Polytene Chromosomal Maps of 11 Drosophila Species: The Order of Genomic Scaffolds Inferred From Genetic and Physical Maps

Stephen W. Schaeffer,[*,1] Arjun Bhutkar,[†,‡] Bryant F. McAllister,[§] Muneo Matsuda,[¶]
Luciano M. Matzkin,[&] Patrick M. O'Grady,[**] Claudia Rohde,[††] Vera L. S. Valente,[‡‡]
Montserrat Aguadé,[§§] Wyatt W. Anderson,[¶¶] Kevin Edwards,[**] Ana C. L. Garcia,[‡‡]
Josh Goodman,[&&] James Hartigan,[***] Eiko Kataoka,[¶] Richard T. Lapoint,[†††]
Elena R. Lozovsky,[‡‡‡] Carlos A. Machado,[&] Mohamed A. F. Noor,[§§§]
Montserrat Papaceit,[§§] Laura K. Reed,[&,2] Stephen Richards,[¶¶¶]
Tania T. Rieger,[&&&] Susan M. Russo,[†,****] Hajime Sato,[¶]
Carmen Segarra,[§§] Douglas R. Smith,[***] Temple F. Smith,[‡]
Victor Strelets,[&&] Yoshiko N. Tobari,[††††] Yoshihiko Tomimura,[‡‡‡‡]
Marvin Wasserman,[§§§§] Thomas Watts,[&] Robert Wilson,[&&]
Kiyohito Yoshida,[¶¶¶¶] Therese A. Markow,[&]
William M. Gelbart[†,****] and Thomas C. Kaufman[&&]

[*]Department of Biology and Institute of Molecular Evolutionary Genetics, The Pennsylvania State University, University Park, Pennsylvania 16802, [†]Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 02138, [‡]BioMolecular Engineering Research Center, Boston University, Boston, Massachusetts 02215, [§]Department of Biological Sciences, University of Iowa, Iowa City, Iowa 52242, [¶]School of Medicine, Kyorin University, Mitaka, Tokyo 181-8611, Japan, [&]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721, [**]Department of Environmental Sciences, Policy and Management, University of California, Berkeley, California 94720, [††]Centro Acadêmico de Vitória, Universidade Federal de Pernambuco, CEP 55608-680, Vitória de Santo Antão/PE, Brazil, [‡‡]Departamento de Genética, Universidade Federal do Rio Grande do Sul, CEP 91501-970, Porto Alegre/RS, Brazil, [§§]Departament de Genètica, Universitat de Barcelona, 08028 Barcelona, Spain, [¶¶]Department of Genetics, University of Georgia, Athens, Georgia 30602, [&&]Department of Biology, Indiana University, Bloomington, Indiana 47405, [***]Agencourt Bioscience, Beverly, Massachusetts 01915, [†††]Department of Biological Sciences, Illinois State University, Normal, Illinois 61790, [‡‡‡]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, [§§§]Department of Biology, Duke University, Durham, North Carolina 27708, [¶¶¶]Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas 77030, [&&&]Departamento de Genética, Centro de Ciências Biológicas, Universidade Federal de Pernambuco, CEP 50670-900, Recife/PE, Brazil ****FlyBase, The Biological Laboratories, Harvard University, Cambridge, Massachusetts 02138, [††††]Institute of Evolutionary Biology, Setagaya-ku, Tokyo 158-0098, Japan, [‡‡‡‡]Shiba Gakuen, Minato-ku, Tokyo 105-0011, Japan, [§§§§]Department of Biology, CUNY-Queens, Flushing, New York 11367 and [¶¶¶¶]Hokkaido University, EESBIO, Sapporo, Hokkaido 060-0810, Japan

ABSTRACT

The sequencing of the 12 genomes of members of the genus Drosophila was taken as an opportunity to reevaluate the genetic and physical maps for 11 of the species, in part to aid in the mapping of assembled scaffolds. Here, we present an overview of the importance of cytogenetic maps to Drosophila biology and to the concepts of chromosomal evolution. Physical and genetic markers were used to anchor the genome assembly scaffolds to the polytene chromosomal maps for each species. In addition, a computational approach was used to anchor smaller scaffolds on the basis of the analysis of syntenic blocks. We present the chromosomal map data from each of the 11 sequenced non-*Drosophila melanogaster* species as a series of sections. Each section reviews the history of the polytene chromosome maps for each species, presents the new polytene chromosome maps, and anchors the genomic scaffolds to the cytological maps using genetic and physical markers. The mapping data agree with Muller's idea that the majority of Drosophila genes are syntenic. Despite the conservation of genes within homologous chromosome arms across species, the karyotypes of these species have changed through the fusion of chromosomal arms followed by subsequent rearrangement events.

[1]Corresponding author: Department of Biology, Pennsylvania State University, 208 Erwin W. Mueller Laboratories, University Park, PA 16802-5301. E-mail: sws4@psu.edu

[2]Present address: Department of Genetics, NC State University, Raleigh, NC 27695.

ONE of the primary strengths of the genus Drosophila as a model system has been the relative ease of generating detailed cytogenetic maps. Indeed, the first definitive mapping of genes to chromosomes
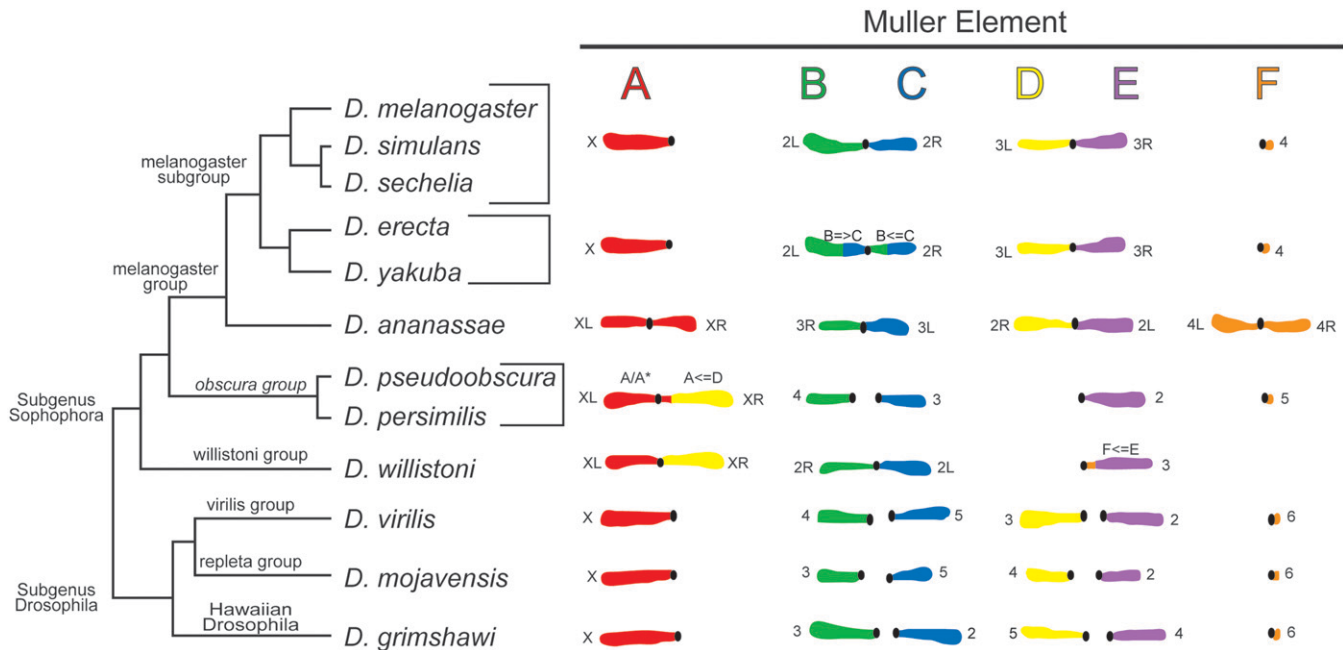
FIGURE 1.—Karyotypic and syntenic relationships of the 12 sequenced species of the genus Drosophila. (Left) The phylogenetic relationships of the 12 species. The members of the two main subgenera, Drosophila and Sophophora, are distinguished by the configuration of their autosomes. The ancestral pattern is shown by the subgenus Drosophila, which has all acrocentric chromosomes. The members of the subgenus Sophophora differ from this pattern in having varying numbers of fusions of these elements. (Right) The chromosomal arms are separated and aligned as Muller syntenic elements. Each element is differently colored and the arms are also designated by their conventional numbering. There is not a simple one-to-one correspondence for all of the species chromosome arms and a single Muller element. This lack of correspondence is associated with identified fusion and/or inversion events that reassociate all or portions of arms/elements in six of the species relative to *D. melanogaster*. The black dots designate the positions of the centromeres.

was performed in *Drosophila melanogaster* (BRIDGES 1916). The subsequent discovery of polytene chromosomes in the salivary glands in this same species (PAINTER 1934) and their codification into fine-structure genetic/cytogenetic maps represents perhaps one of the first forays into "genomics." Polytene maps (BRIDGES 1935; LEFEVRE 1976) provided an important genetic tool for mapping genes, for detecting genetic diversity within populations, and for inferring phylogenies among related species (DOBZHANSKY and STURTEVANT 1938; JUDD *et al.* 1972; ASHBURNER and LEMEUNIER 1976; LEMEUNIER and ASHBURNER 1976). STURTEVANT and TAN (1937) laid the groundwork for comparative genomics when they established that genes within the chromosomal arms are conserved or syntenic among species. In an insightful melding of the gene mapping and evolutionary studies, H. J. MULLER (1940) proposed that the genomes of Drosophila species were subdivided into a set of homologous elements represented by chromosome arms. What MULLER (1940) noted, which was subsequently elaborated on by STURTEVANT and NOVITSKI (1941), was that the presumed homologs of identified mutant alleles within a chromosome arm of *D. melanogaster* were also confined to a single arm in other species within the genus where mapping data were available. Using *D. melanogaster* as a reference, Muller

proposed that each of the five major chromosome arms plus the dot chromosome be given a letter designation (A–F) and that this nomenclature be used to identify equivalent linkage groups within the genus.

The ancestral organization of the Muller elements found in the subgenus Drosophila is six acrocentric rods (POWELL 1997), but a variety of rearrangement events have altered the organization of the Muller elements within the Sophophora subgenus (Figure 1). There is a pericentric inversion in *D. ananassae* on the X or A element that converts the normally acrocentric X into a metacentric chromosome. A further remarkable karyotypic change can also be seen on Muller element F of *D. ananassae*. The F element that is normally a small, dot-like chromosome is a large metacentric that is equivalent in size to the X or A element. The Sophophora species have also accumulated a variety of chromosomal fusions (Figure 1). The large autosomes of *D. melanogaster* are products of centromeric fusions between the B and C elements (symbolized as Muller B·C or chromosome 2L·2R) and D·E elements (3L·3R). The autosomes of *D. willistoni* also resulted from fusions of autosomal Muller elements (B·C and F·E; Figure 1) (PAPACEIT and JUAN 1998). The metacentric X of *D. willistoni, D. pseudoobscura,* and *D. persimilis* was generated via a fusion of A·D elements (the X and autosomal 3L arms of *D. melanogaster*)

that changed the transmission mode of an autosomal element to sex-linked inheritance. The absence of the fusion between A and D in *D. subobscura*, a close relative of *D. pseudoobscura* and *D. persimilis*, suggests that the *obscura* and *willistoni* group fusions are not homologous.

There is not, however, a perfect one-to-one correspondence between the Muller element and chromosome arm among all Drosophila species. Elements have rearranged following chromosome fusions. The specifics of these rearrangements are shown in Figure 1 and some of the highlights are presented here. In *D. erecta* and *D. yakuba*, there is a shared pericentric inversion at the base of the B·C element (2L·2R in *D. melanogaster*). In *D. pseudoobscura* and *D. persimilis*, an apparent pericentric inversion has moved genes from the Muller A to the Muller D element (Segarra *et al.* 1995). Comparative genomic analysis of the sequences involved in these rearrangements may provide valuable clues about the mechanisms driving the reorganization of these genomes.

The wide-ranging conservation of gene content within Muller elements is useful for assigning and organizing genomic scaffolds to specific chromosomal arms, while hybridization of DNA markers to polytene chromosomes is necessary to confirm the placement of assembled sequence on the chromosomes. While the polytene maps provide a valuable tool for orienting large scaffolds (>1 Mb), genome assemblies may yield many small scaffolds that are difficult to map in this manner. Thus, computational methods that take advantage of conservation of gene order among species have been developed to aid in the mapping of smaller scaffolds (Bhutkar *et al.* 2006). In contrast to the strong syntenic conservation in Drosophila, the order of genes along chromosome arms is poorly conserved due to the accumulation of inversions that shuffle gene order (Segarra and Aguadé 1992; Ranz *et al.* 2001). This shuffling of gene order provides strong inference that two scaffolds can also be joined when the genes at the terminal ends are adjacent in a conserved block of genes in the genomes of other species (Bhutkar *et al.* 2008).

The cytogenetic maps of the 11 Drosophila species provide a useful medium for presenting the mapped genome scaffolds using a web-based format such as GBrowse available through FlyBase Consortium (1999). The available cytogenetic maps of the 11 species, however, vary in quality and nomenclature. The *D. pseudoobscura* and *D. persimilis* chromosomes depicted in photographic maps contain several bends, making them less than ideal for web-based presentation of the data (Kastritsis and Crumpacker 1966; Moore and Taylor 1986). Two nomenclature problems exist in how the cytogenetic maps were divided into sections and subsections. The gold standard used in *D. melanogaster* divides the cytogenetic map into numbered sections and each section was subdivided into lettered subsections (Bridges 1935). In the non-*D. melanogaster* species, chromosomal maps have been divided into major sections, but not all sections have been divided into subsections. In all but one species, *D. mojavensis* (Wasserman 1992), numbered sections and lettered subsections were used. Lack of concordance among the cytogenetic maps indicated a need for revision of the maps and standardization of nomenclature where possible. The completion of genomic sequencing and whole-genome assembly of 11 Drosophila species provides an opportunity to update, organize, and synthesize the genetic and physical maps of each species and to standardize the quality, presentation, and nomenclature of cytogenetic maps from each. Assembled scaffolds were aligned to the polytene chromosomes of the newly sequenced species by evaluation of existing and newly obtained map data as well as computational comparisons among species, especially in relation to the organization of genes in the genome of *D. melanogaster.*

Mapping the assembly scaffolds in precise order and orientation may not seem important to Drosophilists interested in the molecular biology of their favorite gene. However, well-supported maps of assembly scaffolds will allow one to address questions about genome rearrangement, a virtual black box in evolutionary biology. Many different genome-level questions are especially suitable for investigation using species in the genus Drosophila.

First, what is the molecular basis of differences among species, including their inability to produce viable and fertile progeny? Genome sequences of closely related species pairs, such as *D. pseudoobscura* and *D. persimilis* and *D. simulans* and *D. sechelia*, will facilitate comprehensive genomic analyses of functional differences between genomes, whereas each of the other species represents a reference for generating genomewide mapping resources for studies within each group.

Second, what is the mechanistic basis for the origin of distinct features of the sex chromosomes and how are selfish versions such as the sex ratio variety of *D. pseudoobscura* and *D. persimilis* born? The X chromosome in the *obscura* and *willistoni* groups has independently acquired an entirely new arm. These genome sequences provide a substrate for determining changes associated with the transformation from autosome into X chromosome and for defining unique DNA signatures of this sex chromosome (Gallach *et al.* 2007; Sturgill *et al.* 2007). A close relative of *D. virilis* has similarly acquired a new arm of the X through a centromere fusion, and comparative genomic studies of this derived X chromosome in *D. americana* will provide insight into unique forms of selection that contribute to the initial divergence between sex chromosomes (McAllister and Evans 2006; Evans *et al.* 2007).

Third, how do new inversions originate? A dichotomy currently exists between inversions that appear to have arisen through ectopic exchange at repetitive sequences and those lacking any evidence of shared repeats at the

breakpoints (RANZ *et al.* 2007). The difference may indicate distinct pathways of origin, or alternatively, the mechanism of origin may be obscured by time. Analyses of the structure at the breakpoints of inversions relative to their age will distinguish between immediate causal mechanisms of origin and subsequent changes within these regions of the genome.

Fourth, what is the molecular basis of gene arrangement polymorphism? The species selected for sequencing and their close relatives exhibit a diverse array of chromosomal rearrangements consisting of both paracentric and pericentric inversions in addition to centromeric fusions between chromosomal elements (HSU 1952; STONE *et al.* 1960; VIEIRA *et al.* 1997a). These naturally occurring rearrangements are excellent substrates to determine the factors that influence change in chromosomal structure (VIEIRA *et al.* 2001; MCALLISTER 2002, 2003). Using the positions of known chromosomal rearrangements and comparative map data, these genome sequences provide a guide for inferring the positions of genes relative to other genome arrangements within these or closely related species (VIEIRA *et al.* 2006; EVANS *et al.* 2007).

Fifth, what is the basis for gene arrangement polymorphism on some chromosome arms and not others using *D. subobscura* and *D. pseudoobscura* as models? These questions are now more tractable, given the complete genome sequence available for two members of the *obscura* species group. These species still have their limitations in that few balancer and mutant strains exist. In addition, transformation systems to introduce mutations and shuttle genes into different strains are also limited, but are likely to be developed in the future.

Here, we present the scaffold maps for 11 species of Drosophila. The data in this article are presented in a series of sections. The MATERIALS AND METHODS describes the bioinformatic methods that were used to aid in the scaffold joining. The RESULTS is divided into a section based on the computational analysis and a section for each species that present (1) the history of the polytene maps; (2) the mapping data used to anchor the scaffolds; (3), the new polytene chromosomal maps; and (4) the problems that were discovered in the assembled sequences. The DISCUSSION presents new insights that emerged from the mapping of genome scaffolds as well as potential pitfalls of the whole-genome shotgun approach.

MATERIALS AND METHODS

Strains, polytene map preparation, and genome sequence: The descriptions of the genome strains for the Drosophila species sequenced for the Comparative Analysis Freeze 1 (CAF1) are described at the *Drosophila* Species Stock Center website (see http://www.flybase.org for the current link). Details on the preparation of polytene maps and the strategy

for anchoring scaffolds to the polytene chromosomes can be found in the supplemental Materials and Methods section. The CAF1 assembly scaffolds were mapped to the polytene chromosomes.
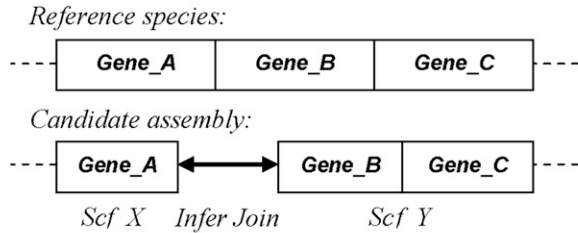
Computational support for anchoring scaffolds: Syntenic analysis of genome assembly data served as the basis of computational predictions for anchoring scaffolds to specific chromosome arms. CAF1-assembled sequences (http://rana.lbl.gov/drosophila/caf1.html) of all 11 genomes were analyzed using Synpipe (BHUTKAR *et al.* 2006), a computational tool for gene homology and synteny inference. Synpipe uses an annotated peptide set from a reference species, in this case the *D. melanogaster* Release 4.3 annotation (FLYBASE CONSORTIUM 1999), and infers an initial set of homology assignments in a candidate genome assembly. A tBLASTn-based approach (ALTSCHUL *et al.* 1997) was utilized for this. Following this, synteny chains consisting of neighboring genes (in the same order as in the reference species) are inferred. Initial homology assignments are then refined with the objective of maximizing the size of syntenic chains in the presence of paralogs. Merging adjacent synteny chains to allow for localized gene order scrambling (presumably due to micro-inversions on the chromosome) results in expanded synteny chains. This is based on various user-defined thresholds for allowable scrambling, with a default of up to 10 genes whose order can be locally scrambled. This approach accommodates contig and scaffold gaps in the candidate assembly by tagging homologs that might lie in unsequenced assembly gaps, on the edges of scaffolds or contigs, or on small assembly fragments. It also provides adjacent gene-pair information between species, which has been shown to be concordant with phylogenetic relationships (BHUTKAR *et al.* 2007a). The output of this algorithm, including gene-pair information, provides a data set that can be used for comparative analysis of syntenic blocks and boundaries to infer chromosomal rearrangements between species (BHUTKAR *et al.* 2008), to refine multi-species alignments and orthologous gene calls, to identify probable assembly errors, and to anchor scaffolds along chromosome arms in the proper place and orientation.

As has been noted, in the genus Drosophila orthologous genes are localized on the same Muller elements across species (STURTEVANT and TAN 1937; STURTEVANT and NOVITSKI 1941; RICHARDS *et al.* 2005), with the exception of a relatively few cases of gene movement across arms (GONZALEZ *et al.* 2004; SCHULZE *et al.* 2006; BHUTKAR *et al.* 2007b). The assignment of homologs across scaffolds then lends itself to mapping scaffolds to specific Muller elements on the basis of the location of the majority of genes within the genome of *D. melanogaster*. Once scaffolds have been assigned to Muller elements in this manner, the order and orientation of various assembly scaffolds along chromosome arms can be predicted using orthologous markers placed on the edges of scaffolds in the context of synteny relationships across species. To predict the adjacency of two scaffolds (a "scaffold join"), the following cases were analyzed: (a) conserved synteny, (b) conserved synteny with assembly gaps, (c) rearrangement supported by adjacent species, (d) rearrangement supported by an inferred ancestral arrangement, and (e) species-specific rearrangement with trace-back. A visual representation of these five inferences is shown in Figure 2.

If one of these cases supported the adjacency requirement for two scaffolds assigned to the same Muller element, a "scaffold join" was predicted. On the basis of the edges that were joined, the scaffolds were mutually oriented in a superscaffold with respect to their assembly orientation. Superscaffold edges were further considered for potential joins with other scaffolds or superscaffolds. This recursive procedure continues until no further predictions can be made on the basis of syntenic block data. Various levels of confidence can be

A Conserved Synteny (*S,LS*)

Reference species:

- - - | Gene_A | Gene_B | Gene_C | - - -

Candidate assembly:

- - - | Gene_A | ←→ | Gene_B | Gene_C | - - -

Scf X   *Infer Join*   Scf Y

B Conserved Synteny with assembly gaps (*SG*)

Reference species:

- - - | Gene_A | Gene_B | Gene_C | - - -

Candidate assembly:

- - - | Gene_A | ←Gene_B in assembly gap→ | Gene_C | - - -

Scf X   *Infer Join*   Scf Y

C Rearrangement supported by closely related species

Reference species:

- - - | Gene_A | Gene_B | Gene_C | - - -

Closely related species:

- - - | Gene_A | Gene_R | Gene_S | - - -

Candidate species:

- - - | Gene_A | ←→ | Gene_R | Gene_S | - - -

Scf X   *Infer Join*   Scf Y
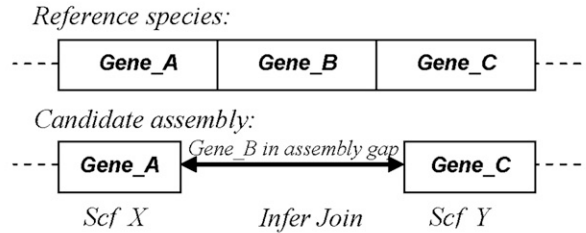
D Rearrangement supported by inferred ancestral arrangement [no support from closely related species]

Reference species:

- - - | Gene_A | Gene_B | Gene_C | - - -

One or more distantly related species:

- - - | Gene_A | Gene_R | Gene_S | - - -

Candidate species:

- - - | Gene_A | ←→ | Gene_R | Gene_S | - - -

Scf X   *Infer Join*   Scf Y
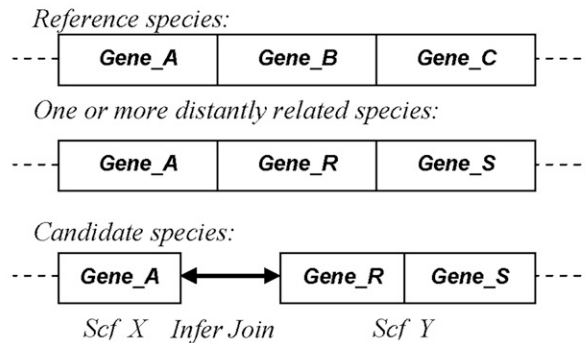
E Species-specific rearrangement with traceback [no support from other species] (*B*)

Reference species or closely related species:

- - - | Gene_A | Gene_B(+) | Gene_C(+) | Gene_D | Gene_E | Gene_F | Gene_G | - - -

Traceback (2 inversions):

- - - | Gene_A | *assembly gap* | Gene_R | Gene_S | Gene_T | - - - | Gene_B(+) | Gene_C(+) | - - -
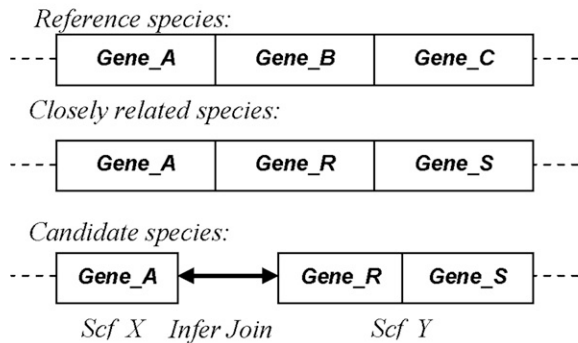
- - - | Gene_A | *assembly gap* | Gene_C(-) | Gene_B(-) | Gene_T | Gene_S | Gene_R | - - -

- - - | Gene_A | *assembly gap* | Gene_B(+) | Gene_C(+) | Gene_T | Gene_S | Gene_R | - - -

Candidate species:

- - - | Gene_A | ←→ | Gene_B(+) | Gene_C(+) | Gene_T | Gene_S | Gene_R | - - -

Scf X   *Infer Join*   Scf Y
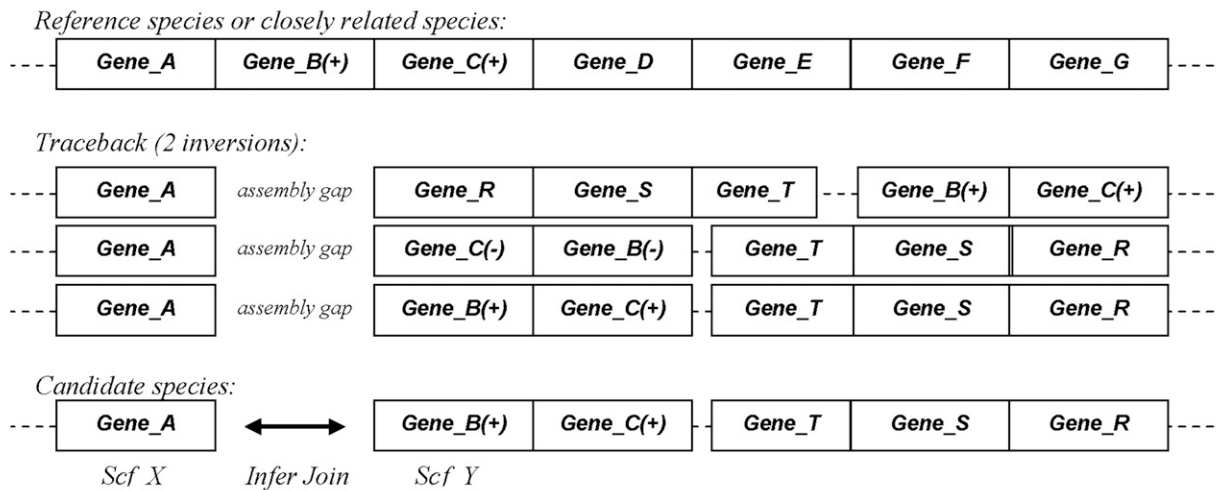
F Example from *D. virilis* Muller A

| S | | S | | S | | S | | S | | LS | | LS | | S | |

| - | | + | | + | | - | | ⊞ | | - | | - | | - | |

| s_13042 (-) | s_12928 (+) | s_12932 (+) | s_12726 (-) | s_12472 (?) | s_12970 (-) | s_12799 | s_12528 | s_13036 |
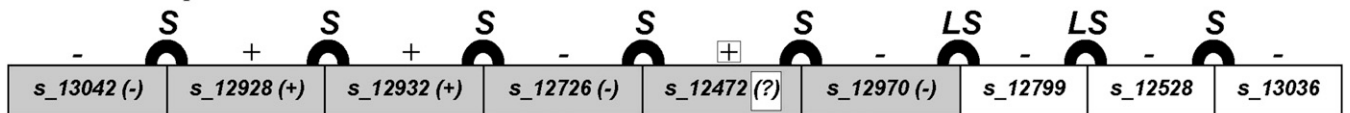
G Example from *D. persimilis* Muller D

| | SG | | S | | S | | S | | S | | B | | SG | | S | |
| + | | + | | + | | - | | + | | + | | + | | + | | - |

| s_12 (+) | s_9 (+) | s_257 (+) | s_294 (-) | s_47 (+) | s_260 (+) | s_48 (+) | s_40 (+) | s_568 (-) |
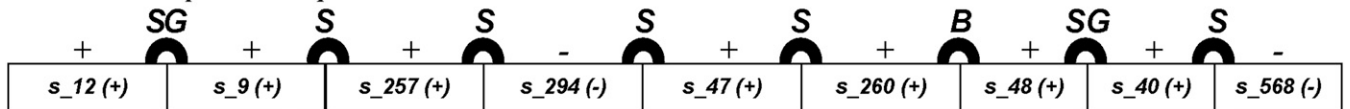
assigned to these predictions on the basis of the level of evidence. For example, joins made in cases where straightforward synteny via sequential gene order was used are assigned a higher confidence than joins made using species-specific breakpoints. At the completion of this process, scaffolds are bucketed either into blocks of superscaffolds, ideally one per chromosome arm with orientation information, or into a bucket of unconnected scaffolds that have no evidence to support their joins.

RESULTS

Assignment of scaffolds to Muller elements and placement in ordered and oriented superscaffold blocks was inferred for each of the 11 species. Table 1 lists the number and lengths of scaffolds assigned to each Muller element of each species. In addition, Table 1 also shows this information for scaffolds whose location and orientation was verified with physical and genetic mapping data. Additionally, syntenic block information was also used to assign assembly scaffolds to heterochromatic regions of the genome based on majority hits from *D. melanogaster* heterochromatic genes. Table 2 summarizes the scaffold assignment data for euchromatic, heterochromatic, and unmapped scaffolds. An important observation is that the amount of DNA and genes in the scaffolds assigned to a Muller element is similar among the 11 species (109.7–153.1 Mb), but the unassigned fraction is more variable among the species (14.7–31.7).

The unassigned scaffolds are expected to be heterochromatic and 90% of these scaffolds are <100 kb in length (see supplemental text). The results of inferring scaffold joins to form superscaffolds (scaffold names, GenBank identifiers, scaffold order, and orientation information) for each species are included as supplemental Tables 2–14. The summary of superscaffold order for the 11 Drosophila species based on the computational- and marker-based analyses are shown in supplemental Table 15. Figure 2 shows an example of scaffold order inferred along one chromosome arm, the kind of evidence used to predict the order and orientation, and how these computational predictions complemented the experimental analysis process.

Inferring the order and orientation of scaffolds allowed the creation of cross-species synteny maps for full-length Muller element comparisons. These maps are based on the annotated protein set of a reference species. One set of synteny relationships was derived using Release 4.3 of the well-annotated *D. melanogaster* peptide set (FLYBASE CONSORTIUM 1999). The prediction of scaffold order information and consensus annotation sets (DROSOPHILA 12 GENOMES CONSORTIUM 2007) allowed the use of other species as reference sets, now that their Muller element-wide gene order could be predicted. Orthology and syntenic block information was derived using Synpipe while allowing for assembly gaps and localized scrambling. Localized gene-

FIGURE 2.—Computational support for anchoring scaffolds. (A) Scaffold joins inferred using conserved synteny. Joins supported by conserved synteny (*S*) and localized gene scrambling with conserved synteny for the larger context (*LS*, not shown) are included here. Boxes represent adjacent genes (in order on a chromosome in the reference species or on one or more scaffolds in a candidate assembly). (B) Scaffold joins inferred using conserved synteny and intervening assembly gaps (*SG*). Gene_B is inferred, by Synpipe, to lie in an assembly gap in the candidate genome assembly. (C) Scaffold joins inferred using rearrangements supported by closely related species. The reference species is not informative in this case. Closely related species inform the joining process if they do not have a scaffold break in their assembly around the same position. This scenario could also be complicated by intermediate markers missing due to an unsequenced assembly gap. (D) Scaffold joins inferred using rearrangements supported by an inferred ancestral arrangement. In such cases, there is no support from closely related species or the reference species. One or more distantly related species allow inference of a conserved ancestral arrangement (especially when they flank the ancestral node in a given phylogeny). In such cases, there is no support from closely related species or the reference species. Such joins are considered to be of lower confidence compared to earlier cases. (E) Scaffold joins inferred using species-specific rearrangement with trace-back (*B*). In such cases, there is lack of direct support from any species in the phylogeny. Through a series of inversions (two in this case), the gene order of a candidate genome assembly, with an assumed scaffold join, could match that of a reference species (or closely related species). Such a scaffold join can then be inferred and the gene order could be interpreted as the result of a species-specific rearrangement. (F and G) Boxes denote scaffolds anchored on a chromosome with relative order and orientation. Scaffolds are identified with their CAF1 scaffold identifiers (DROSOPHILA 12 GENOMES CONSORTIUM 2007). Shaded boxes highlight scaffolds that have experimental support for their placement and orientation (on the basis of the *in situ* localization of markers). Arcs connecting two scaffolds show that these scaffold joins were supported by predictions based on analysis of synteny and rearrangement data. Each arc is identified by abbreviations (*S, LS, SG, B*) identified earlier. Scaffold orientation predicted by synteny analysis is listed above each scaffold. Scaffolds are not drawn to scale. (F) *D. virilis* Muller element A inference. Synteny-based predictions supplement experimental data by helping orient scaffold s_12472 for which there was a single marker. There is complete agreement between these predictions and experimental analysis. Predictions are based on conserved synteny (*S*) with localized scrambling at the edges of scaffolds in two cases (*LS*). The three unshaded scaffolds to the right (s_12799–s_13036) are placed using computational support. These are likely in the pericentromeric heterochromatin. Consistent with this hypothesis, each contains between 35 and 45% of transposable element content in contrast with 3–5% for similarly sized euchromatin scaffolds (B. MCALLISTER, personal communication). (G) *D. persimilis* Muller element D inference. A short fragment of this chromosome is shown here (a total of 46 joins are predicted for this arm). Synpipe analysis provided the primary evidence for inferring scaffold joins for this fragmented assembly. These predictions illustrate joins based on various criteria, including species-specific breakpoints (*B*) at the edges of scaffolds and assembly gaps between scaffolds (*SG*). These predictions agree with the gene order established using the extensive set of known experimental markers in *D. pseudoobscura* (the closest available genome to *D. persimilis*).

TABLE 1

Assembly scaffolds mapped to Muller elements in 11 Drosophila species

Species	Assembly coverage	Mapping status	Muller A	Muller B	Muller C	Muller D	Muller E	Muller F
D. simulans	1×, 2.8×[a]	Mapped[b]	1 (17.0)	1 (22.0)	1 (19.6)	1 (22.6)	1 (27.5)	1 (1.0)
		Assigned[c]	353 (2.4)	46 (0.3)	129 (0.6)	53 (0.2)	71 (0.3)	7 (0.0)
D. sechellia	4.9×	Mapped	22[d] (20.9)	10[d] (22.2)	5[d] (19.6)	7[d] (22.8)	15[d] (27.8)	3 (1.2)
		Assigned	20[d] (0.3)	9[d] (0.1)	14[d] (0.6)	6[d] (0.1)	9[d] (0.1)	0 (0.0)
D. erecta	11×	Mapped	2 (21.3)	1 (26.6)	1 (22.6)	1 (25.8)	2 (28.2)	1 (1.3)
		Assigned	0 (0.0)	6 (0.1)	2 (0.0)	1 (0.0)	1 (0.0)	1 (0.2)
D. yakuba	9.1×	Mapped	1 (21.8)	1 (22.3)	1 (21.1)	1 (24.2)	1 (28.8)	1 (1.4)
		Assigned	10 (0.20)	5 (0.2)	7 (2.9)	75 (2.4)	22 (0.7)	2 (0.1)
D. ananassae	9.6×	Mapped	13 (31.5)	5 (22.6)	1(19.9)	1 (23.3)	4 (33.2)	0 (0.0)
		Assigned	16 (7.0)	8 (0.1)	4 (2.1)	5 (0.1)	1 (0.0)	16 (17.8)
D. pseudoobscura	7×	Mapped	3[d] (20.3)	11 (27.5)	5 (19.9)	7[d] (30.5)	1 (30.8)	0 (0.0)
		Assigned	55 (1.3)	22 (0.4)	18 (0.5)	20 (0.4)	18 (0.4)	25 (1.2)
D. persimilis	4.1×	Mapped	21 (21.1)	12 (28.4)	11[d] (19.9)	52 (29.2)	5 (31.7)	0 (0.0)
		Assigned	55 (5.3)	17 (0.9)	9 (0.9)	32 (0.3)	6 (0.1)	5 (1.5)
D. willistoni	8.4×	Mapped	5 (27.9)	8 (32.3)	8[d] (29.6)	8[d] (29.6)	3 (33.7)	FC
		Assigned	4 (0.1)	20 (14.9)	4 (1.6)	3 (2.4)	10 (0.4)	FC
D. virilis	8.7×	Mapped	6 (30.5)	3 (28.7)	4 (27.3)	4 (26.7)	4 (35.5)	1 (2.0)
		Assigned	9 (5.4)	10 (1.8)	5 (0.0)	8 (0.1)	8 (0.7)	1 (0.1)
D. mojavensis	8.6×	Mapped	5 (32.0)	1 (32.4)	1 (26.9)	2 (27.3)	1 (34.1)	0 (0.0)
		Assigned	5 (2.6)	10 (0.2)	6 (0.1)	6 (0.0)	9 (0.1)	1 (3.4)
D. grimshawi	7.9×	Mapped	3 (26.4)	4 (26.7)	3 (23.5)	1 (24.6)	5 (34.2)	0 (0.0)
		Assigned	14 (0.1)	47 (0.8)	29 (0.6)	34 (1.2)	24 (0.5)	4 (1.3)

The number of genome assembly scaffolds assigned to each Muller element in a species is shown. Scaffold lists and sizes are detailed in the supplemental tables. FC, Muller F in *D. willistoni* has fused to Muller E. Numbers in parentheses indicate size in megabases.

[a] *D. simulans* mosaic assembly.

[b] A scaffold was mapped if genetic or physical mapping data confirmed the assignment of the scaffold to the Muller element.

[c] A scaffold was assigned to a Muller element on the basis of the majority hits from orthologous *D. melanogaster* genes using Synpipe, but the location and orientation was not verified with mapping data.

[d] These assignments include corrections for assembly mis-joins uncovered by Synpipe. They are detailed in the supplemental tables and text.

order scrambling, or microsynteny, was permitted within specified thresholds. Gene duplication was not addressed in this analysis and the best syntenic placement was specified as the ortholog for a given gene from the reference species. In accordance with the phylogeny, synteny maps using either *D. melanogaster* or *D. virilis* as reference sets show longer conserved blocks in more closely related species than in more distantly related species (BHUTKAR *et al.* 2008).

SUBGENUS SOPHOPHORA

D. melanogaster subgroup species maps

The sequenced species most closely related to *D. melanogaster* (*D. simulans*, *D. sechellia*, *D. erecta*, and *D. yakuba*) belong to the *melanogaster* group. All five are nearly identical morphologically and have karyotypes similar to *D. melanogaster*, consisting of an acrocentric rod X chromosome, two large metacentric autosomes (2L·2R and 3L·3R), and a small dot fourth chromosome. Males have a "J-shaped," entirely heterochromatic Y

chromosome. In polytene chromosome preparations, there are five long banded euchromatic elements corresponding to the X and each of the major autosomal arms. The centromeric portion of each chromosome is embedded in the chromocenter, which contains the pericentric heterochromatin. Relative to the banded euchromatin, this heterochromatic compartment of the genome is underreplicated. The fourth chromosome can be seen as a small, banded nub usually associated with the chromocenter. The heterochromatic Y is not banded and lies entirely within the chromocentral region of the nucleus. Each of the six banded euchromatic arms corresponds to each of the Muller elements that are used to define the syntenic relationships among the species of the genus (X, A; 2L, B; 2R, C; 3L, D; 3R, E; 4, F) (MULLER 1940; STURTEVANT and NOVITSKI 1941).

Perhaps one of the earliest forays into the genomics of the genus was made possible by the development of the polytene chromosome maps of *D. melanogaster* by BRIDGES (1935). He proposed an alphanumeric nomenclature by dividing each arm into 20 numbered units and each of these into the six subunits A–F and by

TABLE 2

Summary of assignment of assembly scaffolds to various regions of genomes in various species

Species	Euchromatic mapped[a]		Euchromatic assigned[b]		Heterochromatic assigned[c]		Unassigned[d]		Total		
	n	Size in Mb (%)	n	Size in Mb (%)	n	Size in Mb (%)	n	Size in Mb (%)	n	Size in Mb	Flow cytometry[e]
D. simulans	6	109.7 (79.6)	665	113.7 (82.5)	60	3.29 (2.4)	9,280	20.9 (15.2)	10,005	137.8	162 ± 2
D. sechellia	66	113.7 (68.4)	120	115.6 (69.4)	59	4.76 (2.9)	14,555	46.1 (27.7)	14,734	166.5	171 ± 4
D. erecta	8	125.8 (82.4)	19	126.1 (82.5)	14	1.34 (0.9)	5,091	25.3 (16.5)	5,124	152.7	135 ± 4
D. yakuba	6	119.6 (71.1)	127	126.1 (75.0)	32	9.33 (5.5)	8,182	32.6 (19.4)	8,341	168.2	190 ± 11
D. ananassae	24	130.5 (56.5)	74	157.6 (68.2)	12	7.84 (3.4)	13,663	65.5 (28.4)	13,749	231.0	217 ± 9
D. pseudoobscura	27	127.6 (82.8)	185	131.8 (85.5)	21	0.57 (0.4)	4,692	20.3 (13.2)	4,898	154.1	193 ± 4
D. persimilis	101	130.2 (69.1)	225	139.2 (73.9)	13	0.79 (0.4)	12,600	48.2 (25.6)	12,838	188.4	193 ± 5
D. willistoni	32	153.1 (64.7)	73	172.5 (72.9)	10	0.73 (0.3)	14,845	63.3 (26.8)	14,928	236.6	222 ± 7
D. virilis	22	150.7 (73.2)	63	158.8 (77.1)	4	1.19 (0.6)	13,463	45.9 (22.3)	13,530	206.0	364 ± 13
D. mojavensis	10	152.7 (78.8)	47	159.1 (82.1)	9	0.48 (0.2)	6,785	34.1 (17.6)	6,841	193.8	130 ± 0
D. grimshawi	16	135.4 (67.5)	168	139.9 (69.8)	5	0.03 (0.0)	17,267	60.4 (30.1)	17,440	200.5	231 ± 5

Scaffold lists and sizes are detailed in supplemental tables.

[a] Euchromatic scaffolds assigned and mapped to a Muller element.

[b] Euchromatic scaffolds assigned to a Muller element using Synpipe. This also includes the mapped scaffolds.

[c] Scaffolds assigned to heterochromatic regions of Muller elements on the basis of the majority hits from orthologous D. melanogaster heterochromatic genes, using Synpipe.

[d] Scaffolds that have not been assigned or mapped to a Muller element in the species.

[e] Megabases of DNA determined from flow cytometry (Bosco et al. 2007).

assigning a number designation for each of the bands therein. This codification of the chromosome map allowed the designation of chromosomal aberration breakpoints and the localizations of genes associated with these breaks. While these maps were useful and the standard for many years, they were augmented by the production of mosaic photographic maps by LEFEVRE (1976) and further refined by the electron microscope maps of SORSA (1988a,b,c).

The four sequenced *melanogaster* group species are so similar to *D. melanogaster* that it has been possible to use the aforementioned maps in mapping efforts in these species. Indeed, the mapping of several genes in *D. simulans* by *in situ* hybridization has used the BRIDGES (1935) *D. melanogaster* nomenclature to designate the location of cloned sequences from *D. simulans* (FLYBASE CONSORTIUM 1999 and references cited therein). Moreover, the mapping of inverted sequences that serve to distinguish the gene order in the four species has utilized the BRIDGES (1935) maps and nomenclature. In some cases, the position of the paracentric inversion breakpoints can be observed directly by virtue of the fact that viable, albeit sterile, hybrid progeny can be produced with interspecific mating, *e.g.*, between *D. melanogaster* and *D. simulans* (HORTON 1939; LEMEUNIER and ASHBURNER 1976). In cases where these hybrids cannot be generated, the similarity in the chromosome-banding patterns has allowed the mapping of inversion breakpoints by side-by-side comparisons of photographic maps of the different species (LEMEUNIER and ASHBURNER 1976, 1984). The traditional use of the BRIDGES' (1935) maps and nomenclature in the comparative cytology of these species dictated the use of the combined Bridges ideogram and Lefevre photographic maps for the alignment of the scaffolds to the chromosomal maps. Additionally, the large number of clear orthologies among the species with *D. melanogaster*, the conserved synteny and the accurate assignment of genes to the polytene map in *D. melanogaster*, coupled with the similarities in karyotype and banding pattern, allowed a fairly robust alignment of the assembled scaffolds with the chromosomal maps. This was even the case for *D. simulans* and *D. sechellia*, which were sequenced to lower coverage than the other species and thus had more problematic and fragmented assemblies, respectively. Indeed, the excellent alignment of scaffold to chromosome enabled the discovery of apparent assembly errors that will have to be addressed in future analyses on these species.

Chromosome maps—*D. simulans*: The alignment of *D. melanogaster* and *D. simulans* has revealed some problems with the mosaic *D. simulans* assembly (BEGUN *et al.* 2007) and these will be pointed out below in a presentation of each arm/element alignment. There are also 665 small scaffolds that we have not attempted to align to the chromosomes. However, because these contain *D. melanogaster* orthologs, we have provided a

tentative linkage call on the basis of a possible syntenic conservation between *D. simulans* and *D. melanogaster*. These assignments can be found in the supplemental tables.

Muller element A (chromosome X, sections 1–20): As noted above, there is a single large scaffold associated with the X chromosome (Figure 3). This scaffold is 17.0 Mb in length and contains 1360 called orthologs. Using these and their *D. melanogaster* cytological assignments, it is possible to align the scaffold from a site near the telomere (1A) to the base (20F) of the X. As a cross-check on this alignment, we extracted from FlyBase all of the *D. simulans* genes that had been cytologically localized. Within this group, those that were also found in the orthology calls for the X were identified and their cytology correlated with their position in the scaffold. This process resulted in 11 additional points of alignment that were all consistent with the computational analysis (supplemental Table 16).

Muller element B (chromosome 2L, sections 21–40): There is a single large scaffold of 22.0 Mb assigned to B/2L with 2160 called orthologs. Correlation of the molecular and inferred cytological maps allows the alignment of the scaffold from near the telomere (21A) to the base (40F) of this arm (Figure 3). Again, the cytologically localized genes were extracted from FlyBase and those represented in the ortholog set were identified. Sixteen of these allowed further confirmation of the scaffold alignment (Figure 3, supplemental Table 16).

Muller element C (chromosome 2R, sections 41–60): A single large scaffold of 19.6 Mb is associated with C/2R. This scaffold contains 2328 called orthologs. Alignment of the molecular and cytological maps shows that the scaffold extends from the base (41E) to the telomere (60F) of this element (Figure 3). The alignment also demonstrates a problem in the assembly. At a position approximately in the middle of the scaffold, there is an ~350-kb sequence, which by cytological mapping should reside near the telomere in 3L/D (solid boxes in Figure 3). Because hybrids of *D. simulans* and *D. melanogaster* have been examined cytologically, a transposition of this size would have been noted. Thus, it is likely that the insertion of these sequences into the 2R/C scaffold represents an assembly error. Eleven cytologically localized *D. melanogaster* orthologs were identified and examined. These again are consistent with the computational alignment (Figure 3, supplemental Table 16).

Muller element D (chromosome 3L, sections 61–80): A single large scaffold of 22.6 Mb is associated with D/3L. Using the 2289 called orthologs and their positions in the *D. melanogaster* genome it is possible to align the scaffold from near the telomere (61A) to the base (80F) of this arm (Figure 3). We again extracted from FlyBase the genes that had been previously localized to the D/3L arm and found those that had called orthologs. The locations of this group of 19 were consistent with and
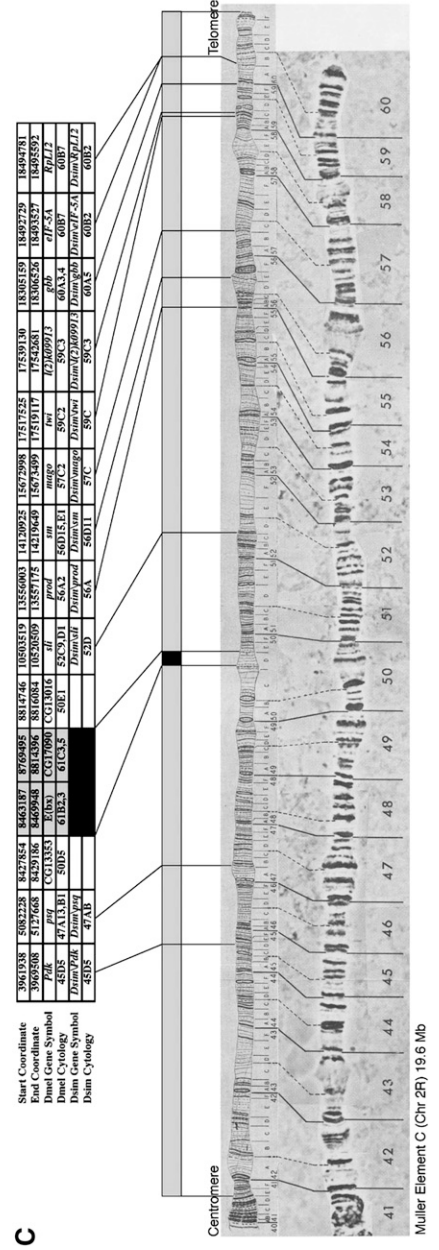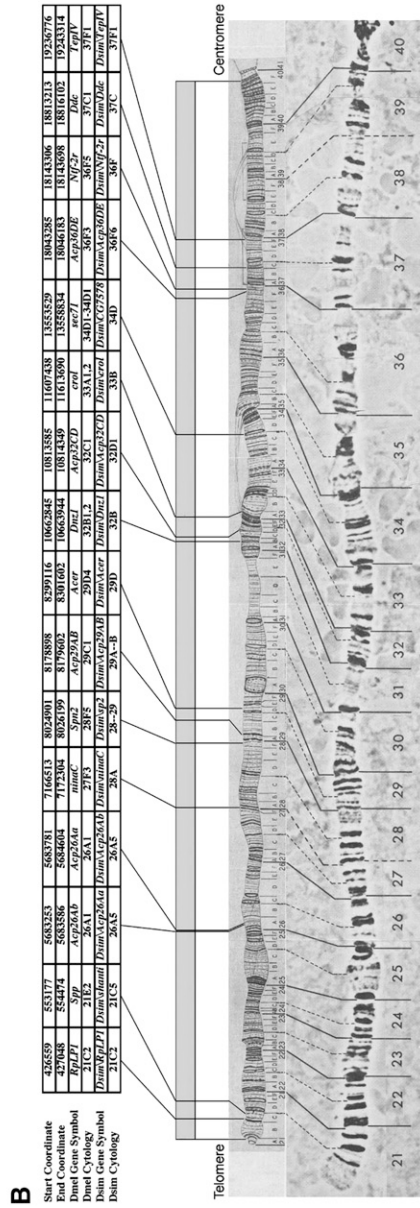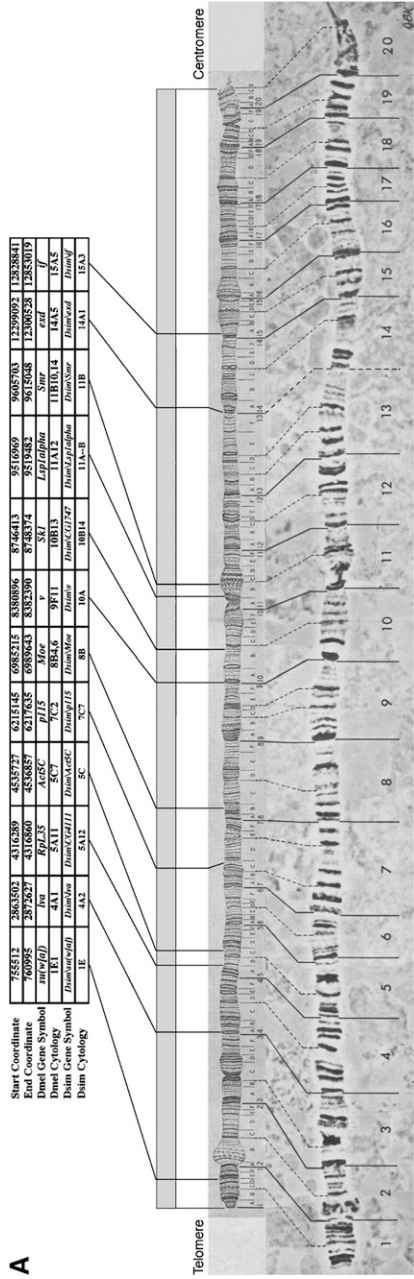
confirmed the computational alignment (Figure 3, supplemental Table 16). The clear gap in the shaded box above the chromosome in Figure 3 indicates the position of the material assembled into the C/2R scaffold.

Muller element E (chromosome 3R, sections 81–100): A single large 27.5-Mb scaffold is associated with D/3R, which contains 2921 called orthologs. The alignment of this scaffold's molecular map and the associated *D. melanogaster* cytological map positions the scaffold from near the base (81F) at one end to near the telomere (100E) at the other (Figure 3). Cytological examination of hybrids between *D. simulans* and *D. melanogaster* has shown that these two species differ in this element by a large paracentric inversion with breakpoints in 84F1 and 93F6-7 (Horton 1939; Lemeunier and Ashburner 1976; Ranz *et al.* 2007). The computational alignment of scaffold to chromosome clearly reveals the presence of the inversion and confirms the cytologically determined breakpoints (Figure 3). The previously localized genes contained within the called ortholog set retrieved from FlyBase again were consistent with and confirmed the computational alignment (Figure 3, supplemental Table 16).

Muller element F (chromosome 4, section 101–102): The small fourth chromosome has a single scaffold of 1.0 Mb mapped to it (Figure 3). There are 59 called orthologs and these allow an alignment that extends from near the base (102A) to the telomere (102F) of the chromosome. There are apparent problems with the assembly of this scaffold in that two genes, *Dyrk3* and *Mitf*, which map near the telomere in *D. melanogaster* and in the other three *melanogaster* group species, are placed in the proximal third of the *D. simulans* assembly. There are also fewer orthology calls in this species for this chromosome as compared to the other *melanogaster* group species. Inspection of the regions where these should be found indicates that there are several gaps in the assembly. This fact and the problem noted above with respect to the C/2R and D/3L assemblies indicates that the assemblies in *D. simulans* should be used with caution.

Chromosome maps—*D. sechellia*: Due to the low level coverage, the *D. sechellia* assembly is the most fragmented of the *melanogaster* group species. Additionally, there are a large number of short scaffolds that contain only one or two orthologs of *D. melanogaster* genes. A list of these can be found in supplemental Table 17. A comparison of those lists reveals that many of these short scaffolds are apparent duplicates of small portions of the longer ones that we were able to align with the chromosomes.

Muller element A (chromosome X, sections 1–20): The X chromosome can be aligned to 22 scaffolds ranging in size from 139 kb to 3.3 Mb. The smallest number of called orthologs in a scaffold is 7 and the largest 329. The total coverage of the aligned scaffolds is 20.9 Mb containing 1936 called orthologs. The aligned scaffolds cover the X from near the telomere (1A) to near the base (20C/D) of the chromosome with at least five significant gaps (Figure

A

Start Coordinate	755512	2363502	4316250	4535727	6215145	6985215	8380896	8746413	9516069	12399092	13828841
End Coordinate	760995	2872627	4316860	4536857	6217635	6980643	8382390	8743374	9219482	12300258	12853019
Dmel Gene Symbol	su(w[a])	Ura	Rpl35	AcSC	pl15	Mor		Snr	erd		
Dmel Cytology	1E	4I.1	5A.11	5C7	7C2	8B4.6	10A	9F11	11A–B	Lsp1alpha	Spr
Dsim Gene Symbol	DsimVsu(w[a])	DsimVau	DsimVCsd111	DsimVAcSC	DsimVl15	DsimVMor	DsimVv	DsimVCG1747	DsimVSnr	11B13.14	14A.5
Dsim Cytology	1E	4A2	5A12	5C	7C7	8B		10B14	11B	DsimVerd	DsimVf
										14A1	15A3

Muller Element A (Chr X) 17 Mb

B

Start Coordinate	426659	5683253	5683781	7166513	8024901	8178898	8399116	10625845	10813585	11607438	13553529	18043285	18143306	18813213
End Coordinate	427048	5684586	5684604	7172304	8026199	8179602	8301602	10663944	10814349	11613690	13553834	18046183	18143698	18816102
Dmel Gene Symbol	RpLP1	Spp	Acp26Ab	Acp26Aa	nimC	Spn2	Acp29AB	Acer	Dnc1	Acp32CD	sec71	Acp36DE	Nf2r	Ddc
Dmel Cytology	21C2	21E2	26A1	26A1	27F3	28F5	29C1	29D4	32B1.2	33A1.2	34D1–34D1	36F5	36F5	37F1
Dsim Gene Symbol	DsimVRpLP1	DsimVahand1	DsimVAcp26Ab	DsimVAcp26Aa	DsimVnimC	DsimVspn2	DsimVAcp29AB	DsimVAcer	DsimVDnc1	DsimVAcp32CD	DsimVCG3578	DsimVAcp36DE	DsimVNf-2r	DsimVDdc
Dsim Cytology	21C2	21C5	26A5	28A	28–29	29A–B	29D	32B	32D1	33B	34D	36F6	36F	37C

19236776		19243314
TepIV		TepIV
37F1		37F1
DsimVTepIV		
37C		

Muller Element B (Chr 2L) 22 Mb

C

Start Coordinate	3961938	5082228	8427854	8463187	8769495	8814746	10503519	13556003	14120925	15672998	17539130	18305159	18492729	18494781
End Coordinate	3905508	5127668	8429186	8469948	8760940	8814084	10520809	13557175	14219649	17519117	17543681	18306526	18493527	18495592
Dmel Gene Symbol	Pdk	Pal	CG13353	CG17090	Pu	tal	prod	mago	Pak	pbb	dl-7-SA	eb1	RpL12	
Dmel Cytology	45D5	47A13.B1	50D5	61B2.3	61C3.5	56A2	52D1	56A	57C	59C2	59C3	60A3.4	60B1	60B2
Dsim Gene Symbol	DsimVPdk	DsimVpq		DsimVprod	DsimVvago	DsimVaft	DsimVdl	DsimVprod		DsimVmago	DsimVvn	DsimVgbb	DsimVdl7-SA	DsimVRpL12
Dsim Cytology	45D5	47A.B	50D11		52D	56A		57C	59C	59C3	60A5	60B2	60B2	

Muller Element C (Chr 2R) 19.6 Mb

FIGURE 3.—Alignments of the *D. simulans* sequence scaffolds with the polytene chromosome maps of *D. melanogaster*. A single major scaffold covers each of the chromosome arms. A listing of all of the identified orthologs, their coordinates in the scaffolds, and their cytological ordering can be seen in supplemental Table 16 along with the smaller scaffolds that were not aligned with the chromosomes. With one exception, the scaffolds align well with the chromosomes and extend for essentially the entire length of each arm. The primary exception to this is seen in Muller element C (2R) where a segment of element D (3L) (solid area) is inserted into the C (2R) scaffold. We interpret this insertion as a mis-assembly of the scaffold. As a check on the alignment, several genes that were previously localized by *in situ* hybridization in *D. simulans* and reported in FlyBase were identified in the list of orthology calls. These genes are listed in the tables above the chromosome maps and their location in the scaffolds and chromosomes is indicated by a line. Both the called cytology and that predicted by the position of their *D. melanogaster* orthologs are also given in the table. At the top of the table are the starting and ending coordinates in the CAF1 assemblies for the gene listed below. In Muller element E (3R), there is a well-documented paracentric inversion that distinguishes this arm from that of *D. melanogaster*. The inversion is clearly evident in the E (3R) scaffold with breakpoints at 84F9 and 93F6-7. This molecular reordering is nicely consistent with the observed cytology in *D. melanogaster/D. simulans* hybrids. There is, however, no evidence for a set of small inversions reported at the telomeric end of A (X), the base of C (2R), and the base of E (3R). The ideograms are from the drawings by BRIDGES (1935, Figure 3) and are reproduced with kind permission from the American Genetics Association and Oxford University Press. The photographic maps are from LEFEVRE (1976, Figures 21–23).

A Drosophila sechellia

Muller Element A (Chr X) 20.9 Mb

B Muller Element B (Chr 2L) 23.9 Mb

C Muller Element C (Chr 2R) 19.6 Mb

D Drosophila sechellia

Muller Element D (Chr 3L) 22.7 Mb

Muller Element F (Chr 4) 1.2 Mb

Muller Element E (Chr 3R) 27.7 Mb

FIGURE 4.—Alignments of the *D. sechellia* sequence scaffolds with the polytene chromosome maps of *D. melanogaster*. *D. sechellia* was sequenced to only three to four times coverage and thus its assembled scaffolds are more numerous. The number of aligned scaffolds for each Muller element are A (X), 21; B (2L), 10; C (2R), 6; D (3L), 7; and E (3R), 15. The lower coverage in this species has also apparently resulted in assembly errors in that there are four scaffolds (0, 3, 4, and 5) that are derived from noncontiguous sequences, which in some cases derive from different Muller elements. Scaffold 0 is a chimera of D (3L) and E (3R); scaffold 3 joins noncontiguous sequences of B (2L); scaffold 4 is a chimera of A (X) and E (3R) and scaffold 5 joins noncontiguous sequences of B (2L), C (2R), and F (3R). The relative position of all of these mis-associations is shaded in the tables. The lines that connect the tables above the chromosome maps indicate the extent of each scaffold and the lines from the solid boxes down to the chromosome ideogram indicate the cytological limits of each scaffold. Each table shows the first and last *D. melanogaster* ortholog identified, the inferred cytological limits, and the beginning and ending molecular coordinates in the CAF1 assemblies for each scaffold. As in *D. simulans*, the known inversion in E (3R) is readily evident in the molecular map. The proximal breakpoint at 84E9 is found in scaffold 6 and the distal break in 93F6,7 is in that portion of scaffold 0 that aligns with the portion of E (3R). The breakpoints seen here are identical to those seen in *D. simulans*. As is the case for *D. simulans*, we can find no evidence for the small inversions in the tip of A (X) or in the bases of C (2R) and E (3R). All of the identified orthologs, their coordinates, and cytological order in *D. melanogaster* are shown in supplemental Table 17 with a listing of small, unaligned scaffolds. The ideograms are from the drawings by BRIDGES (1935, Figure 3) and are reproduced with kind permission from the American Genetics Association and Oxford University Press. The photographic maps are from LEFEVRE (1976, Figures 21–23).

4). Due to the undersampling of the X relative to the autosomes, this chromosome shows the highest fragmentation. Despite these problems, the alignments are good and the gene order within each scaffold is similar to those seen in *D. melanogaster*. There are, however, clear problems with the assemblies revealed by this and other chromosome/scaffold alignments. Scaffold 4 aligns in part with a region just proximal to the telomere of the X (3D-6E). This scaffold is, however, clearly chimeric in that another large and contiguous portion of the scaffold aligns perfectly with a region near the telomere of E/3R (supplemental Table 17). On the basis of the close similarity of the banding patterns of *D. melanogaster* and *D. sechellia* (Lemeunier and Ashburner 1984), a transposition of material between the X and 3R appears highly unlikely and the association of sequences from these two elements in this scaffold is most likely an assembly error. This is not the only case of scaffold chimerism in this species (see below).

Muller element B (chromosome 2L, sections 21–40): The B/2L chromosome arm can be aligned with 10 scaffolds, which comprise 22.2 Mb of sequence and contain a total of 2210 called orthologs. The scaffolds range in size from 53 kb to 4.7 Mb. The lowest ortholog count is 5 (scaffold 63) and the highest 809 (scaffold 3). The observation that makes the alignment odd is that scaffolds 3 and 5 apparently do not represent a contiguous sequence in the B/2L element (Figure 4, supplemental Table 17). Moreover, scaffold 5 is chimeric and contains sequences from this element as well as from C/2R and E/3R (Figure 4, supplemental Table 17). Interestingly, the three chimeric portions are internally contiguous and are not intermingled. There are also two gaps in the coverage at both the distal and proximal ends of the arm. One possible explanation for the observed contiguity problems is that there have been transpositions or inversions of large blocks of material. The size of these blocks makes it likely that they would have been observed cytologically, but this is not the case. The most likely explanation is that they represent assembly errors associated with low coverage.

Muller element C (chromosome 2R, sections 41–60): The C/2R arm can be aligned to five scaffolds comprising 19.6 Mb of sequence and containing 2448 called orthologs. The scaffolds range in size from 168 kb to 14.2 Mb. Scaffold 59 has the lowest ortholog count at 7, while scaffold 1 has the highest at 1748. As noted, a contiguous fragment of scaffold 5 (189 kb; 141 ortholog calls) aligns with this arm (Figure 4, supplemental Table 17).

Muller element D (chromosome 3L, sections 61–80): Arm D/3L can be aligned with seven scaffolds, which comprise 22.8 Mb of sequence and contain 2240 called orthologs (Figure 4). The scaffolds range in size from 100 kb (scaffold 89) to 9.9 Mb (scaffold 0). Scaffold 89 has the smallest number of called orthologs (4) while scaffold 0 contains the highest (1095). Aside from being the largest scaffold, like scaffolds 4 and 5, scaffold 0 is a chimera. Approximately half of scaffold 0 is aligned to D/3L while the other half is associated with E/3R. Again the two halves are represented by contiguous sequences that align well with the *D. melanogaster* cytological map (Figure 4, supplemental Table 17).

Muller element E (chromosome 3R, sections 81–100): The E/3R element can be aligned to 15 scaffolds comprising 27.7 Mb of sequence and containing 3045 called orthologs. Scaffold 166 is the smallest at 38 kb and, not surprisingly, contains the smallest number of called orthologs, 5. The largest is the other half of scaffold 0 at 11.2 Mb, which also contains the highest number of called orthologs at 1345. As noted earlier, there are also portions of two other chimeric scaffolds (4 and 5) that align with this arm (Figure 4, supplemental Table 17). As in the case of *D. simulans*, *D. sechellia* differs from *D. melanogaster* by virtue of a large paracentric inversion on E/3R (Lemeunier and Ashburner 1984). The reported breakpoints are similar to those of the *D. simulans* inversion (Horton 1939; Lemeunier and Ashburner 1976). Scaffolds 0 and 6 of *D. sechellia* show clear evidence of this inversion at the sequence level and also indicate that the breaks are similar if not identical to the *D. simulans* inversion (Ranz *et al.* 2007) (Figure 4, supplemental Table 17).

Muller element F (chromosome 4, section 101–102): The dot chromosome F/4 can be aligned with three scaffolds comprising 1.2 Mb of sequence and containing in aggregate 72 called orthologs (Figure 4, supplemental Table 17). Scaffold 30 is the longest (666 kb) and contains the highest number of called orthologs (41). Scaffold 52 is the shortest (194 kb) and contains the smallest number of called orthologs (12) (supplemental Table 17).

Chromosome maps—*D. erecta*: The coverage of this species is high and consequently the assemblies appear to be better than the two species covered above. The large contiguous scaffolds that essentially include entire arms have also allowed the ready identification of the inversion breakpoints and complexes that serve to differentiate *D. erecta* from *D. melanogaster* (Lemeunier and Ashburner 1976). Interestingly, the molecularly defined breakpoints are in excellent agreement with the previously reported cytological determinations, offering a testament to the accuracy of those results (Lemeunier and Ashburner 1976). *D. erecta* has a pericentric inversion at the base of the B·C element (2L·2R in *D. melanogaster*). The 2L and 2R arms are referred to as the Muller elements B/C· and ·B/C, respectively, where the dot represents the position of the centromere in the rearranged arms. Similar nomenclature is used for the *D. yakuba* pericentric inversion.

Muller element A (chromosome X, sections 1–20): There are two large scaffolds that cover the majority of this arm. The more distal of these is 4644, which extends from the telomere (1A) through section 3A. It contains 2.5 Mb of sequence and 257 called orthologs. The

second scaffold 4690 begins in section 3B and extends to the base (20E/F) of the A/X element. It contains 18.8 Mb of sequence and 1704 called orthologs. Thus the total coverage of this arm is 21.3 Mb of sequence and 1961 called orthologs (Figure 5, supplemental Table 18). As noted in supplemental Table 18, several inversions differentiate the polytene chromosome-banding pattern when comparing *D. erecta* and *D. melanogaster.* The molecular sequence of scaffold 4690 identifies six inversion breakpoints and three overlapping inversions (Figure 5, supplemental Table 18). The molecularly identified breakpoints agree almost entirely with those reported from direct cytological analysis (Lemeunier and Ashburner 1976). The only point of noncongruence is a very small section (7D-7E) that would have been very difficult to discern in normal cytological preparations or with the methodology used to define the extent of the inversions.

Muller element B/C· (chromosome 2L, sections 21–41): The single large scaffold 4929, encompassing 26.6 Mb of sequence and containing 2332 called orthologs, covers essentially the entirety of B-C·/2L (Figure 5, supplemental Table 18). A discussion of this scaffold/arm association as well as that of ·B-C/2R must include the fact that this species possesses a pericentric inversion that reassociates genetic material between the two distinct Muller elements in *D. melanogaster.* Superimposed on this pericentric inversion is a group of overlapping paracentric inversions. The positions and extent of the overlapping inversions is shown in Figure 5 by a series of brackets above the chromosome. The molecular map determined from scaffold 4929 identifies seven inversion breakpoints in this arm (Figure 5, supplemental Table 18). Again, as in the case of the A/X element, the agreement between the positions of the molecularly and cytologically defined breakpoints is excellent (Lemeunier and Ashburner 1976). The only discrepancy is again a very small segment (26B4-26B2) that would have been difficult to detect.

Muller element ·B/C (chromosome 2R, sections 40–60): There is a single large scaffold 4845 encompassing 22.6 Mb of sequence and containing 2338 called orthologs that extends for the length of ·B-C/2R (Figure 5, supplemental Table 18). Here again, the presence of the other portion of the pericentric inversion reassociates sequences that are found in different Muller elements in *D. melanogaster.* The molecular map derived from scaffold 4845 identifies five inversion breakpoints, including the paracentric inversions superimposed on the pericentric (Figure 5, supplemental Table 18). The brackets above the chromosome in Figure 5 show the positions and extent of the inversions. As in the arms discussed above, the cytological and molecular maps are in excellent agreement. The exception is a cytologically small fragment (35F-36B).

Muller element D (chromosome 3L, sections 61–80): The single large scaffold 4784 encompassing 25.8 Mb of sequence and containing 2448 called orthologs extends for the length of D/3L (Figure 5, supplemental Table 18). The molecular map derived from this scaffold identifies seven paracentric inversion breakpoints. These are associated with four inversions. The distal pair of inversions is overlapping while one of the proximal pair is within the larger, and this pair apparently shares a similar proximal breakpoint (Figure 5). The brackets above the chromosome in Figure 5 show the position and sizes of the inversions. A comparison of the cytologically determined breakpoints with those found here in the molecular map shows that again the cytological determination was extremely accurate with, in this case, no missed small segments (Lemeunier and Ashburner 1976).

Muller element E (chromosome 3R, sections 81–100): There are two scaffolds that can be aligned to the E/3R element. The most proximal of these, 4770, comprises 17.7 Mb of sequence and contains 1929 called orthologs. The more distal scaffold 4820 comprises 10.5 Mb of sequence and contains 1137 called orthologs. Thus a total of 28.2 Mb of sequence and 3066 called orthologs can be aligned to E/3R (Figure 5, supplemental Table 18). At the molecular level, there are four identified paracentric inversion breakpoints. These are in general agreement with the previous cytologically determined breaks (Lemeunier and Ashburner 1976). The brackets above the chromosomes in Figure 5 show the positions and extent of the inverted regions. Unfortunately, there is an apparent small gap in the 4770 scaffold that is near the distal end of the smaller included inversion, and the join between the 4770 and 4820 is near the distal end of the larger inversion. This prevents a more precise mapping of the distal ends of both inversions. It appears, however, that the larger of the two inversions is similar, if not identical, to the E/3R inversion seen in *D. simulans* and *D. sechellia.*

Muller element F (chromosome 4, section 101–102): The single scaffold 4512 encompasses nearly the entire banded portion of chromosome F/4 (Figure 5, supplemental Table 18). This scaffold covers 1.3 Mb of sequence and contains 77 called orthologs. A comparison of this scaffold with the gene content and order of *D. melanogaster* indicates that this element is essentially entirely conserved in gene content and order.

Chromosome maps—*D. yakuba*: The assembly of the *D. yakuba* sequence appears to be reasonably good on the basis of few disagreements with the cytological map and has allowed the assignment of a single large scaffold to each of the six Muller elements. As with *D. sechellia,* several of these small scaffolds that we did not align contain duplications of material found in other small scaffolds or in the large scaffolds aligned to the arms. Thus, assuming that these are not true duplications within the genome, for reasons beyond the scope of this report these would appear to be unassembled fragments that could be used to evaluate limitations in the assembly process used for this genome.

A
Drosophila erecta

Muller Element A (Chr X) 21.3 Mb

B

Muller Element B/C (Chr 2L/R) 26.6 Mb

C

Muller Element C/B (Chr 2R/L) 22.6 Mb

D Drosophila erecta

Scaffold	4784	4784	4784	4784	4784	4784	4784	4784
Coordinate Span	369685=>3177513	3179279=>4708343	4766801=>5851180	5852178=>9302130	9326154=>15093651	15095630=>15093...	17399291=>18658152	18673452=>22842407
Gene Span	Lsp1gamma=>CG32280	Atg18=>Glut4EF-4	Glut-RIb=>CG7387	CG32281=>Glur-RI	PGRP-LA=>CG32147	Indy=>CG32159	CG6859=>Aats-tyr	CG6859=>alpha-Cat
Molecular Cytology	61A6=>63B8	66B11=>65C1	67A4,6=>66B11	63BB=>65C1	67B1=>71B5	75E1,2=>72F1-73A1	71B5=>72F1	75E2=>80F1,2
Observed Cytology	61A=>63B	61A=>65C	67A=>66B	63B=>65C	67A=>71B	75E=>73A	71B=>73A	75E=>80F

6C1<=66B11
4708543<=3179279

63B8=>65C1
5852178=>9302130

61A6=>63B8
3698=>3177513

66B11<=67A4
5851180<=4766801

67B1=>71B5
9334544=>15093651

71B5=>71F2
17399291=>17977300

72F1,73A1<=75E1,2
17393432<=15095630

75E2=>80F1,2
18673452=>22842407

Muller Element D (Chr 3L) 25.8 Mb

Telomere — Centromere

F

Scaffold	4512
Coordinate Span	1277071=>1477...
Gene Span	CG17923=>Dyrk3
Molecular Cytology	101F1=>102F8

Telomere — Centromere

Muller Element F (Chr 4) 1.3 Mb

101 102

E

Scaffold	4770	4770	4770	4770	4770	4770	4820	4820
Coordinate Span	363628=>4103023	4119124=>9216646	9236001=>9330935	9332073=>11841980	11843122=>17662462	17696986=>17745035	10438938=>61468	10425868=>61468
Gene Span	CG12581=>Tom34	CG31176c=>CG14892	Scp2=>Cad89D	CG9925=>CG14899	Ipp=>CG7918	CG34034=>Caki	Caki=>Map205	93F13=>100E3
Molecular Cytology	81F5=>84E9	93F2,6=>89D5	89B18=>89B120	88A3=>89B18	88A3=>84E10	93F7=>93F10,12	93F10,12=>100E	93F1,3=>100E3
Observed Cytology	81=>84F	93F =>	89C	88A=>89C	88A=>84F	93F =>	100F	

81F5,6=>84E9
363628=>4103023

84E10<=88A3
17662462<=11843122

89B18<=89B120
9330935<=9232600

88A3=>89B18
9332073=>11841980

93F7=>93F9,10
17696986<=17734980

89D5<=93F2,6
9216646<=4119124

93F13=>100E3
10425868<=61468

Muller Element E (Chr 3R) 28.2 Mb

Centromere — Telomere

82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

FIGURE 5.—Alignments of the D. erecta sequence scaffolds with the polytene chromosome maps of D. melanogaster. The coverage of D. erecta was 12× and the assembly of the scaffolds, thus, quite good. There are two major scaffolds aligned to the A (X) and E (3R) while each remaining element is represented by single scaffolds that extend for essentially the entire length of each arm. The exception to this is scaffold 4770 in E (3R) where there is an apparent gap in the sequence that is indicated by a shaded bar in the table above the chromosome map. The genome of D. erecta differs from that of D. melanogaster in that there are several paracentric inversions in each of the Muller elements as well as a pericentric inversion between the B (2L) and C (2R) arms. In the A (X) element, there are six inversion breakpoints (three inversions) evident in the molecular map. The solid bars in the tables above the chromosome maps indicate the positions of these. The brackets above the chromosomes show the extent of the overlapping inversions. A similar designation is provided in the tables above the remaining elements. In B·C (2LR), there are seven detected breaks: C/B (2RL), 5; D (3L), 7; E (3R), 4; and F (4), 0. The tables also show the predicted molecular cytology of the breakpoints derived from the alignment of the D. melanogaster orthologs in the scaffolds. Below those values in each table are shown the observed cytological breakpoints derived from a comparison of the polytene banding patterns of D. erecta and D. melanogaster chromosomes. The observed and molecular assignments are in most cases in excellent agreement. Exceptions to this are indicated by shading in the tables. In all exceptional cases, the lack of correspondence is associated with very small cytological intervals that would have been difficult, if not impossible, to detect at the resolution provided by squash preparations. The larger of the two inversions in E (3R) appears to be identical to the inversion in the same element in D. simulans and D. sechellia. A listing of all of the identified orthologs, their coordinates in the scaffolds, and their cytological ordering can be seen in supplemental Table 18 along with the smaller scaffolds that were not aligned to the chromosomes. The ideograms are from the drawings by BRIDGES (1935, Figure 3) and are reproduced with kind permission from the American Genetics Association and Oxford University Press. The photographic maps are from LEFEVRE (1976, Figures 21–23).

A Drosophila yakuba

Muller Element A (Chr X) 21.8 MB

B Muller Element B/C (Chr 2L/R) 22.3 Mb

C Muller Element C/B (Chr 2R/L) 21.1 Mb

Muller Element F (Chr 4) 1.4 Mb

Muller Element D (Chr 3L) 24.2 Mb

Muller Element E (Chr 3R) 28.8 Mb

FIGURE 6.—Alignments of the *D. yakuba* sequence scaffolds with the polytene chromosome maps of *D. melanogaster*. Similar to *D. erecta*, the coverage of *D. yakuba* was high (12 times) and there are several large scaffolds that cover each of the Muller element arms essentially entirely. At the resolution of the alignment technique used here there are no large gaps or apparent mis-assemblies. The solid bars in the tables indicated the positions of the breakpoints found in the molecular maps of the various Muller elements. On A (X), there are 12 breaks; B:C (2LR), 9; C:B (2RL), 13; D (3L), 9; E (3R), 13; and F (4), 0. The brackets above each of the arms indicate the extent of the para- and pericentric inversions associated with these breaks. Again, imbedded within the group of inversions seen in the E (3R) element are breakpoints identical to those associated with the large inversion in *D. simulans*, *D. sechellia*, and *D. erecta*. There are significantly more breaks in *D. yakuba* than in any of the other three species in the *melanogaster* subgroup investigated and thus there is significantly more chromosomal rearrangement in this species relative to *D. melanogaster* as compared to the other three species. It is this fact that underlies the less robust agreement between the molecular and observed cytology. As in Figure 3, shading indicates a detected rearranged segment in the molecular alignment that was not seen in a comparison of the polytene banding patterns of *D. melanogaster* and *D. yakuba*. Again, most of these are associated with small molecular and cytological segments that would have been difficult to discern cytologically. A listing of all of the identified orthologs, their coordinates in the scaffolds, and their cytological ordering is in supplemental Table 19 along with the smaller scaffolds that were not aligned to the chromosomes. The ideograms are from the drawings by BRIDGES (1935, Figure 3) and are reproduced with kind permission from the American Genetics Association and Oxford University Press. The photographic maps are from LEFEVRE (1976, Figures 21–23).

Muller element A (chromosome X, sections 1–20): A single large scaffold comprising 21.8 Mb of sequence and containing 1946 called orthologs can be aligned with the A/X element. This scaffold extends from the telomere at 1A to the base of the chromosome at 20F, covering essentially the entire arm (Figure 6, supplemental Table 19). The molecular map of the scaffold indicates that there are 12 breakpoints that define the endpoints of six overlapping paracentric inversions (Figure 6, supplemental Table 19). The position of the breakpoints and the extent of the inversions are indicated by the brackets above the chromosomes in Figure 6. In an earlier cytological mapping of the inversion breakpoints in this element, LEMEUNIER and ASHBURNER (1976) noted that the number of rearrangements made it difficult to precisely map inversion endpoints. They were, however, able to discern some and these are shown above the chromosome map in Figure 6. The data derived from this alignment of the molecular map have allowed further refinement of the earlier map and provide a more complete picture of the inversion complex that distinguishes the gene order in this species relative to *D. melanogaster.*

Muller element B/C· (chromosome 2L, sections 21–41): A single large scaffold comprising 22.3 Mb of sequence and containing 2342 called orthologs can be aligned to B-C·/2L. The distal end of the scaffold lies at 21A near the telomere and the proximal end at 41F (Figure 6, supplemental Table 19). Thus the scaffold encompasses almost the entire arm. As is the case for *D. erecta*, the Muller B and C elements are reassociated by a pericentric inversion, creating a mix of genetic material from both arms. Superimposed on this is a set of overlapping paracentric inversions. The molecular map of the scaffold identifies nine inversion breakpoints in B-C·/2L (supplemental Table 19). Eight of these are associated with four paracentric inversions. The positions of the breakpoints and the extent of the inversions are shown by the brackets above the chromosomes in Figure 6. As was true for the A/X element, the number and extent of the inversion complexes confounded the cytological analysis of this arm. Despite this fact, several of the cytologically determined breaks are in good agreement with the molecular map, and this new ordering should be considered a refinement of that earlier effort (LEMEUNIER and ASHBURNER 1976).

Muller element ·B/C (chromosome 2R, sections 40–60): The ·B-C/2R arm can be aligned with a single large scaffold comprising 21.1 Mb of sequence and containing 2372 called orthologs (Figure 6, supplemental Table 19). The scaffold's proximal end maps to 40F at the base of the arm to 60F at the telomere and thus encompasses essentially the entire arm. The molecular map identifies 13 breakpoints associated with both the peri- and paracentric inversions. Six of these are associated with a set of three nested and overlapping paracentric inversions at the distal end of the arm that are exclusive

of the pericentric breaks that serve to reassociate the genetic material between the two Muller elements (Figure 6, supplemental Table 19). Again, the earlier cytological analysis was hampered by the extent of rearrangement in 2L·2R. Nonetheless, several of the breaks identified there are in agreement with those seen here. Again, the map provided here should be regarded as a refinement of the earlier cytologically derived map (LEMEUNIER and ASHBURNER 1976).

Muller element D (chromosome 3L, sections 61–80): The D/3L element can be aligned with a single large scaffold comprising 24.2 Mb of sequence and containing 2352 called orthologs. The distal end maps to 61A and the proximal end to 80F, essentially encompassing the entire arm (Figure 6, supplemental Table 19). The molecular map identifies nine breakpoints that are associated with five paracentric inversions. The two most distal inversions are nested and exclusive of the proximal three (Figure 6). The latter are overlapping and the distal breakpoint of two of the group appears to be shared or in close proximity. The positions of the breakpoints and the extent of the inverted segments are indicated by the brackets above the chromosome in Figure 6. The fact that the inversion set in D/3L is less complex than that seen in the A/X and B-C·B-C/ 2L·2R chromosomes made the cytological interpretation of the breakpoints more straightforward (LEMEUNIER and ASHBURNER 1976) and a comparison of the molecularly and cytologically determined breaks shows that they are in excellent agreement (Figure 6, supplemental Table 19).

Muller element E (chromosome 3R, sections 81–100): Arm E/3R can be aligned with a single large scaffold comprising 28.8 Mb and containing 3054 called orthologs. The proximal end of the scaffold maps to 81F and the distal end to 100E, thus encompassing almost the entire arm (Figure 6, supplemental Table 19). The molecular map identifies 12 breakpoints that are associated with eight overlapping/nested paracentric inversions proximally and a single distal inversion that is not associated with the others. The number of breaks and inversions is discordant because there are several shared or proximate breakpoints in the more proximal set of inversions. The positions of the breakpoints and the extent of the inverted segments are indicated by the brackets above the chromosome in Figure 6. There is also an apparent small gap in the sequence of the scaffold indicated by a shaded bar above the chromosome map. Despite the complexity of the rearrangements in this arm, the cytologically determined breakpoints and those found in the molecular map are in good agreement (LEMEUNIER and ASHBURNER 1976). The only exceptions to this are a group of very small chromosome intervals that would have been very difficult to discern in cytological preparations or by the methodology used to determine the breakpoints (Figure 6, supplemental Table 19).

Muller element F (chromosome 4, section 101–102): A single scaffold aligns with the F/4 dot chromosome arm (Figure 6). It comprises 1.4 Mb of sequence and contains 77 called orthologs (supplemental Table 19). The proximal end of the scaffold lies at the base of the arm at 101F and distally at 102F near the telomere, essentially covering the entire arm (Figure 6). There is no large rearrangement of the genetic material at the molecular level and the chromosome appears to be largely conserved in genetic content and order in comparison to *D. melanogaster.*

D. ananassae maps

First described by DOLESCHALL (1858), *D. ananassae* belongs to the *melanogaster* group (although not the *melanogaster* subgroup) of the subgenus Sophophora and is one of eight cosmopolitan species (PATTERSON and STONE 1952; TOBARI 1993a). The mitotic chromosomes of *D. ananassae* were first described by METZ (1916). Following this initial investigation, KAUFMANN (1936), KIKKAWA (1938), and TOBARI *et al.* (1993) established the karyotype as three pairs of V-shaped chromosomes, a V-shaped X, and a rod or J-shaped Y in the male complement. The number of arms seen in polytene chromosome preparations was first reported as six by KIKKAWA (1935). KIKKAWA (1936) further determined that the shortest arms of unequal length are XL and XR while the other four represent the left and right arms of the second and third chromosomes. The largely heterochromatic fourth chromosome is embedded in the chromocenter and is generally not seen as a banded euchromatic element.

Most Drosophila species carry the nucleolus organizer (NOs) on the sex chromosomes (ASHBURNER *et al.* 2005), but a unique Y-4 linkage of the NOs has been reported in *D. ananassae* (KAUFMANN 1937; KIKKAWA 1938; ROY *et al.* 2006; SHIBUSAWA *et al.* 2007). Possibly related to this NO localization is the fact that, in the primary spermatocyte, the X, Y, and fourth chromosomes form a tangled multivalent (HINTON and DOWNS 1975; MATSUDA *et al.* 1983; GONI *et al.* 2006). Another unique character of the species is that spontaneous crossing over occurs in males (KIKKAWA 1937; MORIWAKI 1937), albeit at much lower frequencies than in females. Subsequent to this discovery, HINTON (1970), HINTON and DOWNS (1975), Moriwaki's group (MORIWAKI *et al.* 1970), and MATSUDA *et al.* (1993) have all studied the cytogenetic basis of this phenomenon. Additionally, parthenogenetic females have been isolated from South Pacific island (FUTCH 1972) and Papua New Guinea (MATSUDA and TOBARI 1999) populations.

Early linkage maps of the species were constructed by KIKKAWA (1938) and MORIWAKI (1938, 1940). Unfortunately, almost all of the mutant stocks listed on their maps were lost. Therefore, after World War II, Moriwaki and colleagues and Hinton reisolated mutants and reconstructed the linkage maps (HINTON 1983; reviewed by MATSUDA *et al.* 1993; MORIWAKI and TOBARI 1975; HINTON 1991; unpublished data in TOBARI 1993b). On the basis of these linkage maps, several genetic factors (Enhancers, Suppressors, and Modifiers) controlling male crossing over have been mapped to the chromosomes (HINTON 1970). In addition, genes related to parthenogenesis (MATSUDA and TOBARI 2004) and affecting mating behavior have also been mapped on the second chromosome (DOI *et al.* 2001; YAMADA *et al.* 2002). Hinton's genetic map (see TOBARI 1993b) is generally used as the standard for linkage analyses; however, there are many extant unmapped mutants.

Various populations of *D. ananassae* carry many chromosomal rearrangements, translocations, pericentric inversions, and paracentric inversions (DOBZHANSKY and DREYFUS 1943; FREIRE-MAIA 1961; FUTCH 1966) and several reference maps of the six arms of the polytene chromosomes have been prepared. Seecof (STONE *et al.* 1957) drew a cytological map of a "standard" arrangement. These authors subdivided the polytene chromosomes into 161 divisions, with the enumeration starting at the distal (telomeric) end of each arm. This map was utilized by FUTCH (1966) to describe a variety of chromosome rearrangements found in South Pacific island populations. Subsequently, MORIWAKI and ITO (1969) composed photomaps that were used to describe the puffing patterns and HINTON and DOWNS (1975) produced a new ideogram map for use in their cytogenetic analyses. More recently, we have prepared revised photographic maps (TOBARI *et al.* 1993), which are based on those of MORIWAKI and ITO (1969) to aid in the determination of inversion breakpoints. In these maps, the chromosomes are divided into 100 numerical sections, and each of these is further subdivided into subsections denoted by the letters A, B, C, and D. Section numbering is started from the distal end of XL to the distal end of 3R: XL (1–13), XR (14–20), 2L (21–44), 2R (45–63), 3L (64–81), and 3R (82–99). Section 100 is assigned to the fourth chromosome, although no fourth chromosome bands have yet been identified and we cannot eliminate the possibility of the existence of euchromatic bands on this chromosome. The Muller element equivalences for the six chromosome arms of *D. ananassae* and its relatives are (Muller element = chromosome arm) A = XL and XR, B = 3R, C = 3L, D = 2R, E = 2L, and F = 4.

Using the revised photographic map, TOMIMURA *et al.* (1993) described chromosomal polymorphisms of *D. ananassae* and related species from 30 populations covering the species range. They found five pericentric inversions, one translocation, and 52 paracentric inversions. Among the 52 paracentric inversions, 32 were previously undescribed and three of the cosmopolitan inversions, *In(2L)A*, *In(3L)A*, and *In(3R)A*, were found in almost all the localities covering the species range. Many

Drosophila ananassae

A

Muller Element A (Chr. XL) 19.4 Mb

Muller Element A (Chr. XR) 10.9 Mb

E

Muller Element E (Chr. 2L) 30.1 Mb

D

Muller Element D (Chr. 2R) 23.3 Mb
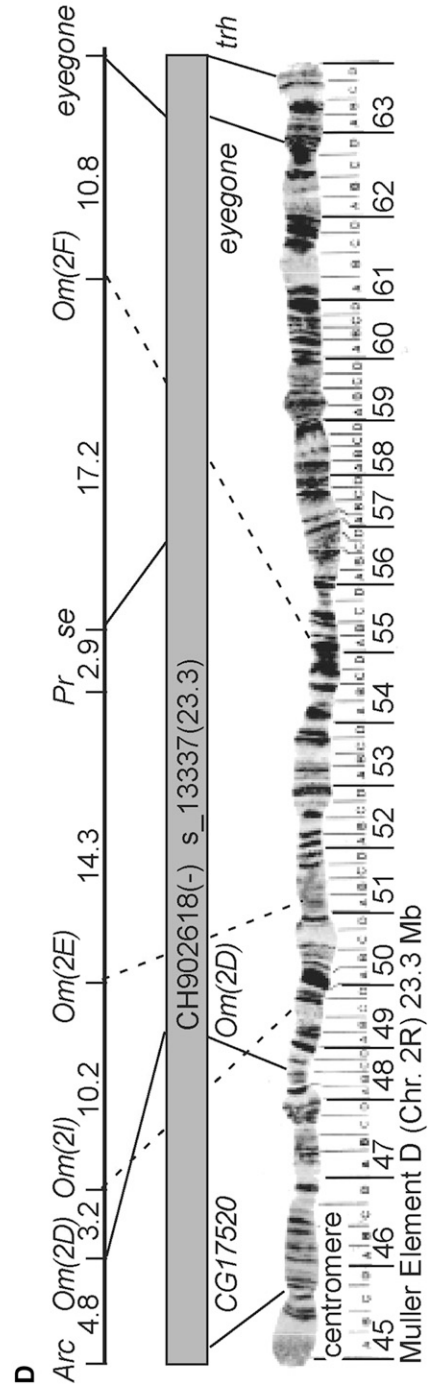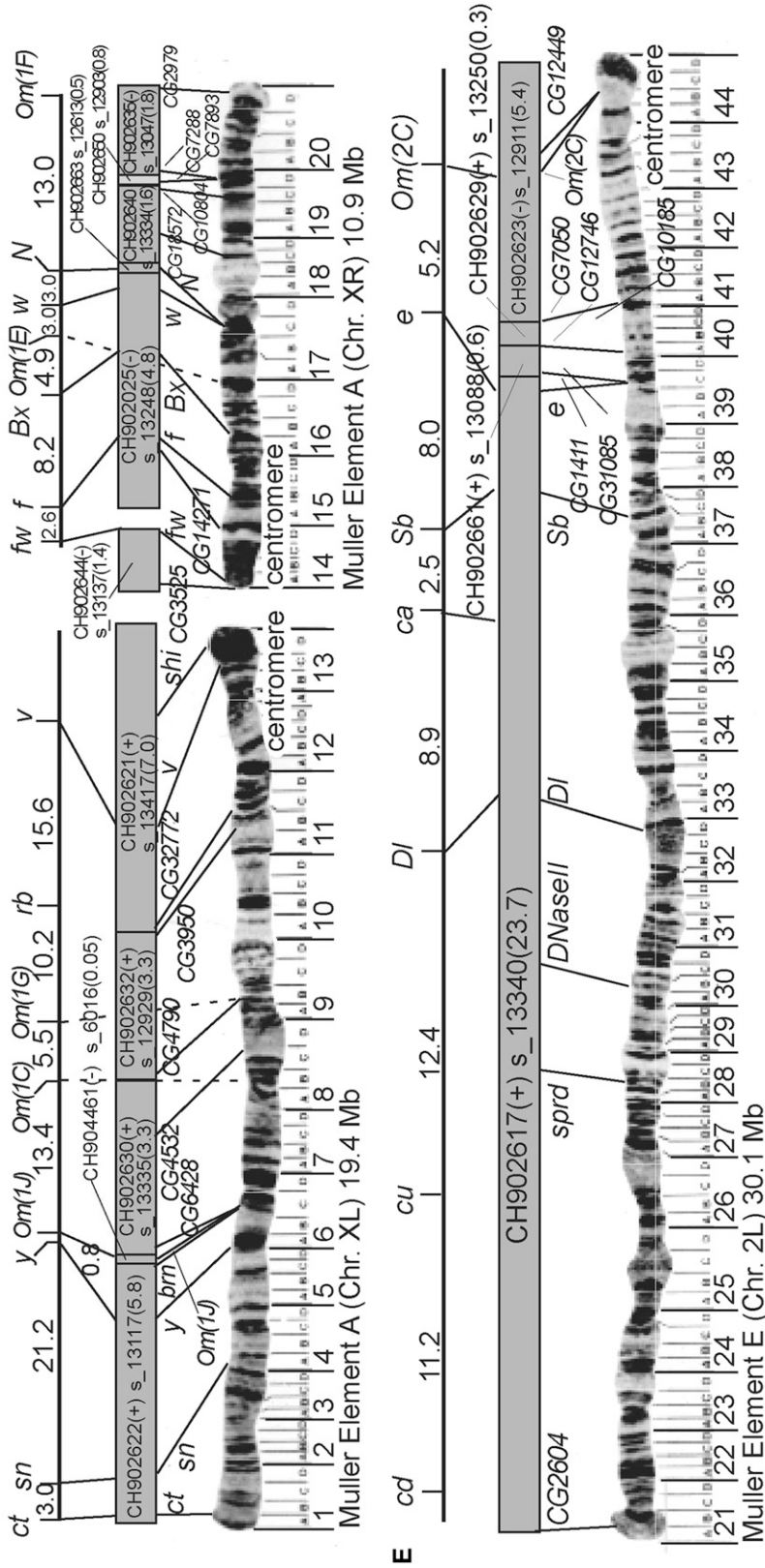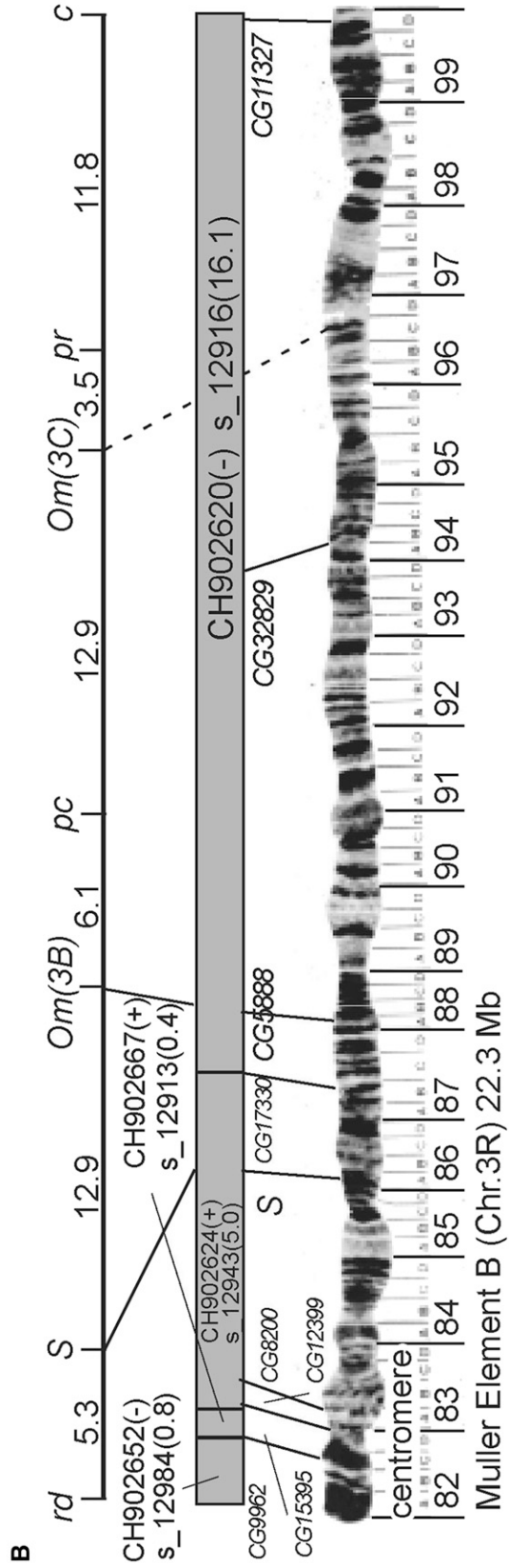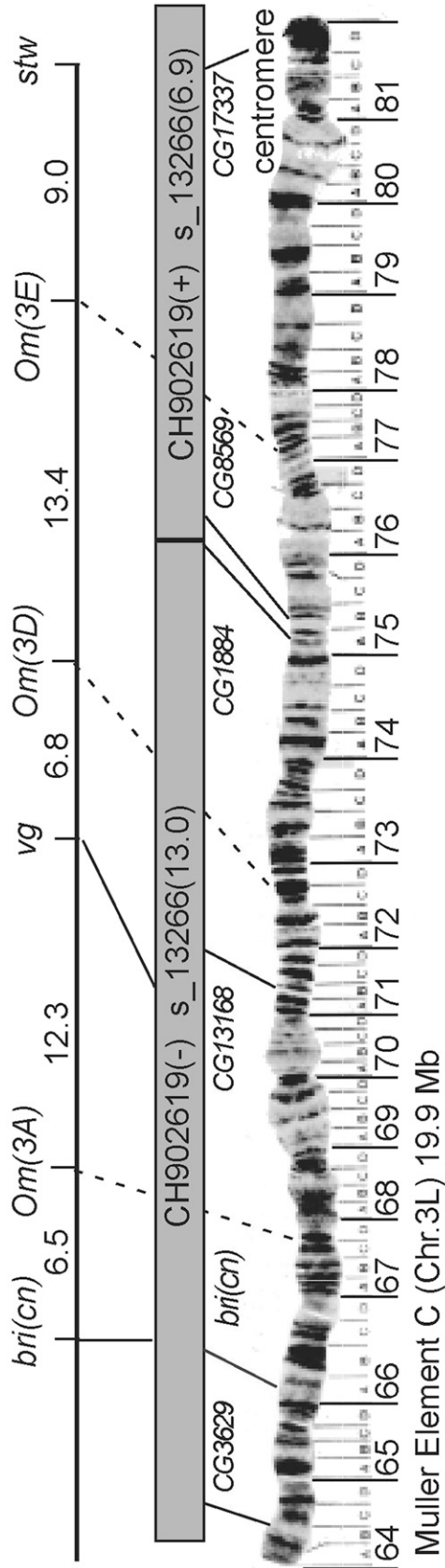
FIGURE 7.—Alignments of the *D. ananassae* sequence scaffolds with the polytene chromosome maps (TOBARI *et al.* 1993). For each chromosome arm, the three maps are the genetic map, the scaffold map showing markers on the genetic and chromosome maps, and the salivary chromosome map with the section and subsection designations. The ideograms and chromosomes from TOBARI *et al.* (1993, Figures 15 and 16) were reproduced with kind permission of the Japan Scientific Societies Press.

overlapping inversions are found in populations of *D. ananassae* and its relatives, making it possible to deduce the phylogenetic relationships of species belonging to the *D. ananassae* complex, including *D. pallidosa, papuensis*-like, *pallidosa*-like, and Taxon K.

Further mapping efforts are related to an unusual *Optic morphology* (*Om*) hypermutability system, which has been used to genetically map lesions at 22 loci (Hinton 1984). It was subsequently discovered that these *Om* mutations were caused by *tom* insertions (Shrimpton *et al.* 1986). Using the *tom* element, 18 of the 22 *Om* loci have been mapped cytologically (Tanda *et al.* 1989; Matsubayashi *et al.* 1991, 1992; Awasaki *et al.* 1994; Juni *et al.* 1996), allowing an alignment of the genetic and polytene maps.

Chromosome maps—*D. ananassae*: The *D. ananassae* strain AABBg1 was used for genomic sequencing instead of the strain used to develop the polytene map (Moriwaki and Ito 1969; Tobari *et al.* 1993). The AABBg1 strain was used because it was the most highly inbred strain available at the time that genome sequencing began. The polytene chromosome map strain (Moriwaki and Ito 1969; Tobari *et al.* 1993) has been used extensively to develop the *D. ananassae* physical map genes using *in situ* hybridization. The karyotypes of the polytene map strain and AABBg1 have the same standard arrangement on all chromosomes except for Muller C (chromosome 3L), where the polytene strain has the standard arrangement and AABBg1 has the terminal inversion *In*(3L)A with breakpoints at 75B and 81A.

Muller element A (chromosome XL, telomere to centromere, sections 1–13, and chromosome XR, centromere to telomere, sections 14–20): A total of 18 scaffolds map to the two arms of the *D. ananassae* X chromosome, accounting for 37.0 Mb (Figure 7, supplemental Table 20). A total of 48 loci have been mapped via *in situ* hybridization in the present experiment (see supplemental Table 20). Five scaffolds map to the euchromatic portion of XL, accounting for 19.4 Mb, while 8 scaffolds map to XR, accounting for 12.0 Mb. There is a gap in the region between scaffold 13335 and scaffold 12929, which corresponds to 8C–9B. Probes derived from sequence in scaffolds 13265, 13333, 13111, 12905 and 12048 hybridized to the chromocenter; however, conserved linkage (CL) analysis places these scaffolds on Muller A between scaffolds 13417 and 13137. Although Stephan (1989) mapped the *fw* gene to 14A by *in situ* hybridization using a probe of the *fw* gene of *D. melanogaster*, in this mapping effort, we could not distinguish between the chromocenter and 14A of XR. Although there are inconsistencies in the order between the molecular and cytological positions within scaffold 13334, this scaffold may be aligned to the proximal region of XR.

Muller element B (chromosome 3R, centromere to telomere, sections 82–99): Five scaffolds cover almost the entire B element and comprise a total of 21.9 Mb of sequence (Figure 7, supplemental Table 20). A total of 20 loci

have been mapped on this chromosome arm, 2 of them by the *tom* transposable element (Matsubayashi *et al.* 1992). Scaffold 12422 is adjacent to scaffold 12913 on the basis of the conserved linkage analysis, but this scaffold may be in the centromeric heterochromatin. We did not use sequences of this scaffold for our *in situ* hybridization analysis to verify this provisional localization.

Muller element C (chromosome 3L, telomere to centromere, section 64–81): A single scaffold (13266) comprising 19.9 Mb of sequence aligned to almost the entire C element (Figure 7, supplemental Table 20). A total of 17 loci have been mapped on this chromosome arm, 3 of them by the *tom* transposable element (Matsubayashi *et al.* 1992). Because Muller C in the genome strain (AABBg1) differs from the polytene map by a single inversion, *In*(3L)A (75B; 81A), it is necessary to divide the scaffold into two parts at the breakpoint, 75B, to align the scaffold with the polytene chromosome. We have not determined the precise sites of the inversion breakpoint at 75B or 81A; however, the length of the inverted segment and position of breakpoints is in good agreement with the scaffold and polytene maps.

Muller element D (chromosome 2R, centromere to telomere, sections 45–63): Only one scaffold, scaffold 13337, aligned with almost the entire D element, and it accounts for 23.3 Mb of the chromosome (Figure 7, supplemental Table 20). A total of 18 loci have been mapped on this chromosome arm, 3 of them using the *tom* transposable element (Matsubayashi *et al.* 1992). The centromeric region 45A-B remains unknown.

Muller element E (chromosome 2L, telomere to centromere, sections 21–44): Four scaffolds cover almost the entire E element from the telomere to centromere for a total length of 33.2 Mb (Figure 7, supplemental Table 20). A total of 34 loci have been mapped on the chromosome. An unusual alignment was found in parts of scaffolds 13250, 13,333, and 13,043, which contain homologous regions of the element A of *D. melanogaster*. The sites of the homologous positions of the element E are 40B, but these could not be distinguished. Further analysis will be necessary to resolve whether these sequence anomalies were due to the transposition or to problems with the assembly.

Muller element F (chromosome 4, section 100): Sixteen scaffolds have F element orthologs that are 17.8 Mb in length (Table 1, supplemental Table 20). We do not find polytene bands associated with the fourth chromosome, which corresponds to Muller's F element. Mitotic preparations show that it is as long as the Y chromosome and is also entirely heterochromatic and likely resides entirely within the chromocenter in polytene preparations. These data are consistent with the increased size of the F element on the basis of cytology. Unfortunately, we could not determine the location of the corresponding scaffolds. We are currently applying FISH on mitotic chromosomes to locate these scaffolds.

As a preliminary analysis to determine whether repetitive DNA has contributed to the expansion of Muller F, we tested the 17.8-Mb scaffold sequences with Repeat Masker using the *D. melanogaster* settings for known repeats. The scaffolds from element F of *D. ananassae* contain an average of 32.5% interspersed repeats that is composed of 27.1% retro element and 5.4% DNA-based repeats. This is similar to the scaffolds assigned to the centromeric regions of the *D. virilis* genome.

D. pseudoobscura and *D. persimilis* maps

D. pseudoobscura and *D. persimilis* are Nearctic members of the *obscura* group of the genus Drosophila. The ancestral configuration of the six Muller elements in the *obscura* group is six acrocentric rods (POWELL 1997), as seen in the Palaearctic *obscura* group species *D. subobscura* (KRIMBAS 1992). *D. pseudoobscura* and *D. persimilis* have evolved from this ancestral condition to have five chromosomes: X, 2, 3, 4, and 5. The metacentric X chromosome is composed of the left and right arms XL and XR that resulted from the fusion of Muller's A and D, respectively. The Y chromosome of *D. pseudoobscura* is composed of genes from Muller's element D, suggesting that the Y is the degenerate copy of the ancestral autosome (CARVALHO and CLARK 2005). Chromosomes 2, 3, 4, and 5 are acrocentric rods. The Muller element equivalences for the six chromosomal arms of the *obscura* group species (Muller element = chromosome arm) are A = XL; B = 4; C = 3; D = XR; E = 2; and F = 5.

The *D. pseudoobscura* salivary chromosomal maps were originally developed by TAN (1935, 1937). The salivary chromosomes were divided into 100 sections across the six chromosomal arms in the following order: XL (1–17), XR (18–42), 2 (43–62), 3 (63–81), 4 (82–99), and 5 (100). The original section assignments were based on camera lucida drawings of the chromosomes and not on photomicrographs. Sections 38–42 were clearly delineated for the tip of chromosome XR; however, sections 18–37 were not because the drawings were based on polytene chromosomes from *D. pseudoobscura*/*D. miranda* hybrids (DOBZHANSKY and TAN 1936). Only chromosome 3 had its 19 sections subdivided into lettered subsections (DOBZHANSKY and STURTEVANT 1938). KASTRITSIS and CRUMPACKER (1966) published photomicrographs of complete *D. pseudoobscura* salivary chromosomes; however, not all sections were labeled in their figures. MOORE and TAYLOR (1986) published photomicrographs of complete *D. persimilis* salivary chromosomes with all sections indicated. In addition, MOORE and TAYLOR included photomicrographs of the chromosomes of *D. pseudoobscura*–*D. persimilis* hybrids, allowing one to precisely map the breakpoints of the inversion differences between the two species.

The salivary chromosomal maps for the genome strains of *D. pseudoobscura* and *D. persimilis* are quite similar except for four inversion differences. Four of the six chromosomal arms have gene arrangement differences within or between the two species (DOBZHANSKY 1944). While these arrangement differences add complexity to the comparison of the *D. pseudoobscura* and *D. persimilis* genomic sequences, they act as valuable physical markers for orienting the genome scaffolds to the salivary chromosomes. Two chromosomes have fixed differences between the two species. Chromosome XL differs by a single fixed inversion that is a diagnostic character that distinguishes the two species (ANDERSON *et al.* 1977). Chromosome 2 also has a fixed inversion difference.

Two chromosomes are segregating for gene arrangement polymorphisms in natural populations of the two species (DOBZHANSKY 1944). Chromosome XR is segregating for two different gene arrangements in each species. *D. pseudoobscura* has two arrangements called standard and sex ratio that differ by three nonoverlapping inversions (STURTEVANT and DOBZHANSKY 1936). The sex ratio chromosome is a meiotic drive element causing males that carry the chromosome to sire ~95% daughters. *D. persimilis* populations also segregate for standard and sex ratio chromosomes; however, the two gene arrangements differ by a single inversion. The sex ratio chromosome of *D. persimilis* appears homosequential with the standard arrangement of *D. pseudoobscura*. The genome strains of *D. pseudoobscura* and *D. persimilis* carry the standard arrangements of XR. Thus, their XR maps differ by a single inversion difference.

In contrast to the other chromosomes, chromosome 3 has >30 different gene arrangements that segregate in populations of *D. pseudoobscura* and *D. persimilis*, making it a central focus of study in these species for >70 years (DOBZHANSKY and STURTEVANT 1938; DOBZHANSKY 1944; ANDERSON *et al.* 1991). The *D. persimilis* genome strain carries the standard arrangement while the *D. pseudoobscura* genome strain carries the Arrowhead arrangement (RICHARDS *et al.* 2005), which was derived from the standard arrangement by a single inversion (DOBZHANSKY 1944).

The Freeze 1 assembly of *D. pseudoobscura* used 755 scaffolds to create 16 supercontigs or ultrascaffolds that were anchored to five of the six chromosomal arms (RICHARDS *et al.* 2005). Chromosomes 2 and 3 (Muller C and E) each had a single supercontig assigned to them, while chromosome arms XL, XR, and 4 (Muller A, D, and B) had 4, 5, and 5 supercontigs, respectively. At the time of the first annotation of *D. pseudoobscura*, no scaffold assignments were made for chromosome 5, the dot (Muller F). By virtue of the new genomes including the *D. persimilis* genome in the CAF1 assemblies, we are now able to assign scaffolds to the *D. pseudoobscura* Muller F.

We now present updated salivary chromosome maps of the *D. pseudoobscura* and *D. persimilis* genome strains. The maps present the boundaries of the sections and

A *Drosophila pseudoobscura*

Muller Element A (Chr XL) 20.3 Mb

D Muller Element A/D (Chr XR) 29.2 Mb

E Muller Element E (Chr 2) 30.8 Mb

C

Drosophila pseudoobscura

FIGURE 8.—Alignments of the *D. pseudoobscura* sequence scaffolds with the polytene chromosome maps. For each chromosome, the four maps are the genetic map, the scaffold map showing markers on the genetic and chromosomal maps, the ideogram of the chromosome showing section and subsection designations, and the salivary chromosome map with the section and subsection designations. "CL" drawn across a scaffold junction indicates that the join is supported by conserved linkage information in at least one other species. Supplemental Table 15 provides the annotation for the numbered scaffolds. The ideograms were reproduced based on the images by DOBZHANSKY and TAN (1936, Plate 1) and TAN (1937, Plate 1) with kind permission of Springer Science and Business Media.

also now include lettered subsections. We have integrated all physical and genetic map data that allow one to orient the sequence scaffolds from the CAF1 assemblies of the two species. We present the data for *D. pseudoobscura* first because this species has the vast majority of mapping data available and indicate changes from the original assembly and annotation (RICHARDS *et al.* 2005). The *D. persimilis* data are presented second because the whole-genome shotgun had 3× coverage and the larger number of scaffolds was oriented to the physical map using the *D. pseudoobscura* assembly.

Chromosome maps—*D. pseudoobscura*: *Muller element A (chromosome XL, sections 1–17):* Three scaffolds map to Muller A for a total of 20.3 Mb of the *D. pseudoobscura* genome (Figure 8, supplemental Table 21). The original assembly and annotation of *D. pseudoobscura* mapped four supercontigs to XL because the genes within the scaffold mapped to Muller A in *D. melanogaster*: ChXL_group1a, ChXL_group1e, ChXL_group3a, and ChXL_group3b. On the basis of the present analysis, two of the XL supercontigs, ChXL_group1a and ChXL_group3a, need to be split and remapped. SEGARRA and AGUADÉ (1992) and SEGARRA *et al.* (1995) used *in situ* hybridization to map *phosphogluconate dehydrogenase* (*Pgd*) and *zeste* (*z*) from Muller A to the base of XR in *D. pseudoobscura*. ChXL_group1a and ChXL_group3a were split into four and two sequences, respectively, to accommodate these results, and one segment from each scaffold was moved to XR. In addition, this move required that ChXL_group3b also be mapped to the base of XR. The junctions of the rejoined sequences on the two arms of the X are supported by conserved linkage in at least one other species.

The movement of genes from Muller A to Muller D is supported by additional evidence. First, the size of the XL and XR arms is not equivalent in *D. pseudoobscura*. Each euchromatic portion of Muller A and D in *D. melanogaster* has 25 Mb of DNA for a total of 50 Mb of sequence. The relative length of the two arms in *D. pseudoobscura* determined from measurements of salivary chromosomes shows that XL and XR are 41.2 and 58.8% of the complete X chromosome. Thus, XL and XR would be expected to have 20.6 and 29.4 Mb, respectively, assuming that DNA content on the two arms is conserved between *D. melanogaster* and *D. pseudoobscura*. The mapped scaffolds on XL and XR are 20.3 and 30.5 Mb, consistent with the relative proportions of the two polytenized chromosomal arms. Second, the Freeze 1 assembly of *D. pseudoobscura* had a scaffold (Contig6811_Contig7852) that included a junction between Muller A and D genes. The CAF1 assembly of *D. persimilis* also has a scaffold (scaffold_12) with a similar A/D junction. Although the *D. persimilis* assembly used the backbone of *D. pseudoobscura* to assist the assembly, a fresh assembly of *D. pseudoobscura* was generated for the backbone sequence. In both cases, the

original assemblies support a join of scaffolds from Muller A and D within the XR chromosome arm.

The distances to the breakpoints for the fixed inversion difference between *D. pseudoobscura* and *D. persimilis* are found in supplemental Table 21 and Table 22. The length of the inverted region (6.8 Mb in *D. pseudoobscura* and 7.3 Mb in *D. persimilis*) is proportional to the size of the inversion inferred from salivary chromosomes (MOORE and TAYLOR 1986). The inferred inversion lengths are similar, but not identical, which may reflect differences in the quality of the assemblies between the two species. *D. persimilis* had only 3× coverage, while *D. pseudoobscura* had 8× coverage.

Muller element B (chromosome 4, sections 82–99): Eleven scaffolds map to Muller B in *D. pseudoobscura* for a total of 27.5 Mb of the genome (Figure 8, supplemental Table 21). The original assembly mapped five scaffolds to chromosome 4: Ch4_group1, Ch4_group2, Ch4_group3, Ch4_group4, and Ch4_group5 (RICHARDS *et al.* 2005). The exact orientation and location of these scaffolds were not completely determined at the time of that publication. PAPACEIT *et al.* (2006) has increased the density of markers on the physical map of Muller B in *D. pseudoobscura* using five known loci and 19 anonymous DNA probes. No sequence information was available for these anonymous probes; however, these DNAs were hybridized to *D. pseudoobscura*, *D. subobscura*, and *D. melanogaster*, providing valuable clues about where these clones are found in the *D. pseudoobscura* sequence. The sequences for the anonymous probes in *D. pseudoobscura* were inferred from the region of hybridization observed in *D. melanogaster*. We identified the cytological coordinates for all *D. melanogaster* genes using the map conversion table in FlyBase (FLYBASE CONSORTIUM 1999). From these data, we mapped the *D. pseudoobscura* orthologs of *D. melanogaster* genes that hybridized to each of the anonymous probes. For instance, the anonymous probe P67 hybridized to region 21C of *D. melanogaster* and 20 genes mapped within the approximate coordinates of 21C. The *D. pseudoobscura* orthologs of these 20 genes are linked within a single conserved linkage block found in a single sequence scaffold, Ch4_group3. P67 maps to 94B and was near the location of known gene marker *Mhc*, which was also located in scaffold Ch4_group3 and was mapped to cytological location 95B. In some cases, the block of genes from *D. melanogaster* mapped to two or more conserved linkage blocks within the scaffolds of *D. pseudoobscura*. In these cases, each region was considered as the potential location for the anonymous hybridization probe. The locations of known probes and unambiguously mapped anonymous probes were used to triangulate the locations of anonymous probes. Once an anonymous probe was mapped to a single sequence region, all the other locations were ruled out. These additional anonymous physical markers were able to place the five sequence scaffolds for Muller B in *D. pseudoobscura*.

Three scaffolds, Ch4_group1, Ch4_group2, and Ch4_group4, had to be split and reoriented on the basis of the physical map data and conservation of linkage relationships among the different Drosophila species. In addition, eight additional scaffolds were mapped to Muller B by virtue of the Synpipe analysis (Bhutkar *et al.* 2006).

Muller element C (chromosome 3 arrowhead arrangement, sections 63–81): A single supercontig (Ch3) was generated in the original *D. pseudoobscura* assembly that accounted for 19.8 Mb of the genome (Figure 8, supplemental Table 21). Four new scaffolds were integrated on the physical map, three at the centromeric end and one inserted in the middle of chromosome 3 (supplemental Table 15). These changes were made on the basis of conserved linkage information among the 11 Drosophila species.

The distances to the breakpoints for the standard-to-Arrowhead inversion event between *D. pseudoobscura* and *D. persimilis* are found in supplemental Table 21 and Table 22. The size of the inverted region (5.9 Mb in *D. pseudoobscura* and 5.9 Mb in *D. persimilis*) is proportional to the size of the inversion inferred from salivary chromosomes (Dobzhansky and Sturtevant 1938).

Muller element A/D (chromosome XR, sections 18–42): Seven scaffolds map to Muller A/D in *D. pseudoobscura* for a total of 30.5 Mb of the genome (Figure 8, supplemental Table 21). The original assembly mapped five scaffolds to Muller A/D in *D. pseudoobscura*: ChXR_group6, ChXR_group9, ChXR_group8, ChXR_group3a, and ChXR_group5 (Richards *et al.* 2005). Scaffold ChXR_group9 was merged with ChXR_group8 because the two scaffolds were adjacent; however, the orientation of the half of the original ChXR_group8 appears reversed on the basis of the configuration of the two inversion breakpoints between *D. pseudoobscura* and *D. persimilis*, analysis of *D. persimilis* original sequence traces with associated mate pairs, and linkage mapping data that are all inconsistent with the original assembly (Ortiz-Barrientos *et al.* 2006). Additionally, three scaffold segments that were incorrectly assigned to XL were added to XR (see above): ChXL_group1a, ChXL_group3a, and ChXL_group3b.

The distances to the breakpoints for the fixed inversion between *D. pseudoobscura* and *D. persimilis* are found in supplemental Table 21 and Table 22. The size of the inverted region (13.2 Mb in *D. pseudoobscura* and 12.2 Mb in *D. persimilis*) is proportional to the size of the inversion inferred from salivary chromosomes (Moore and Taylor 1986).

Analysis of a 25,989-bp nucleotide sequence within the A/D junction region reveals a 1039-bp repeat sequence that generates 11,457 hits in a BLASTN (Altschul *et al.* 1997) search of *D. pseudoobscura* genomic scaffolds. These BLAST hits are overrepresented in Muller F scaffolds as well as in the set of unplaced scaffolds, two scaffold sets likely to be enriched for heterochromatic DNA. This sequence was also found in *D. persimilis* in roughly the same location.

Muller element E (chromosome 2, sections 43–62): A single supercontig (Ch2) was generated in the original *D. pseudoobscura* assembly that accounted for 30.8 Mb of the genome (Figure 8, supplemental Table 21). The distances to the breakpoints for the fixed inversion difference between *D. pseudoobscura* and *D. persimilis* are found in supplemental Table 21 and Table 22. The size of the inverted region (7.6 Mb in *D. pseudoobscura* and 7.7 Mb in *D. persimilis*) is proportional to the size of the inversion inferred from salivary chromosomes (Moore and Taylor 1986).

Muller element F (chromosome 5, section 100): A total of 25 scaffolds were assigned to Muller F in *D. pseudoobscura* for a total of 1.2 Mb of the genome (Figure 8, supplemental Table 21). Five additional scaffolds can be assigned on the basis of comparison to the *D. persimilis* assembly. The scaffold map for Muller F was not well defined in the original assembly and annotation of *D. pseudoobscura* (Richards *et al.* 2005), largely because few genetic and physical markers exist for the dot chromosome. The availability of sequence information from the other species of Drosophila provides valuable clues about the order of the Muller F scaffolds using the Synpipe analysis (Bhutkar *et al.* 2006). The genes of the dot assemble into one supercontig in *D. virilis, D. mojavensis,* and *D. willistoni* and into four supercontigs in *D. persimilis.* These data allowed the scaffolds to be ordered but not oriented in *D. pseudoobscura* on the basis of conserved linkage analysis.

Chromosome maps—*D. persimilis*: The coverage of the *D. persimilis* genome was 3×, which led to a more fragmented genome assembly. The *D. persimilis* assembly was assisted with a new ARACHNE assembly of the *D. pseudoobscura* genome that differed from the ATLAS assembly done by the Human Genome Sequencing Center at Baylor College of Medicine. This process introduced some errors in the *D. persimilis* assembly of Muller C scaffolds. The scaffolds of *D. persimilis* were oriented to the salivary chromosomes on the basis of homology with *D. pseudoobscura,* which had a more detailed genetic and physical map. The karyotype of *D. persimilis* is standard on Muller C.

Muller element A (chromosome XL, sections 1–17): Twenty-one scaffolds map to Muller A in *D. persimilis* for a total of 21.2 Mb of the genome (Figure 9, supplemental Table 22). The distances to the breakpoints for the fixed inversion difference between *D. pseudoobscura* and *D. persimilis* are found in supplemental Table 21 and Table 22. The map of Muller A in *D. persimilis* can be found in Figure 9 and supplemental Table 22.

Muller element B (chromosome 4, sections 82–99): A total of 12 scaffolds map to Muller B in *D. persimilis* for a total of 28.4 Mb of the genome (Figure 9 and supplemental Table 22).
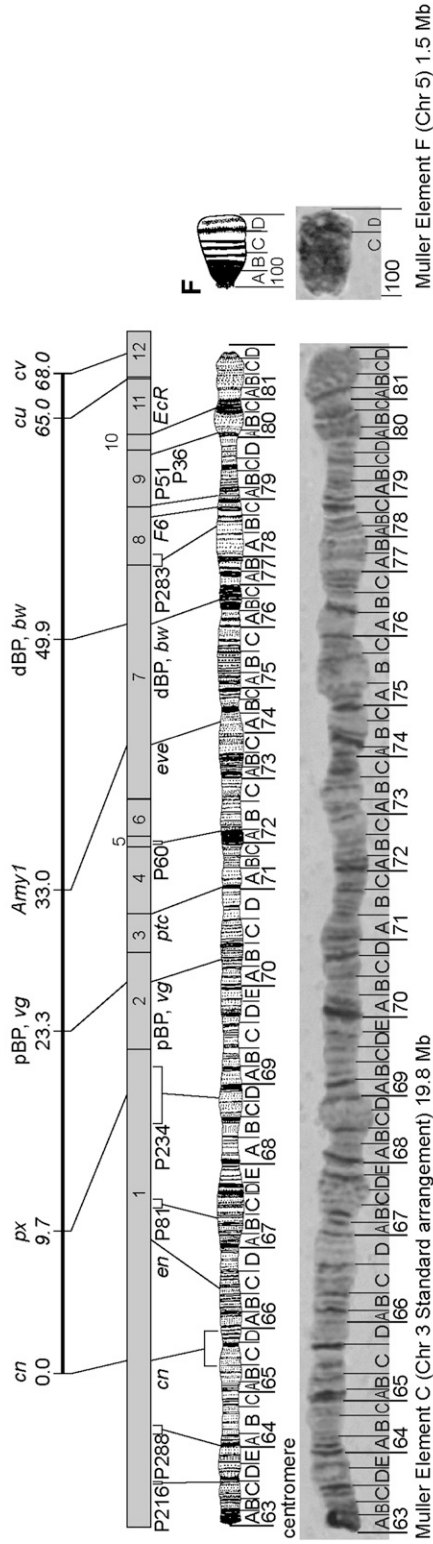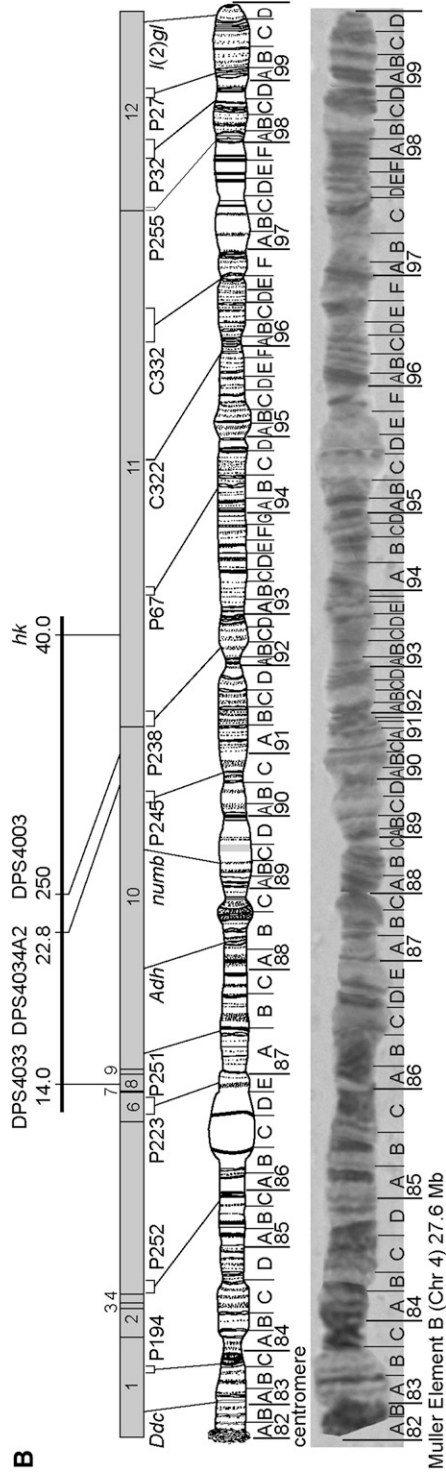
FIGURE 9.—Alignments of the *D. persimilis* sequence scaffolds with the polytene chromosome maps. For each chromosome, the four maps are the genetic map, the scaffold map showing markers on the genetic and chromosomal maps, the ideogram of the chromosome showing section and subsection designations, and the salivary chromosome map with the section and subsection designations. "CL" drawn across a scaffold junction indicates that the join is supported by conserved linkage information in at least one other species. Supplemental Table 15 provides the annotation for the numbered scaffolds. The ideograms were reproduced based on the images by DOBZHANSKY and TAN (1936, Plate 1) and TAN (1937, Plate 1) with kind permission of Springer Science and Business Media.

Muller element C (chromosome 3 standard arrangement, sections 63–81): Eleven scaffolds map to Muller C in *D. persimilis* for a total of 19.9 Mb of the genome (Figure 9 and supplemental Table 22). Two scaffolds (scaffold_2 and scaffold_4) required breakage and rejoining to account for the distance between the standard and Arrowhead inversion breakpoints. The cytogenetic evidence shows that the distance between breakpoints is 5.9 Mb or 29.8% of Muller C in the two species. Both breakpoints for this inversion that were identified by Richards *et al.* (2005) map within scaffold_4 of the *D. persimilis* assembly, but are separated by only 1.4 Mb within the scaffold. Polytene chromosomes in hybrids between the *D. pseudoobscura* and *D. persimilis* genome strains confirm that there is a single inversion difference between the two species (S. W. Schaeffer, unpublished data). The junctions at the scaffold breaks and rejoins are supported by conserved linkage relationships among the 12 Drosophila species (S. W. Schaeffer, unpublished data).

Muller element A/D (chromosome XR, sections 18–42): Fifty-two scaffolds map to Muller A/D in *D. persimilis* for a total of 29.2 Mb of the genome (Figure 9, supplemental Table 22). The distances to the breakpoints for the fixed inversion difference between *D. pseudoobscura* and *D. persimilis* are found in supplemental Tables 21 and 22.

Muller element E (chromosome 2, sections 43–62): Five scaffolds map to Muller E in *D. persimilis* for a total of 31.7 Mb of the genome (Figure 9, supplemental Table 22). The distances to the breakpoints for the fixed inversion difference between *D. pseudoobscura* and *D. persimilis* are in supplemental Table 22.

Muller element F (chromosome 5, section 100): Five scaffolds are assigned to Muller F in *D. persimilis* for a total of 1.5 Mb of the genome (Figure 9, supplemental Table 22), but the scaffolds are not oriented to the map because of the lack of genetic or physical markers.

D. willistoni maps

Members of the *D. willistoni* species group are found in the biomes of the neotropical region and account for >80% of the drosophilid fauna collected from fruit baits. This is particularly true if we consider the ubiquity of some species of the *D. willistoni* subgroup: *D. willistoni, D. paulistorum, D. tropicalis,* and *D. equinoxialis,* especially in the Amazonian region (Martins 1987). The *willistoni* species group is a clade of ~25 described species (Bächli 2006). This group is restricted to the neotropical region and is basal to the Palaearctic and Nearctic members of the *melanogaster* and *obscura* groups. *D. willistoni* is one of the most widely distributed Drosophila species in the New World and can be found in the southern United States, throughout Central America and the Caribbean, and in southern South America, in Argentina (Patterson and Stone 1952; Spassky *et al.* 1971). Along with their sister taxon, the *saltans* group, the *willistoni* species

lack some of the defining secondary sexual characters (*e.g.,* sex combs) found in these more derived groups (O'Grady and Kidwell 2002). Gleason *et al.* (1998) have provided a phylogeny of the *willistoni* species group based on the nucleotide sequences of one mitochondrial and two nuclear genes. On the basis of their analyses, *D. willistoni* is most closely related to *D. tropicalis,* and these two taxa compose the sister group to a clade containing *D. equinoxialis, D. paulistorum,* and *D. pavlovskiana.* Phylogenies inferred from chromosome rearrangement events on Muller B, however, showed that *D. willistoni* is the sister group to *D. tropicalis, D. equinoxialis,* and *D. paulistorum* (Rohde *et al.* 2006).

The first reference map of the polytene chromosomes of *D. willistoni* was drawn by Dobzhansky (1950), who defined the standard karyotype from a strain collected from the Belém population in northern Brazil. Strains from the Belém population were relatively free from inversion variation from which a standard polytene map could be developed (Dobzhansky 1950). The salivary chromosomes were divided into 100 sections across the five chromosomal arms in the following order: XL (Muller A, 1–16), XR (Muller element D, 17–36), IIL (Muller element C, 37–55), IIR (Muller element B, 56–77), and III (Muller element F·E, 78–100).

Spassky and Dobzhansky (1950) isolated mutant strains from the Belém, Brazil, population to develop a genetic map. These genetic maps provide some useful markers for orienting the sequence scaffolds from *D. willistoni,* but these data must be viewed with caution because it is not always obvious if the new photographic maps developed for this project are homosequential with the Dobzhansky (1950) maps. *D. willistoni* has extensive gene arrangement polymorphism on all chromosomes, which has been described from chromosomal variability in natural populations (da Cunha *et al.* 1950, 1959; da Cunha and Dobzhansky 1954; Valente and Araujo 1985, 1986; Valente *et al.* 1993, 2001, 2003; Rohde 2000; Rohde *et al.* 2005).

Regner *et al.* (1996) developed the first photographic maps for *D. willistoni* using the Dobzhansky (1950) drawings to demarcate the major chromosomal sections. In parallel with those studies, *in situ* hybridization mapping studies were used to establish the homology among the different arms and delineate rearrangement breakpoints of *D. willistoni.* Bonorino *et al.* (1993) mapped the *Hsp70* locus in seven species within the *D. willistoni* subgroup and in *D. nebulosa.* The *CuZn Sod* (Rohde *et al.* 1994) and *Alcohol dehydrogenase* (*Adh*) (Rohde *et al.* 1995) genes were also mapped in this species group. Rieger (1999) mapped nine genes in *D. willistoni* (*Hsp8, Hsr-omega, Hsp 27, Ubi, BRC, E74, E75, 71E,* and *Sgs5*) using heterologous probes from the *D. melanogaster* genome.

Rohde (2000) improved the photographic maps of Regner *et al.* (1996) after the analysis of at least 10 chromosomal arms per strain from 22 isofemale lines

from populations representative of almost all geographical regions of the species. The standard arrangement of the X chromosome presented by Regner *et al.* (1996) was changed because this chromosome had a less common karyotype found throughout the geographical distribution of the species. First, autosome (chromosomes II and III) patterns were maintained according to Regner *et al.* (1996), following the order presented by Dobzhansky (1950). The maps were modified to redefine the boundaries of the numbered sections in each chromosomal arm within interband regions. Second, each section was divided into lettered subsections (starting with A in the proximal region), which allow higher precision in the description of inversion breakpoints and *in situ* localization of probes. Third, sections 81 and 80 of the third chromosome, III, were returned to their original order as described by Dobzhansky (1950). These modifications made by Rohde (2000) improved the photographic map and allowed higher precision in the description of inversion breakpoints and the *in situ* localizations.

Chromosome maps—*D. willistoni*: In the photomap presented in Figure 10, all chromosomes are oriented in the same way, with the proximal (centromeric) regions at the left and the distal (telomeric) regions at the right. Despite some modifications, the section boundaries established by Dobzhansky (1950) and Rohde (2000) were preserved. At this time, there are limited numbers of genetic and physical markers for *D. willistoni* so the orientation of the scaffolds should be viewed as provisional.

Muller element A (chromosome XL, sections 1–16): Five scaffolds map to Muller A in *D. willistoni*, comprising a total of 27.9 Mb of genomic sequence (Figure 10, supplemental Table 23). The five scaffolds were joined on the basis of the conserved synteny at the ends of the scaffolds. The orientation was determined on the basis of the location of the *br* locus, which maps close to the telomere (Spassky and Dobzhansky 1950). The order of *sn, f, sc, y, lz,* and *w* on the genetic map is consistent with their order within scaffold 181096. The genetic locations of *Notch* and *cut*, however, are at opposite ends of the scaffold map. This may result from an inversion difference between the genetic mapping strains and the genome strain. The orientation of these scaffolds should be viewed with caution.

Muller element B (chromosome 2R, sections 56–77): Eight scaffolds map to Muller B in *D. willistoni*, comprising a total of 32.3 Mb of genomic sequence (Figure 10, supplemental Table 23). The eight scaffolds were joined on the basis of the conserved synteny at the ends of the scaffolds. Two markers, *Adh* and *Cl*, provide a tentative orientation for the scaffold map, although *Adh* is more distal on the genetic map (Lakovaara and Saura 1972), again suggesting an inversion difference between the genome strain and the mapping strain.

Muller element C (chromosome 2L, sections 37–55): Eight scaffolds map to Muller C in *D. willistoni*, comprising a

total of 29.6 Mb of genomic sequence (Figure 10, supplemental Table 23). The eight scaffolds were joined on the basis of the conserved synteny at the ends of the scaffolds. One assembly error was detected for one of the scaffolds on this arm, scaffold 181009. Nucleotides 1–2,748,112 had 231 orthologous gene calls on Muller C and nucleotides 2,841,197–3,492,693 had 80 orthologous gene calls on Muller D. This scaffold was split between the two Muller elements. Two distal genetic markers *px* and *bw* suggest the orientation of the scaffold map to the polytene chromosomes.

Muller element D (chromosome XR, sections 17–36): Eight scaffolds map to Muller D in *D. willistoni*, comprising a total of 29.6 Mb of genomic sequence (Figure 10, supplemental Table 23). The eight scaffolds were joined on the basis of the conserved synteny at the ends of the scaffolds. Bases 2,841,197–3,585,778 in scaffold 181009 map to Muller D. The genetic and physical markers do not clearly resolve the orientation of the scaffolds to the cytogenetic map. A tentative placement is shown in Figure 10.

Muller element F·E (chromosome 3, sections 78–100): Three scaffolds map to Muller E/F in *D. willistoni*, comprising a total of 33.7 Mb of genomic sequence (Figure 10, supplemental Table 23). The three scaffolds were joined on the basis of the conserved synteny at the ends of the scaffolds. Base 1 of scaffold 181130 was anchored to the centromeric region of chromosome 3 on the basis of the three genes from the fused dot chromosome that hybridize to section 78 on the polytene map (Papaceit and Juan 1998). Probes for the *Xdh* locus map to the central region of the chromosome, which is consistent with the location of *Xdh* within scaffold 181089. Thus, the orientation of the F·E scaffolds is the best supported of the five major chromosomal arms of *D. willistoni*.

The junction between Muller F and E genes is located in scaffold 181130 and is defined between genes CG34036-PA of F and CG17119-PA of E and between nucleotides 2,014,728 and 2,029,101. The 14-kb sequence has no assembly gaps and is 64.7% A + T. This region was used in a BLASTN search to determine if the sequence provides any clues about the fusion and rearrangement process. The BLASTN comparison of the E·F junction region with the *D. willistoni* genome reveals a repeat of 107 bp (2,022,283–2,022,362) and 141 bp (2,024,912–2,025,052) that are found ubiquitously throughout the genome. The majority of hits within *D. willistoni* match unplaced scaffolds, suggesting that this region may have been heterochromatic at one time. These repeats are also found in the Drosophila genomes, but the length of the repeats is 47 bp. A BLASTN search of the sequence between the two repeats finds a match to a *P* element of *D. sturtevanti* (GenBank accession no. AY578784). This could suggest a role for transposable elements in the F·E fusion and subsequent rearrangement.

Muller A (Chr XL) 27.9 Mb

Muller D (Chr XR) 29.6 Mb

FIGURE 10.—Alignments of the *D. willistoni* sequence scaffolds with the polytene chromosome maps. For each chromosome, the four maps are the genetic map, the scaffold map showing markers on the genetic and chromosomal maps, the ideogram of the chromosome showing section designations from the salivary chromosomes reference map of DOBZHANSKY (1950), and the salivary chromosome map with the section and subsection designations and with the photomap of ROHDE (2000). "CL" drawn across a scaffold junction indicates that the join is supported by conserved linkage information in at least one other species. Supplemental Table 15 provides the annotation for the numbered scaffolds. The ideograms from DOBZHANSKY (1950, Figure 9) were reproduced with kind permission of Oxford University Press.

A Drosophila virilis

Muller Element A (Chr X) 30.4 Mb

E Muller Element E (Chr 2) 35.3 Mb

D Muller Element D (Chr 3) 26.6 Mb

FIGURE 11.—Alignments of the *D. virilis* sequence scaffolds with the polytene chromosome maps of *D. virilis*. Scaffolds derived from the whole-genome assembly are indicated as blocks positioned on the basis of *in situ* localization (below) and linkage relationships (above) of markers, which serve as anchor points for the sequence. Comparative genomic analyses also oriented the scaffolds on the basis of conserved linkage relationships. Each scaffold is identified by a unique element ID corresponding to information in supplemental Table 15. Darker shading of a scaffold indicates agreement between the physical and comparative genomic approaches, whereas lighter shading of a scaffold indicates that the orientation is based solely on conserved linkage or solely on marker data. Adjacency of scaffolds represents conserved linkage blocks. Chromosomal positions of terminal markers in each scaffold are indicated and labeled with their identity. Shading represents regions with additional anchor positions within the scaffold with only a subset indicated. Gene symbols are based on existing nomenclature for *D. virilis* with the corresponding gene symbol for *D. melanogaster* in parentheses. Each anchor position is included as supplemental Table 24. The ideograms are from the drawings by KRESS (1993, Figure 1) and the photographic maps are from (GUBENKO and EVGEN'EV (1984, Figure 2); both are reproduced with permission from Springer Science Business and Media.

SUBGENUS DROSOPHILA

D. virilis maps

Analyses of genes and genomes are a primary focus of research efforts on *D. virilis* and its close relatives. The arrangement of chromosomes in *D. virilis* represents the inferred karyotype of the common ancestor of the genus Drosophila (reviewed by CLAYTON and GUEST 1986), and therefore, it has been used frequently as an outgroup for a variety of studies of the Sophophora lineage, especially for comparative sequence analyses of genes in *D. melanogaster*. Each chromosomal element in the *D. virilis* genome exists as an independent acrocentric arrangement (or rod) with a near-terminal centromere, so the diploid number is 12, including the small pair of dot chromosomes. Homology between these individualized elements and the chromosomal arms of the metacentric chromosomes of *D. melanogaster* was demonstrated by early studies using cross-species hybridization (LOUKAS and KAFATOS 1986; WHITING *et al.* 1989), thus providing molecular confirmation of the classically inferred homology among Muller elements. The relationships between the Muller element and the numbering of the six chromosomes of *D. virilis* are (Muller element = chromosome) A = X; B = 4, C = 5, D = 3; E = 2; and F = 6.

Several investigators produced the original drawn maps of the polytene chromosomes of *D. virilis* and developed different nomenclatures for the banding patterns (FUJII 1936, 1942; HUGHES 1936, 1939; PATTERSON *et al.* 1940b; HSU 1952). However, the photographic map and nomenclature developed by GUBENKO and EVGEN'EV (1984) has been adopted as the standard for physical mapping on the polytene chromosomes of *D. virilis*. Under their system, the X chromosome is divided into 19 sections numbered consecutively with section 1 beginning at the telomere and section 19 ending at the proximal junction between the polytene chromosome and chromocenter. Each section of the X is further subdivided into regions A–D. The major autosomes are divided into 10 sections each with further subdivision designated by lettering, and the dot chromosome consists of a single section, 60, with lettered subdivisions. This demarcation of the polytene bands and associated nomenclature was further refined in the graphic map developed by KRESS (1993). The nomenclature for the banding pattern in the polytene chromosomes contained in these maps has been used almost exclusively in physical mapping studies of *D. virilis*. Thus it is unnecessary to consider older representations of the polytene chromosomes with alternative nomenclatures; also, it would create unnecessary confusion to develop a new nomenclature.

Over 75 different *D. virilis* gene sequences have been obtained in studies of conservation within and among genes to identify functional constraints. Many investigators have used a widely shared λ-phage genomic library (BLACKMAN and MESELSON 1986) to isolate and sequence a large portion of these genes. In addition to revealing patterns of sequence conservation and divergence within and flanking single genes, conservation of nested gene structures has also been demonstrated (KAYMER *et al.* 1997). However, early studies of gene arrangement quickly revealed that the appearance of highly integrated overlapping gene structures in *D. melanogaster* is not necessarily an indication of evolutionary conservation, given the plethora of genome rearrangements between these genomes (NEUFELD *et al.* 1991; VON ALLMEN *et al.* 1996). Performing comparative analyses to identify conserved regions of genes and gene clusters in *D. melanogaster* is now a straightforward endeavor, given the availability of complete genome sequences of *D. virilis* and 11 other Drosophila species (DROSOPHILA 12 GENOMES CONSORTIUM 2007).

Emergence of a sparse physical map of loci distributed throughout the genome of *D. virilis* was one indirect outcome of these comparative sequence analyses that frequently involved *in situ* localization of the genes. In addition to these individually mapped genes, a dense physical map was constructed by D. Hartl's lab using *in situ* hybridization to localize large-insert P1 clones at single chromosomal positions (LOZOVSKAYA *et al.* 1993; VIEIRA *et al.* 1997b). Although sequences were originally obtained from only select clones containing a few targeted genes, microsatellite loci and putative gene regions have since been identified from partial sequences of the inserts of other mapped P1 clones (SCHLÖTTERER 2000; HUTTUNEN and SCHLÖTTERER 2002; MCALLISTER and EVANS 2006; VIEIRA *et al.* 2006). Comparative genomic analyses using *D. virilis* as one reference species have also added loci to the physical map (VIEIRA *et al.* 1997a; RANZ *et al.* 1999; PÄÄLLYSAHO *et al.* 2001). Physically mapped positions within the genome and their associated sequences represent a critical link for associating an assembled genome sequence with the underlying arrangement of chromosomes. The dense physical map of *D. virilis* provides a unique opportunity to independently evaluate the accuracy of the assembly and to orient the sequence along each chromosome arm.

Availability of a well-resolved physical map made *D. virilis* a logical choice for genome sequencing, but this species could have presented unique challenges in whole-genome sequencing because of a much larger genome size relative to other Drosophila species. The total genome size of *D. virilis* is estimated to be about twice that of *D. melanogaster* (KAVENOFF and ZIMM 1973; LAIRD 1973). However, ~40% of this enlarged genome is composed of highly repetitive satellite DNAs (GALL *et al.* 1971; SCHWEBER 1974) that are likely sequestered in the heterochromatic regions adjacent to the centromeres of its acrocentric chromosomes (BECK 1977). Consistent with an unequal expansion of heterochromatic and euchromatic genomic regions, a comparison

of a limited number of introns showed their mean length to be only 39% longer in *D. virilis* relative to *D. melanogaster* (Moriyama *et al.* 1998). Enlargement of both heterochromatic and euchromatic genomic regions persists despite there being a deletion bias that causes a high intrinsic rate of loss of repetitive sequences (Petrov *et al.* 1996; Petrov and Hartl 1998). Availability of the sequence of the considerably enlarged genome of *D. virilis* provides a substrate for gaining insight into the regulation of genome size and the asymmetric portioning of this excess into heterochromatic and euchromatic domains.

Although the distinction between heterochromatin and euchromatin is quite strong in the Drosophila genome, and changes in gene position relative to these domains leads to the well-known phenomenon of position-effect variegation, initial comparative studies indicate little conservation in gene content of heterochromatin between *D. virilis* and *D. melanogaster.* Both *light* and *Dbp80* are heterochromatic in *D. melanogaster,* but are located within euchromatic regions of *D. virilis* (Yasuhara *et al.* 2005; Schulze *et al.* 2006). The *RpL15* gene adjacent to *Dbp80* in *D. melanogaster* does appear to be a conserved heterochromatic gene; however, it has moved between chromosomal elements in the *melanogaster* group, and independently, it has moved to a euchromatic position in *D. pseudoobscura* (Schulze *et al.* 2006). Flux in the gene content of the heterochromatin of the major chromosomes contrasts with conservation in the content of the unique dot chromosome (Riddle and Elgin 2006; Slawson *et al.* 2006). Another conserved chromosomal feature is the presence of TART elements at the telomeres (Casacuberta and Pardue 2003). Organization of the assembled genome sequence of *D. virilis* relative to markers in the euchromatic regions of the polytene chromosomes will provide initial guidance for identifying further movement of genes between these chromatin domains and will serve as a reference for conducting comparative genomic analyses in the subgenus Drosophila.

Chromosome maps—*D. virilis:* Availability of a dense physical map, coupled with the identification of conserved syntenic groups at scaffold edges, suggests that most of the euchromatic coding genes are present in 22 scaffolds containing ~150.7 Mb of sequence organized along each chromosomal arm. Furthermore, 10 scaffolds containing ~12.2 Mb of sequence and likely representing centromeric regions of specific chromosomal elements are also identified (supplemental material), but these scaffolds are not oriented along the chromosomes.

Muller element A (chromosome X, sections 1–19): The X chromosome of *D. virilis* has a high marker density due to the large number of individually mapped genes. A total of 105 physically mapped markers anchor six scaffolds along the X chromosome (Figure 11), and 14 additional markers are present in the genome

sequence at positions that are inconsistent with their reported chromosomal location (supplemental Table 24). Because the small subset of markers showing inconsistencies are distributed throughout the chromosome, these disagreements most likely arise from incorrect assignment of map positions—not from errors in the assembled sequence. The set of markers that order and orient the scaffolds span cytological divisions 1A at the telomere to 19C near the centromere, which corresponds to sequences in scaffold 13042 (A1) and scaffold 12970 (A6), respectively. Only 5 markers are represented in the centrally positioned 0.8-Mb scaffold 12472 (A5). Although these markers position the scaffold at cytological positions 13B and 13C, the orientation of the scaffold is not clearly delimited with these mapped positions. The analysis of conserved syntenic gene arrangement orients this scaffold by inferring joins with the flanking scaffolds. Position and orientation of the other five scaffolds is strongly supported by physically mapped markers, by loci within the linkage map, and by conserved syntenic gene arrangement at the junctions between scaffolds (Figure 11). Conservation of syntenic groups identified in the other sequenced genomes indicates no missing genes in the gaps that currently exist at the junctions between the scaffolds. These six scaffolds positioned on the X chromosome contain 30.5 Mb of the assembled genome sequence.

Correspondence between the linkage map and the assembled genome sequence is the strongest for the X chromosome. Six intervals in the linkage map are represented by loci identified within the same scaffold; thus, these intervals provide an estimate of the relationship between physical distance and recombination rate. An average 131 kb/cM was measured from an approximately twofold range from 75 to 158 kb/cM. This high level of recombination had been previously inferred from the overall length of the 170-MU linkage map of element A. Interestingly, the highest recombination rate is estimated for the interval flanked by *yellow* and *scute* near the telomere of the X, which is a region that exhibits an unusually low level of recombination in populations of *D. melanogaster* collected in North America (Aguadé *et al.* 1989).

Muller element B (chromosome 4, sections 40–49): A total of 90 physically mapped markers anchor three scaffolds on chromosome 4 of *D. virilis* (Figure 11; supplemental Table 24). The most distal marker, corresponding to a P1 clone with sequence in scaffold 13246 (B1), is located in the telomeric cytological band 40A. The most proximal marker is the tandemly duplicated *Adh* gene at 49B (Nurminsky *et al.* 1996; Charlesworth *et al.* 1997), which is present in scaffold 12723 (B3). Twelve additional markers have reported map positions along chromosome 4, but their position in the assembly is inconsistent with flanking markers and they likely represent cases of erroneous map positions (supplemental Table 24). Conservation of syntenic blocks

orients these scaffolds in the same manner as the mapping data, and the conservation of gene order at scaffold edges indicates that known coding genes are not missing in the gaps. Therefore, the three scaffolds (13246, B1; 12963, B2; and 12723, B3), which represent 28.7 Mb of the assembly, appear to contain the entire euchromatic region of chromosome 4 (Figure 11).

Interestingly, loci on the linkage map of chromosome 4 are not congruent with the arrangement of the scaffolds inferred with physical markers and conserved synteny. One endpoint of the linkage map, *plexus* (*Dvir\px*) at the zero position, is located near the distal end of the genome sequence within scaffold 13246 (B1) using the putative ortholog *Dmel\net* as a BLAST query (Figure 11). At the proximal end, the terminal locus *black* (*Dvir\bl*) corresponds to the appropriate position within scaffold 12723 (B3), using the putative homolog *Dmel\b* to locate the corresponding gene within the genome sequence. Intermediate positions on the linkage map, however, are not consistent with this arrangement (supplemental Table 24). This inconsistency has several possible causes, such as errors in the assembly, incorrect assignment of homology between mutant phenotypes, or errors in the linkage map. Appearance of systematic disagreement between positions in the linkage map and the oriented sequence indicates that reported positions on the linkage map of chromosome 4 may comprise two linkage groups constructed in opposite directions along this chromosome.

Muller element C (chromosome 5, sections 50–59): Four scaffolds of the assembled genome sequence are oriented on chromosome 5 using physical map data representing 68 markers (Figure 11). Mapped positions span the entire polytene chromosome map from the telomere at cytological band 50A, which corresponds with the sequence of scaffold 12823 (C1), to the pericentromeric heterochromatin at cytological band 59F, which corresponds to the sequence of scaffold 13324 (C4). Conserved syntenic gene arrangement identified only a single scaffold join on chromosome 5, represented by the telomeric scaffold (12823, C1) and the adjacent subtelomeric scaffold 10324 (C2).

As indicated in supplemental Table 24, five markers distributed throughout chromosome 5 are located in the genome sequence at positions that are inconsistent with their reported position on the physical map. More importantly, a block of markers contained in the large scaffold positioned at the center of chromosome 5 exhibits a shared inconsistency with their reported map positions, thus indicating a likely case of an error in the assembly (Figure 11). A block of markers localized on the cytological map at cytological subdivisions 51B/C corresponds with the interval between positions 636 and 906 kb in the sequence of scaffold 12875. However, starting at position 1732 kb–20,532 kb, this scaffold is oriented, respectively, from cytological band 59E to 51F. This disagreement between the

sequence and groups of markers indicates that scaffold 12875 does not represent the correct organization of chromosome 5. Furthermore, conservation of a syntenic group consisting of putative orthologs of *Dmel\Rya-r44F* at position 1735 kb in scaffold 12875 and *Dmel\CG8740* near the end of scaffold 13324 is consistent with an erroneous join within or between contigs in scaffold 12875 near position 1730 kb. Thus, base 1–~1700 kb of scaffold 12875 is likely to be oriented in the minus direction around 51 B/C (designated as C3′ in Figure 11). This corrects the apparent discrepancy between widely separated positions in the sequence corresponding with markers located within cytological section 59. One of these markers is a P1 clone from a library with an average insert size of 65.8 kb (Lozovskaya *et al.* 1993). This P1 clone has one end sequence located at position 1732 kb in scaffold 12875 (C3) and the other near the edge of scaffold 13325 (C4), which further suggests a join between an internal position within scaffold 12875 and scaffold 13324.

Although this apparent error in the assembly of scaffold 12875 affected the inference of junctions between scaffolds on the basis of the conserved syntenic groups, no other scaffolds (>5 kb) contain putative orthologs of genes on element C. Therefore, these four scaffolds, which contain a total of 27.3 Mb, apparently represent the sequence of chromosome 5 without any large gaps in the regions between the organized scaffolds.

Muller element D (chromosome 3, sections 30–39): The largest scaffold in the assembled genome sequence (scaffold 13049 at 25.1 Mb) is anchored onto chromosome 3 with 107 markers distributed between cytological bands 30B and 39F (D4 in Figure 11). Two additional scaffolds contain markers mapping to the telomeric subdivisions 30A and 30B; however, the single marker in the most distal scaffold (10322, D1) does not orient the sequence, and the markers in the subtelomeric scaffold (12758, D3) provide only a tentative arrangement of this sequence. Conservation of syntenic groups, however, resolves the order and orientation as a single block consisting of these scaffolds and an additional small scaffold in the subtelomeric region (D2). Genes present in the syntenic group identified from the other genome sequences are missing at the inferred join between D3 and D4, which is represented by the gap in Figure 11. Overall, four scaffolds encompassing a total of 26.7 Mb of the assembled genome contain the sequence of chromosome 3.

The linkage group representing chromosome 3 is associated with the genome sequence through two loci with clear homologies to genes in *D. melanogaster* (Figure 11). One of the loci used to position the linkage map is the *cinnabar* locus of *D. virilis*, which was previously considered by Sturtevant and Novitski (1941) to be homologous with the *scarlet* locus of *D. melanogaster*. The reverse homology (*i.e., Dvir\st* with *Dmel\cn*) is evident

on the linkage map of element C (Figure 11). Homology between the *short veins* (*sv*) locus of *D. virilis* and the *rhomboid* (*Dmel\rho*) locus of *D. melanogaster* provides a second anchor for the linkage map. These two loci are located ∼2.1 Mb apart in the assembly and separated by 15.5 MU, which is consistent with the length of a centimorgan estimated from the X chromosome.

Muller element E (chromosome 2, sections 20–29): Chromosome 2 is the longest, and the 35.5 Mb of sequence contained in the scaffolds placed on this chromosome reflects this feature. Three scaffolds are oriented along chromosome 2 using 134 physically mapped markers spanning the region between 20C and 29E (Figure 11). An additional 15 markers are located in the sequence at positions that are inconsistent with their reported chromosomal location (supplemental Table 24). The syntenic analysis indicated that genes are missing at the inferred scaffold joins between scaffold 12822 (E1) and 13047 (E2) and between scaffold 13047 (E2) and scaffold 12855 (E3). However, the comparative analysis places an additional scaffold (12954, E4) in the proximal euchromatin at the base of chromosome 2 without any indication of missing genes between these scaffold edges.

Muller element F (chromosome 6, section 60): Extensive mapping and sequencing of the dot chromosome of *D. virilis* by S. Elgin and colleagues (SLAWSON *et al.* 2006) provides a unique resource for identifying and orienting the sequence of this chromosome. An ∼2-Mb single scaffold (13052, F1) appears to contain the entire assembled sequence of chromosome 6 (Figure 11). Furthermore, this is the only scaffold identified as containing genes located on element F of *D. melanogaster.*

D. mojavensis maps

D. mojavensis is a member of the *mulleri* complex of the highly speciose *repleta* group that underwent an explosive radiation in the New World (WASSERMAN 1982). All of the >24 *mulleri* complex species are cactophilic, inhabiting regions and hosts that are inhospitable to many other Drosophila species. The polytene map of *D. repleta* was employed as the standard for all *repleta* group species (WASSERMAN 1962). To create maps for other members of the *repleta* group, Wharton's original drawings were physically cut and rearranged to suit the banding patterns observed microscopically and in photomicrographs, and the original nomenclature for the *D. repleta* map was retained in other species. It then was possible to tabulate the nature and number of chromosomal variations observed in the *repleta* radiation. Of 240 chromosomal rearrangements in the *repleta* group, 235 were found to be paracentric inversions and 4 were centric fusions (WASSERMAN 1982).

On the basis of the changes in the *repleta* map, WASSERMAN reconstructed the mutational events ancestral to the *mulleri* complex (WASSERMAN 1960) and *D. mojavensis* (WASSERMAN 1962). The closest relative of *D.*

mojavensis, D. arizonae, differs from it by three fixed inversions in the X (Muller element A), second (Muller E), and third (Muller B) chromosomes. In *D. mojavensis,* polymorphism for chromosomal inversions exists in chromosomes 2 (five inversions) and 3 (two inversions) and varies among populations (METTLER 1963; JOHNSON 1980). Although chromosomes 4, 5, and 6 are colinear between the two species, and reproductive isolation is incomplete, there is no evidence for introgression in either direction (COUNTERMAN and NOOR 2006; MACHADO *et al.* 2007b).

D. mojavensis was described by PATTERSON *et al.* (1940a) from a specimen found in southern California. Since then, three other populations have been discovered (Sonora, Mexico; Baja California, Mexico; Santa Catalina Island, CA) breeding in different host cactus species at the regional level. The four geographic host races show genetic differentiation from one another (REED and MARKOW 2004; ROSS and MARKOW 2006; MACHADO *et al.* 2007a) along with varying levels of reproductive isolation among themselves and with their sibling species, *D. arizonae* (VIGNEAULT and ZOUROS 1986; MARKOW 1991; REED and MARKOW 2004). These features of *D. mojavensis* have made it a popular and useful model species for studies of adaptation (MATZKIN *et al.* 2006) and speciation (MARKOW and HOCUTT 1998).

The strain of *D. mojavensis* utilized in the genome-sequencing project was derived from an isofemale strain collected on Santa Catalina Island that is fixed for the standard gene arrangements for chromosomes 2 and 3. The gene arrangements in the salivary chromosomes of this strain form the basis for the current physical and genetic maps for *D. mojavensis*. We constructed a standardized polytene chromosome map for the karyotype of the *D. mojavensis* strain from Catalina Island with the aid of our own photomicrographs as well as the modified (RUIZ *et al.* 1990) original drawings of WHARTON (1942). Chromosome numbers were maintained as in Wharton's original *D. repleta* maps: chromosome 1 = X or element A; 2 = E; 3 = B; 4 = D; 5 = C; and 6 = F (dot). At this time, we have omitted chromosome 6 from the map. New nomenclature was created by dividing each element into 20 sections of roughly similar size. The sections were demarcated by sharp bands. The 20 sections were numbered from the distal to the proximal end for each chromosome. Each numbered section then was further subdivided (distal to proximal) into four lettered sections, again demarcated by prominent bands in photographs.

Chromosome maps—D. mojavensis: *Muller element A (chromosome X, sections 1–20):* Five scaffolds map to Muller A, comprising 32.0 Mb of sequence that localizes to cytological regions 1–20. Scaffold 6473, 16.9 Mb, was anchored to the proximal end of the chromosome (19B) (Figure 12, supplemental Table 25). Computational approaches were used to determine the positions of scaffolds 6328, 6308, and 6359, although they have
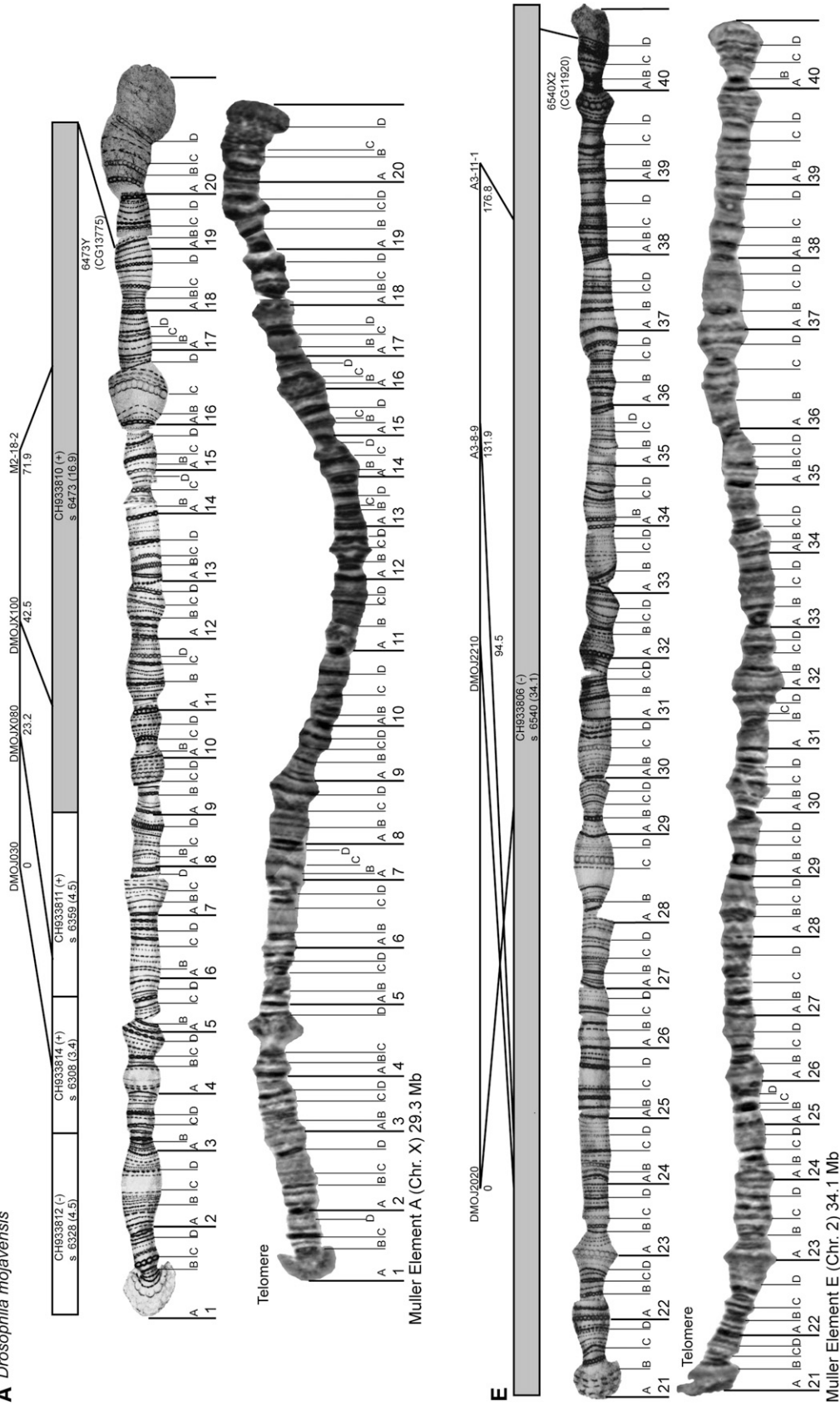
FIGURE 12.—Alignments of the *D. mojavensis* sequence scaffolds with the polytene chromosome maps of *D. mojavensis*. Organization of assembled genome sequence on polytene chromosomes of *D. mojavensis*. Scaffolds derived from the whole-genome assembly are indicated as blocks positioned on the basis of *in situ* localization (below) and linkage relationships (above) of markers (recombination units shown), which serve as anchor points for the sequence. With the exception of linkage locus A3-10-13, all linkage loci were found in the assembly. Comparative genomic analyses also oriented the scaffolds on the basis of conserved linkage relationships. Each scaffold is identified by a unique element ID with GenBank accession number; direction (+ or − relative to telomere and a "?" if direction is unknown), scaffold ID, and overall size (in megabases) indicated. Shading of a scaffold indicates agreement between the physical and comparative genomic approaches and open areas of a scaffold indicate that the orientation is based solely on conserved linkage. Adjacent scaffolds represent conserved linkage blocks. Chromosomal positions of terminal markers in each scaffold are indicated and labeled by their probe name, and the corresponding gene symbols (based on the nomenclature for *D. melanogaster*) are shown below in parentheses.

FIGURE 12.—*Continued.*

Figure showing Drosophila grimshawi polytene chromosome maps. A: Muller Element A (ChX), 24.5Mb. B: Muller Element B (Ch3), 26.7Mb. C: Muller Element C (Ch2), 23.5Mb.
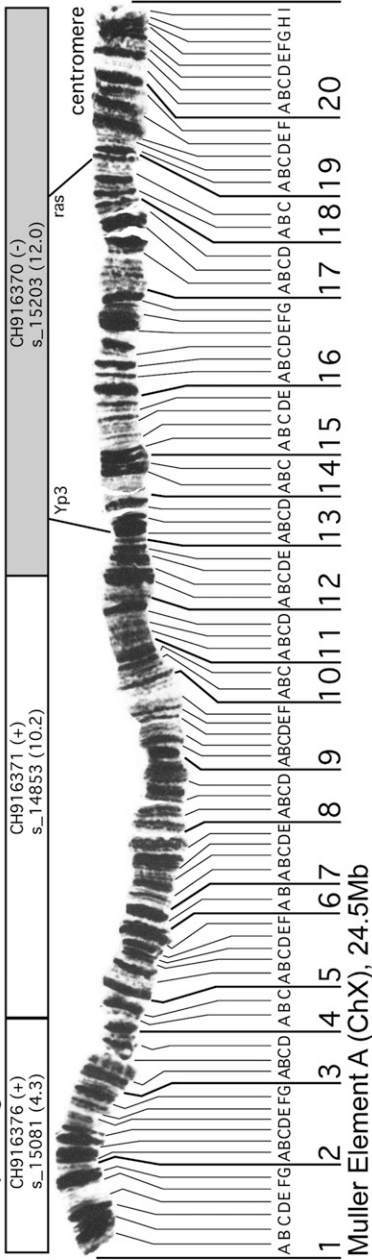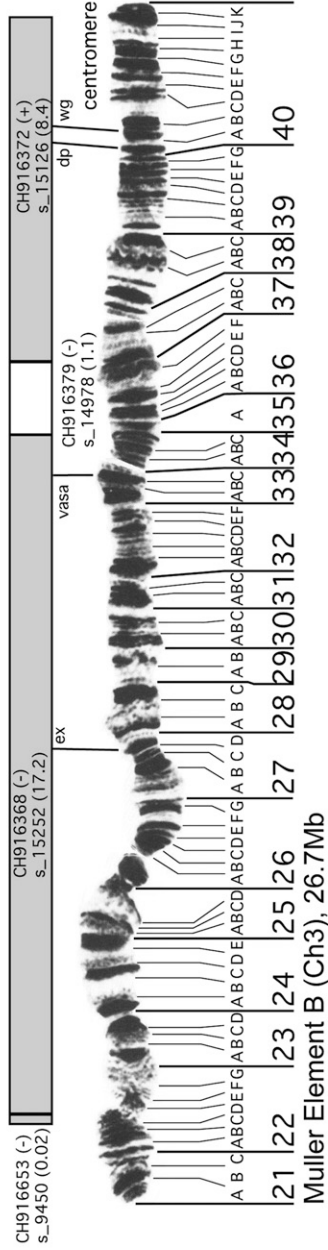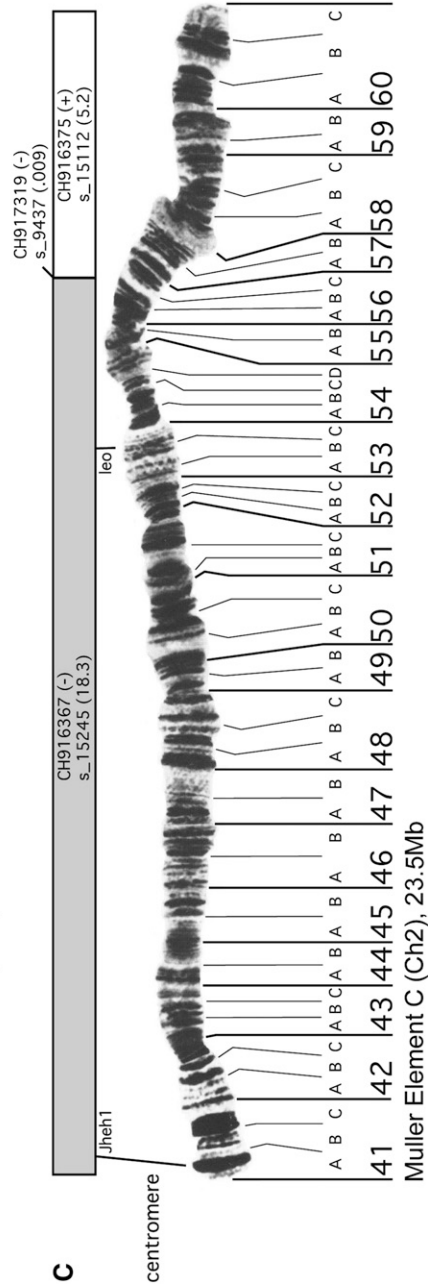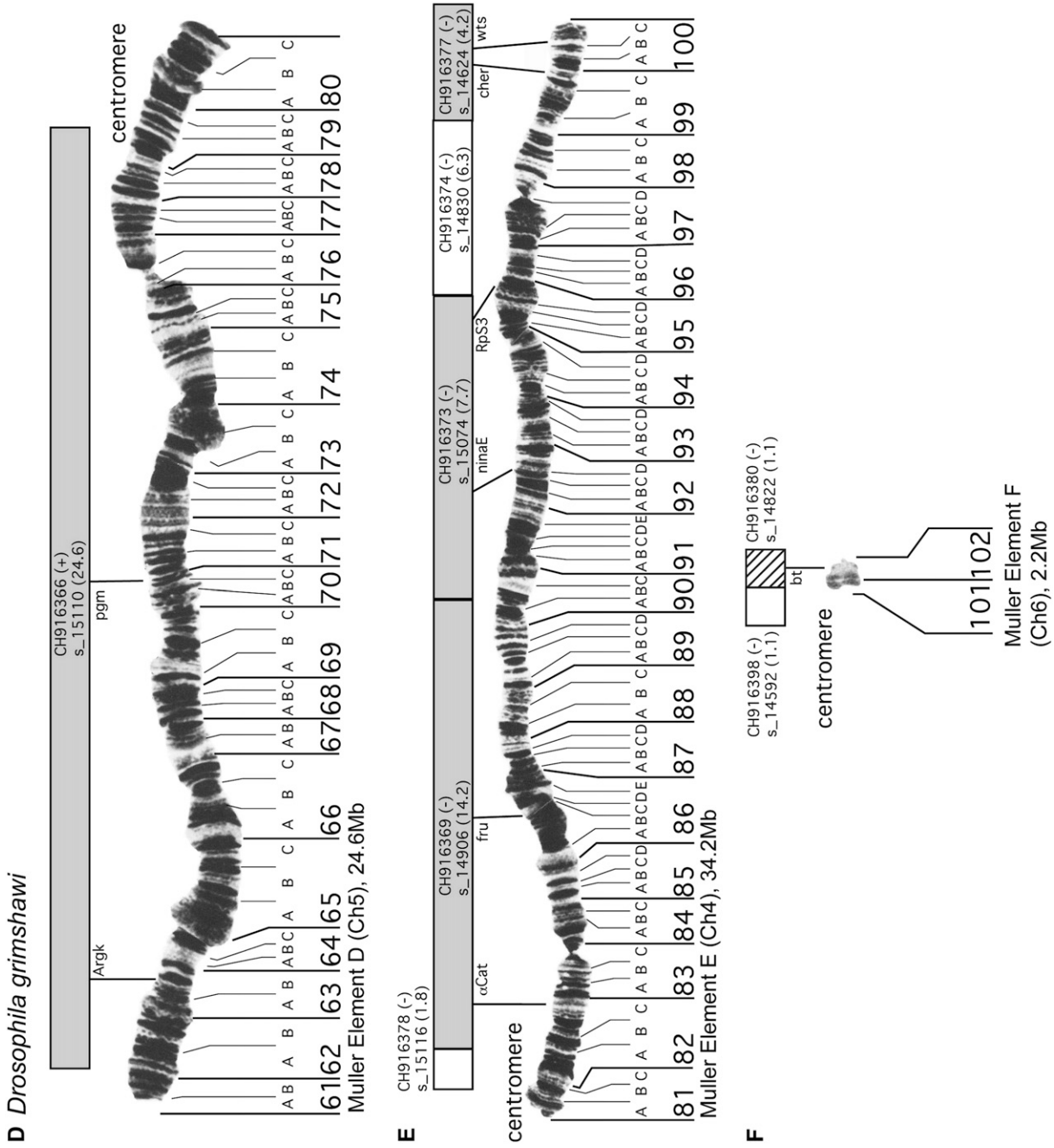
FIGURE 13.—Alignments of the *D. grimshawi* sequence scaffolds with the polytene chromosome maps of *D. grimshawi*. For each chromosome, the two maps are the scaffold map showing markers on the genetic and chromosomal maps and the salivary chromosome map with the section and subsection designations. The photographic map is from CARSON *et al.* (1992, Figures 1–5) and is reproduced with kind permission of CRC Press.

not been anchored yet to the distal end of element A. The linkage markers further support the proposed relationship between the scaffolds. Two of the four loci, DMOJX030 and DMOJX080, are found within scaffolds 6308 and 6359, respectively.

Muller element B (chromosome 3, sections 41–60): One scaffold maps to Muller B, comprising 32.4 Mb of sequence that localizes to cytological regions 41–60 (Figure 12, supplemental Table 25). The scaffold, 6500, was anchored near its distal end to 44B. The relationships of five markers were established for chromosome 3. No BLAST hits were found for marker A3-10-13 and hence its relationship on the scaffold cannot be determined. Marker A1-2-1 (not shown) is located 36.7 cM units proximal to M2-17-5 and has a BLAST hit to scaffold 6499, a small (421 kb) scaffold of unknown orientation and placement relative to scaffold 6500.

Muller element C (chromosome 5, sections 81–100): One scaffold maps to Muller C, comprising 26.9 Mb of sequence that localizes to cytological regions 81–100 (Figure 12, supplemental Table 25). The single scaffold 6496 was anchored at the telomeric (distal) end to 82A. The genetic map consisted of five markers. Marker A3-12-6 does not occur in the same order between the linkage map and the genome sequence. This discrepancy cannot be checked on an independent map because the previous linkage map (STATEN *et al.* 2004) lacks sufficient information about linkage on the fifth chromosome.

Muller element D (chromosome 4, sections 61–80): Two scaffolds map to Muller D, comprising 27.3 Mb of sequence (Figure 12, supplemental Table 25). Scaffold 6680 was 24.7 Mb and was localized at the proximal end by two different probes to 80C and at the distal end to 63A. Seven genetic markers were localized in chromosome 4. The linkage map of chromosome 4 is composed of seven markers. The linkage discrepancy on the proximal end of chromosome 4 (DMOJ4040 and DMOJ4050) is likely due to statistical variance in the map over fairly long genetic intervals. The STATEN *et al.* (2004) linkage map agrees with the genome sequence in that region. It is possible that the discrepancy at the distal end (A2-13-1 and M2-19-2) reflects inaccurate orientation of scaffold 6654.

Muller element E (chromosome 2, sections 21–40): Chromosome 2 has a single scaffold, 6540, which comprises 34.1 Mb of sequence and contains cytological regions 21–40 (Figure 12, supplemental Table 25). The orientation of scaffold 6540 was determined by the *in situ* hybridization of probe 6540X2 to position 40D at the proximal end of the assembled sequence. Four markers were used in the genetic map. On chromosome 2, the DMOJ2020 marker maps to the central portion of scaffold 6540, but to the distal end of the linkage group. We reanalyzed a previous *D. mojavensis* linkage map (STATEN *et al.* 2004), which contains some common markers and found comparable alignment discrepan-

cies with the genome sequence in that region of chromosome 2. Thus, the scaffold sequence for Muller E should be viewed with caution.

Muller element F: Muller element F is not included in our analysis because we failed to visualize the chromosome in any of the polytene chromosome preparations. A single scaffold comprising 3.4 Mb of sequence is assigned to Muller F on the basis of conserved linkage.

D. grimshawi maps

Members of the Hawaiian Drosophila are a premiere example of adaptive radiation in nature. This group contains nearly 1000 species, most of which display extreme sexual dimorphism, elaborate courtship and mating displays, and a high degree of host plant specificity (O'GRADY 2002). This Hawaiian Drosophila lineage is placed in the subgenus Drosophila where it is sister to the *virilis–repleta* radiation (MARKOW and O'GRADY 2006). *D. grimshawi*, a species of Hawaiian Drosophila, is placed within the "picture wing" species group, so named because of the spectacularly pigmented wings that are used in courtship and mating displays (EDWARDS *et al.* 2007).

There are currently 112 described picture wing species, the majority of which have been included in Hampton Carson's polytene chromosome phylogeny (CARSON 1992). *D. grimshawi* carries the standard chromosomal arrangement for this phylogeny; all other picture wing gene arrangements can be reached via a combination of 228 naturally occurring inversions (CARSON 1992; CARSON *et al.* 1992). CARSON *et al.* (1992) provide detailed maps of polytene chromosome inversions in several species of Hawaiian Drosophila, including *D. grimshawi*.

Chromosome maps—*D. grimshawi*: The dearth of *in situ* hybridization data for *D. grimshawi* posed a unique problem for anchoring genomic scaffolds to the polytene chromosomes. The extensive chromosome phylogeny work of Carson and colleagues (CARSON 1992; CARSON *et al.* 1992), coupled with a low frequency of between-species chromosomal variation, allowed *in situ* hybridizations in related species (*D. heteroneura, D. silvestris,* and *D. nigribasis*) to assist in localizing genes within *D. grimshawi*. Polytene chromosome band assignments were made for a total of 29 loci (supplemental Table 26). *D. melanogaster* homologs (supplemental Table 26) for these taxa were used to query the *D. grimshawi* genome (ALTSCHUL *et al.* 1997) and to assign scaffolds to a specific chromosome (supplemental Tables 15 and 26).

Muller element A (chromosome X, sections 1–20): Scaffold 15203 localized to Muller's A on the basis of three loci (Figure 13, supplemental Table 26). The yolk protein gene (*Yp3*) maps to position 13B in *D. grimshawi*. The *ras* locus maps to position 18C on the X chromosome of *D. silvestris* and the *per* locus is near position 17B in

D. nigribasis. The X chromosome of *D. grimshawi* differs from *D. nigribasis* by 9 inversions and from *D. silvestris* by 10 inversions. Of these, 4 occur outside the regions being mapped and therefore are not important for the purposes of placing these scaffolds. The five (*D. nigribasis*) or six (*D. silvestris*) inversions that do occur within the region mapped alter the arrangement of the polytene bands between the *D. grimshawi* standard and the other two taxa; these species are not homosequential within this region. However, the relative order of the *Yp, per,* and *ras* markers are the same in all three species. We can therefore conclude that the orientation of the contig will be the same, even though gene order between 13B and 18C is not conserved. Two additional scaffolds, 14851 and 15081, were anchored by synteny. Together, these account for a total of 26.4 Mb of sequence using these three markers.

Muller element B (chromosome 3, sections 21–40): Two scaffolds, 15252 and 15126, which together account for a total of 25.5 Mb, can be placed and oriented using six markers (Figure 13, supplemental Table 26). Two markers were mapped by *in situ* hybridization in *D. grimshawi* (*ex* and *vasa*). An additional four markers were mapped in *D. silvestris* (*Pgk, Pez*) and in *D. heteroneura* (*dp, wg*). Element B differs in two chromosome inversions between *D. grimshawi* and these other taxa. The 3m inversion is between bands 23A and 25D. The 3d inversion spans 35A to 38C. None of the markers mapped in *D. silvestris* or *D. heteroneura* localize to these regions so the band identities are identical. Computational analysis of synteny places joins scaffolds 15252 and 15126 via scaffold 14978. A fourth scaffold, 9450, is joined to the distal end of scaffold 15252, creating a superscaffold 26.7 Mb in length.

Muller element C (chromosome 2, sections 41–60): Scaffold 15242, accounting for 18.3 Mb of sequence, has been placed and oriented to Muller's C (Figure 13, supplemental Table 26). Although there are no markers mapped in *D. grimshawi*, a total of seven loci have been mapped by *in situ* hybridization. Three loci (*leo, lola, vg*) have been placed in *D. heteroneura*, and four regions (*spin, Khc, Jheh1, αAmy*) are known from *D. silvestris*. In natural populations, element C differs in a single inversion, 2m, between bands 41A and 46A, the other species relative to *D. grimshawi*. The *D. silvestris* lines in which *Khc, spin,* and *Jheh1* were localized were homosequential with the *D. grimshawi* standard. Two additional scaffolds, 9437 and 15112, are joined to scaffold 15245 via syntenic analysis, yielding a 23.5-Mb superscaffold.

Muller element D (chromosome 5, sections 61–80): Scaffold 15110, 24.6 Mb in length, maps to element D (Figure 13, supplemental Table 26). Six loci have been mapped to Muller's D, three from *D. silvestris* (*Argk, ATPsynb, Pgm*), two from *D. heteroneura* (*lark, Cp18*), and one from *D. grimshawi* (*αCat*). These two species are homosequential with *D. grimshawi* for Muller's D so localizations are directly comparable between taxa. One

gene located on element D of *D. melanogaster* was previously shown to map on element E of *D. grimshawi* (*αCat*), indicating movement of the gene between these elements.

Muller element E (chromosome 4, sections 81–100): Three scaffolds have been localized to this chromosome via *in situ* hybridization (Figure 13, supplemental Table 26). A total of eight loci have been mapped to this element, two in *D. grimshawi* (*Hr96, αCat*), two in *D. heteroneura* (*wts, fru*), and four in *D. silvestris* (*RpS3, ninaE, Sryβ, cher*). There is a single fixed inversion, 4b, between *D. silvestris–D. heteroneura* and *D. grimshawi* spanning the region from 93A to 97D. The *RpS3* locus is involved in this inversion and has been localized to band 95D in *D. grimshawi*, a position equivalent to the mapped position 95B in *D. silvestris*. The remainder of the chromosome is unaffected by this inversion. Two scaffolds, 14830 and 15116, can be joined to the mapped scaffolds, generating a single 34.2-Mb superscaffold for this chromosome.

Muller F (chromosome 6 sections): A single marker, *bt*, localized in *D. grimshawi*, anchors scaffold 14822 to the microchromosome (Muller's F). Computational analysis joins scaffold 14592 to this, creating a 1.3-Mb superscaffold for the dot chromosome. Figure 13 shows Muller's F from *D. silvestris* (CARSON 1992); there are no inversions or band differences in this element known from any Hawaiian Drosophila.

Unplaced or orphan scaffolds for the 11 species: The supplemental text describes general properties of scaffolds that were not assigned or mapped to one of the Muller elements in each of the species. These scaffolds are also referred to as orphan scaffolds.

DISCUSSION

The combined approaches of comparative syntenic block analysis and direct physical mapping have collectively informed the orientation of these genome sequences relative to the polytene chromosomal maps. One of the more remarkable features of this analysis is the relatively low impact that genome size and structure had on the ability to assemble gene-rich components of each chromosome arm. Differences in genome size appear to arise primarily from variation in the heterochromatic or unassigned DNA of centromeric regions (Table 2) and the large differences in genome size among these species had little impact on the "quality" of the assemblies using the whole-genome shotgun (WGS) method. For example, measures of DNA content differ quite dramatically among species in the subgenus Drosophila (Table 2) (BOSCO *et al.* 2007); *D. virilis* has the largest estimated genome size of all 12 sequenced species, whereas *D. mojavensis* has the smallest estimated genome size (Table 2). Virtually identical sequencing and assembly methods were used in these two species, yet the sequence localized to the chromosome arms of *D. virilis* is only slightly more fragmented at the level of scaffolds compared with *D. mojavensis*.

The sequence of *D. pseudoobscura* represents an interesting contrast because this species also has a relatively small genome, yet the assembled genome sequence of this species is quite fragmented. Whether this higher level of fragmentation is attributable to biological *vs.* methodological causes is unclear due to differences in methods used to generate the assembly of *D. pseudoobscura* (RICHARDS *et al.* 2005). Different assembly programs can lead to different levels of fragmentation because the algorithms are more or less conservative about joining contigs together to form scaffolds. For instance, an assembly will be more fragmented if the number of BAC paired end reads required to join two contigs together is large. The relative proportion of small to large insert clones used in the whole-genome shotgun sequencing can also influence fragmentation. The use of fewer BAC clones will result in more fragmentation. With more repetitive sequences in the genome, assemblies will be more fragmented. We are tempted to conclude that differences in the assembler were responsible for the higher level of fragmentation in *D. pseudoobscura*. The Freeze 1 assembly of *D. pseudoobscura* was done with ATLAS (HAVLAK *et al.* 2004), while the *D. persimilis* assembly was done as assisted assembly using an ARACHNE assembly of the *D. pseudoobscura* genome. The *D. persimilis* assembly appears to be less fragmented through a more aggressive joining of contigs, which led to mis-joins. This suggests that a more conservative assembly, while more fragmented, may have fewer mis-joins. The data presented here suggest that assembly algorithms that integrate paired end reads and conserved linkage information of close relatives may be a useful strategy for reducing fragmentation of assemblies.

Computational analysis: Synteny-based predictions for "scaffold joins" were used to complement the placement of scaffolds with genetic and physical anchors in a given species. The computational predictions of scaffold order and orientation were critical for filling gaps where experimental markers are not currently available. Although the orientation of the complete sequence of each chromosome arm should be viewed as a first draft, the overall synergism between the computational and marker-based organization of these chromosome-sized scaffolds generates confidence in the inferences.

Computational predictions and marker data proved to be mutually informative. The analysis of the genome sequence of *D. virilis* is a good example of computational predictions that were useful in orienting small scaffolds with respect to flanking scaffolds because marker data indicated the order of scaffolds along the chromosome, but the resolution of standard physical mapping was insufficient to accurately orient the scaffolds along the chromosome (*e.g.*, A5 internally with Muller element A and D1 at the distal end of element D, Figure 11). Small scaffolds not containing any experi-

mental markers were also inserted and oriented using synteny information (*e.g.*, D2 within element D and E4 at the proximal end of element E, Figure 11). The computational analysis also provided a comprehensive screen for orthologs of *D. melanogaster* genes, and on the basis of these orthology calls, almost all of the gene-containing regions appear to be organized along the chromosome arms of *D. virilis* and the other species.

The computational analysis is dependent upon having an accurate assembly, whereas the marker data provide a completely independent assessment of the assembled sequence. One caveat of marker data is the need for accuracy in the ordering of reference positions. These analyses uncovered several instances where cytological positions and linkage relationships did not match the assembled sequences. These discrepancies are due to either errors in the assembled sequences or errors in the ordering of the reference markers. One particular discrepancy for a set of mapped positions on chromosome 5 (element C) of *D. virilis* revealed an apparent error in the assembly of the largest scaffold mapped to this chromosome arm. The assembly error made it impossible to detect syntenic groups with the edges of this scaffold because the erroneous mis-join within the scaffold occurred between two real scaffold edges, thus hiding them from the Synpipe analysis (BHUTKAR *et al.* 2006).

Assembly mis-joins also influenced the computational approach in *D. pseudoobscura*. A number of assembly superscaffolds were broken apart on the basis of experimental data, indicating an alternative alignment along the chromosomes. On the other hand, computational predictions performed well in the case of *D. persimilis*, despite a more fragmented assembly because of lower sequence coverage. The major exception was on Muller C where the assembly rearranged segments within the inverted region in the central part of the chromosome. A large number (90) of scaffold joins were inferred across five of six Muller elements, and there was perfect agreement with the experimental data derived from the backbone species *D. pseudoobscura*, as shown in Table 3.

The syntenic analysis is quite effective in identifying errors in the assembly arising from joins between sequences of different Muller elements. Errors are clearly apparent in scaffolds containing adjacent blocks of genes that belong to different Muller elements, but there is no indication from chromosomal rearrangements that this composite gene order is expected, and the genes show the expected associations with Muller elements in closely related species. A number of these mis-joins, in the case of *D. sechellia*, for example, have been confirmed. A list of probable assembly mis-joins is provided in the supplemental material. These corrections of the initial freezes of the assemblies will provide more accurate data for downstream analysis. For example, knowledge of Muller element-wide scaffold and

gene order and misassembly information enabled a fine-scale analysis of chromosome-wide rearrangements between species (BHUTKAR *et al.* 2008).

Patterns of chromosome evolution in the genus Drosophila: A wide spectrum of gross changes in overall chromosome form is represented by species of Drosophila. For example, the curious dot chromosome, which has maintained a common form in all three species of the subgenus Drosophila and in most members of the subgenus Sophophora, exhibits two derived configurations in these species. A dot chromosome is not evident in the karyotype of *D. willistoni*. This small chromosome fused with the proximal end of the E element to generate a composite chromosome. The fusion is clearly demarcated in the genome sequence because genes from element F are mostly contained in a localized block within a single scaffold that has a distinct transition into genes from element E (supplemental Table 15). In this case, the composite nature of this scaffold captures the visible change in the karyotype (PAPACEIT and JUAN 1998). The junction between the two elements is relatively precise, being separated by 14 kb in length without gaps in the sequence. The junction sequence has repeats with at least one signature of a *P* element, suggesting that these repeats could have played a role in the fusion process. The other change in the dot chromosome entails the expansion of this chromosome into a large metacentric chromosome in *D. ananassae*. The scaffolds associated with element F contain the same syntenic genes seen in the dot chromosome of other species. The expansion of the F element in *D. ananassae* occurred through the addition of 32.5% more retro and DNA element sequence without major changes in coding capacity. Even with this expansion, the structure of the assembled scaffolds for the "dot" of *D. ananassae* does not differ substantially from some of the other species in the Sophophora lineage where the genes are also distributed among several scaffolds. By contrast, element F is represented by one or two scaffolds for species in the subgenus Drosophila. These differences in assembly of the dot in the two subgenera are correlated with differences in chromatin structure. The dot of *D. melanogaster* is heterochromatic, while a heterochromatic structure is not evident for the dot of *D. virilis* (SLAWSON *et al.* 2006). It is unclear how chromatin structure could influence the WGS approach, but the differences in efficiency of assembling the sequence of element F, despite a relatively uniform level of transposable element content, suggest a possible impact on the effectiveness of genome sequencing and assembly.

The complete characterization of conserved linkage groups between these genome sequences further supports previous observations of limited exchange of genetic content between Muller's elements (RICHARDS *et al.* 2005). However, exchange events have occurred during the evolution of these Drosophila species, and

TABLE 3

Comparison of *D. persimilis* scaffold joins predicted by Synpipe synteny and rearrangement analysis *vs.* experimental markers in the backbone *D. pseudoobscura* assembly

Muller element	Total joins	Synpipe inferred joins	Agreement with experimental analysis no. (%)
A	20	19	19 (100)
A/D	52	46	46 (100)
E	4	4	4 (100)
C	14	10	10 (100)
B	11	11	11 (100)
All	101	90	90 (100)

The *D. persimilis* assembly was fairly fragmented (especially on Muller A/D). However, Synpipe was able to predict most of the scaffold joins, and these were in complete agreement with experimental analysis from *D. pseudoobscura*, which was used to anchor the *D. persimilis* assembly.

these instances of gene movement not only are limited to members of multigene families, but also include single-copy genes (RANZ *et al.* 2003, 2007; BHUTKAR *et al.* 2007b). This analysis reveals that large-scale genome rearrangement has played a limited role in the movement of genes between chromosome arms. For example, *D. erecta* and *D. yakuba* share a pericentric inversion at the base of the B·C element (2L·2R in *D. melanogaster*). Thus, relative to *D. melanogaster*, the B and C elements are now mixed from telomere to centromere. The new order is B → C and C → B.

Elements A and D were fused in *D. pseudoobscura* and *D. persimilis* to form a metacentric X chromosome that contains genes homologous to the X and autosomal 3L arms of *D. melanogaster*. Genes from the A element have moved to the base of the D element in the two *obscura* group species. This interchromosomal change was first explained by a pericentric inversion between elements A and D after the centric fusion that generated the metacentric X chromosome of *D. pseudoobscura* (SEGARRA and AGUADÉ 1992; SEGARRA *et al.* 1995). However, a pericentric inversion in the fused A and D elements would be expected to cause a reciprocal exchange of gene content between elements. No evidence for such an exchange has been detected in the assembled genome of *D. pseudoobscura*. Thus, the pericentric inversion responsible for the exchange between the A and D elements might be asymmetric with one of the breakpoints in the heterochromatic region very close to the centromere of element D. The accumulation of transposable elements in heterochromatic regions (RIZZON *et al.* 2002) and their putative role in the origin of inversions (CÁCERES *et al.* 1999) might account for this possibility.

Alternatively, after the fusion of both elements, the centromere might have been repositioned toward a more interstitial region of element A. Centromere repositioning (CR) has been proposed to explain the karyotype

evolution of diverse mammalian species (FERRERI *et al.* 2005; CARBONE *et al.* 2006), but up to now CR has not been described either in Drosophila or in other insects. Centromere repositioning events would cause, like asymmetric pericentric inversions, a relocation of gene content between elements (with the shortening of one element and the lengthening of the other one), but with no inversion of gene order relative to the centromere. However, it is not straightforward to distinguish between the pericentric inversion and the centromere repositioning alternatives in the Drosophila case described here. The A·D fusion occurred following the divergence between the lineage that gave rise to *pseudoobscura* and the lineage currently represented by *D. subobscura* and close relatives. Given the ancient split between these lineages (19–22 MYA according to GAO *et al.* 2007), the gene order around the centromere of the fused and nonfused elements might have completely changed as a result of paracentric inversions fixed in both lineages.

Rearrangement of elements in the Drosophila genome also occurs through pericentric inversion. A pericentric inversion of the A element in *D. ananassae* converted the normally acrocentric X into a metacentric chromosome, thus representing a change in karyotype other than fusion between Muller's elements. The scaffolds in the genome sequence map separately to the two chromosome arms generated by this inversion. In addition to the relatively rare cases of major chromosomal rearrangements (fusions, translocations, and pericentric inversions), these species possess a rich spectrum of paracentric inversions and their presence in the sequenced genomes has been readily detected and confirmed (SPERLICH and PFRIEM 1986; BHUTKAR *et al.* 2008).

Evaluation of these assembled genome sequences relative to the chromosome maps yields initial insight into the distinction between euchromatic and heterochromatic genomic regions as represented in polytenized chromosomal regions and the chromocenter of salivary glands. Scaffolds mapping to centromeric regions comprise initial glimpses into heterochromatic domains of these genomes (supplemental Text). The ~10-fold increase in the abundance of interspersed repeats in the "orphan" scaffolds located in the pericentromeric regions in the genome of *D. virilis* is consistent with the genome of *D. melanogaster* (SMITH *et al.* 2007). Gene content within the pericentric regions appears quite variable among species, because *D. mojavensis* and *D. grimshawi* are essentially devoid of these "orphan" scaffolds, whereas *D. virilis* and *D. ananassae* have quite extensive regions of assembled pericentric sequence containing a modest number of genes. Gene content of these regions, however, exhibits a high rate of flux among species consistent with previous studies (YASUHARA *et al.* 2005; SCHULZE *et al.* 2006). Genes in the heterochromatin of *D. melanogaster* are generally distributed throughout the genomes of the other species, and from a different

perspective, genes residing in pericentromeric scaffolds of *D. virilis* are generally present in euchromatic regions of *D. melanogaster.* This movement between chromosomal domains also generates the appearance of much more intra-element movement using the position of genes in the genome of *D. melanogaster* as a reference. For example, two scaffolds of *D. virilis* contain a mosaic of 17 genes from the D and E elements of *D. melanogaster.* Each of these genes is located within large scaffolds mapped to the E element of *D. mojavensis, D. grimshawi,* and *D. willistoni.* Therefore, the ancestral location of these genes appears to be element E, which has been observed as directly for *RpL15* of *D. virilis* and αCat of *D. grimshawi.* These comparisons support a previous inference that the location of a small number (∼7) of genes on element D of *D. melanogaster* is a derived feature, possibly resulting from a small pericentric inversion(s) within the centromeric region of chromosome 3 in the *melanogaster* group that exchanged genes from element E to element D (SCHULZE *et al.* 2006).

Mapped assembly scaffolds add value to the comparative genome data: The data presented here should be viewed as the starting point for experiments designed to understand how genomes evolve. These experimental studies are now more tractable, given the complete genome sequence of 12 representative species in the genus Drosophila. Availability of BAC libraries for nine additional Drosophila species further enables comparative genomic analyses (see http://www.genome.arizona.edu/BAC_special_projects/#Drosophila) (MARKOW *et al.* 2003). Studies of these species will be limited initially due to the availability of few balancer and mutant strains. In addition, transformation systems to introduce mutations and shuttle genes into different strains are also limited, but are likely to be developed in the future.

Conclusion: The scaffold maps presented here represent the first pass at anchoring the assembled sequences to the physical map and one should view these assignments with caution. The ordering and orientation of scaffolds revealed by this analysis will be represented in FlyBase (http://www.flybase.org) as a sequence of each chromosome arm in each species. The amount of genetic and physical mapping that was used in the different species varied, which leads to different levels of confidence in the maps. Gaps between scaffolds will be indicated by insertion of ambiguities. Although each chromosome will appear as one contiguous sequence, several real or possible limitations should be recognized when using these sequences for downstream analyses: (1) distances are not accurate for the intervals between positions that span scaffold joins; (2) a substantial amount of sequence, including coding regions, may be missing from each scaffold join; (3) scaffolds designated in the unplaced bin may contain sequence corresponding to gaps between oriented contigs or scaffolds, and the scaffolds in the unplaced bin are completely unordered; and (4) unrecognized assembly errors may exist within

scaffolds. With these caveats in mind, the sequences do provide a comprehensive resource for comparison among the genomes of these species of Drosophila and represent a starting point for additional experiments that evaluate these uncertainties.

The original motivation for the 12 genomes project was to develop new bioinformatic tools to aid with genome assembly using whole-genome shotgun methods. The power of Drosophila is the availability of a sophisticated polytene chromosome map that allows for careful verification of new computational approaches for ordering and orienting scaffolds together. This approach shows that there are added benefits of using a set of related species in genome biology. Comparison of gene order among close relatives will help to order and orient scaffolds in species groups where physical map data may be limited.

There are additional challenges to the computational approaches described here. The Drosophila genome is relatively well behaved with respect to rearrangements. About 95% of the genes are syntenic and the major mechanism of genome change is inversion events. Mammalian genomes have the added complication of many more transposition events. Thus, new computational methods must take these rearrangements into account as new tools are developed.

LITERATURE CITED

AGUADÉ, M., N. MIYASHITA and C. H. LANGLEY, 1989   Reduced variation in the *yellow-achaete-scute* region in natural populations of *Drosophila melanogaster*. Genetics **122:** 607–615.

ALTSCHUL, S. F., T. L. MADDEN, A. A. SCHAFFER, J. ZHANG, Z. ZHANG et al., 1997   Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:** 3389–3402.

ANDERSON, W. W., F. J. AYALA and R. E. MICHOD, 1977   Chromosomal and allozymic diagnosis of three species of *Drosophila*, *Drosophila pseudoobscura*, *Drosophila persimilis*, and *Drosophila miranda*. J. Hered. **68:** 71–74.

ANDERSON, W. W., J. ARNOLD, D. G. BALDWIN, A. T. BECKENBACH, C. J. BROWN et al., 1991   Four decades of inversion polymorphism in *Drosophila pseudoobscura*. Proc. Natl. Acad. Sci. USA **88:** 10367–10371.

ASHBURNER, M., and F. LEMEUNIER, 1976   Relationships within the melanogaster species subgroup of the genus Drosophila (Sophophora). I. Inversion polymorphisms in *Drosophila melanogaster* and *Drosophila simulans*. Proc. R. Soc. Lond. B Biol. Sci. **193:** 137–157.

ASHBURNER, M., K. G. GOLIC and R. S. HAWLEY, 2005   *Drosophila: A Laboratory Handbook*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

AWASAKI, T., N. JUNI, T. HAMABATA, K. YOSHIDA, M. MATSUDA et al., 1994   Retrotransposon-induced ectopic expression of cut causes the Om(*1A*) mutant in *Drosophila ananassae*. Genetics **137:** 165–174.

BÄCHLI, C., 2006   Taxodros: the database on taxonomy of Drosophilidae. http://www.taxodros.unizh.ch.

BECK, M. L., 1977   Localization of constitutive heterochromatin in the chromosomes of *Drosophila virilis*. Chromosoma **42:** 53–55.

BEGUN, D. J., A. K. HOLLOWAY, K. STEVENS, L. W. HILLIER, Y.-P. POH et al., 2007   Population genomics: whole-genome analysis of

polymorphism and divergence in *Drosophila simulans*. PLoS Biol. **5:** 2534–2559.

Bhutkar, A., S. Russo, T. F. Smith and W. M. Gelbart, 2006 Techniques for multi-genome synteny analysis to overcome assembly limitations. Genome Inform. **17:** 152–161.

Bhutkar, A., W. M. Gelbart and T. F. Smith, 2007a Inferring genome-scale rearrangement phylogeny and ancestral gene order: a Drosophila case study. Genome Biol. **8:** R236.

Bhutkar, A., S. M. Russo, T. F. Smith and W. M. Gelbart, 2007b Genome scale analysis of positionally relocated genes. Genome Res. **17:** 1880–1887.

Bhutkar, A., S. W. Schaeffer, S. Russo, M. Xu, T. F. Smith *et al.*, 2008 Chromosomal rearrangement inferred from comparisons of 12 Drosophila genomes. Genetics **179:** 1657–1680.

Blackman, R. K., and M. Meselson, 1986 Interspecifc nucleotide sequence comparisons used to identify regulatory and structural features of the *Drosophila hsp82* gene. J. Mol. Biol. **188:** 499–515.

Bonorino, C. B. C., M. Pereira, C. E. V. Alonso, V. L. S. Valente and E. Abdelhay, 1993 *In situ* mapping of the *hsp70* locus in seven species of the *willistoni* group of *Drosophila*. Rev. Bras. Genet. **16:** 561–571.

Bosco, G., P. Campbell, J. Leiva-Neto and T. A. Markow, 2007 Analysis of Drosophila species genome size and satellite DNA content reveals significant differences among strains as well as between species. Genetics **177:** 1277–1290.

Bridges, C. B., 1916 Non-disjunction as proof of the chromosome theory of heredity. Genetics **1:** 1–52.

Bridges, C. B., 1935 Salivary chromosome maps with a key to the banding of the chromosomes of *Drosophila melanogaster*. J. Hered. **26:** 60–64.

Cáceres, M., J. M. Ranz, A. Barbadilla, M. Long and A. Ruiz, 1999 Generation of a widespread *Drosophila* inversion by a transposable element. Science **285:** 415–418.

Carbone, L., S. G. Nergadze, E. Magnani, D. Misceo, M. Francesca Cardone *et al.*, 2006 Evolutionary movement of centromeres in horse, donkey, and zebra. Genomics **87:** 777–782.

Carson, H. L., 1992 Inversions in Hawaiian *Drosophila*, pp. 407–439 in *Drosophila Inversion Polymorphism*, edited by C. Krimbas and J. R. Powell. CRC Press, Boca Raton, FL.

Carson, H. L., J. Tonzetich and L. T. Doescher, 1992 Polytene chromosome maps for Hawaiian *Drosophila*, pp. 441–453 in *Drosophila Inversion Polymorphism*, edited by C. Krimbas and J. R. Powell. CRC Press, Boca Raton, FL.

Carvalho, A. B., and A. G. Clark, 2005 Y chromosome of *D. pseudoobscura* is not homologous to the ancestral Drosophila Y. Science **307:** 108–110.

Casacuberta, E., and M. L. Pardue, 2003 Transposon telomeres are widely distributed in the Drosophila genus: TART elements in the virilis group. Proc. Natl. Acad. Sci. USA **100:** 3363–3368.

Charlesworth, B., D. Charlesworth, J. Hnilicka, A. Yu and D. S. Guttman, 1997 Lack of degeneration of loci on the neo-Y chromosome of *Drosophila americana americana*. Genetics **145:** 989–1002.

Clayton, F. E., and W. C. Guest, 1986 Overview of chromosomal evolution in the family Drosophilidae, pp. 1–38 in *The Genetics and Biology of Drosophila*, edited by M. Ashburner, H. L. Carson and J. N. Thompson. Academic Press, New York.

Counterman, B. A., and M. A. Noor, 2006 Multilocus test for introgression between the cactophilic species *Drosophila mojavensis* and *Drosophila arizonae*. Am. Nat. **168:** 682–696.

da Cunha, A. B., and T. Dobzhansky, 1954 A further study of chromosomal polymorphism in *Drosophila willistoni* in its relation to the environment. Evolution **8:** 119–134.

da Cunha, A. B., H. Burla and T. Dobzhansky, 1950 Adaptive chromosomal polymorphism in *Drosophila willistoni*. Evolution **4:** 212–235.

da Cunha, A. B., T. Dobzhansky, O. A. Pavlovsky and B. Spassky, 1959 Genetics of natural populations. XXVIII. Supplementary data on the chromosomal polymorphism in *Drosophila willistoni* in its relation to the environment. Evolution **13:** 389–404.

Dobzhansky, T., 1944 Chromosomal races in *Drosophila pseudoobscura* and *Drosophila persimilis*. Carnegie Inst. Washington Publ. **554:** 47–144.

Dobzhansky, T., 1950 The chromosomes of *Drosophila willistoni*. J. Hered. **41:** 156–158.

Dobzhansky, T., and A. Dreyfus, 1943 Chromosomal aberrations in Brazilian *Drosophila ananassae*. Proc. Natl. Acad. Sci. USA **29:** 301–305.

Dobzhansky, T., and A. H. Sturtevant, 1938 Inversions in the chromosomes of *Drosophila pseudoobscura*. Genetics **23:** 28–64.

Dobzhansky, T., and C. C. Tan, 1936 Studies of hybrid sterility III. A comparison of the gene arrangement in two species, *Drosophila pseudoobscura* and *Drosophila miranda*. Z. Indukt. Abstamm. Vererbungsl. **72:** 88–114.

Doi, M., M. Matsuda, M. Tomaru, H. Matsubayashi and Y. Oguma, 2001 A locus for female discrimination behavior causing sexual isolation in Drosophila. Proc. Natl. Acad. Sci. USA **98:** 6714–6719.

Doleschall, C. L., 1858 Derde bijdrage tot de kennis der dipteren fauna van Nederlandsch Indie. Natuurkundige Vereeniging in Nederlandsch-Indië **17:** 73–128.

Drosophila 12 Genomes Consortium, 2007 Evolution of genes and genomes on the *Drosophila* phylogeny. Nature **450:** 203–218.

Edwards, K. A., L. T. Doescher, K. Y. Kaneshiro and D. Yamamoto, 2007 A database of wing diversity in the Hawaiian Drosophila. PLoS One **2:** 1–12.

Evans, A. L., P. A. Mena and B. F. McAllister, 2007 Positive selection near an inversion breakpoint on the neo-X chromosome of *Drosophila americana*. Genetics **177:** 1303–1319.

Ferreri, G. C., D. M. Liscinsky, J. A. Mack, M. D. Eldridge and R. J. O'Neill, 2005 Retention of latent centromeres in the mammalian genome. J. Hered. **96:** 217–224.

FlyBase Consortium, 1999 The FlyBase database of the Drosophila Genome Projects and community literature. Nucleic Acids Res. **27:** 85–88.

Freire-Maia, N., 1961 Peculiar gene arrangements in Brazilian natural populations of *Drosophila ananassae*. Evolution **15:** 486–495.

Fujii, S., 1936 Salivary gland chromosomes of *Drosophila virilis*. Cytologia **7:** 272–275.

Fujii, S., 1942 Further studies on the salivary gland chromosomes of *Drosophila virilis*. Cytologia **12:** 435–459.

Futch, D. G., 1966 A study of speciation in South Pacific populations of *Drosophila ananassae*. Univ. Texas Publs. Stud. Genet. **6615:** 79–120.

Futch, D. G., 1972 A preliminary note on parthenogenesis in *D. ananassae*. Drosoph. Inf. Serv. **48:** 78.

Gall, J. G., E. H. Cohen and M. L. Polan, 1971 Repetitive DNA sequences in Drosophila. Chromosoma **33:** 319–344.

Gallach, M., V. Arnau and I. Marin, 2007 Global patterns of sequence evolution in Drosophila. BMC Genomics **8:** 408.

Gao, J. J., H. A. Watabe, T. Aotsuka, J. F. Pang and Y. P. Zhang, 2007 Molecular phylogeny of the Drosophila obscura species group, with emphasis on the Old World species. BMC Evol. Biol. **7:** 87.

Gleason, J. M., E. C. Griffith and J. R. Powell, 1998 A molecular phylogeny of the *Drosophila willistoni* group: Conflicts between species concepts? Evolution **52:** 1093–1103.

Goni, B., M. Matsuda and Y. N. Tobari, 2006 Chiasmata and chromosome breakages are related to crossing over in *Drosophila ananassae* males. Genome **49:** 1374–1383.

Gonzalez, J., F. Casals and A. Ruiz, 2004 Duplicative and conservative transpositions of larval serum protein 1 genes in the genus Drosophila. Genetics **168:** 253–264.

Gubenko, I. S., and M. B. Evgen'ev, 1984 Cytological and linkage maps of *Drosophila virilis* chromosomes. Genetica **65:** 127–139.

Havlak, P., R. Chen, K. J. Durbin, A. Egan, Y. Ren *et al.*, 2004 The Atlas genome assembly system. Genome Res. **14:** 721–732.

Hinton, C. W., 1970 Identification of two loci controlling crossing over in males of *Drosophila ananassae*. Genetics **66:** 663–676.

Hinton, C. W., 1983 Relations between factors controlling crossing over and mutability in males of *Drosophila ananassae*. Genetics **104:** 95–112.

Hinton, C. W., 1984 Morphogenetically specific mutability in *Drosophila ananassae*. Genetics **106:** 631–653.

Hinton, C. W., and J. E. Downs, 1975 The mitotic, polytene and meiotic chromosomes of *Drosophila ananassae*. J. Hered. **66:** 353–361.

Horton, I. H., 1939 A comparison of the salivary gland chromosomes of *Drosophila melanogaster* and *D. simulans*. Genetics **24:** 234–243.

Hsu, T. C., 1952 Chromosomal variation and evolution in the virilis group of Drosophila. Univ. Texas Publs. **5204:** 35–72.

HUGHES, R. D., 1936 The morphology of the normal salivary chromosomes of *Drosophila virilis*. J. Hered. **27**: 305–306.

HUGHES, R. D., 1939 An analysis of the chromosomes of the two subspecies *Drosophila virilis virilis* and *Drosophila virilis americana*. Genetics **24**: 811–834.

HUTTUNEN, S., and C. SCHLÖTTERER, 2002 Isolation and characterization of microsatellites in *Drosophila virilis* and their cross species amplification in members of the *D. virilis* group. Mol. Ecol. Notes **2**: 593–597.

JOHNSON, W. R., 1980 Chromosomal polymorphism in natural populations of the desert adapted species, *Drosophila mojavensis*. Ph.D. Thesis, University of Arizona, Tucson, AZ.

JUDD, B. H., M. W. SHEN and T. C. KAUFMAN, 1972 The anatomy and function of a segment of the *X* chromosome of *Drosophila melanogaster*. Genetics **71**: 139–156.

JUNI, N., T. AWASAKI, K. YOSHIDA and S. H. HORI, 1996 The *Om*(*1E*) mutation in *Drosophila ananassae* causes compound eye overgrowth due to *tom* retrotransposon-driven overexpression of a novel gene. Genetics **143**: 1257–1270.

KASTRITSIS, C. D., and D. W. CRUMPACKER, 1966 Gene arrangements in the third chromosome of *Drosophila pseudoobscura*. I. Configurations with tester chromosomes. J. Hered. **57**: 150–158.

KAUFMANN, B. P., 1936 The chromosomes of *Drosophila ananassae*. Science **83**: 39.

KAUFMANN, B. P., 1937 Chromosome studies on *Drosophila ananassae*. Genetics **22**: 197–198.

KAVENOFF, R., and B. H. ZIMM, 1973 Chromosome-sized DNA molecules from Drosophila. Chromosoma **41**: 1–27.

KAYMER, M., A. DEBES, H. KRESS and U. KURZIK-DUMKE, 1997 Sequence, molecular organization and products of the *Drosophila virilis* homologs of the *D. melanogaster* nested genes lethal(2) tumorous imaginal discs [*l*(*2*) *tid*] and lethal(2) neighbour of tid [*l*(*2*) *not*]. Gene **204**: 91–103.

KIKKAWA, H., 1935 An inference as to the constitution of X-chromosome in *Drosophila*. Proc. Imp. Acad. Japan **11**: 62–65.

KIKKAWA, H., 1936 Chromosomes of *Drosophila ananassae*. Drosoph. Inf. Serv. **5**: 25.

KIKKAWA, H., 1937 Spontaneous crossing-over in the male of *Drosophila ananassae*. Zool. Mag. Tokyo **49**: 159–160.

KIKKAWA, H., 1938 Studies on the genetics and cytology of *Drosophila ananassae*. Genetica **20**: 458–516.

KRESS, H., 1993 The salivary gland chromosomes of *Drosophila virilis*: a cytological map, pattern of transcription and aspects of chromosome evolution. Chromosoma **102**: 734–742.

KRIMBAS, C. B., 1992 The inversion polymorphism of *Drosophila subobscura*, pp. 127–220 in *Drosophila Inversion Polymorphism*, edited by C. B. KRIMBAS and J. R. POWELL. CRC Press, Boca Raton, FL.

LAIRD, C. D., 1973 DNA of Drosophila chromosomes. Annu. Rev. Genet. **7**: 177–204.

LAKOVAARA, S., and A. SAURA, 1972 Location of enzyme loci in chromosomes of *Drosophila willistoni*. Experientia **28**: 355–356.

LEFEVRE, G., 1976 A photographic representation and interpretation of the polytene chromosomes of Drosophila melanogaster salivary glands, pp. 31–66 in *The Genetics and Biology of Drosophila*, edited by M. ASHBURNER and E. NOVITSKI. Academic Press, New York.

LEMEUNIER, F., and M. ASHBURNER, 1976 Relationships within the melanogaster species subgroup of the genus *Drosophila* (Sophophora). II. Phylogenetic relationships between six species based upon polytene chromosome banding sequences. Proc. R. Soc. Lond. B Biol. Sci. **193**: 275–294.

LEMEUNIER, F., and M. ASHBURNER, 1984 Relationships within the melanogaster species subgroup of the genus Drosophila (Sophophora). IV. The chromosomes of two new species. Chromosoma **89**: 343–351.

LOUKAS, M., and F. C. KAFATOS, 1986 The actin loci in the genus Drosophila: establishment of chromosomal homologies among distantly related species by *in situ* hybridization. Chromosoma **94**: 297–308.

LOZOVSKAYA, E. R., D. A. PETROV and D. L. HARTL, 1993 A combined molecular and cytogenetic approach to genome evolution in *Drosophila* using large-fragment DNA cloning. Chromosoma **102**: 253–266.

MACHADO, C. A., T. S. HASELKORN and M. A. NOOR, 2007a Evaluation of the genomic extent of effects of fixed inversion differences on intraspecific variation and interspecific gene flow in *Drosophila pseudoobscura* and *D. persimilis*. Genetics **175**: 1289–1306.

MACHADO, C. A., L. M. MATZKIN, L. K. REED and T. A. MARKOW, 2007b Multilocus nuclear sequences reveal intra- and interspecific relationships among chromosomally polymorphic species of cactophilic Drosophila. Mol. Ecol. **16**: 3009–3024.

MARKOW, T. A., 1991 Sexual isolation among populations of *Drosophila mojavensis*. Evolution **45**: 1525–1529.

MARKOW, T. A., and G. D. HOCUTT, 1998 Reproductive isolation in Sonoran Desert Drosophila: testing the limits of the rules, pp. 234–244 in *Endless Forms: Species and Speciation*, edited by D. HOWARD and S. H. BERLOCHER. Oxford University Press, Oxford.

MARKOW, T. A., and P. M. O'GRADY, 2006 *Drosophila: A Guide for Species Identification and Use*. Academic Press, London.

MARKOW, T. A., B. F. MCALLISTER and T. C. KAUFMAN, 2003 A white paper requesting BAC library construction: Drosophila as a model for comparative genomics. http://www.genome.gov/10001852.

MARTINS, M., 1987 Variacao espacial e temporal de algumas especies e grupos de Drosophila (Diptera) em duas reservas de matas isoladas, nas vizinhancas de Manaus (Amazonas, Brasil). Bolm Mus. para. Emilio Goeldi Ser. Zool. **3**: 195–218.

MATSUBAYASHI, H., N. JUNI, K. USUI, S. H. HORI and Y. N. TOBARI, 1991 Molecular and histological characterizations of the *Om*(*2D*) mutants in *Drosophila ananassae*. Mol. Gen. Genet. **227**: 165–172.

MATSUBAYASHI, H., M. MATSUDA, Y. TOMIMURA, M. SHIBATA and Y. N. TOBARI, 1992 Cytological mapping of *Om* mutants of *Drosophila ananassae*. Jpn. J. Genet. **67**: 259–264.

MATSUDA, M., and Y. N. TOBARI, 1999 A parthenogenetic strain of *D. pallidosa*-like in the *D. ananassae* complex. Drosoph. Inf. Serv. **82**: 49–50.

MATSUDA, M., and Y. N. TOBARI, 2004 Genetic analyses of several *Drosophila ananassae*-complex species show a low-frequency major gene for parthenogenesis that maps to chromosome 2. Genet. Res. **83**: 83–89.

MATSUDA, M., H. T. IMAI and Y. N. TOBARI, 1983 Cytogenetic analysis of recombination in males of *Drosophila ananassae*. Chromosoma **88**: 286–292.

MATSUDA, M., H. SATO and Y. N. TOBARI, 1993 Crossing over in males, pp. 53–71 in *Drosophila ananassae: Genetical and Biological Aspects*, edited by Y. N. TOBARI. Japan Scientific Society Press, Tokyo.

MATZKIN, L. M., T. D. WATTS, B. G. BITLER, C. A. MACHADO and T. A. MARKOW, 2006 Functional genomics of cactus host shifts in *Drosophila mojavensis*. Mol. Ecol. **15**: 4635–4643.

MCALLISTER, B. F., 2002 Chromosomal and allelic variation in *Drosophila americana*: selective maintenance of a chromosomal cline. Genome **45**: 13–21.

MCALLISTER, B. F., 2003 Sequence differentiation associated with an inversion on the neo-*X* chromosome of *Drosophila americana*. Genetics **165**: 1317–1328.

MCALLISTER, B. F., and A. L. EVANS, 2006 Increased nucleotide diversity with transient Y linkage in *Drosophila americana*. PLoS One **1**: e112.

METTLER, L. E., 1963 *Drosophila mojavensis baja*, a new form in the mulleri complex. Drosoph. Inf. Serv. **38**: 57–58.

METZ, C. W., 1916 Chromosome studies on the Diptera. III. Additional type of chromosome groups in the *Drosophilidae*. Am. Nat. **50**: 587–599.

MOORE, B. C., and C. E. TAYLOR, 1986 *Drosophila* of southern California. III. Gene arrangements of *Drosophila persimilis*. J. Hered. **77**: 313–323.

MORIWAKI, D., 1937 A high ratio of crossing over in *Drosophila ananassae*. Z. Indukt. Abstamm. Vererbungsl. **74**: 17–23.

MORIWAKI, D., 1938 Enhanced crossing-over in the second chromosome of *Drosophila ananassae* (a preliminary note). Jpn. J.Genet. **14**: 283–284.

MORIWAKI, D., 1940 Enhanced crossing over in the second chromosome of *Drosophila ananassae*. Jpn. J.Genet. **16**: 37–48.

MORIWAKI, D., and S. ITO, 1969 Studies on puffing in the salivary gland chromosomes of *Drosophila ananassae*. Jpn. J.Genet. **44**: 129–138.

MORIWAKI, D., and Y. N. TOBARI, 1975 *Drosophila ananassae*, pp. 513–535 in *Handbook of Genetics*, edited by R. C. KING. Plenum Press, New York.

MORIWAKI, D., Y. N. TOBARI and Y. OGUMA, 1970   Spontaneous crossing-over in the male of *Drosophila ananassae*. Jpn. J.Genet. **45:** 411–420.

MORIYAMA, E. N., D. A. PETROV and D. L. HARTL, 1998   Genome size and intron size in Drosophila. Mol. Biol. Evol. **15:** 770–773.

MULLER, H. J., 1940   Bearings of the 'Drosophila' work on systematics, pp. 185–268 in *The New Systematics*, edited by J. HUXLEY. Clarendon Press, Oxford.

NEUFELD, T. P., R. W. CARTHEW and G. M. RUBIN, 1991   Evolution of gene position: chromosomal arrangement and sequence comparison of the *Drosophila melanogaster* and *Drosophila virilis sina* and *Rh4* genes. Proc. Natl. Acad. Sci. USA **88:** 10203–10207.

NURMINSKY, D. I., E. N. MORIYAMA, E. R. LOZOVSKAYA and D. L. HARTL, 1996   Molecular phylogeny and genome evolution in the *Drosophila virilis* species group: duplications of the *alcohol dehydrogenase* gene. Mol. Biol. Evol. **13:** 132–149.

O'GRADY, P. M., 2002   Species to genera: phylogenetic inference in the Hawaiian Drosophilidae, pp. 17–30 in *Molecular Systematics and Evolution: Theory and Practice*, edited by R. DESALLE, G. GIRIBET and W. WHEELER. Birkhauser Verlag, Berlin.

O'GRADY, P. M., and M. G. KIDWELL, 2002   Phylogeny of the subgenus *Sophophora* (Diptera: drosophilidae) based on combined analysis of nuclear and mitochondrial sequences. Mol. Phylogenet. Evol. **22:** 442–453.

ORTIZ-BARRIENTOS, D., A. S. CHANG and M. A. NOOR, 2006   A recombinational portrait of the *Drosophila pseudoobscura* genome. Genet. Res.. **87:** 23–31.

PÄÄLLYSAHO, S., S. HUTTUNEN and A. HOIKKALA, 2001   Identification of X chromosomal restriction fragment length polymorphism markers and their use in a gene localization study in *Drosophila virilis* and *D. littoralis*. Genome **44:** 242–248.

PAINTER, T. S., 1934   A new method for the study of chromosomal aberrations and the plotting of chromosomal maps in *Drosophila melanogaster*. Genetics **19:** 175–188.

PAPACEIT, M., and E. JUAN, 1998   Fate of dot chromosome genes in *Drosophila willistoni* and *Scaptodrosophila lebanonensis* determined by *in situ* hybridization. Chromosome Res. **6:** 49–54.

PAPACEIT, M., M. AGUADÉ and C. SEGARRA, 2006   Chromosomal evolution of elements B and C in the Sophophora subgenus of Drosophila: evolutionary rate and polymorphism. Evolution **60:** 768–781.

PATTERSON, J. T., and W. S. STONE, 1952   *Evolution in the Genus Drosophila*. MacMillan, New York.

PATTERSON, J. T., M. S. BROWN, W. S. STONE and J. F. CROW, 1940a   Hybridization in the Mulleri group of Drosophila. Univ. Texas Publs. **4032:** 251–256.

PATTERSON, J. T., W. S. STONE and A. B. GRIFFEN, 1940b   Evolution of the virilis group in Drosophila. Univ. Texas Publs. **4032:** 218–250.

PETROV, D. A., and D. L. HARTL, 1998   High rate of DNA loss in the *Drosophila melanogaster* and *Drosophila virilis* species groups. Mol. Biol. Evol. **15:** 293–302.

PETROV, D. A., E. R. LOZOVSKAYA and D. L. HARTL, 1996   High intrinsic rate of DNA loss in *Drosophila*. Nature **384:** 346–349.

POWELL, J. R., 1997   *Progress and Prospects in Evolutionary Biology: The Drosophila Model*. Oxford University Press, New York.

RANZ, J. M., M. CACERES and A. RUIZ, 1999   Comparative mapping of cosmids and gene clones from a 1.6 Mb chromosomal region of *Drosophila melanogaster* in three species of the distantly related subgenus Drosophila. Chromosoma **108:** 32–43.

RANZ, J. M., F. CASALS and A. RUIZ, 2001   How malleable is the eukaryotic genome? Extreme rate of chromosomal rearrangement in the genus Drosophila. Genome Res. **11:** 230–239.

RANZ, J. M., J. GONZALEZ, F. CASALS and A. RUIZ, 2003   Low occurrence of gene transposition events during the evolution of the genus Drosophila. Evolution **57:** 1325–1335.

RANZ, J. M., D. MAURIN, Y. S. CHAN, M. VON GROTTHUSS, L. W. HILLIER *et al.*, 2007   Principles of genome evolution in the *Drosophila melanogaster* species group. PLoS Biol. **5:** 1366–1381.

REED, L. K., and T. A. MARKOW, 2004   Early events in speciation: polymorphism for hybrid male sterility in Drosophila. Proc. Natl. Acad. Sci. USA **101:** 9009–9012.

REGNER, L. P., M. S. PEREIRA, C. E. ALONSO, E. ABDELHAY and V. L. VALENTE, 1996   Genomic distribution of *P* elements in *Drosophila willistoni* and a search for their relationship with chromosomal inversions. J. Hered. **87:** 191–198.

RICHARDS, S., Y. LIU, B. R. BETTENCOURT, P. HRADECKY, S. LETOVSKY *et al.*, 2005   Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene and *cis*-element evolution. Genome Res. **15:** 1–18.

RIDDLE, N. C., and S. C. R. ELGIN, 2006   The dot chromosome of Drosophila: insights into chromatin states and their change over evolutionary time. Chromosome Res. **14:** 405–416.

RIEGER, T. T., 1999   *Mapeamento por hibridação in situ e expressão no desenvolvimento de genes de reposta a estresses e ao hormônio ecdisona em espécies do grupo willistoni de Drosophila*. Ph.D. Thesis, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.

RIZZON, C., G. MARAIS, M. GOUY and C. BIEMONT, 2002   Recombination rate and the distribution of transposable elements in the *Drosophila melanogaster* genome. Genome Res. **12:** 400–407.

ROHDE, C., 2000   *Polimorfismo cromossômico e elementos transponíveis em Drosophila willistoni*. Ph.D. Thesis, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.

ROHDE, C., H. PINTO, JR., V. H. VALIATI, A. SCHRANK and V. L. VALENTE, 1994   Localization of the Cu/Zn superoxide dismutase gene in the *Drosophila willistoni* species group by *in situ* hybridization. Cytobios **80:** 193–198.

ROHDE, C., E. ABDELHAY, H. PINTO JUNIOR, A. SCHRANK and V. L. VALENTE, 1995   Analysis and *in situ* mapping of the *Adh* locus in species of the willistoni group of Drosophila. Cytobios **81:** 37–47.

ROHDE, C., T. H. DEGRANDI, D. C. DE TONI and V. L. S. VALENTE, 2005   *Drosophila willistoni* polytene chromosomes. I. Pericentric inversion on X chromosome. Caryologia **58:** 249–254.

ROHDE, C., A. C. GARCIA, V. H. VALIATI and V. L. VALENTE, 2006   Chromosomal evolution of sibling species of the *Drosophila willistoni* group. I. chromosomal arm IIR (Muller's element B). Genetica **126:** 77–88.

ROSS, C. L., and T. A. MARKOW, 2006   Microsatellite variation among diverging populations of *Drosophila mojavensis*. J. Evol. Biol. **19:** 1691–1700.

ROY, V., L. MONTI-DEDIEU, N. CHAMINADE, S. SILJAK-YAKOVLEV, S. AULARD *et al.*, 2006   Evolution of the chromosomal location of *rDNA* genes in two *Drosophila* species subgroups: *ananassae* and *melanogaster*. Heredity **94:** 388–395.

RUIZ, A., W. B. HEED and M. WASSERMAN, 1990   Evolution of the *mojavensis* cluster of cactophilic *Drosophila* with descriptions of two new species. J. Hered. **81:** 30–42.

SCHLÖTTERER, C., 2000   Microsatellite analysis indicates genetic differentiation of the neo-sex chromosomes in *Drosophila americana americana*. Heredity **85:** 610–616.

SCHULZE, S. R., B. F. MCALLISTER, D. A. R. SINCLAIR, K. A. FITZPATRICK, M. MARCHETTI *et al.*, 2006   Heterochromatic genes in Drosophila: a comparative analysis of two genes. Genetics **173:** 1433–1445.

SCHWEBER, M., 1974   The satellite bands of the DNA of *Drosophila virilis*. Chromosoma **44:** 371–382.

SEGARRA, C., and M. AGUADÉ, 1992   Molecular organization of the X chromosome in different species of the *obscura* group of Drosophila. Genetics **130:** 513–521.

SEGARRA, C., E. R. LOZOVSKAYA, G. RIBO, M. AGUADÉ and D. L. HARTL, 1995   P1 clones from *Drosophila melanogaster* as markers to study the chromosomal evolution of Muller's A element in two species of the obscura group of Drosophila. Chromosoma **104:** 129–136.

SHIBUSAWA, M., Y. N. TOBARI and M. MATSUDA, 2007   Variations in the chromosomal locations of the *rRNA* genes and pairing ability during male meiosis in the *D. ananassae* complex. Annual Drosophila Research Conference, Vol. 702C, Philadelphia.

SHRIMPTON, A. E., E. A. MONTGOMERY and C. H. LANGLEY, 1986   *Om* mutations in *Drosophila ananassae* are linked to insertions of a transposable element. Genetics **114:** 125–135.

SLAWSON, E. E., C. D. SHAFFER, C. D. MALONE, W. LEUNG, E. KELLMANN *et al.*, 2006   Comparison of dot chromosome sequences from *D. melanogaster* and *D. virilis* reveals an enrichment of DNA transposon sequences in heterochromatic domains. Genome Biol. **7:** R15.

SMITH, C. D., S. SHU, C. J. MUNGALL and G. H. KARPEN, 2007   The release 5.1 annotation of *Drosophila melanogaster* heterochromatin. Science **316:** 1586–1591.

SORSA, V., 1988a *Chromosome Maps of Drosophila*, Vol. I. CRC Press, Boca Raton, FL.

SORSA, V., 1988b *Chromosome Maps of Drosophila*, Vol. II. CRC Press, Boca Raton, FL.

SORSA, V., 1988c *Polytene Chromosomes in Genetic Research.* John Wiley & Sons, New York.

SPASSKY, B., and T. DOBZHANSKY, 1950 Comparative genetics of *Drosophila willistoni.* Heredity **4:** 201–215.

SPASSKY, B., R. C. RICHMOND, S. PEREZ-SALAS, O. A. PAVLOVSKY, C. A. MOURAO *et al.*, 1971 Geography of sibling species related to *Drosophila willistoni*, and the semi-species of the *Drosophila paulistorum* complex. Evolution **25:** 129–143.

SPERLICH, D., and P. PFRIEM, 1986 Chromosomal polymorphism in natural and experimental populations, pp. 257–309 in *The Genetics and Biology of Drosophila*, edited by M. ASHBURNER, H. L. CARSON and J. N. THOMSON. Academic Press, New York.

STATEN, R., S. D. SCHULLY and M. A. NOOR, 2004 A microsatellite linkage map of *Drosophila mojavensis.* BMC Genet. **5:** 1–8.

STEPHAN, W., 1989 Molecular genetic variation in the centromeric region of the X chromosome in three *Drosophila ananassae* populations. II. The *Om (1D)* locus. Mol. Biol. Evol. **6:** 624–635.

STONE, W. S., M. R. WHEELER, W. P. SPENCER, F. D. WILSON, J. T. NEUENSCHWANDER *et al.*, 1957 Genetic studies of irradiated natural populations of Drosophila. Univ. Texas Publs. **5721:** 260–316.

STONE, W. S., W. C. GUEST and F. D. WILSON, 1960 The evolutionary implications of the cytological polymorphism and phylogeny in the virilis group of Drosophila. Proc. Natl. Acad. Sci. USA **46:** 350–361.

STURGILL, D., Y. ZHANG, M. PARISI and B. OLIVER, 2007 Demasculinization of X chromosomes in the Drosophila genus. Nature **450:** 238–241.

STURTEVANT, A. H., and T. DOBZHANSKY, 1936 Geographical distribution and cytology of "sex ratio" in *Drosophila pseudoobscura* and related species. Genetics **21:** 473–490.

STURTEVANT, A. H., and E. NOVITSKI, 1941 The homologies of the chromosome elements in the genus Drosophila. Genetics **26:** 517–541.

STURTEVANT, A. H., and C. C. TAN, 1937 The comparative genetics of *Drosophila pseudoobscura* and *D. melanogaster.* J. Genet. **34:** 415–432.

TAN, C. C., 1935 Salivary gland chromosomes in the two races of *Drosophila pseudoobscura.* Genetics **20:** 392–402.

TAN, C. C., 1937 The cytological maps of the autosomes in *Drosophila pseudoobscura.* Zeitschrift fur Zellforschung u. mikr. Anatomie **26:** 439–462.

TANDA, S., A. E. SHRIMPTON, C. W. HINTON and C. H. LANGLEY, 1989 Analysis of the *Om(1D)* locus in *Drosophila ananassae.* Genetics **123:** 495–502.

TOBARI, Y. N., 1993a *Drosophila ananassae: Genetic and Biological Aspects.* Japan Scientific Society Press, Tokyo.

TOBARI, Y. N., 1993b Linkage maps, pp. 49–51 in *Drosophila ananassae: Genetical and Biological Aspects*, edited by Y. N. TOBARI. Japan Scientific Society Press, Tokyo.

TOBARI, Y. N., B. GONI, Y. TOMIMURA and M. MATSUDA, 1993 Chromosomes, pp. 23–48 in *Drosophila ananassae: Genetical and Biological Aspects*, edited by Y. N. TOBARI. Japan Scientific Society Press, Tokyo.

TOMIMURA, Y., M. MATSUDA, Y. N. TOBARI, M. L. CARIOU, J. L. DA LAGE *et al.*, 1993 Population genetics, pp. 139–198 in *Drosophila ananassae: Genetical and Biological Aspects*, edited by Y. N. TOBARI. Japan Scientific Society Press, Tokyo.

VALENTE, V. L. S., and A. M. ARAUJO, 1985 Observations on the chromosomal polymorphism of natural populations of *Drosophila willistoni* and its association with the choice of feeding and breeding sites. Rev. Bras. Genet. **8:** 271–284.

VALENTE, V. L. S., and A. M. ARAUJO, 1986 Chromosomal polymorphism, climatic factors, and variation in population size of *Drosophila willistoni* in southern Brazil. Heredity **57:** 149–159.

VALENTE, V. L. S., A. RUSZCZYK and R. A. DOS SANTOS, 1993 Chromosomal polymorphism in urban *Drosophila willistoni.* Rev. Bras. Genet. **16:** 307–319.

VALENTE, V. L. S., C. ROHDE, V. H. VALIATI, N. B. MORALES and B. GONI, 2001 Chromosome inversions occurring in Uruguayan populations of *Drosophila willistoni.* Drosoph. Inf. Serv. **84:** 55–59.

VALENTE, V. L. S., B. GONI, V. H. VALIATI, C. ROHDE and N. B. MORALES, 2003 Chromosomal polymorphism in *Drosophila willistoni* populations from Uruguay. Genet. Mol. Biol. **26:** 163–173.

VIEIRA, J., C. P. VIEIRA, D. L. HARTL and E. R. LOZOVSKAYA, 1997a Discordant rates of chromosome evolution in the *Drosophila virilis* species group. Genetics **147:** 223–230.

VIEIRA, J., C. P. VIEIRA, D. L. HARTL and E. R. LOZOVSKAYA, 1997b A framework physical map of *Drosophila virilis* based on P1 clones: applications in genome evolution. Chromosoma **106:** 99–107.

VIEIRA, J., B. F. McALLISTER and B. CHARLESWORTH, 2001 Evidence for selection at the *fused1* locus of *Drosophila americana.* Genetics **158:** 279–290.

VIEIRA, C. P., A. ALMEIDA, J. D. DIAS and J. VIEIRA, 2006 On the location of the gene(s) harbouring the advantageous variant that maintains the X/4 fusion of *Drosophila americana.* Genet. Res. **87:** 163–174.

VIGNEAULT, G., and E. ZOUROS, 1986 The genetics of asymmetrical male sterility in *Drosophila mojavensis* and *Drosophila arizonensis* hybrids: interactions between the Y-chromosome and autosomes. Evolution **40:** 1160–1170.

VON ALLMEN, G., I. HOGGA, A. SPIERER, F. KARCH, W. BENDER *et al.*, 1996 Splits in fruitfly Hox gene complexes. Nature **380:** 116.

WASSERMAN, M., 1960 Cytological and phylogenetic relationships in the repleta group of the genus Drosophila. Proc. Natl. Acad. Sci. USA **46:** 842–859.

WASSERMAN, M., 1962 Cytological studies of the repleta group of the genus Drosophila: V. The mulleri subgroup. Univ. Texas Publs. Stud. Genet. **2:** 85–117.

WASSERMAN, M., 1982 Evolution of the *repleta* group, pp. 61–139 in *The Genetics and Biology of Drosophila*, edited by M. ASHBURNER, H. L. CARSON and J. N. THOMPSON. Academic Press, New York.

WASSERMAN, M., 1992 Cytological evolution of the *Drosophila repleta* species group, pp. 455–552 in *Drosophila Inversion Polymorphism*, edited by C. B. KRIMBAS. CRC Press, Boca Raton, FL.

WHARTON, L., 1942 Analysis of the repleta group of Drosophila. Univ. Texas Publs. **4228:** 23–52.

WHITING, J. H., M. D. PLILEY, J. L. FARMER and D. E. JEFFERY, 1989 *In situ* hybridization analysis of chromosomal homologies in *Drosophila melanogaster* and *Drosophila virilis.* Genetics **122:** 99–109.

YAMADA, H., M. MATSUDA and Y. OGUMA, 2002 Genetics of sexual isolation based on courtship song between two sympatric species: *Drosophila ananassae* and *D. pallidosa.* Genetica **116:** 225–237.

YASUHARA, J. C., C. H. DeCREASE and B. T. WAKIMOTO, 2005 Evolution of heterochromatic genes of Drosophila. Proc. Natl. Acad. Sci. USA **102:** 10958–10963.

Communicating editor: R. S. HAWLEY