# A BAC Contig Map of the *Ly49* Gene Cluster in 129 Mice Reveals Extensive Differences in Gene Content Relative to C57BL/6 Mice

Andrew P. Makrigiannis,[1] Amanda T. Pau,[1] Pamela L. Schwartzberg,[2] Daniel W. McVicar,[1] Thomas W. Beck,[3] and Stephen K. Anderson[1,3,*]

[1]*Laboratory of Experimental Immunology, Center for Cancer Research, National Cancer Institute–Frederick, Frederick, Maryland 21702-1201, USA*
[2]*National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA*
[3]*Intramural Research Support Program, SAIC–Frederick, National Cancer Institute–Frederick, Frederick, Maryland 21702-1201, USA*

**To whom correspondence and reprint requests should be addressed. Fax: (301) 846-1673. E-mail: andersn@ncifcrf.gov.*

**The murine *Ly49* gene family is functionally analogous to the human killer cell Ig-like receptor (KIR) family of class I major histocompatibility complex (MHC) receptors. The number of *KIR* genes varies dramatically between individuals; however, the organization of the *Ly49* genes has only been determined for the C57BL/6 (B6) mouse. The organization of the 129 *Ly49* loci was determined from a BAC contig map by PCR and Southern blot analysis. In addition to the 10 *Ly49* genes known from previous studies of the 129/J strain, 8 new genes were localized to the 129 *Ly49* cluster. A gene order of *Ly49*$q_1$, *e*, (*v*, $q_2$), *e*/$c_2$, *l*/*r*, *s*, *t*, *e*/$c_1$, *r*, *u*, *w*/*i*, $i_1$, *g*, *p*/*d*, ($i_2$, *p*), and *o* was determined. The 129 *Ly49* gene cluster is predicted to span ~ 600 kb. These results indicate that *Ly49* gene numbers can be significantly different between inbred mouse strains, analogous to the haplotype differences observed in the human *KIR* genes.**

**Key Words: mouse, NK cells, MHC receptor, Ly49**

## INTRODUCTION

The murine Ly49 family of class I MHC receptors represent the functional analogs of the human KIR family. Although they represent structurally distinct protein families, the C-type lectin-related Ly49 proteins and the Ig-related KIR proteins associate with identical signaling molecules to achieve either activation or inhibition of natural killer (NK) cells in response to specific MHC class I ligands [1], indicating convergent evolution of function. There is a high degree of variability in *KIR* expression patterns between individuals [2]. The *KIR* locus of a given haplotype seems to be composed of framework genes present in most humans, interspersed with regions of variable gene content [3].

The *Ly49* gene cluster has been mapped in the B6 mouse genome to a region called the NK complex (NKC) on chromosome 6, and as is the case for the human *KIR* cluster, the genes are arranged in tandem with an identical transcriptional orientation. All known genes, except *Ly49b*, are clustered in a region predicted to be ~ 420 kb in length [4–6]. *Ly49b* is the most divergent member of the family and is situated 750 kb telomeric of *Ly49a*. Resistance to ectromelia and murine cytomegalovirus (MCMV) infection has been genetically linked to the NKC [7,8]. Recent reports have shown that the activating receptor, Ly49H, is required for resistance to MCMV infection [9]. Identification of this viral resistance gene was facilitated by the observation of major genomic polymorphisms in the *Ly49* cluster between B6 and DBA/2 mice [10,11].

Of the 14 *Ly49* genes currently identified in the B6 genome, 10 have been shown to produce mRNAs with a complete coding region (*Ly49a–j*), three represent transcribed pseudogenes (*Ly49k*, *m*, and *n*), and the remaining gene (*Ly49l*) does not produce a transcript in the B6 strain [12–16]. Gene silencing as a mechanism of repertoire diversity is evidenced by the reports that pseudogenes such as *Ly49l* and *Ly49m* in B6 mice produce functional transcripts in other mouse strains [15,16]. Ly49-specific monoclonal antibody staining of NK cells from different mouse strains has indicated that a high degree of polymorphism exists in the *Ly49* loci [17]. Southern analysis of a variety of mouse strains with *Ly49*-specific probes indicated highly variable restriction fragment sizes and numbers [5]. Recent studies have shown that the presence of the *Ly49h* gene is strain-dependent [10,11].

To gain a better understanding of differences in *Ly49* expression between mouse strains, extensive screening of a 129/J NK-cell cDNA library was undertaken and resulted in the identification of 10 different 129 *Ly49* genes, including



0888-7543/02 $35.00

**TABLE 1:** 129 *Ly49* gene sequence characteristics

| *Ly49* gene | Arginine for DAP12 assoc.[a] | exon 3/intron 3 boundary | intron 3 size (bp) | intron 3/exon 4 boundary |
|---|---|---|---|---|
| *e* | no | aaaga]gtgag | 869 | tttag[ttttt |
| *g* | no | aacga]gtgag | 1205 | cttag[ttttt |
| $i_1$ | no | aaaga]gtgag | 850 | tttag[ttttt |
| *o* | no | aaaaa]gtgag | 1141 | cttag[ttttt |
| *p* | yes | aaaca]gtgag | 883 | cttag[ttttt |
| *r* | yes | aaaca]gtgag | 881 | cttag[ttttt |
| *s* | no | aaaga]gtgag | 857 | tttag[ttttt |
| *t* | no | aaaaa]gtgag | 1204 | ttaag[tcaac |
| *u* | yes | aaaga]gtaag | 917 | tttag[ttttt |
| *v* | no | aaaaa]gtgag | 1146 | cttag[ttttt |
| *b* | no | acaca]gtgag | 881 | tttag[ttttc |
| $e/c_1$ | no | aaaga]gtgag | 871 | tttag[ttttt |
| $e/c_2$ | no | aaaga]gtgag | 870 | tttag[ttttt |
| $i_2$ | no | aaaga]gtgag | 853 | tttag[atttt |
| *l/r* | yes | aaaca]gtgag | 888 | cttag[ttttt |
| *p/d* | yes | aaaca]gtgag | 865 | cttag[ttttt |
| $q_1$ | no | aaaca]gtgag | 873 | tttag[ttttg |
| $q_2$ | no | aaaca]gtaag | 879 | tttag[ttttg |
| *u/i* | yes | aaaga]gtaag | 947 | tttag[ttttt |

[a]All 129 sequences were aligned with PileUp (SeqWeb, GCG) using progressive pairwise alignment and the relevant codon was compared to previously characterized 129 *Ly49* genes.

several novel family members (*Ly49e*, *g*, *i*, *o*, *p*, *r*, *s*, *t*, *u*, and *v*) [18]. Some 129/J *Ly49* cDNAs were clearly alleles of B6 genes (*Ly49e*, *g*, and *i*), whereas others appeared to be new genes (*Ly49t*, *v*, and *p*). The remaining 129 genes were given new designations, as it was not clear whether they should be assigned as alleles [18]. To resolve this question, and to gain a better understanding of the total *Ly49* gene content in the 129 strain, a physical map of the 129 *Ly49* gene cluster was constructed using restriction fingerprint analysis combined with Southern hybridization and PCR screening of bacterial artificial chromosomes (BACs). The data presented demonstrate that plasticity in gene content between haplotypes represents another parallel between the *Ly49* and *KIR* gene families.

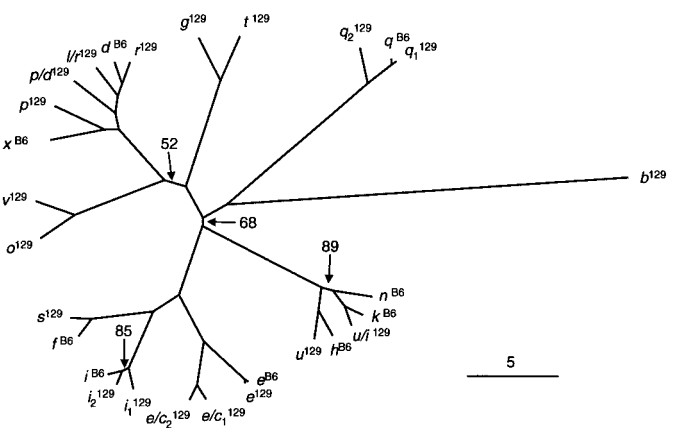



**FIG. 1.** Phylogram of 129 and B6 *Ly49* intron 3 sequences. An unrooted phylogenetic tree of the intron 3 sequence from 28 *Ly49* genes is shown. The intron 3 sequences for *Ly49c*, *i*, *h*, *n*, *k*, and *d* for the B6 strain were discovered in GenBank (acc. no. AC090127). B6-*Ly49e*, *f*, *q*, and *x* intron 3 sequences were found in GenBank (acc. no. AC090563; *Ly49x* is located in the region listed as 136,346–142,211 bp). GenBank accession numbers for 129 *Ly49* genes are provided in the Materials and Methods, as are tree construction details. Numbers at branch nodes represent bootstrap confidence-level percentages of 1000 bootstrap replications. Unlabeled branch nodes have a bootstrap value of > 90%. The scale bar indicates the number of substitutions per 100 nt.

## RESULTS AND DISCUSSION

### Isolation of 129 Ly49 BAC Clones

To investigate the genomic organization of the *Ly49* gene locus in the 129 inbred mouse strains, the 129/SvJ BAC library, ES1-RW4, was screened by PCR using primers specific for several *Ly49* exon 4 sequences. We isolated four different clones, allowing the initial characterization of the region containing $Ly49s^{129}$ to $u^{129}$. To characterize the remaining *Ly49* loci and to gain sufficient numbers of overlapping BAC clones to construct a contig, a new genomic BAC library, RPCI-22 (Segment I) from 129/SvEvTac, with greater genomic coverage was then screened using all known 129/J Ly49 cDNAs as probes to hybridize to library filters. A total of 42 clones were initially identified and 36 positives were confirmed by PCR, partial nucleotide sequencing, and Southern analysis.

There is considerable genetic variation among 129 substrains, especially the 129/SvJ mouse [19]. However, a genome-wide analysis of simple sequence length polymorphisms has shown that the region of the 129/SvJ genome on which the NKC resides is of 129 origin [20]. We have previously isolated the $Ly49r^{129/SvJ}$, $s^{129/SvJ}$, and $v^{129/SvJ}$ cDNAs and found them to be identical to the 129/J sequences (A.P.M. and S.K.A., unpublished data). Collectively, this strongly suggests that 129/SvJ mice possess a 129-derived *Ly49* gene cluster. 129/SvEvTac mice are much more closely related to the 129/J strain than is 129/SvJ [19].

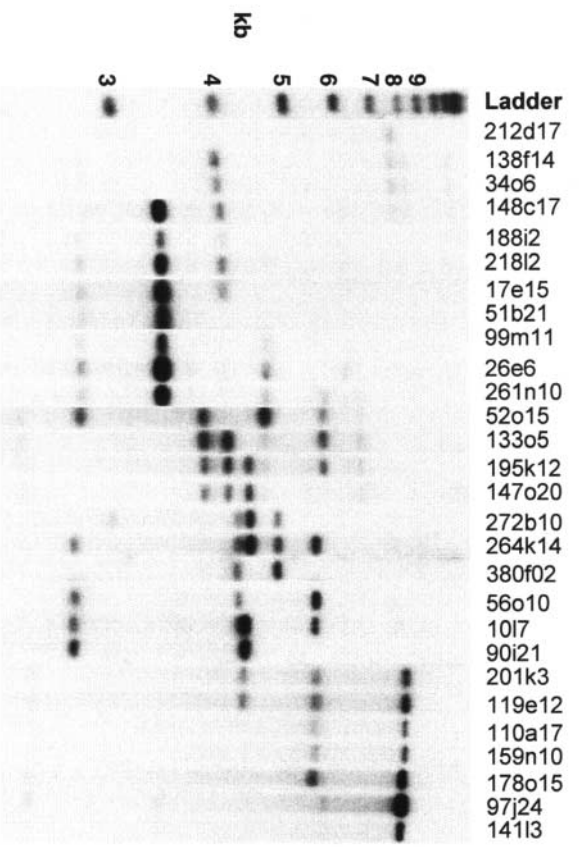



**FIG. 2.** Southern analysis of 129 BAC clones. A representative experiment showing hybridizing fragments of an ordered set of BAC clones. BACs were digested with *Eco*RI, separated in a 1% agarose gel, and then stained with SYBR-Green. After transfer, an *Ly49o*-exon 4 probe was hybridized to the membrane at 65°C.

**BAC Gene Content Deduced by PCR**

A PCR-based assay was developed to characterize each BAC clone with respect to the types and numbers of *Ly49* genes present. Primers specific for exons 3 and 4 of each known 129 *Ly49* gene were designed and used to amplify *Ly49* gene fragments from each BAC. All PCR products that were in the correct size range for intron 3 (0.9 kb) [21] were cloned and sequenced. The 3′ end of exon 3 and 5′ end sequence of exon 4 allowed the identification of the *Ly49* gene from which the clone was derived. All known 129/J *Ly49* cDNAs found identical matches in the exon 3/exon 4 partial sequences from the 129/SvEvTac-derived BAC clones. This suggests that the 129/SvEvTac, 129/SvJ (discussed above), and 129/J substrains all possess a "129" *Ly49* haplotype.

The *Ly49*-specific primers also amplified many new genes in addition to the genes for which they were designed. The total number of different *Ly49* genes discovered among all the 129 BAC clones tested was 19 and are summarized in Table 1. It was decided to temporarily name the new 129 genes based on exon 3/exon 4 homologies and give subscripts to those for which the same name was already in use. Nine new genes (*Ly49b*, $e/c_1$, $e/c_2$, $i_2$, *l/r*, *p/d*, *u/i*, $q_1$, and $q_2$) were discovered in addition to the 10 previously known for this strain. Two different sequences were found that were highly homologous to a novel B6 transcript deposited in GenBank and known as *Ly49q* (GenBank acc. no. AB033769). The new 129 *Ly49* genes are not unique to the 129/SvEvTac strain as the $Ly49e/c_1$ and *u/i* partial genomic sequences were also detected in 129/SvJ BAC clones.

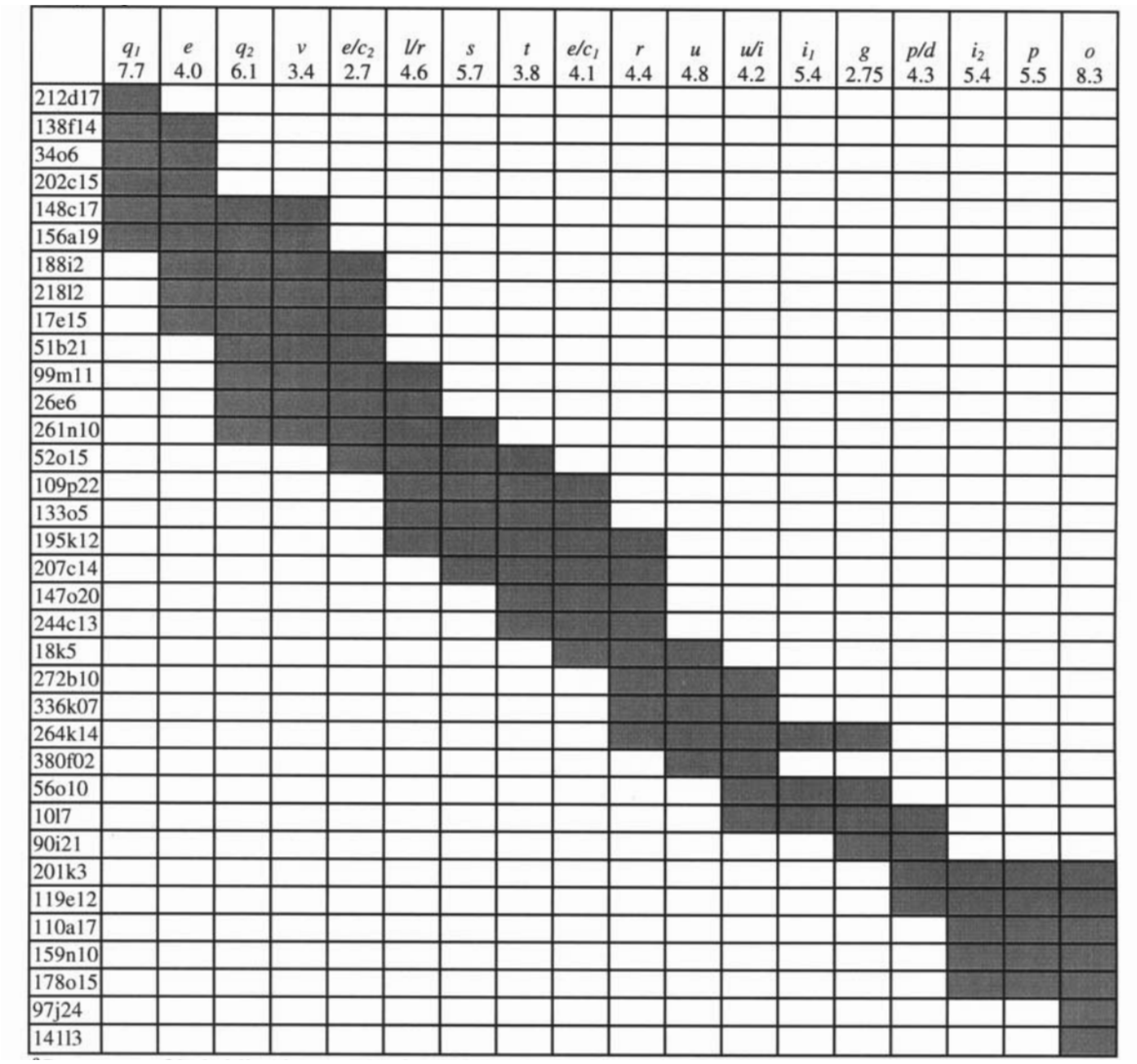
To further characterize the new 129 *Ly49* genes, we analyzed the exon 3 coding sequence for an arginine codon in the putative transmembrane domain characteristic of activating Ly49 receptors. In addition to the previously known activators (*Ly49p*, *r*, and *u*), *Ly49l/r*, *p/d*, and *u/i* also possessed an arginine codon (Table 1). Thus, 129 mice may possess up to six activating *Ly49* genes. However, whether any of the latter three express functional transcripts is unknown. To gain further insight into the origin of these new genes, the full sequence of the exon 3/intron 3/exon 4 regions was determined. Table 1 summarizes the observed length of intron 3 as well as the predicted exon 3/intron 3 and intron 3/exon 4 boundaries. The placement of the intron sequence between exon 3 and 4 of the 10 previously characterized 129 *Ly49* cDNAs (*Ly49e–v*) confirmed the predicted intron 3 boundaries for these genes and suggests that the exon/intron structure of all *Ly49* genes is similar to that of *Ly49a* [21]. The average intron 3 size for these genes was ~ 875 bp. Phylogenetic analysis was done on these 129 gene fragments along with all available B6 intron 3 sequences. Figure 1 shows the intron 3 relationships of the new 129 *Ly49* genes to each other and to most B6 *Ly49* genes. The clustering shown (Fig. 1) for the 10 "old" 129 genes reflect the homologies we previously reported for the whole coding region [18]. Of the new 129 genes, *Ly49u/i* clustered with the B6-*Ly49h*, *n*, and *k* group, $Ly49e/c_1$ and $e/c_2$ clustered with *Ly49e*, *Ly49l/r* and *p/d* grouped with the *Ly49d* subgroup. $Ly49i_2$, $q_1$, and $q_2$ were most homologous to their B6 counterparts. B6-'*x*' is a gene discovered in GenBank (AC090563) and is most similar to *Ly49d* over the putative coding region.

**Characterization of 129 BAC Clones by Southern Hybridization**

To confirm the gene content of each BAC predicted by the PCR analysis described above and to generate a gene order, BAC clones were digested with *Eco*RI and analyzed by Southern hybridization (Fig. 2). Initial hybridization was carried out at low stringency to detect all *Ly49*-related fragments. Once a catalog of fragment sizes for each BAC was completed, specific probing at higher stringency was done with the exon 3–exon 4 PCR fragments that were generated and sequenced for the PCR analysis described above. This matched a specific restriction fragment to each 129 *Ly49* gene. The gene content

doi:10.1006/geno.2002.6724, available online at http://www.idealibrary.com on IDEAL

**TABLE 2:** Ordering of 129 *Ly49* genes by comparision of BAC clone restriction fragments hybridizing to *Ly49* probes[a]




[a]Summary of hybridization results for 129 BAC clones. The left column shows the relative order of BAC clones. All clones listed are from RPCI-22 (Segment I), except for 207c14, 380f02, 272b10, and 336k07 which belong to ES1-RW4.

determined by Southern versus PCR analysis was almost identical, but eliminated some false positives due to slight contamination detected by PCR.

Grouping genes together on a given BAC allowed a gene order to be established (Table 2). The gene order deduced for the *Ly49* locus of 129 mice is *Ly49*$q_1$, *e*, (*v*, $q_2$), $e/c_2$, *l/r*, *s*, *t*, $e/c_1$, *r*, *u*, *u/i*, $i_1$, *g*, *p/d*, ($i_2$, *p*), and *o*. It should be noted that *Ly49v* and $q_2$ were always found on the same BACs so their individual order with respect to the rest of the genes is not known. This is also true for *Ly49*$i_2$ and *p*. *Ly49*$i_1$ and $i_2$ exon 4 probes hybridized to similarly sized *Eco*RI restriction fragments, thus PCR cloning and sequencing were used to identify the correct gene in each BAC. In addition, five BACs contained only the *Ly49b* gene so no conclusion could be made as to its position with this cluster. The data further suggest that the *Ly49q* gene in B6 mice will be found closely associated with *Ly49e*. This is supported by the discovery of the *Ly49*$q^{B6}$ genomic coding sequence in a B6-derived BAC clone (GenBank acc. no. AC090563) that also possesses *Ly49*$e^{B6}$ and $f^{B6}$. Finally, three sets of hybridizing *Eco*RI fragments could not be assigned to

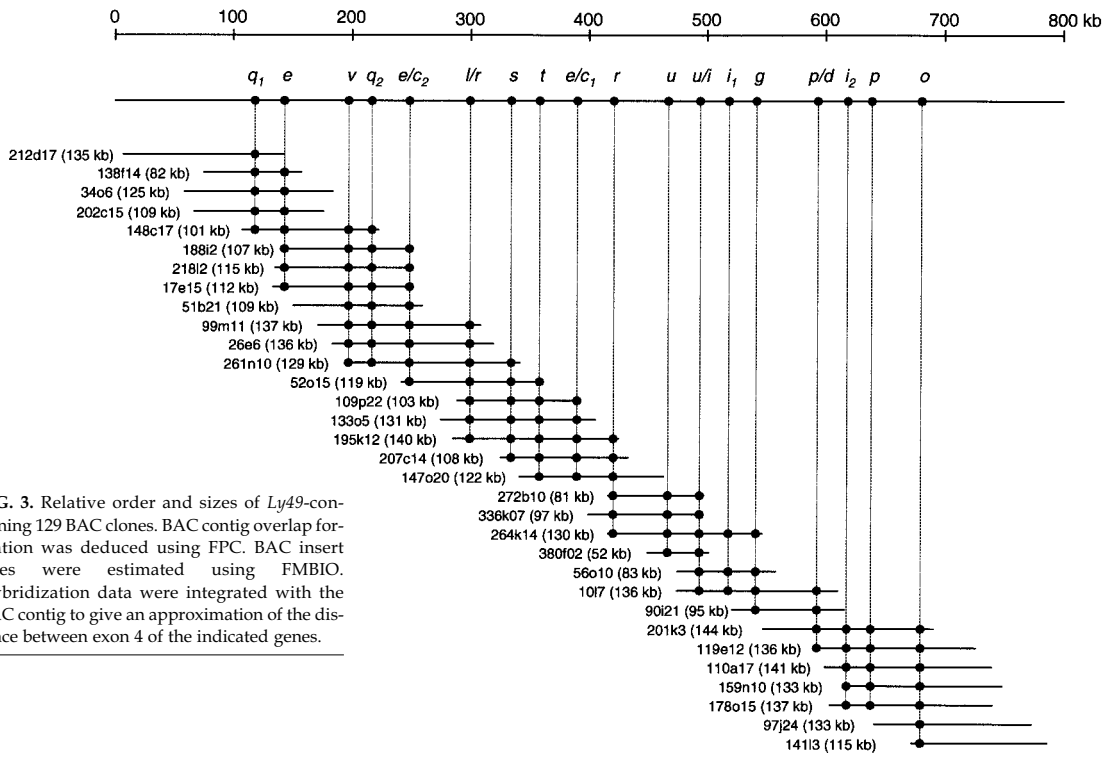



**FIG. 3.** Relative order and sizes of *Ly49*-containing 129 BAC clones. BAC contig overlap formation was deduced using FPC. BAC insert sizes were estimated using FMBIO. Hybridization data were integrated with the BAC contig to give an approximation of the distance between exon 4 of the indicated genes.

any genes. A 2.9-kb fragment was localized between *Ly49e* and $q_2$, a 2.3-kb fragment was found to reside between *Ly49t* and $e/c_1$, and a 2.35-kb fragment always colocalized with *Ly49o*. Whether these unidentified fragments represent new *Ly49* genes is unknown at this time.

### Fingerprint Analysis of 129 Murine BAC Clones

To generate a BAC contig, characterize the amount of sequence overlap between BACs, and determine BAC insert sizes, a restriction fragment fingerprint analysis was undertaken. Before transferring for Southern analysis, agarose gels containing separated *Eco*RI digests of each BAC clone along with a size standard were fluorescently stained. Gel images were collected and analyzed for restriction fragment mobility. Computer-assisted analysis using Fingerprinting Contigs (FPC) software allowed the determination of the probability of coincidence scores for each clone relative to all other clones and so identified overlapping segments. A single BAC contig containing 32 clones was assembled using a pairwise probability of coincidence cut-off score of $10^{-12}$ (Fig. 3). The five BACs containing the *Ly49b* gene did not form part of the main contig, a result consistent with its more distant location in the B6 genome (~ 750 kb) being greater than the average BAC length of this library. The automated BAC contig assembly using the total number of matching restriction fragments between overlapping clones is shown in Fig. 3, along with the estimated locations of the *Ly49* genes detected on each clone.

To determine the size of the BAC contig and thus the size of the 129 *Ly49* locus, we used FMBIO software to estimate the size of each restriction fragment of all clones in base pairs. This allowed an estimation of the size of each BAC insert (Fig. 3). The estimated average BAC insert size from genomic library RPCI-22 (Segment I) was 121 kb. Integration of the estimated size data with the BAC contig assembled by FPC indicates that the 129 *Ly49* locus spans ~ 600 kb.

### New Genes or Alleles?

Previous studies comparing coding region homology of *Ly49* genes between B6 and 129 mouse strains suggested some probable alleles [18]. The finding that 129 mice possess at least 19 genes makes it even more likely that alleles can be identified between the two strains, especially with the order of these genes now elucidated. Figure 4 is a schematic comparison of the B6 and 129 gene maps. The following characteristics were used to assign a gene as an allele, but were not necessarily exclusive: 1) coding region homology [18]; 2) intron 3 homology (Fig. 1); 3) length of predicted *Eco*RI fragment containing exon 4; and 4) gene order. Analysis of B6 *Ly49* genomic sequences found in GenBank predicted the following sizes for *Eco*RI fragments containing exon 4: *Ly49k*, 4186 bp; *Ly49h*, 4793 bp; *Ly49n*, 3907 bp; *Ly49d*, 4384 bp; *Ly49f*, 5629 bp; and

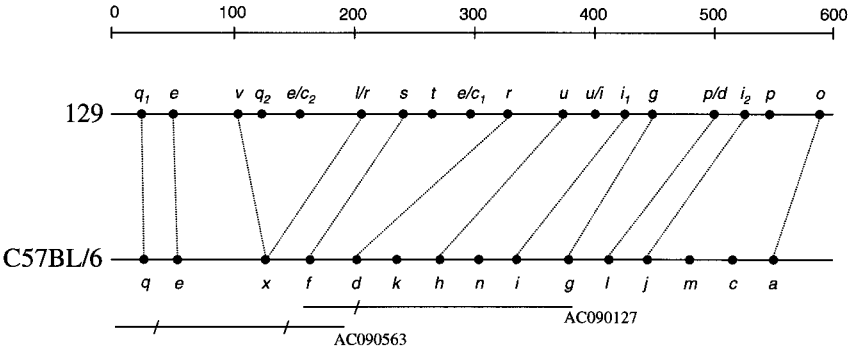



**FIG. 4.** Comparison of the 129 and B6 *Ly49* gene maps. The B6 gene map is reproduced from available GenBank BAC sequence data and a previous report [6]. Segment breakpoints within the B6-derived BACs are indicated by a solidus (/). The most similar genes between the two strains, which may represent alleles, are indicated by a dashed line. All gene markers were based on exon 4 position, with the exception of $Ly49g^{B6}$, which was based on exon 7. The maps are drawn so that the centromere and telomere are to the right and left, respectively. The scale bar is in kilobases.

*Ly49e*, 4073 bp (from GenBank acc. nos. AC090563 and AC090127). By these criteria, $Ly49e^{129}$, $g^{129}$, and $i_1{}^{129}$ are identified as the alleles of the same name in B6 mice, as are $Ly49f^{B6}$ and $Ly49s^{129}$, $Ly49d^{B6}$ and $Ly49r^{129}$, and $Ly49h^{B6}$ and $Ly49u^{129}$. Less convincingly, a case can be made for $Ly49l^{B6}$ and $Ly49p/d^{129}$, $Ly49j^{B6}$ and $Ly49i_2{}^{129}$, and $Ly49a^{B6}$ and $Ly49o^{129}$ as being alleles based on available coding sequence homology and gene location. The *Ly49x* gene indicated on the B6 gene map was also discovered in GenBank (acc. no. AC090563) and, based on sequence homology, is most similar to the *Ly49l/r* and *v* 129 genes in the 5′- and 3′-coding regions, respectively. Thus, $Ly49x^{B6}$ may be the result of a recombination between these two genes that deleted the $Ly49q_2$ and $e/c_2$ genes found in the 129 haplotype. There is complete sequence linking $Ly49e^{B6}$ and $x^{B6}$, allowing a placement of $Ly49x^{B6}$ within the B6 cluster. At this time, there is not a complete sequence available for the B6 *Ly49* cluster. However, two nearly complete BAC sequences from the telomeric end of the cluster account for ~ 400 kb (Fig. 4). Considering that there are at least five additional genes (*Ly49l*, *j*, *m*, *c*, and *a*), the B6 *Ly49* cluster should be > 500 kb in length.

$Ly49q^{B6}$ appears on the same BAC as $Ly49e^{B6}$ but in an unordered segment (GenBank acc. no. AC090563). It most likely resides centromeric of $Ly49e^{B6}$, as is the case for $Ly49q_1{}^{129}$, with which it shares near-sequence identity over intron 3 and the partial exon 3 and 4 coding region and so is the best candidate for an allele. Similarly, as $Ly49b^{B6}$ is the only known member of its own *Ly49* subfamily, the high homology of the 129 *Ly49b* exon 3 and 4 fragments to the B6 and BALB *Ly49b* cDNA (99.5% identity) suggests that it represents the 129 allele of the same gene. The most curious allele assignment is $Ly49u/i^{129}$ and $Ly49k^{B6}$. Both genes share a very high degree (> 98%) of sequence identity in intron 3 and possess an ~ 4.2-kb *Eco*RI fragment containing exon 4. If both B6 and 129 maps are correct, then this may be an instance of gene inversion as has been documented for the H-2 region in some mouse strains [22]. However, the transcriptional orientation of *Ly49u/i* is unknown at this time. Alternatively, *Ly49n* may be the allele of *Ly49u/i*, and the higher sequence conservation to *Ly49k* may be a coincidence. $Ly49q_2$, $e/c_2$, *l/r*, *t*, $e/c_1$, and *p* appear to have no identifiable B6 counterparts. Indeed, no alleles may exist for *Ly49t* and $e/c_1$ in B6 mice, as there is a contiguous genomic sequence available for the intergenic region between *Ly49f* and *d* (GenBank acc. no. AC090127). Likewise, *Ly49n*, *m*, and *c* from B6 mice have no apparent allele in 129 mice. It should be noted that the term allele might not be applicable to rapidly evolving gene families in which duplications, deletions, and recombinations are haplotype-specific.

The 129 mice possess at least six activating *Ly49* genes (*Ly49p*, *r*, *u*, *l/r*, *p/d*, and *u/i*), but complete transcripts have only been found for three (*Ly49p*, *r*, and *u*). B6 mice also possess six activating *Ly49* genes (*Ly49d*, *h*, *k*, *n*, *l*, and *m*), but

**TABLE 3:** Primers used to amplify *Ly49* gene fragments from BAC clones

| *Ly49* | Sense primer (exon 3) | Antisense primer (exon 4) |
|---|---|---|
| *e* | 5′-CCCTGGCAGCTCACTGTGAG | 5′-GAAGTTCCTCACCTGGACTG |
| *g* | 5′-GAAGCTCATTGTGATAGCTTG | 5′-TGTGTGCTGTGAGGAATCTG |
| *i* | 5′-GTGTTCAGTATCCTGGCAAC | 5′-CCTGTTCTCTTTTGATGTATTCCAGAAT |
| $i_2$ | 5′-GTGTTCAGTATCCTGGCAAC | 5′-CTGTTCTCTTTTAATGTATTCCAGAA |
| *o* | 5′-GTTCATTGTGATAGCTCTTGGC | 5′-GTTGAGGGATTCCAGAAGATC |
| *p* | 5′-TCATTGTGATAGCTCTCGGG | 5′-CTGTTCTGATCCCTGTTGAGG |
| *q* | 5′-GTGCTCTATCCCCTGGCATC | 5′-CCTGTACTACACTCTAGAGG |
| *r* | 5′-GTGTTCAGTTCCCTGGCAGTT | 5′-GTGTGCTGTAAGGAATCTAAAACAGTC |
| *s* | 5′-TGTTCAGGAAGCTGGCAGCT | 5′-GTGTCCTGTGAGGAATCTACG |
| *t* | 5′-GTATTCAGTCCCCTGGAAGCT | 5′-TGTGTGCTGTGAGGAATCTAAAAGAG |
| *u* | 5′-GAATCCTCTGTTCCCTTCGG | 5′-TGTGTTCTGTGTGGAATG |
| *v* | 5′-GTGTTCAGTCCCCTGGAAGT | 5′-CTTGGTTTTATTATACAATCTGTTCTT |

only two produce complete and functional transcripts (*Ly49d* and *h*). Thus, it is possible that *Ly49l/r*, *p/d*, and *u/i* may also represent pseudogenes. It is interesting to note that pseudogenes have been observed only for activating receptors in B6 mice. This may reflect a rapid evolution wherein such genes are created and discarded as a result of constantly changing selective pressures. The recent finding that the activating *Ly49h* is required for suppressing viral growth during MCMV infection is consistent with such a theory [9–11]. Viruses are constantly evolving, and the immune system must also change rapidly to defend against new pathogen variants. The purpose of the other activating *Ly49* genes is unknown, but they also may play a role in resistance to viral infection.

In general, nucleotide homology of alleles between B6 and 129 was ~ 98%. The exception to this was the $Ly49q_1^{129}$, $e^{129}$, and $b^{129}$ genes, which possessed > 99.5% nucleotide homology to $Ly49q^{B6}$, $e^{B6}$, and $b^{B6}$, respectively. This suggests the existence of a common ancestor to B6 and 129, and that a recombination may have taken place between *Ly49e* and *Ly49v* (and *Ly49a* and *Ly49b*) during the generation of either B6 or 129 mice. Alternatively, the exceptional homology of the *Ly49e*, *q*, and *b* alleles may reflect a unique function requiring sequence conservation for these genes relative to the other Ly49 receptors.

These findings suggest that the *Ly49* genomic locus evolves at a very high rate like the genes that code for their ligands (class I MHC). Furthermore, this study shows that the *Ly49* genes share another characteristic with the functionally analogous *KIR* genes in humans, namely gene plasticity. Analogous to the observed structure of two *KIR* haplotypes [3], the *Ly49* gene cluster in 129 versus B6 strains is composed of common genes interspersed by variable regions where some genes may be absent. Final conclusions as to *Ly49* gene content and haplotype variability await complete sequencing of these loci from multiple mouse strains.

## MATERIALS AND METHODS

***BACs.*** BAC clones containing genomic segments of the *Ly49* locus were obtained from two libraries: 1) Mouse ES (Release I) was produced from the RW4 embryonic cell line of the 129/SvJ substrain and was cloned into the *Hin*dIII site of pBeloBACII with an average insert size of 120 kb and 3× genomic coverage (Incyte Genomics, Palo Alto, CA); and 2) RPCI-22 (Segment I) was constructed from female 129/SvEvTac splenocyte genomic DNA cloned into the *Eco*RI site of pBACe3.6 displaying an average insert size of 155 kb and 5.3× genomic coverage (Roswell Park Cancer Institute, Buffalo, NY).

***Restriction fragment fingerprint mapping.*** Restriction fragment fingerprint analysis was carried out essentially as described [23]. Briefly, BAC DNA was purified using standard commercial plasmid/BAC isolation kits (Wizard Plus Minipreps, Promega, Madison, WI; Nucleobond BAC Maxi Kit, Clontech, Palo Alto, CA). A total of 250 ng of BAC DNA was digested with *Eco*RI for 4 hours at 37°C and then separated in 1% agarose for 17 hours at 70 V in recirculating 1× TAE maintained at 14°C using a CHEF DRII Chiller System (Bio-Rad Laboratories). Agarose gels were stained with SYBR-Green (Molecular Probes) and imaged using a Hitachi Fluorimager set at 300 dpi, 16 bits, reading sensitivity 100%, and a 505-nm filter. Insert sizes were estimated from the *Eco*RI fingerprints using FMBIO Analysis software package (Hitachi). Gel images were transferred to a DEC alpha (FCRDC Supercomputing Facility) using the raw data setting in Fetch and imported into IMAGE V3.9a. The normalized band positions from each clone were extracted and sent to Fingerprinting Contigs (FPC Version 3.2.1) [24]. Automated contig construction was performed essentially as described using FPC Eq. [25] and a probability of coincidence score of $10^{-12}$, tolerance of 8, diffbury at 0.1, and minbands at 3 [25,26]. Contig verification included reanalysis of ES1-RW4-derived BACs using *Bam*HI and *Hin*dIII fingerprints, side-by-side comparison of *Eco*RI fingerprints predicted to be overlapping, Southern blot hybridization, and cloning and sequencing of PCR amplification products.

***Southern hybridization, PCR, and DNA sequencing.*** ES1-RW4 clones were identified by PCR using *Ly49* exon 4-specific primers: forward, 5′-ATAACTGCAGCAACAYGCAAAG-3′, and reverse, 5′-GTGTGTSCTGTRAGGAATG-3′. Positive clones from RPCI-22 were identified by Southern hybridization of high-density spotted library filters using random prime-labeled (Gibco BRL, Rockville, MD) cDNAs of 129 *Ly49e*, *g*, *i*, *o*, *p*, *r*, *s*, *t*, *u*, and *v* excised with *Bst*XI from the pEF6/V5-His-TOPO constructs described [18]. Primers that are shown in Table 3 were designed to include intron 3 and partial sequence of exons 3 and 4 (average of 200 bp) to aid in identification of genes. Cycling parameters used for all PCR experiments in this study are as follows: 94°C for 30 seconds, 57°C for 30 seconds, and 72°C for 60 seconds, for a total of 35 cycles. Platinum PCR Supermix (Gibco BRL) was used along with 0.5 μg of forward and reverse primers and 0.1 μg BAC DNA for each amplification. PCR products were cloned into pCR2.1-TOPO using the TOPO TA cloning kit (Invitrogen, Carlsbad, CA) and sequenced with vector-specific primers by the Laboratory of Molecular Technology (NCI–Frederick, Frederick, MD).

After imaging, restriction fragment fingerprints were transferred to Duralon UV (Stratagene) using standard protocols for hybridization studies. Southern analysis was done using random prime-labeled probes made from *Eco*RI-excised inserts of the exon 3–4 clones described above.

***Phylogenetic analysis.*** Intron sequences were determined from clones of multiple PCR reactions and using several BACs as templates. Sequences were analyzed using the GCG SeqWeb Version 2 online sequence analysis utility (University of Wisconsin, Madison, WI). Specifically, sequences were aligned by PileUp and the dendrogram was constructed in GrowTree by employing the unweighted pair-group method using arithmetic averages. The Nexus output file generated was then used with TreeView V1.66 to create phylograms [27]. The bootstrap values in 1000 replications were calculated using the aligned sequences from PileUp and PHYLIP Version 3.5c. The GenBank accession codes of the 129 *Ly49* intron 3 and partial exon 3/exon 4 sequences are as follows: $Ly49q_1$ (AF425085), *Ly49e* (AF425086), $Ly49q_2$ (AF425087), $Ly49e/c_2$ (AF425088), *Ly49l/r* (AF425089), *Ly49s* (AF425090), $Ly49e/c_1$ (AF425091), *Ly49r* (AF425092), *Ly49u* (AF425093), *Ly49u/i* (AF425094), $Ly49i_1$ (AF425095), *Ly49p/d* (AF425096), $Ly49i_2$ (AF425097), *Ly49p* (AF425098), *Ly49b* (AF425099), *Ly49o* (AF432858), *Ly49v* (AF432859), *Ly49t* (AF432860), and *Ly49g* (AF432861).

### ACKNOWLEDGMENTS

*This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. N01-C0-12400.*

**RECEIVED FOR PUBLICATION OCTOBER 31;**
**ACCEPTED DECEMBER 17, 2001.**

### REFERENCES

1. McVicar, D. W., and Burshtyn, D. N. Intracellular signaling by the killer immunoglobulin-like receptors (KIR) and Ly49. *Science*'s STKE, http://stke.sciencemag.org/cgi/content/full/OC_sigtrans;2001-75/re1, 27 March 2001.
2. Uhrberg, M., *et al.* (1997). Human diversity in killer cell inhibitory receptor genes. *Immunity* **7:** 753–763.
3. Wilson, M. J., *et al.* (2000). Plasticity in the organization and sequences of human KIR/ILT gene families. *Proc. Natl. Acad. Sci. USA* **97:** 4778–4783.
4. Brown, M. G., *et al.* (1997). A 2-Mb YAC contig and physical map of the natural killer gene complex on mouse chromosome 6. *Genomics* **42:** 16–25.
5. McQueen, K. L., Freeman, J. D., Takei, F., and Mager, D. L. (1998). Localization of five new Ly49 genes, including three closely related to Ly49c. *Immunogenetics* **48:** 174–183.
6. Depatie, C., *et al.* (2000). Sequence-ready BAC contig, physical, and transcriptional map of a 2-Mb region overlapping the mouse chromosome 6 host-resistance locus cmv1. *Genomics* **66:** 161–174.
7. Delano, M. L., and Brownstein, D. G. (1995). Innate resistance to lethal mousepox is

genetically linked to the NK gene complex on chromosome 6 and correlates with early restriction of virus replication by cells with an NK phenotype. *J. Virol.* **69:** 5875–5877.

**8.** Scalzo, A. A., *et al.* (1992). The effect of the Cmv-1 resistance gene, which is linked to the natural killer cell gene complex, is mediated by natural killer cells. *J. Immunol.* **149:** 581–589.

**9.** Daniels, K. A., *et al.* (2001). Murine cytomegalovirus is regulated by a discrete subset of natural killer cells reactive with monoclonal antibody to Ly49H. *J. Exp. Med.* **194:** 29–44.

**10.** Lee, S. H., *et al.* (2001). Susceptibility to mouse cytomegalovirus is associated with deletion of an activating natural killer cell receptor of the C-type lectin superfamily. *Nat. Genet.* **28:** 42–45.

**11.** Brown, M. G., *et al.* (2001). Vital involvement of a natural killer cell activation receptor in resistance to viral infection. *Science* **292:** 934–937.

**12.** Smith, H. R., Karlhofer, F. M., and Yokoyama, W. M. (1994). Ly-49 multigene family expressed by IL-2-activated NK cells. *J. Immunol.* **153:** 1068–1079.

**13.** Brennan, J., Mager, D., Jefferies, W., and Takei, F. (1994). Expression of different members of the Ly-49 gene family defines distinct natural killer cell subsets and cell adhesion properties. *J. Exp. Med.* **180:** 2287–2295.

**14.** McQueen, K. L., Lohwasser, S., Takei, F., and Mager, D. L. (1999). Expression analysis of new Ly49 genes: most transcripts of Ly49j lack the transmembrane domain. *Immunogenetics* **49:** 685–691.

**15.** Makrigiannis, A. P., *et al.* (2000). Identification of the Ly49L protein: evidence for activating counterparts to inhibitory Ly49 proteins. *J. Leukoc. Biol.* **68:** 765–771.

**16.** Silver, E. T., Gong, D. E., Hazes, B., and Kane, K. P. (2001). Ly-49W, an activating receptor of nonobese diabetic mice with close homology to the inhibitory receptor Ly-49G, recognizes H-2Dk and H-2Dd. *J. Immunol.* **166:** 2333–2341.

**17.** Ortaldo, J. R., *et al.* (1999). Ly-49 receptor expression and functional analysis in multiple mouse strains. *J. Leukoc. Biol.* **66:** 512–520.

**18.** Makrigiannis, A. P., *et al.* (2001). Class I MHC-binding characteristics of the 129/J Ly49 repertoire. *J. Immunol.* **166:** 5034–5043.

**19.** Simpson, E. M., *et al.* (1997). Genetic variation among 129 substrains and its importance for targeted mutagenesis in mice. *Nat. Genet.* **16:** 19–27.

**20.** Threadgill, D. W., Yee, D., Matin, A., Nadeau, J. H., and Magnuson, T. (1997). Genealogy of the 129 inbred strains: 129/SvJ is a contaminated inbred strain. *Mamm. Genome* **8:** 390–393.

**21.** Kubo, S., *et al.* (1993). The gene encoding mouse lymphocyte antigen Ly-49: structural analysis and the 5′-flanking sequence. *Gene* **136:** 329–331.

**22.** Shin, H. S., Flaherty, L., Artzt, K., Bennett, D., and Ravetch, J. (1983). Inversion in the H-2 complex of t-haplotypes in mice. *Nature* **306:** 380–383.

**23.** Beck, T. W., *et al.* (2001). Comparative feline genomics: a BAC/PAC contig map of the major histocompatibility complex class II region. *Genomics* **71:** 282–295.

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

**25.** Marra, M., *et al.* (1999). A map for sequence analysis of the *Arabidopsis thaliana* genome. *Nat. Genet.* **22:** 265–270.

**26.** Sulston, J., *et al.* (1988). Software for genome mapping by fingerprinting techniques. *Comput. Appl. Biosci.* **4:** 125–132.

**27.** Page, R. D. (1996). TreeView: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* **12:** 357–358.