Radiation hybrid mapping of the zebrafish genome

NEIL A. HUKRIEDE*, LUCILLE JOLY†, MICHAEL TSANG*, JENNIFER MILES†, PATRICIA TELLIS‡,
JONATHAN A. EPSTEIN*, WILLIAM B. BARBAZUK§, FRANK N. LI§, BARRY PAW¶, JOHN H. POSTLETHWAIT‖,
THOMAS J. HUDSON‡, LEONARD I. ZON¶, JOHN D. MCPHERSON§, MARIO CHEVRETTE‡, IGOR B. DAWID*,
STEPHEN L. JOHNSON§, AND MARC EKKER†**

*Laboratory of Molecular Genetics and Unit on Biological Computation, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892; †Loeb Health Research at the Ottawa Hospital, Department of Medicine, University of Ottawa, Ottawa, Canada, K1Y 4E9; ‡Montreal General Hospital Research Institute and Department of Surgery, McGill University, Montreal, Canada, H3G 1A4; §Department of Genetics, Washington University Medical School, St. Louis, MO 63110; ¶Howard Hughes Medical Institute and Department of Hematology, Children's Hospital, Boston, MA 02115; and ‖Institute of Neuroscience, University of Oregon, Eugene, OR 97403

ABSTRACT The zebrafish is an excellent genetic system for the study of vertebrate development and disease. In an effort to provide a rapid and robust tool for zebrafish gene mapping, a panel of radiation hybrids (RH) was produced by fusion of irradiated zebrafish AB9 cells with mouse B78 cells. The overall retention of zebrafish sequences in the 93 RH cell lines that constitute the LN54 panel is 22%. Characterization of the LN54 panel with 849 simple sequence length polymorphism markers, 84 cloned genes and 122 expressed sequence tags allowed the production of an RH map whose total size was 11,501 centiRays. From this value, we estimated the average breakpoint frequency of the LN54 RH panel to correspond to 1 centiRay = 148 kilobase. Placement of a group of 235 unbiased markers on the RH map suggests that the map generated for the LN54 panel, at present, covers 88% of the zebrafish genome. Comparison of marker positions in RH and meiotic maps indicated a 96% concordance. Mapping expressed sequence tags and cloned genes by using the LN54 panel should prove to be a valuable method for the identification of candidate genes for specific mutations in zebrafish.

Somatic-cell hybrids and radiation hybrids (RHs) have played a key role in the mapping of human and mouse genes (1–7). Cell hybrids constitute one of the most expedient methods for assigning genes to chromosomes or chromosome segments, because mapping with cell hybrids does not require gene polymorphism. RHs are generated by irradiating cells from a donor species, causing random chromosomal breaks, and fusing these to a cell line from a different species. Donor-cell chromosome fragments are retained to different extents in the ensuing hybrid cells. Typing a panel of RHs with PCR-based sequence-tagged sites creates an RH map in which the frequency of breakpoints between two markers is proportional to the distance between them.

The large collection of mutations produced in the zebrafish constitutes a valuable resource for the study of vertebrate developmental mechanisms (8–10). The efficient identification of the genes disrupted by mutation in zebrafish requires dense maps of the genome. Meiotic maps based on rapid-amplified polymorphic DNA sequences and microsatellite markers have been produced (11–16). Since localization of cDNAs and expressed sequence tags (ESTs) on meiotic maps requires the identification of polymorphisms, the use of RH mapping is a valuable complementary method suitable for high-throughput cDNA/EST mapping projects to identify candidate genes for available mutants.

We have previously shown that stable transfer of zebrafish chromosomes or chromosome segments to a rodent cell line was possible (17). Markers from the simple sequence-length polymorphism (SSLP) meiotic map could be anchored on a panel of zebrafish/mouse somatic-cell hybrids (14). Furthermore, Kwok *et al.* (18) demonstrated that RH technology could be used for nonmammalian vertebrates. In the present study, we report characterization of LN54, a zebrafish RH panel composed of 93 cell lines. We characterized the panel for 1,053 markers, including 84 genes and 122 ESTs, generating a map that we compared with a meiotic map by using a set of common markers.

MATERIALS AND METHODS

Production of RHs. We fused irradiated zebrafish fin AB9 cells to mouse B78 melanoma cells. The B78 recipient cell line is not deficient in an enzyme that could be used to select for zebrafish chromosomal elements in hybrids. Therefore, zebrafish chromosomes were tagged with the aminoglycoside phosphotransferase gene that confers resistance to G418, as described (17). More than 400 independent G418-resistant AB9 clones were pooled for fusion experiments. Briefly, 3×10^7 G418-resistant cells were irradiated with x-ray doses between 2,000 and 9,000 rad, mixed with an equal number of B78 cells, and fused in the presence of polyethylene glycol as described (17). G418 (800 μg/ml) was added 24 h after fusion. No colonies were observed in the controls (irradiated AB9 cells; unfused B78 cells; and irradiated AB9 and B78 cells mixed in the absence of polyethylene glycol). Approximately 3 weeks after fusion, colonies were picked, and the cells were grown for DNA extraction or frozen for future expansion.

The radiation dose applied to break the donor chromosomes is a critical factor because it influences the mapping resolution and retention frequency of a hybrid panel (19). We were able to obtain hybrids at all radiation doses (2,000, 3,000, 4,000, 5,000, 7,000, and 9,000 rad) that were tested. A subset of the lines obtained at 3,000 rad (91 lines), 4,000 rad (125 lines), and 5,000 rad (141 lines) were typed with 18 microsatellite markers (Research Genetics, Huntsville, AL) from linkage group 14 (LG 14), 25 markers from LG25, and 25 markers chosen at random from other LGs. The average retention rates were 21%, 18%, and 21% for the three radiation doses, respectively, with variation between clones within each panel. Preliminary determination of average breakpoint frequency indicated centiRay (1 centiRay (cR) = 1% frequency of a breakage occur-
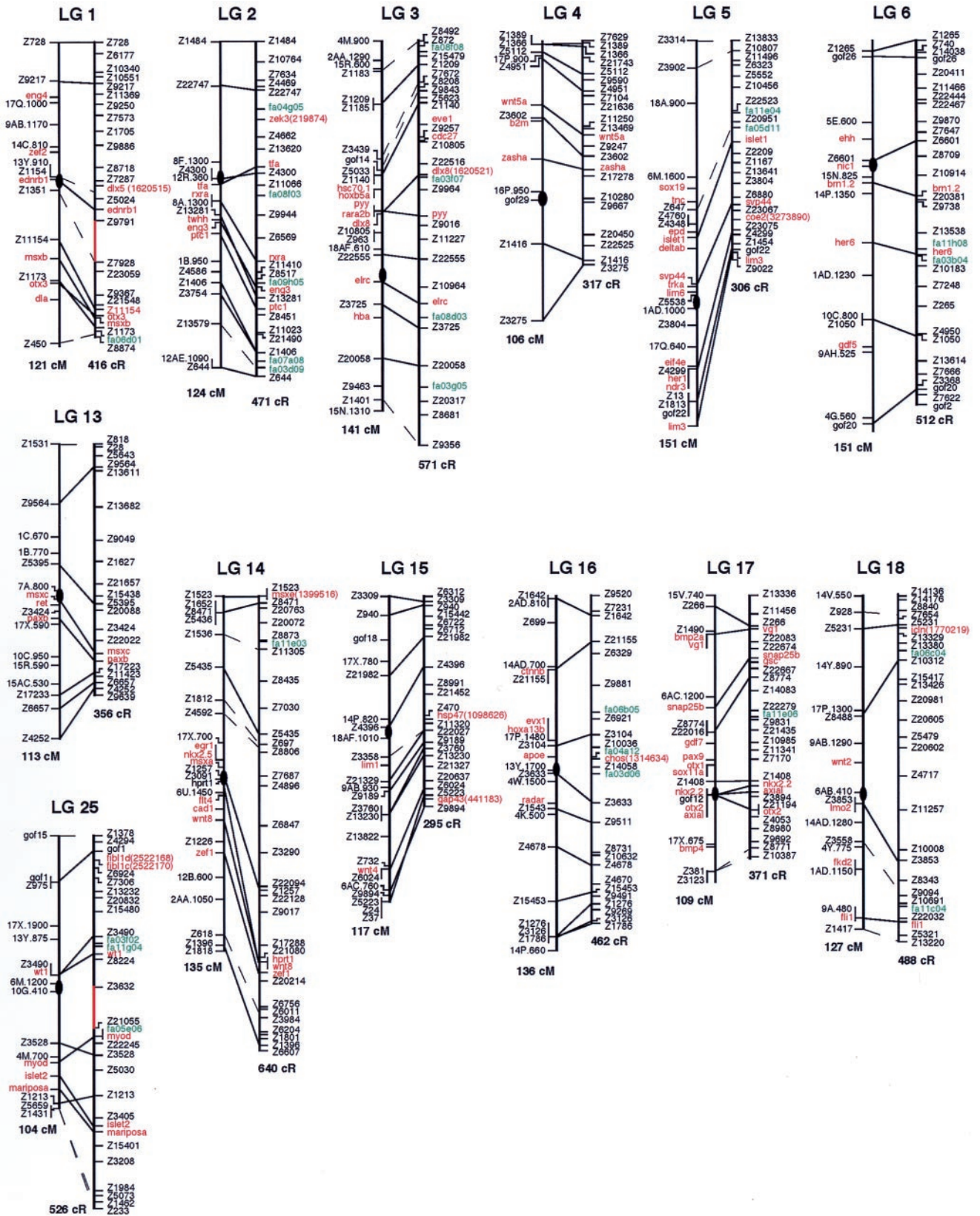

A

FIG. 1.   Comparison of RH and meiotic maps. For each linkage group (LG), the meiotic map is on the left and the RH map is on the right. Solid lines connect the same marker placed on both maps; dashed lines connect markers that fall into the same high-confidence bin. Markers depicted in black are SSLP markers, markers in red are cloned genes, and markers in green are ESTs. Below the map of each LG is the total distance estimate, in centiMorgans (cM) for the meiotic maps and cR for the RH maps. Four gaps remain on the RH maps, regions not linked at lods of five or higher; these gaps are depicted in red (LG 1, 20, 23, and 25). All placement markers are italicized. (*Figure continues on opposite page.*)

B
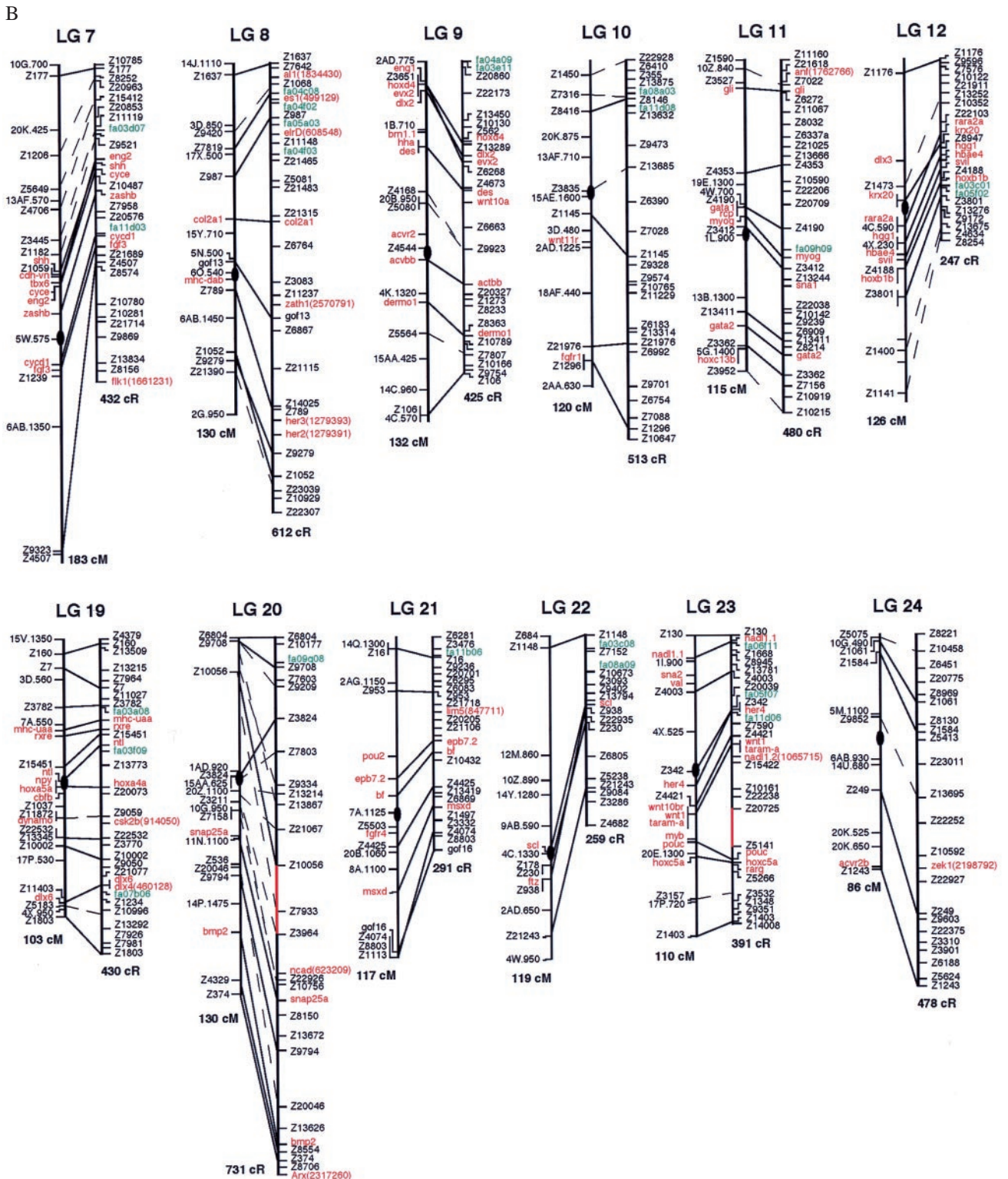
Fig. 1. (*A and legend appear on opposite page.*)

ring between two markers after exposure to a specific radiation dose) values of 181 kilobase (kb), 140 kb, and 99 kb, respectively, for the three radiation doses.

Expansion of a Reference Panel. A total of 93 lines, 81 from the 5,000-rad irradiation and 12 from the 4,000-rad dose, were chosen for expansion, primarily on the basis of a suitable retention rate. An average of 15.6 mg of genomic DNA was harvested from each line, enough for >75,000 assays in duplicate for the protocol outlined below. Expansion did not

result in major changes in retention rate for individual clones. The resulting panel of 93 lines was named LN54 (**L**oeb, **N**IH, **5**,000 rad, **4**,000 rad).

Characterization of the LN54 Panel. The RH panel was characterized with SSLP markers and sequence-tagged sites developed from cloned cDNAs and ESTs. Oligonucleotide primer sequences for SSLP markers were obtained from the M. Fishman lab (http://zebrafish.mgh.harvard.edu/) and synthesized by Midland Certified Reagent (Midland, TX) or by Life

Technologies (Rockville, MD). Oligonucleotides for cDNAs were according to Postlethwait *et al.* (15) or designed by using Hillier's oligonucleotide selection program (20) from deposited GenBank or dbEST sequences. Oligonucleotide sequences for ESTs are available at the web site: http://zfish.wustl.edu.

PCR Reactions. Each PCR reaction contained 100 ng of hybrid-cell DNA, 100 ng of DNA from each of the parental cell lines, or 100 ng of a 1:10 mixture of zebrafish AB9 DNA and mouse B78 DNA/ 0.25 μM the two oligonucleotide primers/10 mM Tris·HCl, pH 8.3/50 mM KCl/1.5 mM MgCl$_2$/0.2 mM each dATP, dCTP, dGTP, and dTTP/1 unit of *Taq* DNA polymerase, in a total volume of 20 μl. PCR was performed for 32 cycles: 30 sec at 94°C, 30 sec at the appropriate annealing temperature for a given primer set, and 30 sec at 72°C; plus a pre-dwell of 4 min at 94°C and a post-dwell of 7 min at 72°C. PCR products were separated on 1.5% agarose gels, and the images were captured with an IS-1000 Digital Imaging System (version 2.02, Alpha Innotech, San Leandro, CA), an Eagle-EyeII (Stratagene), or a Gel Doc 1000 (Bio-Rad). PCR assays were performed in duplicate and, when necessary, up to four times.

Analysis and Map Construction. The RH panel was scored according to Hudson *et al.* (6). Caution is required when scoring the panels, because the donor chromosomal fragments are present at different molarities among the hybrid cell lines; thus, levels of PCR products may vary (6). Therefore, all assays were performed at least twice. Discordance between the multiple runs of a marker was kept at or below 5%; markers with higher discordance were eliminated. The average discordance for all markers in this analysis was 1.5%.

The maps were constructed by using the RHMAPPER program (http://waldo.wi.mit.edu/ftp/distribution/software/rhmapper/), which was first used to create the Whitehead/MIT human RH map by using the Genebridge4 human radiation hybrid panel (6). The method used to build the maps will be described in more detail elsewhere, but the methodology used is essentially as found in the RHMAPPER manual (http://www-genome.wi.mit.edu/ftp/distribution/software/rhmapper/doc/rhmapper.html). The parameters used for generating the maps were similar to those established for the Whitehead/MIT human maps, except for the following parameters: alpha = 0.01, beta = 0.01, retention__frequency = 0.20, placement__to__far = 30 cR.

RESULTS

Mapping the Panel. With meiotic maps well established in zebrafish (11–16), a set of well characterized markers existed with which to establish framework RH maps. Because our preliminary analysis indicated a break frequency of the selected zebrafish radiation hybrids that is comparable to those of RH panels from other species (human, mouse), we decided to assay an initial set of 1,000 markers (SSLPs, ESTs, and cloned genes). We initially selected 25 SSLP markers from each linkage group (13, 14, 21). The choice of markers used for the initial data set was based on reproducible PCR results and spacing over each linkage group. We also tried to incorporate as many anchor markers from the meiotic maps as possible. Additional SSLP markers were chosen as needed to fill in gaps that arose when computationally mapping the genome (see below). Additionally, to incorporate markers from the other established meiotic maps, we used 84 cloned genes from the Mother of Pearl map (11, 12, 15). Finally, we used a set of 122 ESTs. In total, 1,053 markers were used to establish the maps.

The first step in building the zebrafish RH map was to construct a set of high-quality framework maps across each linkage group with the 25 SSLP markers. Analysis of the RH vectors by using RHMAPPER allowed us to create an initial set of RH linkage groups with threshold lod scores for linkage of

10 or greater (the lod score is the logarithm of the likelihood ratio for linkage). Connectivity between each of these high-confidence linkage groups was determined by finding markers within two different groups that linked at a lod score of five or better. If a significant linkage between groups, predicted to be linked based on the meiotic maps, could not be established by using the above criterion, additional SSLP markers, localized within the gap, were assayed. The framework map was subjected to the "ripple test," which evaluated local permutations, and selected the optimal order. Regions that contained gaps when all of the SSLP markers had been used were either typed with sequence-tagged sites from cloned genes or ESTs, which fell within the defined gap. If none existed in the area, the region was left as a gap; four gaps remain over the entire genome (shown in red in Fig. 1). The resulting framework maps consist of 703 SSLP, EST, and cloned genes.

We attempted to localize an additional 350 markers relative to the framework map. These markers were positioned on placement maps because (*i*) they are redundant in the sense that they map to nearly identical positions as markers from the framework or (*ii*) they caused map expansion, as determined by the "exclusion test" of RHMAPPER. Any potential placement marker must be placed within 30 cR of a framework marker in order for it to be added to the placement map. Placement markers do not contribute to the total size of the map (in cR) and do not affect the spacing between adjacent framework markers. Of the 350 potential placement markers, 250 were positioned in this way. Among them, 43 were positioned on a different linkage group when compared with their meiotic map position. At present, it is uncertain whether these markers are incorrectly linked on our RH maps or on the meiotic maps; additional analysis will be needed to determine their correct positions. In total, there are 953 markers on either the framework maps or the placement maps.

The remaining 100 markers did not fall within the above parameters used for creating placement maps. Possibilities to explain the failure to map these markers include (*i*) the markers may fall into chromosomal regions that currently have too low of a density of markers for proper mapping and (*ii*) the markers were typed incorrectly.

The RH maps were compared with meiotic maps (Fig. 1) generated as described by Johnson *et al.* (12). Distances between adjacent markers and the local order of markers were compared. The maps display the meiotic order on the left and the RH order on the right. Solid lines are direct comparisons between markers that were placed on both panels, whereas the dashed lines are comparisons between markers that fall into the same high-confidence bin. The RH maps do not display all markers because of space constraints. The markers depicted were chosen for two reasons: (*i*) the markers are found on both maps and (*ii*) the markers are evenly spaced across the linkage group. RH maps showing all available markers can be seen at the following web sites: http://dir.nichd.nih.gov/lmg/lmgdev.htm; http://zfish.uoregon.edu/ZFIN/; or http://zfish.wustl.edu.

Retention Rate. The overall retention rate for the expanded RH panel was ≈22%. For individual linkage groups, retention rates varied between 13% for LG12 to 36% and 37% for LG14 and LG20, respectively (Fig. 2). Selection for the presence of zebrafish chromosome fragments in hybrids was based on resistance to G418 conferred by the transfected *neo* gene. Thus, the relatively high retention of LG14 and LG20 could be an indication that they contain the sites of integration of the drug-resistance marker. We consider this unlikely, though, because hybrid fusions were made with a pool of more than 400 G418-resistant colonies. Thus, we believe that *neo* resistance in hybrids is likely based on random integration events, and no one area of the genome should be more likely than any other to confer resistance to G418.
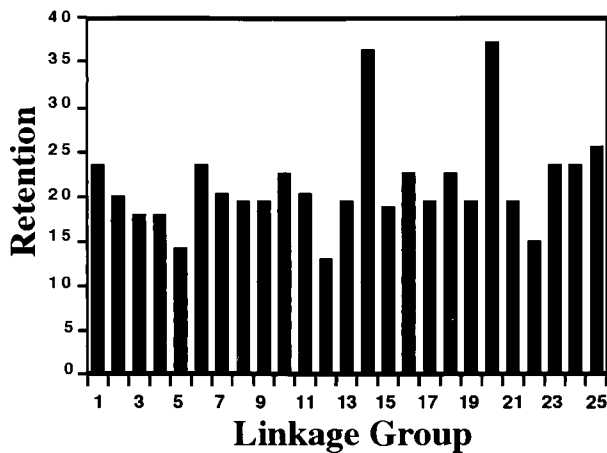
FIG. 2.   Average retention of the LN54 Panel.

Resolution and Coverage of the Genome. Estimates of average resolution of a panel are based on average retention and total length, in cR, for all of the framework maps. Adding the total length for all 25 linkage groups results in a total genome size estimate of 11,501 cR. The estimated size of the zebrafish genome of 1.7×10^9 bp (22) leads to a value of 1 cR = 148 kb. As a means of comparison, the human GeneBridge4 RH panel, which has been widely used, has a retention rate of 30% and a breakage frequency corresponding to 1 cR = 300 kb (5, 6).

At this time, the effective resolution of the map is largely determined by the interval size between framework markers ($\approx$2.4 megabase on average). In RH mapping, framework markers can only be ordered with a high likelihood if they are separated by several breakpoints. Based on the analysis of other RH panels, such as Genebridge4, we estimate that a high-likelihood framework map can grow to a density of 1 marker every 3–4 breakpoints. With a value of 1 cR = 148 kb, we thus estimate an effective potential resolution of the LN54 panel to be $\approx$500 kb.

To obtain an estimate of total genome coverage, a group of 235 additional markers was chosen. This group consisted of ESTs, SSLP markers, and cloned genes. The choice of the group was twofold (*i*) The ESTs were chosen randomly as an unbiased group of markers (122 ESTs). (*ii*) The SSLP markers and cloned genes (112 total) were chosen so that their placements could be compared with the meiotic maps. The probability of mapping any given marker with the LN54 panel, which can be an estimate of genome coverage, was found to be 88% for this set of markers. Breaking the markers into two groups, ESTs versus cloned genes and SSLP markers, indicates 87% coverage for the ESTs and 89% coverage for the cloned genes and SSLP markers. Comparing the RH-mapped positions of the cloned genes and SSLP markers (those 89% that could be mapped) with their positions on meiotic maps resulted in 96% concordance.

DISCUSSION

The LN54 RH panel is one of two RH panels available for mapping zebrafish genes (18); these panels can form the basis for large-scale EST mapping of the zebrafish genome. We compared the LN54 panel to panels in other species. The Genebridge 4 panel has a potential resolution of 1 megabase, whereas our panel has a potential resolution of about 500 kb (5, 6). The T31 panel (mouse) is also a first-generation panel and has a potential resolution of 378 kb (23). Multiple factors can influence the resolution of an RH panel and explain differences in resolution between the LN54, Genbridge 4, and T31 (three panels made by using roughly similar doses of

radiation). Genomes from different species may have different sensitivities to radiation doses, and differences in equipment might have resulted in differences in the true dose of radiation for one or more of the panels. Moreover, the possibility exists that the total RH map lengths generated by different programs can vary significantly. For example, Hudson *et al.* (6) used RHMAPPER to map the Genebridge 4 panel, whereas Gyapay *et al.* (5) used the RHMAP package to map an earlier, expanded version of the Genebridge 4 panel. In this case, RHMAPPER generated an average value of 300 kb/cR, whereas RHMAP generated an average value of 208 kb/cR. This demonstrates that two different mapping programs could yield differences as great as 31% for the same panel. RHMAPPER was used for the LN54 and Genebridge 4 maps, whereas RHMAP was used to map the T31 panel and it is, therefore, possible that the estimated potential resolution of the LN54 panel would be influenced by the mapping program used. The other RH panel presently available for the zebrafish was mapped with the SAMAPPER program (R. Geisler, personal communication), and it is therefore difficult to compare the estimated resolution of the two panels at this time.

If the LN54 panel is to be a viable option for rapid mapping of ESTs, the panel needs to produce a map with relatively good coverage of the entire zebrafish genome. Estimates for coverage of the zebrafish RH map generated with the LN54 panel is 88%; the genome coverage of the map obtained with the Genebridge 4 panel was estimated to be 100%, whereas a second human panel, G3, provided a map with 74% coverage (6, 24).

Two possible factors might explain the <100% coverage of the genome by the LN54 panel. First, there may not be enough zebrafish DNA retained in each hybrid to cover the entire genome. The overall retention of the LN54 panel is 22%, which is barely within the parameters (25) suggested for full coverage of a genome (20–50% retention for 100 hybrid lines). This is especially true if one considers that some linkage groups have even lower retention (Fig. 2). Second, we may not currently have enough markers on the panel to map at 100% coverage. There are four gaps found on our framework maps. It is not clear if these gaps are true representations of regions of the zebrafish genome that are not retained in the LN54 panel or if these regions simply contain too few markers to be mapped at present. By typing more markers on the LN54 panel and finding more markers that improve the framework maps, we may eventually increase the current 88% estimated coverage.

These results demonstrate that the LN54 RH panel will be an effective tool for the mapping of an extensive collection of zebrafish ESTs, and hence, for the molecular characterization of mutations that disrupt genes essential for vertebrate development, organ morphogenesis, physiology, behavior, and health.

1. Ringertz, N. R. & Savage, R. E. (1976) *Cell Hybrids* (Academic, London).
2. Ruddle, F. H. (1981) *Nature (London)* **294,** 115–119.
3. Goss, S. & Harris, H. (1975) *Nature (London)* **255,** 680–684.
4. Cox, D. R., Burmeister, M., Price, E. R., Kim, S. & Myers, R. M. (1990) *Science* **250,** 245–250.

5. Gyapay, G., Schmitt, K., Fizames, C., Jones, H., Vega-Czarnay, N., Spillet, D., Muselet, D., Prud'homme, J.-F., Di, C., Auffray, C., *et al.* (1996) *Hum. Mol. Genet.* **5,** 339–346.
6. Hudson, T. J., Stein, L. D., Gerety, S. S., Ma, J., Castle, A. B., Silva, J., Slonim, D. K., Baptista, R., Kruglyak, L., Xu, S. H., *et al.* (1995) *Science* **270,** 1945–1954.
7. Schuler, G. D., Boguski, M. S., Stewart, E. A., Stein, L. D., Gyapay, G., Rice, K., White, R. E., Rodriguez-Tome, P., Aggarwal, A., Bajorek, E., *et al.* (1996) *Science* **274,** 540–546.
8. Kimmel, C. B. (1989) *Trends Genet.* **5,** 283–288.
9. Driever, W., Solnica-Krezel, L., Schier, A. F., Neuhauss, S. C. F., Malicki, J., Stemple, D. L., Stainier, D. Y. R., Zwartkruis, F., Abdelilah, S., Rangini, Z., *et al.* (1996) *Development* (Cambridge, U.K.) **123,** 37–46.
10. Haffter, P., Granato, M., Brand, M., Mullins, M. C., Hammerschmidt, M., Kane, D. A., Odenthal, J., van Eeden, F. J. M., Jianmg, Y.-J., Heisenberg, C.-P., *et al.* (1996) *Development (Cambridge, U.K.)* **123,** 1–36.
11. Postlethwait, J. H., Johnson, S. L., Midson, C. N., Talbot, W. S., Gates, M., Ballinger, E. W., Africa, D., Andrews, R., Carl, T., Eisen, J. S., *et al.* (1994) *Science* **264,** 699–703.
12. Johnson, S. L., Gates, M. A., Johnson, M., Talbot, W. S., Horne, S., Baik, K., Rude, S., Wong, J. R. & Postletwait, J. H. (1996) *Genetics* **142,** 1277–1288.
13. Knapik, E. W., Goodman, A., Atkinson, O. S., Roberts, C. T., Shiozawa, M., Sim, C. U., Weksler-Zangen, S., Trolliet, M. R., Futrell, C., Innes, B. A., *et al.* (1996) *Development (Cambridge, U.K.)* **123,** 451–460.
14. Knapik, E. W., Goodman, A., Ekker, M., Chevrette, M., Delgado, J., Neuhaus, S., Shimoda, N., Driever, W., Fishman, M. C. & Jacob, H. J. (1998) *Nat. Genet.* **18,** 338–343.
15. Postlethwait, J. H., Yan, Y.-L., Gates, M. A., Horne, S., Ekker, M., Amores, A., Brownlie, A., Donovan, A., Egan, E. S., Force, A., *et al.* (1998) *Nat. Genet.* **18,** 345–349, and erratum (1999) **19,** 303.
16. Shimoda, N., Kanpik, E. W., Ziniti, J., Sim, C., Yamada, E., Kaplan, S., Jackson, D., de Sauvage, F., Jacob, H. & Fishman, M. C. (1999) *Genomics*, in press.
17. Ekker, M., Speevak, M. D., Martin, C. C., Joly, L., Giroux, G. & Chevrette, M. (1996) *Genomics* **33,** 57–64.
18. Kwok, C., Korn, R. M., Davis, M. E., Burt, D. W., Critcher, D. W., McCarthy, L., Paw, B. H., Zon, L. I., Goodfellow, P. N. & Schmitt, K. (1998) *Nucleic Acids Res.* **26,** 3562–3566.
19. Siden, T. S., Kumlien, J., Schwartz, C. E. & Rohme, D. (1992) *Somaic. Cell Mol. Genet.* **18,** 33–44.
20. Hillier, L. & Green, P. (1991) *PCR Methods Appl.* **2,** 124–128.
21. Goff, D. J., Galvin, K., Katz, H., Westerfield, M., Lander, E. S. & Tabin, C. J. (1992) *Genomics* **14,** 200–202.
22. Hinegardner, R. (1968) *Am. Nat.* **102,** 517–523.
23. McCarthy, L. C., Terret, J., Davis, M. E., Knights, C. J., Smith, A. L., Critcher, R., Schmitt, K., Hudson, J., Spurr, N. K. & Goodfellow, P. N. (1997) *Genome Res.* **7,** 1153–1161.
24. Stewart, E. A., McKusick, K. B., Aggarwal, A., Bajorek, E., Brady, S., Chu, A., Fang, N., Hadley, D., Harris, M., Hussain, S., *et al.* (1997) *Genome Res.* **7,** 422–433.
25. Walter, M. A. & Goodfellow, P. N. (1993) *Trends Genet.* **9,** 352–356.