A BAC-based physical map of the *Drosophila buzzatii* genome

Josefa González[1], Michael Nefedov[2], Ian Bosdet[3], Ferran Casals[1*], Oriol Calvete[1], Alejandra Delprat[1], Heesun Shin[3], Readman Chiu[3], Carrie Mathewson[3], Natasja Wye[3], Roger A. Hoskins[4], Jacqueline E. Schein[3], Pieter de Jong[2], and Alfredo Ruiz[1,2]

[1] Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona) Spain.

[2] Children's Hospital Oakland Research Institute, Oakland, California 94609, USA.

[3] Genome Sciences Centre, B.C. Cancer Research Centre, Vancouver, Canada V5Z-4E6.

[4] Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.

*Present address: Unitat de Biologia Evolutiva, Facultat de Ciències de la Salut i de la Vida, Universitat Pompeu Fabra, 08003 Barcelona, Spain.

Corresponding author: Alfredo Ruiz, Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona) Spain. Tel. 34-93-581-2729, Fax 34-93-581-2387, e-mail: Alfredo.Ruiz@uab.es.

Running title: Physical map of the *Drosophila buzzatii* genome.

Keywords: Drosophila, BAC library, physical map, fingerprinting, *in situ* hybridization.

ABSTRACT

Large-insert genomic libraries facilitate cloning of large genomic regions, allow the construction of clone-based physical maps and provide useful resources for sequencing entire genomes. *Drosophila buzzatii* is a representative species of the *repleta* group in the Drosophila subgenus, which is being widely used as a model in studies of genome evolution, ecological adaptation and speciation. We constructed a Bacterial Artificial Chromosome (BAC) genomic library of *D. buzzatii* using the shuttle vector pTARBAC2.1. The library comprises 18,353 clones with an average insert size of 152 kb and a ~18X expected representation of the *D. buzzatii* euchromatic genome. We screened the entire library with six euchromatic gene probes and estimated the actual genome representation to be ~23X. In addition, we fingerprinted by restriction digestion and agarose gel electrophoresis a sample of 9,555 clones, and assembled them using FingerPrinted Contigs (FPC) software and manual editing into 345 contigs (mean of 26 clones per contig) and 670 singletons. Finally, we anchored 181 large contigs (containing 7,788 clones) to the *D. buzzatii* salivary gland polytene chromosomes by *in situ* hybridization of 427 representative clones. The BAC library and a database with all the information regarding the high coverage BAC-based physical map described in this paper are available to the research community.

[Supplemental material is available online at www.genome.org. The following individuals kindly provided reagents, samples, or unpublished information as indicated in the paper: S. Celniker, B. Negre and B. Pfeiffer.]

INTRODUCTION

A variety of genomic resources have been developed as part of the Drosophila Genome Project, including the high quality sequence and annotation of the *D. melanogaster* genome (Adams et al. 2000, Celniker and Rubin 2003). Comparatively few genomic resources have been available for other species within the genus Drosophila. Recently, the genome sequence of *D. pseudoobscura* became available (Richards et al. 2005), and whole genome shotgun sequences of ten other Drosophila species are available or in progress (http://rana.lbl.gov/drosophila/multipleflies.html). Phylogenetic analyses indicate that two main lineages exist within the Drosophila genus, which diverged ~60 myr ago (Powell 1997; Tamura et al. 2004). One lineage leads to the *Sophophora* subgenus with ~300 species (including *D. melanogaster* and *D. pseudoobscura*), whereas the other one leads to the subgenera *Drosophila* (including *D. virilis* and *D. buzzatii*) and *Idiomyia* (Hawaiian species), with ~ 700 and 375 described species, respectively (Powell 1997, http://taxodros.unizh.ch). Thus many *Drosophila* species are relatively distantly related to *D. melanogaster* and so genomic resources developed for this species have a somewhat limited applicability to them (Segarra et al. 1995, Ranz et al. 2001, Podemski et al. 2001, González et al. 2002). Fosmid and BAC libraries for some Drosophila species have been produced or are currently in production (http://www.genome.gov; http://tdgc.arl.arizona.edu/baclibraries.htm).

Here, we describe the construction of a BAC library and a BAC-based physical map of the *D. buzzatii* genome. *D. buzzatii* belongs to the *repleta* species group of the *Drosophila* subgenus (Wasserman 1992), a group comprising ~100 species that has been used for studies of ecological adaptation and speciation for over sixty years (Spencer 1941, Crow 1942, Wharton 1942, Barker and Starmer 1982, Barker et al. 1990). Efforts to map the genome of *D.*

buzzatii began fifty years ago with the comparative analysis of its salivary gland chromosomes to establish the phylogenetic relationships between *repleta* group species (Wasserman 1954, 1962; Ruiz et al. 1982, Ruiz and Wasserman 1993). This was followed by the linkage mapping of a small number of visible mutants (Schafer et al. 1993). In the last ten years, around 400 DNA markers have been mapped by *in situ* hybridization to the *D. buzzatii* chromosomes (Ranz et al. 1997, 2003; Laayouni et al. 2000, Casals et al. 2003). No large-insert genomic libraries or clone-based physical maps were previously available for this species.

RESULTS AND DISCUSSION

Construction of *D.buzzatii* BAC library

We constructed a BAC library from the *D. buzzatii* st-1 strain. High molecular-weight (HMW) DNA was prepared from adults, partially digested with *EcoRI* and *EcoRI* methylase, size fractioned and cloned into the pTARBAC2.1 shuttle vector (Hoskins et al. 2000, Osoegawa et al. 2004). The *D. buzzatii* BAC library comprises 18,353 clones arrayed in 48 microtiter plates (see Methods). We determined the average insert size to be 152 kb, by *EcoRI* restriction fingerprinting of 9,555 clones (Figure 1A). The size distribution is somewhat skewed to the right which results in a very high proportion (98.6 %) of cloned inserts larger than 100 kb. The size of the genomes of the *repleta* group species is ~220 Mb, with ~70% in the euchromatin (Ranz et al. 2001), so the expected redundancy of the library is ~18X. We hybridized two gridded filters containing the entire library with six euchromatic gene probes. The average number of positive clones per probe was 23, which provides an estimate of the actual representation of the euchromatin in the library (see Supplemental Material).

Fingerprinting and automatic contig assembly

To build a physical map of the *D. buzzatii* genome, we first fingerprinted and assembled into contigs 9,555 BAC clones using high throughput methods (Marra et al. 1997, Schein et al. 2004). The fingerprint data were automatically assembled using the FPC software (Soderlund et al. 1997, 2000) with a cut-off score of 10^{-11}. This threshold value represents the maximum allowable probability of a chance match between any two clones. The automated assembly produced 634 contigs and 516 unmatched clones (i.e. singletons, see Supplemental Material for further details).

Hybridization of BAC clones to salivary gland chromosomes

We hybridized to the *D. buzzatii* chromosomes 552 clones representing 443 contigs. The information from 427 clones giving one primary hybridization signal was used for map construction. We also hybridized a subset of 163 BAC clones to the chromosomes of *D. repleta*, another species of the repleta group whose cytological maps (Wharton 1942) have been used as the standard reference for all species in this group (Wasserman 1992). The results allowed us to revise the homology between chromosomes and chromosomal segments of *D. buzzatii* and *D. repleta* (Ruiz and Wasserman 1993) and to reconstruct the *D. buzzatii* chromosomes using the *D. repleta* cytological maps (Wharton 1942).

An integrated physical map of the *D. buzzatii* genome

Information from the fingerprint assembly, the cytological localization of BACs, and the library screening with gene probes was merged to produce an integrated physical map (see Methods). Manual editing and merging allowed us to reduce the number of contigs from the initial set of 634 to a final set of 345. Figure 1B shows the distribution of clones within

contigs. The mean number of clones per contig is 26, and the largest number of clones in a contig is 351. The fingerprints of a subset of overlapping clones within each contig were compared, and the size of the genomic region covered by each contig was estimated. The average contig size is estimated to be 338 kb (Figure 1C). Some of the contigs are quite large (30 contigs are larger than 800 kb) although many (216) are relatively small (100-300 kb). The largest contig is ~1.9 Mb.

Using the cytological data, we anchored 181 contigs to the *D. buzzatii* chromosomes. These contigs contain 7,788 (81.5%) of the fingerprinted clones (Supplemental Table S1). Maps of the *D. buzzatii* chromosomes with the cytological localization of the 427 markers and the 181 contigs they represent are shown in Figure 2. A complete list of clones and *in situ* hybridization results is given in Suplemental Table S2.

The size and cytological span of 15 of the largest contigs were used to estimate the DNA content per cytological band in the salivary gland chromosome map. Taking into account the total number of bands and the total size of the contigs included in our integrated map (Figure 2), we estimate that the physical map covers ~89% of the euchromatic portion of the *D. buzzatii* chromosomes. The cytological data indicate that BAC coverage extends nearly to the telomeres while pericentric regions are less well represented, probably due to the high content of repetitive DNA in these regions (Figure 2).

Unrestricted access to the resources described in this paper is provided. A database containing all of the fingerprint images and analyses, clone sizes, contig composition, library screenings, and *in situ* hybridization images can be accessed using iCE (Fjell et al. 2003) at http://www.bcgsc.ca/bioinfo/ice. The *D. buzzatii* BAC library (CHORI-225) is available from BACPAC Resources (http://bacpac.chori.org). We expect that the BAC library and high-

coverage BAC-based physical map will be highly useful resources not only for those working in *D. buzzatii* as a model system but also to all those interested in the comparative analysis of genomes. The usefulness of this BAC-based physical map extends to many repleta group species, because their cytological relationships have been determined using *D. repleta* chromosomes as a reference (Wasserman 1992). In addition, the library has already been used to successfully sequence part of the *Hox* gene complex of *D. buzzatii* (Negre et al. 2005). Finally, the *D. buzzatii* map may help in the assembly of some of the Drosophila genomes currently being sequenced particularly that of *D. mojavensis* which also belongs to the repleta species group.

METHODS

Flies

The *D. buzzatii* strain used to construct the BAC library and to map BACs by *in situ* hybridization (st-1) is fixed for the standard arrangement in all chromosomes and was produced by Betrán et al. (1998). The *D. repleta* stock used for *in situ* hybridization was no. 1611.6 from the National Drosophila Species Resource Center (Bowling Green, OH).

BAC library construction

The library was constructed according to the improved methods described in detail in Frengen et al. (1999) and Osoegawa et al. (1999, 2004). HMW DNA was prepared from 3 g of adult flies, including equal numbers of females and males, as described in Hoskins et al. (2000). The partially digested HMW DNA was size fractionated by Pulse Field Gel Electrophoresis and

fractions corresponding to 150-250 kb DNA fragments were recovered by electroelution and cloned in pTARBAR2.1. See Supplemental Material for further details.

In situ hybridization of BAC clones

In situ hybridizations were carried out as in González *et al*. (2002). Probes were labelled with biotin-16-dUTP . The hybridization temperatures were 37º for *D. buzzatii* chromosomes and 25º for *D. repleta* chromosomes. We hybridized 552 BAC clones to *D. buzzatii* salivary gland chromosomes; 506 gave positive results, and 427 producing a single primary hybridization signal were used in physical map construction (Supplemental Tables S1 and S2). Nine clones gave two signals; these may represent chimeric clones or mixtures of two clones due to cross-well contamination. This low rate (1.6%) is in agreement with the low level of chimerism observed in other BAC libraries (Osoegawa *et al.* 2001). In total, 70 clones gave more than two hybridization signals and/or hybridized to the pericentromeric regions and the microchromosome, probably due to repetitive DNA content. The density of transposable elements increases near Drosophila centric heterochromatin (Kaminker et al. 2002). Clones from such repeat-rich regions cannot be assigned to a particular chromosomal site. Thus, there is a relative scarcity of markers in our physical map near the centric heterochromatin, especially on the X chromosome (Figure 2). Examples of the different types of hybridization results are shown in Supplemental Figure S1.

ACKNOWLEDGMENTS

We thank Susan Celniker, Bàrbara Negre, Barret Pfeiffer, Theresa Ren and Chung Li Shu, for help. This work was supported by grant BMC2002-01708 from the Dirección General de Enseñanza Superior e Investigación Científica (MEC, Spain) awarded to A.R.

FIGURE LEGENDS

Figure 1. Size distribution of the 9,555 *D. buzzatii* BAC clones analyzed by fingerprinting (A). Distribution of clones in contigs (B) and contig sizes for the 345 contigs in the fingerprint map (C).

Figure 2. Integrated BAC-based physical map of the *D. buzzatii* genome. [We consider the cytological map to be a kind of physical map.] Vertical lines indicate the relative position of the 427 BAC clones which produced a primary hybridization signal and represent 181 contigs. Singletons are represented as discontinuous vertical lines. Clone names are shown above the chromosomes. Clone names separated by a bar were hybridized individually. Clone names separated by a plus sign were hybridized as a mixture. The contigs to which the hybridized clones belong are represented by short horizontal segments below the chromosomes along with the contig number. The length of these segments is roughly proportional to contig size.

REFERENCES

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., et al. 2000. The genome sequence of *Drosophila melanogaster*. *Science* **287:** 2185-2195.

Barker, J.S.F., and Starmer, W.T. (eds). 1982. *The cactus-yeast*-Drosophila *model system*. Academic Press, Sydney.

Barker, J.S.F., Starmer, W.T., and MacIntyre, R. (eds). 1990. *Ecological and evolutionary genetics of* Drosophila. Plenum Press, New York.

Betrán, E., Santos, M., and Ruiz, A. 1998. Antagonistic pleiotropic effect of second-chromosome inversions on body size and early life-history traits in *Drosophila buzzatii*. *Evolution* **52**: 144-154.

Casals, F., Cáceres, M., and Ruiz, A. 2003. The foldback-like transposon *Galileo* is involved in the generation of two different natural chromosomal inversions of *Drosophila buzzatii*. *Mol. Biol. Evol*. **20:** 674-685.

Celniker, D.E., and Rubin, G. M. 2003. The *Drosophila melanogaster* genome. *Annu. Rev. Genomics Hum Genet.* **4**: 89-117.

Crow, J.F. 1942. Cross fertility and isolating mechanisms in the *Drosophila mulleri* group. *Univ. Texas Publ*. **4228:** 53-67.

Fjell, C.D., Bosdet, I., Schein, J.E., Jones, S.J., and Marra, M.A. 2003. Internet Contig Explorer (iCE) a tool for visualizing clone fingerprint maps. *Genome Res.* **13**:1244-1249.

Frengen, E., Weichenhan, D., Zhao, B., Osoegawa, K., van Geel, M., and de Jong, P. 1999. A modular, positive selection bacterial artificial chromosome vector with multiple cloning sites. *Genomics* **58:** 250-253.

González, J., Ranz, J.M., and Ruiz, A. 2002. Chromosomal elements evolve at different rates in the *Drosophila* genome. *Genetics* **161:** 1137-1154.

Hoskins, R.A., Nelson, C.R., Berman, B.P., Laverty, T.R., George, R.A., Ciesiolka, L., Naeemuddin, M., Arenson, A.D., Durvin, J., David, R.G., et al. 2000. A BAC-based physical map of the major autosomes of *Drosophila melanogaster*. *Science* **287:** 2271-2274.

Kaminker, J.S., Bergman, C.M., Kronmiller, B., Carlson, J., Svirskas, R., Patel, S., Frise, E., Wheeler, D.A., Lewis, S.E., Rubin, G.M., et al. 2002. The transposable elements of the Drosophila melanogaster euchromatin: a genomic perspective. *Genome Biology* **3:** research0084.

Laayouni, H., Santos, M., and Fontdevila, A. 2000. Toward a physical map of *Drosophila buzzatii*: use of randomly amplified polymorphic DNA polymorphisms and sequence-tagged site landmarks. *Genetics* **156:** 1797-1816.

Marra, M.A., Kucaba, T.A., Dietrich, N.L., Green, E.D., Brownstein, B., Wilson, R.K., McDonald, K.M., Hillier, L.W., McPherson, J.D., and Waterson, R.H. 1997. High throughput fingerprint analysis of large-insert clones. *Genome Res.* **7:** 1072-1084.

Negre, B., Casillas, S., Suzanne, M., Sánchez-Herrero, E., Akam, M., Nefedov, M., Barbadilla, A., de Jong, P., and Ruiz, A. 2005. Conservation of regulatory sequences and gene expression patterns in the disintegrating *Drosophila Hox* gene complex. *Genome Res*. (in press).

Osoegawa, K., de Jong, P., Frengen, E., and Ioannou, P.A. 1999. Construction of bacterial artificial chromosome (BAC) libraries. In *Current Protocols in Human Genetics* (eds. N.C. Dracopoli et al.), pp. 5.15.1-5.15.33. John Wiley & Sons, New York.

Osoegawa, K, Mammoser, A.G., Wu, C., Frengen, E., Zeng, C., Catanese, J.J., and de Jong, P.J. 2001. A bacterial artificial chromosome library for sequencing the complete human genome. *Genome Res.* **11:** 483-496.

Osoegawa, K., Zhu, B., Shu, C.L., Ren, T., Cao, Q., Vessere, G.M., Lutz, M.M., Jensen-Seaman, M.I., Zhao, S., and de Jong, P.J. 2004. BAC resources for the rat genome project. *Genome Res*. **14**:780-785.

Podemski, L., Ferrer, C., and Locke, J. 2001. Whole arm inversions of chromosome 4 in Drosophila species. *Chromosoma* **110:** 305-312.

Powell, J.R. 1997. *Progress and Prospects in Evolutionary Biology. The* Drosophila *Model*. Oxford Univ. Press, New York.

Ranz, J.M., Segarra, C., and Ruiz, A. 1997. Chromosomal homology and molecular organization of Muller's elements D and E in the *Drosophila repleta* group. *Genetics* **145:** 281-295.

Ranz, J.M., Casals, F., and Ruiz, A. 2001. How malleable is the eukaryotic genome? Extreme rate of chromosomal rearrangements in the genus *Drosophila*. *Genome Res*. **11:** 230-239.

Ranz, J.M., González, J., Casals, F., and Ruiz, A. 2003. Low occurrence of gene transposition events during the evolution of the genus Drosophila. *Evolution* **57:** 1325-1335.

Richards, S., Liu, Y., Bettencourt, B. R., Hradecky, P., Letovsy, S., Nielsen, R., Thornton, K., Hubisz, M. J., Chen, R., Meisel, R. P., Couronne, O., et al. 2005. Comparative genome sequencing of *Drosophila pseudoobscura*: Chromosomal, gene, and *cis*-element evolution. *Genome Res.* **15:** 1-18.

Ruiz, A., Fontdevila, A., and Wasserman, M. 1982. The evolutionary history of *Drosophila buzzatii*. III. Cytogenetic relationships between two species of the *buzzatii* cluster. *Genetics* **101:** 503-518.

Ruiz, A., and Wasserman, M. 1993. Evolutionary cytogenetics of the *Drosophila buzzatii* species complex. *Heredity* **70:** 582-596.

Schafer, D.J., Fredline, D.K., Knibbs, W.R., Green, M.M., and Barker, J.S.F. 1993. Genetics and linkage mapping of *Drosophila buzzatii*. *J. Hered*. **84:** 188-194.

Schein, J., Kucaba, T., Sekhon, M., Smailus, D., Waterston, R., and Marra, M. 2004. High-throughput BAC fingerprinting. In: Methods in Molecular Biology (eds. S. Zhao and M. Stodolsky) **255**:143-156. Humana Press: Totowa, New Jersey.

Segarra, C., Lozovskaya, E.R., Ribó, G., Aguadé, M., and Hartl, D.L. 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.

Soderlund, C., Longden, I., and Mott, R. 1997. FPC: a system for building contigs from restriction fingerprinted clones. *Comput. Appl. Biosci*. **13:** 523-535.

Soderlund, C., Humphray, S., Dunham, A., and French, L. 2000. Contigs built with fingerprints, markers, and FPC V4.7. *Genome Res*. **10**: 1772-1787.

Spencer, W.P. 1941. Ecological factors and Drosophila speciation. *Ohio Jour. Sci*. **41:** 190-200.

Tamura, K., Subramanian, S., and Kumar, S. 2004. Temporal patterns of fruit fly (Drosophila) evolution revealed by mutation clocks. Mol. Biol. Evol. 21: 36-44.

Wasserman, M. 1954. Cytological studies of the *repleta* group. *Univ. Texas Publ.* **5422:** 130-152.

Wasserman, M. 1962. Cytological studies of the *repleta* group of the genus *Drosophila*. V. The *mulleri* subgroup. *Univ. Texas Publ*. **6205:** 85-117.

Wasserman, M. 1992. Cytological evolution of the *Drosophila repleta* species group. In Drosophila *Inversion Polymorphism* (eds. C.B. Krimbas and J.R. Powell), pp. 455-552. CRC Press, Boca Raton, Florida.

Wharton, L.T. 1942. Analysis of the repleta group of Drosophila. *Univ. Texas Publ*. **4228:** 23-52.

WEB SITE REFERENCES

http://rana.lbl.gov/drosophila/multipleflies.html; Assembly/Alignment/Annotation of 12 Drosophila Genomes.

http://taxodros.unizh.ch; The database on Taxonomy of Drosophilidae

http://www.genome.gov; National Human Genome Research Institute

http://www.bcgsc.ca/bioinfo/ice; Internet Contig Explorer (iCE) download from the Genome Sciences Centre (British Columbia Cancer Agency) server.

http://bacpac.chori.org/; BACPAC Resources Center at Children's Hospital Oakland Research Institute.

http://tdgc.arl.arizona.edu/baclibraries.htm; Tucson Drosophila Genomics Consortium.