



ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in
Insect Science

Genomes of Diptera

1 Brian M Wiegmann¹ and Stephen Richards²

3Q1

4 Diptera (true flies) are among the most diverse holometabolan
5 insect orders and were the first eukaryotic order to have a
6 representative genome fully sequenced. 110 fly species have
7 publically available genome assemblies and many hundreds of
8 population-level genomes have been generated in the model
9 organisms *Drosophila melanogaster* and the malaria mosquito
10 *Anopheles gambiae*. Comparative genomics carried out in a
11 phylogenetic context is illuminating many aspects of fly
12 biology, providing unprecedented insight into variability in
13 genome structure, gene content, genetic mechanisms, and
14 rates and patterns of evolution in genes, populations, and
15 species. Despite the rich availability of genomic resources in
16 flies, there remain many fly lineages to which new genome
17 sequencing efforts should be directed. Such efforts would be
18 most valuable in fly families or clades that exhibit multiple
19 origins of key fly behaviors such as blood feeding, phytophagy,
20 parasitism, pollination, and mycophagy.

Addresses

21^{Q2} ¹ Department of Entomology & Plant Pathology, North Carolina State
22 University, Raleigh, NC 27695, United States

23 ² Human Genome Sequencing Center, Department of Molecular and
24 Human Genetics, Baylor College of Medicine, Houston, TX 77006,
25 United States

Corresponding author: Wiegmann, Brian M (bwiegman@ncsu.edu)

26 **Current Opinion in Insect Science** 2018, **25**:xx–yy

27 This review comes from a themed issue on **Insect genomics**

28 Edited by **Stephen Richards**, **Anna Childers**, and **Christopher**
29 **Childers**

30 [doi:10.1016/j.cois.2018.01.007](https://doi.org/10.1016/j.cois.2018.01.007)

31 2214-5745/© 2018 Published by Elsevier Inc.

Introduction

33 Diptera (true flies) are among the most diverse insect
34 orders, in both species (approx. 158 000 described species
35 in 180+ families), and in the diversity of their life histo-
36 ries, behaviors, trophic habits, human impacts and mor-
37 phological adaptations [1]. Because of the unquestioned
38 importance of *Drosophila melanogaster* as the premier
39 genetic model organism; the tremendous harm caused
40 to human health by vectoring disease (e.g., mosquitoes,
41 *tsetse*) and to agriculture as pests (e.g., Mediterranean fruit
42 fly: *Ceratitits capitata*; Hessian fly: *Mayetiola destructor*;
sheep blow fly: *Lucilia sericata*), flies were among the first

43 animals to have a fully sequenced genome and, as an
44 order, include the largest number of insect species with
45 full and draft genomes currently completed and publicly
46 available (110 genomes). This number is growing as
47 deeper genomic sampling opens up phylogenetic com-
48 parisons to provide frameworks for evaluating genetic
49 mechanisms and evolutionary processes at species-level
50 and population-level. The breadth of genomic sampling
51 is also increasing throughout the order into non-model
52 groups [2,3**]. This wider taxonomic coverage is espe-
53 cially valuable for increasing the evolutionary scope of
54 assessments outside of *Drosophila* and mosquitoes. Phy-
55 logenetic comparisons both within and between fly fami-
56 lies have led to landmark studies in gene regulation and
57 repair [4–7], development [8,9], neurobiology [10,11], sex
58 determination [12], insecticide resistance [13,14], trophic
59 specialization [15] and ecological adaptation [16–18],
60 among many others. Without doubt, fly science —
61 advanced by NextGen sequencing and the growth of
62 more complete genomic resources — has driven the
63 genomics revolution in insects. Here, we briefly summa-
64 rize the taxonomic distribution of available dipteran
65 genomes, review some of the most exciting findings
66 generated by a ‘comparative genomics’ approach and
67 provide some broad guidelines for choosing additional
68 taxa to sequence in the new era of rapid, low cost, draft
69 genomes.

Diptera genomes — current status

70 Molecular, phylogenetic, and fossil evidence place the
71 origin of flies in the late Permian ~260 mya, beginning at
72 about the same time of the other major holometabolan
73 orders [19,20]. Their prodigious morphological and
74 genetic diversity and an episodic history of rapid diversi-
75 fication have challenged efforts to fully resolve the fly tree
76 of life, but phylogenomic studies are providing a major
77 new source of evidence and the expansion of more deeply
78 sampled genomic resources are rapidly improving our
79 understanding of the fly tree [1,21]. Most recent studies
80 support classification of flies into five major infraorders,
81 Tipulomorpha, Culicomorpha, Psychodomorpha, Bibio-
82 nomorpha, and Brachycera [1,19,22]. The Brachycera
83 (short-horned Diptera) with about 20 suprafamilial groups
84 and over 80 000 described species contains major radi-
85 ations both among the early diverging, Lower Brachycera
86 lineages (180 mya) and in the schizophoran Cyclorrhapha
87 or ‘higher flies’ (65 mya). Schizophoran Cyclorrhapha
88 are diverse taxonomically and behaviorally, with over 78 fam-
89 ilies and a wide range of habits including multiple inde-
90 pendent origins of phytophagy, parasitoidism, mycoph-
91 agy, mammal parasitism and myiasis, blood-feeding, and
92 the widespread larval feeding habit of saprophagy on all
93

2 Insect genomics

manner of decaying organic substrates. Important pollinators occur throughout the order and repeated close associations with flowering plants seem to have been an important driver of fly morphological-diversity, species-diversity, and ecological diversity [23–25]. Genomes are now available in all five infraorders, but of the 110 genomes currently available across Diptera, most of these are represented by the many closely related species sampled in just two families: Culicidae (mosquitoes, Culicomorpha; 27 genomes) and Drosophilidae (Brachycera: Cyclorrhapha: Schizophora: Ephydroidea; 33 genomes) (Table 1).

Taxonomic coverage of genome sequences is increasing across Diptera through new rapid sequencing technologies, ‘informed’ informatics pipelines that capitalize on the existing knowledge base, and concerted efforts to expand resources for well-studied clades [2,3^{**},26^{*},27,28^{**}]. Logically, the available assemblies have been completed primarily in medically or agriculturally important species where these data provide important information for vast research communities to facilitate exploration of potential control mechanisms. This includes genomes in multiple non-mosquito, blood-feeding disease vectors, including sand flies (three species, Psychodidae), *tsetse* (six species; Glossinidae), and stable fly (one species; Muscidae); and, in crop or fruit pests like the wheat pest Hessian fly and 10 species of true fruit flies (Tephritidae), or the blow flies, *Lucilia sericata* and *Lucilia cuprina*, which are myiasis-causing pests of sheep [29]. Vicoso and Bachtrog [3^{**}] and Dikow *et al.* [2] recently expanded genomic knowledge in flies by adding draft genomes in 35 species distributed throughout the order and notably increasing coverage for non-model flies in the Lower Diptera and Lower Brachycera.

Characteristics of the fly genome

Fly genomes are highly variable in size when compared throughout the order ranging from 89.58 Mb (Antarctic midge, *Belgica antarctica*) to 1.18 Gb (horn fly, *Haematobia irritans*) (Animal Genome Size Database, accessed 1/2/2018). Genome size can also vary significantly within a family, ranging widely in sampled mosquitoes (210 Mb to 1.9 Gb) and 2–3-fold within the relatively small sized genomes of Drosophilidae (117–386 Mb) [30]. Variation in genome size is likely the result of large differences in transposable element (TE) and other repetitive non-coding DNA [31,32]. Studies of genome architecture, gene content, synteny, inversions, and the evolution of non-coding elements are important areas of comparative genomic study in flies [33–35]. Extraordinary coverage in *Drosophila* has led to groundbreaking work on how new genes are formed [36,37], how genes and genomic regions interact and are regulated [38], and how genomic events can shape organismal history [39,40]. Gene content is also highly variable in flies. Drosophilids and mosquitoes can vary by 5000 genes with 13 920 genes estimated in

Drosophila melanogaster, and 18 955 in *Culex quinquefasciatus*, but the extremes are as few as 13 517 in *Belgica antarctica* [41^{*}] and as many as 23 884 in the house fly, *Musca domestica* [42]. Despite this, the dynamics of gene family evolution are increasingly resolved through the availability of species-level phylogenetic comparisons. Major changes in copy number associated with physiological and behavioral adaptations have been demonstrated for P450s, chemoreceptors, and odorant binding proteins, for example in the house fly, presumably due to their trophic behaviors, exposure to insecticides, and associations with animal pathogens [14]. Expansion and contraction in the overall gene copy number has been shown to be as much as 5× faster in *Anopheles* than in *Drosophila* [43^{*}]. The large number of high quality assemblies available for multiple species of mosquitoes and *Drosophila* has also enabled the discovery, exploration and evolutionary analysis of small regulatory elements such as microRNA [44,45], PIWI and Aubergine [46], and functional small reading frames (smORF) [47].

Genomic sequencing for flies began in Cyclorrhapha with the completion of *Drosophila melanogaster* [48] and in the Culicomorpha with malaria mosquito, *Anopheles gambiae* [49] and yellow fever mosquito *Aedes aegypti* [50]. These genomes catalyzed the successful establishment and growth of genome databases, annotation reference libraries, and analysis tools and pipelines; but also invigorated large international research consortia that produced excellent assemblies for 12 *Drosophila* species [33] and 16 Anopheline mosquitoes [51]. For this reason, most of our knowledge of fly genome variability comes at two widely differing levels of divergence. Mosquitoes and *Drosophila* shared a common ancestor approximately 240 mya [19] thereby representing one of the oldest splits among extant fly lineages. Among *Drosophila* and *Anopheles* species, splits are likely to span divergences from only 0.24 mya to as many as 22–55 mya for sampled *Drosophila* [52] and from 0.54 through 1.8–100 mya among anophelines [43^{*}]. Genomic comparisons of mosquito and *Drosophila*, and between flies and other insect orders, reveal a marked genome-wide evolutionary rate acceleration in Diptera [53], making flies a truly ‘long-branched’ taxon relative to other insects. Mosquito/*Drosophila* comparisons show that these two flies have diverged at a significantly greater rate from each other as can be seen in comparisons among ancient diverging vertebrates (450 mya) [54,55].

The impact of phylogenetically-informed dipteran comparative genomics

Diptera provide some of the best examples available in insects as to the power of comparative genomics to advance understanding of ecological and evolutionary patterns and mechanisms, adaptations and physiological functions, and the linkage of genotype and phenotype. A striking example is the identification of patterns of

Download English Version:

<https://daneshyari.com/en/article/8878551>

Download Persian Version:

<https://daneshyari.com/article/8878551>

[Daneshyari.com](https://daneshyari.com)