

Rearrangements like those observed for the social parasite bumblebees appear to be infrequent in some well-studied groups such as Diptera and Lepidoptera where global genome architectures are generally conserved. Therefore, models from other diverse insect groups are needed to investigate different modes of genome structure evolution. Indeed, analyses of high-quality chromosome-level assemblies of aphids (Table 1) show that their autosomes have undergone dramatic reorganisations in contrast to their sex chromosomes where gene content of the X chromosome has remained highly stable [37,38]. As a model system to investigate the evolution of resistance to insecticides, reference-quality aphid genomes are also enabling comprehensive assessments of within-species variation to understand genomic responses to strong selective forces [39].

The pea aphid was one of the first insects to be sequenced and has served as a valuable model for understanding genomic consequences of host-symbiont interactions. However, genomic resources for new systems are needed to explore the many types of endosymbioses found across different insects. The genome of the rice weevil, *Sitophilus oryzae*, is not yet assembled to chromosome level but shows high BUSCO completeness (Table 1) thereby providing a confident basis from which to investigate how key metabolic processes might be partitioned between host and endosymbiont [40]. Quality and completeness are also particularly critical when tracing cases of horizontal gene transfer, e.g. duplicated bacterial-origin mannosidases in the 1150 Mbp genome assembly of the stink bug *Halyomorpha halys* [41], and bacterial cell wall hydrolase genes acquired by Coccinellinae ladybird beetles identified in the high-quality genome of *Cryptolaemus montrouzieri* [42].

Amongst the most well-known of the Coccinellinae, the harlequin ladybird *Harmonia axyridis* is widely considered to be one of the world's most invasive insects. Many insects are, or have the potential to become, invasives that can cause great damage to natural ecosystems or agricultural crops. Accumulating genomics resources from a variety of insect groups are helping to diversify the models used to study invasion biology and potentially develop new genetic control measures. Hi-C data helped to build a chromosome-level assembly for the two-spot harlequin morph, but with lower BUSCO completeness than prior to Hi-C scaffolding [43] (Table 1). These data, along with assemblies for other morphs e.g. [44], also offer new opportunities to develop the use of these ladybirds, which display more than 200 described colour forms, as an important model system for investigating the genetics of colour pattern polymorphisms [45,46].

Being laboratory-tractable is a key feature of the most versatile model species. For example, the painted lady butterfly, *Vanessa cardui*, can be easily reared in the laboratory and is amenable to CRISPR/Cas9 genome editing, making this widespread, generalist species with complex wing patterns an excellent model. The genome assembly, recently upgraded to chromosome level [47], with transcriptomics data from multiple tissues and developmental stages provides the framework to employ genetic manipulations and functional genomics data for studying migration, host-plant coevolution, and colour patterning [48]. CRISPR/Cas9 has also been established for the tea geometrid moth, *Ectropis grisescens*, which, along with its relevance as an agricultural pest, presents an interesting system for studying insect interactions with plant allelochemicals as well as shape and colour adaptations for effective camouflage. Hi-C scaffolding of PacBio data placed 97.8% of the assembly on 31 chromosomes with an assembly span of 785 Mbp (Table 1) and 18,746 annotated protein-coding genes. The genome maintains the ancestral lepidopteran karyotype (n=31), and separate resequencing of male (ZZ) and female (ZW) individuals allowed for the identification of the Z chromosome and several W candidate scaffolds [49].

While still often challenging, long reads are proving particularly useful for assembling such repeat-rich insect sex chromosomes. For example, the *Pieris macdunnoughii* assembly (Table 1) was built using ONT long reads, where polishing with additional short-read data increased complete lepidopteran BUSCOs by almost 3%. Comparing the resolved sex chromosomes in *Pieris* butterflies of European and North American lineages shows that the fusion event that created the neo-Z chromosome occurred prior to their divergence [50]. These genome resources support this emerging model system for studying maladaptation in plant-insect interactions, where the North American butterflies lay their eggs on invasive Eurasian mustard plants that are lethal to the larvae.

These examples of emerging models with reference genome assemblies show how technological advances are supporting new models by delivering high-quality data. The reference genomes themselves provide a framework onto which new knowledge can be mapped, from comparative genomics analyses, molecular biology experiments, as well as functional and population genomics datasets (transcriptomics, proteomics, metabolomics, resequencing, etc.). Using the number of NCBI BioProjects as a proxy to gauge the extent of genome-enabled research activities shows how the classical model insect species, *Drosophila melanogaster*, is associated with an order of magnitude more registered projects than the other most represented species (Figure 2). Amongst the others are well-known species that are economically important, vectors of human diseases, or agricultural pests, all of which have had publicly available draft assemblies for more than five years and almost all of which now have published high-quality assembly upgrades, including most recently for the fall armyworm [51], the tiger mosquito [52], the brown planthopper [53], and the red flour beetle [54]. Species representing emerging model systems such as the examples outlined above are expected to similarly build genome-anchored knowledgebases that support and enrich the exploration of the diversity of insect biology and evolution.

Table 1. Selected examples of emerging models supported by high-quality genome resources. Completeness assessments with BUSCO v4.1.4 and assembly statistics sourced from the A3Cat: Arthropoda Assembly Assessment Catalogue [10], or (*) directly from [38]. C=Complete , [S=Complete Single, D=Complete Duplicated], F=Fragmented, M=Missing. Ph=Principal Haplotype, Ah=Alternative Haplotype.

Taxon	Assembly Size (Mbp)	Scaffold N50 (Mbp)	Insecta BUSCO % C,[S,D],F,M	Arthropoda BUSCO % C,[S,D],F,M
Mayfly: *Cloeon dipterum*	180	0.46	96.1,[94.1,2.0],0.9,3.0	97.2,[95.0,2.2],1.2,1.6
Bumblebees: *Bombus haemorrhoidalis*	241	15.09	99.7,[99.4,0.3],0.2,0.1	99.4,[99.3,0.1],0.3,0.3
Bumblebees: *Bombus ignitus*	243	15.19	98.3,[98.1,0.2],0.7.1.0	97.6,[97.5,0.1],1.3,1.1
Bumblebees: *Bombus turneri*	243	9.70	99.6,[99.3,0.3],0.2,0.2	99.2,[99.2,0.0],0.5,0.3
Bumblebees: *Bombus breviceps*	248	14.71	99.6,[99.4,0.2],0.1,0.3	99.1,[99.1,0.0],0.4,0.5
Bumblebees: *Bombus pyrosoma*	255	15.22	99.7,[99.5,0.2],0.1,0.2	99.6,[99.6,0.0],0.1,0.3
Bumblebees: *Bombus hortorum*	296	17.02	99.6,[99.2,0.4],0.1,0.3	99.5,[99.2,0.3],0.3,0.2
Aphids: *Myzus persicae* *	395	69.48	NA	97.1,[94.2,2.9],0.5,2.4
Aphids: *Acyrthosiphon pisum* *	526	126.60	NA	97.6,[94.7,2.9],0.4,2.1
Aphids: *Rhopalosiphum maidis*	326	93.30	97.0,[94.8,2.2],0.7,2.3	98.3,[95.4,2.9],0.5,1.2
Weevil: *Sitophilus oryzae*	770	2.86	97.8,[95.8,2.0],0.7,1.5	98.5,[97.1,1.4],0.3,1.2
Stink bug: *Halyomorpha halys*	1150	0.80	97.4,[96.0,1.4],1.0,1.6	96.7,[95.2,1.5],1.4,1.9
Ladybird: *Cryptolaemus montrouzieri*	988	10.38	97.1,[96.0,1.1],0.6,2.3	97.0,[96.4,0.6],1.1,1.9
Ladybird: *Harmonia axyridis*	417	2.05	92.4,[90.0,2.4],1.2,6.4	91.7,[89.3,2.4],1.4,6.9
Butterfly: *Vanessa cardui* (Ph)	425	14.62	98.9,[98.8,0.1],0.4,0.7	98.9,[98.6,0.3],0.4,0.7
Butterfly: *Vanessa cardui* (Ah)	401	2.75	96.1,[96.0,0.1],0.4,3.5	95.6,[95.4,0.2],0.5,3.9
Moth: *Ectropis grisescens*	785	26.91	96.4,[95.7,0.7],1.2,2.4	95.6,[95.2,0.4],1.9,2.5
Butterfly: *Pieris macdunnoughii*	317	5.20	97.2,[96.3,0.9],0.4,2.4	97.2,[96.5,0.7],0.9,1.9

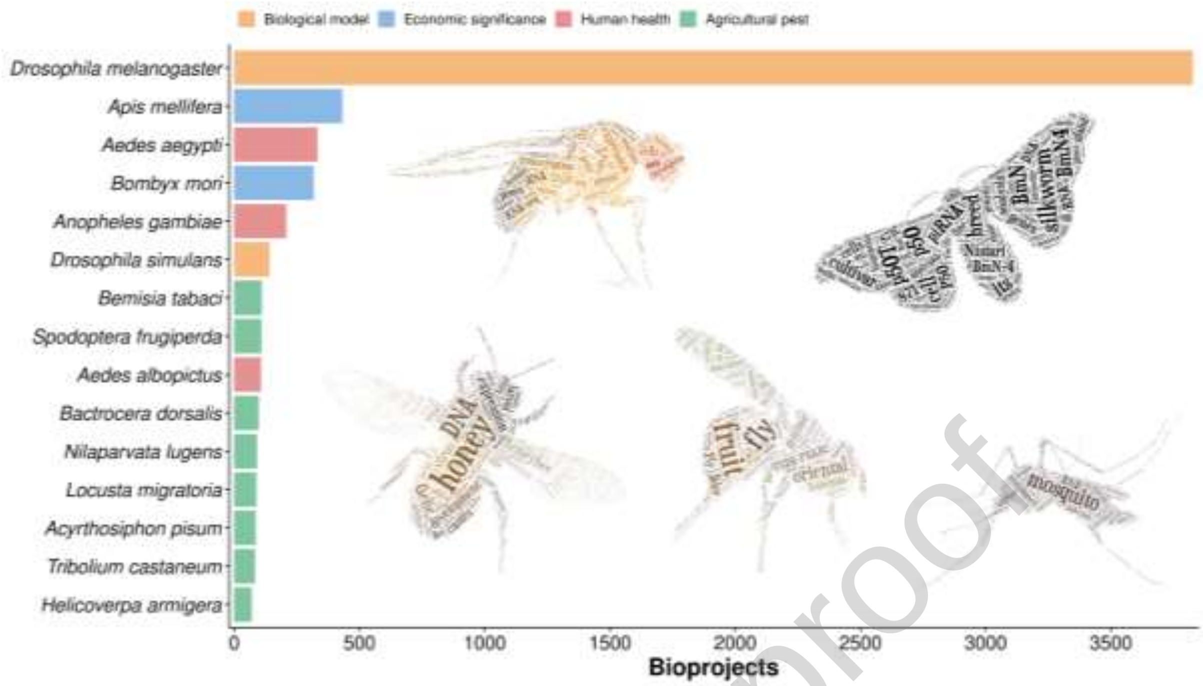

Figure 2. Number of BioProject entries for the fifteen most represented insect species. Counts of BioProjects sourced from the United States National Centre for Biotechnology Information (NCBI, January 2022) show that the classical model species, *Drosophila melanogaster*, is associated with an order of magnitude more registered projects than the other most represented species. Bar colours represent a simplified 'principal research interest/relevance' category for each species. 'Word Clouds' for selected species are built from the collated titles of all their available BioProjects. *Drosophila melanogaster* (Diptera): fruit fly; *Apis mellifera* (Hymenoptera): western honey bee; *Aedes aegypti* (Diptera): yellow fever mosquito; *Bombyx mori* (Lepidoptera): domestic silk moth; *Anopheles gambiae* (Diptera): African malaria mosquito; *Drosophila simulans* (Diptera): fruit fly; *Bemisia tabaci* (Hemiptera): silverleaf whitefly; *Spodoptera frugiperda* (Lepidoptera): fall armyworm; *Aedes albopictus* (Diptera): tiger mosquito; *Bactrocera dorsalis* (Diptera): oriental fruit fly; *Nilaparvata lugens* (Hemiptera): brown planthopper; *Locusta migratoria* (Orthoptera): migratory locust; *Acyrthosiphon pisum* (Hemiptera): pea aphid; *Tribolium castaneum* (Coleoptera): red flour beetle; *Helicoverpa armigera* (Lepidoptera): cotton bollworm.

Conclusions

New technologies are helping to greatly expand the diversity of insect species for which genome resources are being generated across Insecta [9,10], presenting opportunities to develop new model systems for studying a large variety of biological phenomena. Within-genus sampling is also reaching new levels of resolution, exemplified by the genome assemblies for 101 lines of 93 drosophilid species spanning 14 species groups and 35 sub-groups [55]. Nevertheless, challenges such as working with large repeat-rich genomes or very small specimens from which to extract high-molecular-weight DNA mean that achieving reference-quality standards can still be arduous [12]. The active participation of the arthropod genomics community in the development of standards and provision of guidelines and protocols through initiatives coordinating the scaling up of reference genome generation are helping to overcome many of these challenges [22,56,57]. Gene content completeness and other quality assessments during production and of the resulting chromosome-level assemblies will therefore continue to play a key role in establishing genome resources that best support the development of new model systems and advance understanding of insect biology and evolution.

Declaration of Competing Interest

The authors certify that they have no affiliations with or involvement in any organisation or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

• of special interest

Mathers TC, Wouters RHM, Mugford ST, Swarbreck D, van Oosterhout C, Hogenhout SA: ***Chromosome-Scale Genome Assemblies of Aphids Reveal Extensively Rearranged Autosomes and Long-Term Conservation of the X Chromosome***. Mol Biol Evol 2021, 38:856–875.

This study highlights the importance of examining genome evolutionary features in different insect clades, especially chromosome rearrangements as they can lead to rapid evolution and

speciation. By developing and comparing high-quality genome resources for these aphids, dramatic reorganisations of their autosomes was found to sharply contrast the much more stably maintained gene content of the aphid sex chromosome.

• of special interest
Singh KS, Cordeiro EMG, Troczka BJ, Pym A, Mackisack J, Mathers TC, Duarte A, Legeai F, Robin S, Bielza P, et al.: ***Global patterns in genomic diversity underpinning the evolution of insecticide resistance in the aphid crop pest Myzus persicae***. Commun Biol 2021, 4:847.
Using a chromosome-level genome assembly of the aphid, *Myzus persicae,* and 40X short-read sequencing of 127 clones derived from 19 countries, this study shows how a high-quality reference enables comprehensive characterisations of genome-wide patterns of global genetic variation. This system helps understand genomic responses of insects to insecticide exposures and is particularly relevant for the control of agricultural pests.

• of special interest
Parisot N, Vargas-Chávez C, Goubert C, Baa-Puyoulet P, Balmand S, Beranger L, Blanc C, Bonnamour A, Boulesteix M, Burlet N, et al.: ***The transposable element-rich genome of the cereal pest Sitophilus oryzae***. BMC Biol 2021, 19:241.
Beyond building resources to guide progress towards novel pest control tools for this beetle, this study establishes the groundwork for developing a new genome-enabled model for investigating the relationships between endosymbionts and their insect hosts. Despite their relatively recent association, the intracellular symbiotic bacterium appears to rely on the rice weevil for several key amino acids and nucleotides.

•• of outstanding interest
Almudi I, Vizueta J, Wyatt CDR, de Mendoza A, Marlétaz F, Firbas PN, Feuda R, Masiero G, Medina P, Alcaina-Caro A, et al.: ***Genomic adaptations to aquatic and aerial life in mayflies and the origin of insect wings***. Nat Commun 2020, 11:2631.
This study exemplifies the coming together of establishing a new laboratory-tractable system with the generation of genome resources and extensive functional genomics data to support novel biological investigations. The mayfly genome provided the framework to explore patterns of gene expression throughout its aquatic and aerial life cycle and across different organs, and to identify a core set of genes involved in insect wing development.

•• of outstanding interest
Sun C, Huang J, Wang Y, Zhao X, Su L, Thomas GWC, Zhao M, Zhang X, Jungreis I, Kellis M, et al.: ***Genus-wide characterization of bumblebee genomes provides insights into their evolution and variation in ecological and behavioral traits***. Mol Biol Evol 2021, 38:486–501.
Chromosome-level assemblies generated for five of the 17 bumblebee species in this study allowed tracing of the rearrangements that created the unusual 25-chromosome karyotype in social parasites. The high-quality genome resources from sampling species across the genus supported the quantification of genetic and genomic variation across the *Bombus* phylogeny, where high levels of gene tree discordance are likely driven by incomplete lineage sorting.

•• of outstanding interest
Chen M, Mei Y, Chen X, Chen X, Xiao D, He K, Li Q, Wu M, Wang S, Zhang F, et al.: ***A chromosome- level assembly of the harlequin ladybird Harmonia axyridis as a genomic resource to study beetle and invasion biology***. Mol Ecol Resour 2021, 21:1318–1332.
In this study, the generation of a chromosome-level genome assembly of this charismatic ladybird additionally identified the X chromosome and Y-linked scaffolds by separately resequencing males and females. These resources support the development of the harlequin as a model for studying invasion biology in insects, and, with more than 200 described colour forms, for investigating the genetics of colour pattern polymorphisms.

Acknowledgements

The authors thank Sagane Dind, Giulia Campli, Livio Ruzzante, and Antonin Thiébaut for providing useful suggestions for improvements and valuable feedback on the manuscript. This work was supported by the Swiss National Science Foundation [grant numbers PP00P3_170664 and PP00P3_202669].

References

1. Ankeny R, Leonelli S: **Model organisms**. Cambridge University Press; 2020.
2. Bilder D, Irvine KD: **Taking stock of the *Drosophila* research ecosystem**. *Genetics* 2017, **206**:1227–1236.
3. Adams MD: **The genome sequence of *Drosophila melanogaster***. *Science* 2000, **287**:2185–2195.
4. Matthews BJ, Vosshall LB: **How to turn an organism into a model organism in 10 'easy' steps**. *J Exp Biol* 2020, **223**:jeb218198.
5. Campbell JF, Athanassiou CG, Hagstrum DW, Zhu KY: ***Tribolium castaneum*: A model insect for fundamental and applied research**. *Annu Rev Entomol* 2022, **67**:347–365.
6. Averof M, Sinigaglia C: **Introduction to emerging systems**. *EvoDevo* 2020, **11**:8, s13227-020-00153-y.
7. Duffy MA, García-Robledo C, Gordon SP, Grant NA, Green DA, Kamath A, Penczykowski RM, Rebolleda-Gómez M, Wale N, Zaman L: **Model systems in ecology, evolution, and behavior: a call for diversity in our model systems and discipline**. *Am Nat* 2021, **198**:53–68.
8. Hotaling S, Kelley JL, Frandsen PB: **Aquatic insects are dramatically underrepresented in genomic research**. *Insects* 2020, **11**:601.
9. Hotaling S, Sproul JS, Heckenhauer J, Powell A, Larracuente AM, Pauls SU, Kelley JL, Frandsen PB: **Long-reads are revolutionizing 20 years of insect genome sequencing**. *Genome Biol Evol* 2021, doi:10.1093/gbe/evab138.
10. Feron R, Waterhouse RM: **Assessing species coverage and assembly quality of rapidly accumulating sequenced genomes**. *Gigascience*; 2022, doi: 10.1093/gigascience/giac006.
11. González VL, Devine AM, Trizna M, Mulcahy DG, Barker KB, Coddington JA: **Open access genomic resources for terrestrial arthropods**. *Curr Opin Insect Sci* 2018, **25**:91–98.
12. Li F, Zhao X, Li M, He K, Huang C, Zhou Y, Li Z, Walters JR: **Insect genomes: progress and challenges**. *Insect Mol Biol* 2019, **28**:739–758.
13. Waterhouse RM, Tegenfeldt F, Li J, Zdobnov EM, Kriventseva EV: **OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs**. *Nucleic Acids Res* 2013, **41**:D358-65.
14. Kriventseva EV, Tegenfeldt F, Petty TJ, Waterhouse RM, Simão FA, Pozdnyakov IA, Ioannidis P, Zdobnov EM: **OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software**. *Nucleic Acids Res* 2015, **43**:D250–D256.
15. Zdobnov EM, Tegenfeldt F, Kuznetsov D, Waterhouse RM, Simão FA, Ioannidis P, Seppey M, Loetscher A, Kriventseva EV: **OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs**. *Nucleic Acids Res* 2017, **45**:D744–D749.
16. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM: **BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs**. *Bioinformatics* 2015, **31**:3210–3212.
17. Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM: **BUSCO applications from quality assessments to gene prediction and phylogenomics**. *Mol Biol Evol* 2018, **35**:543–548.
18. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM: **BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes**. *Mol Biol Evol* 2021, **38**:4647–4654.
19. Kriventseva EV, Kuznetsov D, Tegenfeldt F, Manni M, Dias R, Simão FA, Zdobnov EM: **OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs**. *Nucleic Acids Res* 2019, **47**:D807–D811.
20. Waterhouse RM, Seppey M, Simão FA, Zdobnov EM: **Using BUSCO to assess insect genomic resources**. In *Insect Genomics*. Edited by Brown SJ, Pfrender ME. Springer New York; 2019:59–74.
21. Manni M, Berkeley MR, Seppey M, Zdobnov EM: **BUSCO: assessing genomic data quality and beyond**. *Curr Protoc* 2021, **1**.
22. Lawniczak MKN, Durbin R, Flicek P, Lindblad-Toh K, Wei X, Archibald JM, Baker WJ, Belov K, Blaxter ML, Marques Bonet T, et al.: **Standards recommendations for the Earth BioGenome Project**. *Proc Natl Acad Sci* 2022, **119**:e2115639118.
23. Feron R, Waterhouse RM: **Supporting data for "Assessing species coverage and assembly quality of rapidly accumulating sequenced genomes."** *GigaDB* 2022,

doi:10.5524/100974.
24. Favreau E, Martínez-Ruiz C, Rodrigues Santiago L, Hammond RL, Wurm Y: **Genes and genomic processes underpinning the social lives of ants**. *Curr Opin Insect Sci* 2018, **25**:83–90.
25. Branstetter MG, Childers AK, Cox-Foster D, Hopper KR, Kapheim KM, Toth AL, Worley KC: **Genomes of the Hymenoptera**. *Curr Opin Insect Sci* 2018, **25**:65–75.
26. Panfilio KA, Angelini DR: **By land, air, and sea: hemipteran diversity through the genomic lens**. *Curr Opin Insect Sci* 2018, **25**:106–115.
27. McKenna DD: **Beetle genomes in the 21st century: prospects, progress and priorities**. *Curr Opin Insect Sci* 2018, **25**:76–82.
28. Wiegmann BM, Richards S: **Genomes of Diptera**. *Curr Opin Insect Sci* 2018, **25**:116–124.
29. Ruzzante L, Reijnders MJMF, Waterhouse RM: **Of genes and genomes: mosquito evolution and diversity**. *Trends Parasitol* 2019, **35**:32–51.
30. Triant DA, Cinel SD, Kawahara AY: **Lepidoptera genomes: current knowledge, gaps and future directions**. *Curr Opin Insect Sci* 2018, **25**:99–105.
31. Richards S, Childers A, Childers C: **Editorial overview: Insect genomics: Arthropod genomic resources for the 21st century: It only counts if it's in the database!** *Curr Opin Insect Sci* 2018, **25**:iv–vii.
32. Almudi I, Vizueta J, Wyatt CDR, de Mendoza A, Marlétaz F, Firbas PN, Feuda R, Masiero G, Medina P, Alcaina-Caro A, et al.: **Genomic adaptations to aquatic and aerial life in mayflies and the origin of insect wings**. *Nat Commun* 2020, **11**:2631.
33. Sun C, Huang J, Wang Y, Zhao X, Su L, Thomas GWC, Zhao M, Zhang X, Jungreis I, Kellis M, et al.: **Genus-wide characterization of bumblebee genomes provides insights into their evolution and variation in ecological and behavioral traits**. *Mol Biol Evol* 2021, **38**:486–501.
34. Christmas MJ, Jones JC, Olsson A, Wallerman O, Bunikis I, Kierczak M, Peona V, Whitley KM, Larva T, Suh A, et al.: **Genetic barriers to historical gene flow between cryptic species of alpine bumblebees revealed by comparative population genomics**. *Mol Biol Evol* 2021, **38**:3126–3143.
35. Crowley L, University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding collective, Wellcome Sanger Institute Tree of Life programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective, Tree of Life Core Informatics collective, Darwin Tree of Life Consortium: **The genome sequence of the garden bumblebee, *Bombus hortorum* (Linnaeus, 1761)**. *Wellcome Open Res* 2021, **6**:270.
36. Maebe K, Hart AF, Marshall L, Vandamme P, Vereecken NJ, Michez D, Smagghe G: **Bumblebee resilience to climate change, through plastic and adaptive responses**. *Glob Change Biol* 2021, **27**:4223–4237.
37. Li Y, Zhang B, Moran NA: **The aphid X chromosome is a dangerous place for functionally important genes: diverse evolution of hemipteran genomes based on chromosome-level assemblies**. *Mol Biol Evol* 2020, **37**:2357–2368.
38. Mathers TC, Wouters RHM, Mugford ST, Swarbreck D, van Oosterhout C, Hogenhout SA: **Chromosome-scale genome assemblies of aphids reveal extensively rearranged autosomes and long-term conservation of the X chromosome**. *Mol Biol Evol* 2021, **38**:856–875.
39. Singh KS, Cordeiro EMG, Troczka BJ, Pym A, Mackisack J, Mathers TC, Duarte A, Legeai F, Robin S, Bielza P, et al.: **Global patterns in genomic diversity underpinning the evolution of insecticide resistance in the aphid crop pest *Myzus persicae***. *Commun Biol* 2021, **4**:847.
40. Parisot N, Vargas-Chávez C, Goubert C, Baa-Puyoulet P, Balmand S, Beranger L, Blanc C, Bonnamour A, Boulesteix M, Burlet N, et al.: **The transposable element-rich genome of the cereal pest *Sitophilus oryzae***. *BMC Biol* 2021, **19**:241.
41. Sparks ME, Bansal R, Benoit JB, Blackburn MB, Chao H, Chen M, Cheng S, Childers C, Dinh H, Doddapaneni HV, et al.: **Brown marmorated stink bug, *Halyomorpha halys* (Stål), genome: putative underpinnings of polyphagy, insecticide resistance potential and biology of a top worldwide pest**. *BMC Genomics* 2020, **21**:227.
42. Li H-S, Tang X-F, Huang Y-H, Xu Z-Y, Chen M-L, Du X-Y, Qiu B-Y, Chen P-T, Zhang W, Ślipiński A, et al.: **Horizontally acquired antibacterial genes associated with adaptive radiation of ladybird beetles**. *BMC Biol* 2021, **19**:7.
43. Chen M, Mei Y, Chen X, Chen X, Xiao D, He K, Li Q, Wu M, Wang S, Zhang F, et al.: **A chromosome- level assembly of the harlequin ladybird *Harmonia axyridis* as a genomic resource to study beetle and invasion biology**. *Mol Ecol Resour* 2021, **21**:1318–

1332.
44. Boyes D, Crowley L, University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding collective, Wellcome Sanger Institute Tree of Life programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective, Tree of Life Core Informatics collective, Darwin Tree of Life Consortium: **The genome sequence of the harlequin ladybird, *Harmonia axyridis* (Pallas, 1773)**. *Wellcome Open Res* 2021, **6**:300.
45. Gautier M, Yamaguchi J, Foucaud J, Loiseau A, Ausset A, Facon B, Gschloessl B, Lagnel J, Loire E, Parrinello H, et al.: **The genomic basis of color pattern polymorphism in the harlequin ladybird**. *Curr Biol* 2018, **28**:3296-3302.e7.
46. Ando T, Matsuda T, Goto K, Hara K, Ito A, Hirata J, Yatomi J, Kajitani R, Okuno M, Yamaguchi K, et al.: **Repeated inversions within a pannier intron drive diversification of intraspecific colour patterns of ladybird beetles**. *Nat Commun* 2018, **9**:3843.
47. Lohse K, Wright C, Talavera G, García-Berro A, Darwin Tree of Life Barcoding collective, Wellcome Sanger Institute Tree of Life programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective, Tree of Life Core Informatics collective, Darwin Tree of Life Consortium: **The genome sequence of the painted lady, *Vanessa cardui* Linnaeus 1758**. *Wellcome Open Res* 2021, **6**:324.
48. Zhang L, Steward RA, Wheat CW, Reed RD: **High-quality genome assembly and comprehensive transcriptome of the painted lady butterfly *Vanessa cardui***. *Genome Biol Evol* 2021, **13**:evab145.
49. Pan Y, Fang G, Wang Z, Cao Y, Liu Y, Li G, Liu X, Xiao Q, Zhan S: **Chromosome-level genome reference and genome editing of the tea geometrid**. *Mol Ecol Resour* 2021, **21**:2034–2049.
50. Steward RA, Okamura Y, Boggs CL, Vogel H, Wheat CW: **The genome of the margined white butterfly (*Pieris macdunnoughii*): sex chromosome insights and the power of polishing with PoolSeq data**. *Genome Biol Evol* 2021, **13**:evab053.
51. Zhang L, Liu B, Zheng W, Liu C, Zhang D, Zhao S, Li Z, Xu P, Wilson K, Withers A, et al.: **Genetic structure and insecticide resistance characteristics of fall armyworm populations invading China**. *Mol Ecol Resour* 2020, **20**:1682–1696.
52. Palatini U, Masri RA, Cosme LV, Koren S, Thibaud-Nissen F, Biedler JK, Krsticevic F, Johnston JS, Halbach R, Crawford JE, et al.: **Improved reference genome of the arboviral vector *Aedes albopictus***. *Genome Biol* 2020, **21**:215.
53. Ye Y, Zhang H, Li D, Zhuo J, Shen Y, Hu Q, Zhang C: **Chromosome-level assembly of the brown planthopper genome with a characterized Y chromosome**. *Mol Ecol Resour* 2021, **21**:1287–1298.
54. Herndon N, Shelton J, Gerischer L, Ioannidis P, Ninova M, Dönitz J, Waterhouse RM, Liang C, Damm C, Siemanowski J, et al.: **Enhanced genome assembly and a new official gene set for *Tribolium castaneum***. *BMC Genomics* 2020, **21**:47.
55. Kim BY, Wang JR, Miller DE, Barmina O, Delaney E, Thompson A, Comeault AA, Peede D, D'Agostino ER, Pelaez J, et al.: **Highly contiguous assemblies of 101 drosophilid genomes**. *eLife* 2021, **10**:e66405.
56. i5K Consortium: **The i5K initiative: advancing arthropod genomics for knowledge, human health, agriculture, and the environment**. *J Hered* 2013, **104**:595–600.
57. Formenti G, Theissinger K, Fernandes C, Bista I, Bombarely A, Bleidorn C, Ciofi C, Crottini A, Godoy JA, Höglund J, et al.: **The era of reference genomes in conservation genomics**. *Trends Ecol Evol* 2022, doi:10.1016/j.tree.2021.11.008.

Declaration of Competing Interest

The authors certify that they have no affiliations with or involvement in any organisation or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Highlights

- New genome resources are expanding possibilities for developing insect model systems
- The quality of genome assemblies varies in contiguity and gene content completeness

- Technological advances are supporting new models by delivering high-quality data
- Emerging model systems advance understanding of insect biology and evolution