Tetrahymena Micronuclear Genome Mapping: A High-Resolution Meiotic Map of Chromosome 1L

Steve Wickert[1] and Eduardo Orias

Department of Molecular, Cellular, and Developmental Biology, University of California, Santa Barbara, California 93106

Manuscript received July 1, 1999
Accepted for publication November 11, 1999

ABSTRACT

The ciliate *Tetrahymena thermophila* is a useful model organism that combines diverse experimental advantages with powerful capabilities for genetic manipulation. The genetics of Tetrahymena are especially rich among eukaryotic cells, because it possesses two distinct but related nuclear genomes within one cytoplasm, contained separately in the micronucleus (MIC) and the macronucleus (MAC). In an effort to advance fulfillment of Tetrahymena's potential as a genetic system, we are mapping both genomes and investigating the correspondence between them. With the latter goal especially in mind, we report here a high-resolution meiotic linkage map of the left arm of chromosome 1, one of Tetrahymena's five chromosomes. The map consists of 40 markers, with an average spacing of 2.3 cM in the Haldane function and a total length of 88.6 cM. This study represents the first mapping of any large region of the Tetrahymena genome that has been done at this level of detail. Results of a parallel mapping effort in the macronucleus, and the correspondence between the two genomes, can be found in this issue as a companion to this article.

MODEL organisms have been extremely effective research tools in the biological sciences, and the pace of discovery continues to accelerate due to advances in the technology and scale of genome mapping and sequencing. One eukaryotic organism with a proven track record of important contributions and with particular promise is the ciliate *Tetrahymena thermophila*. Tetrahymena represents a very powerful genetic system, coupled with a host of other experimental advantages (reviewed in Orias 1998).

Tetrahymena possesses two distinct but related genomes, called the micronuclear (MIC) and macronuclear (MAC) genomes. The MIC genome is transcriptionally inactive and functions as the germline during sexual reproduction, following a classical Mendelian genetic model. The MAC genome is derived from the MIC genome during the process of sexual reproduction and functions as the somatic genome. It is highly expressed, in contrast to the MIC (reviewed in Bruns 1986; Karrer 1999).

During differentiation of the MAC, the five (pairs of) chromosomes derived from the germline are fragmented in a site-specific way, generating an estimated 200–300 acentromeric fragments that are randomly distributed at MAC division. These fragments are called autonomously replicating pieces (ARPs) or macronuclear chromosomes. The bulk of these pieces are amplified to the average level of 45 copies per cell.

Cells with a heterozygous MAC assort, after many fissions, into clonal descendant lines that have become pure for a single allele at each genetic locus. This process is called phenotypic assortment and represents a non-Mendelian genetic segregation model that is completely distinct from that of the MIC (and of most other eukaryotic cells). For a more thorough discussion of macronuclear genetics and mapping, see the companion to this article (Wickert *et al.* 2000, this issue).

We are mapping both genomes in anticipation of a genomic sequencing initiative for this organism. This three-article series describes recent progress toward that goal, as follows:

1. The first (this article) focuses on MIC mapping and presents a high-resolution genetic linkage map of chromosome 1L (of Tetrahymena's five micronuclear chromosomes). Although we have previously reported some preliminary micronuclear genetic maps (Lynch *et al.* 1995; Brickner *et al.* 1996; Longcor *et al.* 1996), this map represents the first high-resolution mapping that has been done in this organism on this scale and therefore sets standards and methods for the analysis of the rest of the Tetrahymena MIC genome.
2. The second article (Wickert *et al.* 2000, this issue) describes MAC genetics and mapping of the same region and focuses on the correspondence between the MIC and MAC genomes on chromosome 1L. In the MAC, phenotypic assortment gives rise to "coassortment groups" (CAGs; Longcor *et al.* 1996),

Corresponding author: Eduardo Orias, Department of Molecular, Cellular, and Developmental Biology, University of California, Santa Barbara, CA 93106. E-mail: orias@lifesci.lscf.ucsb.edu

[1]*Present address:* Protein Pathways, Inc., 1145 Gayley Ave., Suite 304, Los Angeles, CA 90024.

which are roughly the equivalent of MIC linkage groups, but have a completely different mechanism and kinetics of assortment. The second article presents the first systematic mapping of CAGs over a region of this size.

3. A manuscript submitted for publication (L. Wong, L. Klionsky, S. Wickert, V. Merriam, E. Orias and E. Hamilton, unpublished results) provides additional molecular genetic evidence that MAC pieces (ARPs) are the physical basis of CAGs.

MATERIALS AND METHODS

Strains, crosses, and genetic markers: Strains used, culture conditions, crosses, DNA preparation, and assignment of markers to chromosomes by use of monosomic strains have been previously described (Lynch *et al.* 1995; Brickner *et al.* 1996). The majority of genetic markers (Table 1) are randomly amplified polymorphic DNA (RAPD) polymorphisms between Tetrahymena inbred strains B and C3 that had been previously assigned to chromosome 1L. Preliminary linkage maps of a small subset of these markers have been previously published (Lynch *et al.* 1995; Brickner *et al.* 1996).

Meiotic segregant panels: The meiotic segregants used in this study were generated from three F_1 clones (SB983, SB990, and SB1804; Bleyman *et al.* 1992), obtained by crossing to one another cells of inbred strain B and C3. The meiotic segregants were obtained from these F_1's by genomic exclusion as described in detail in Lynch *et al.* (1995). Both the MIC and MAC of each of these meiotic segregants are homozygous for the germline genome of an independent meiotic product of the B/C3 MIC; thus each is a whole genome homozygote. A total of 197 meiotic segregants were used, comprising subsets of panels 1, 2, and 3 described in Table 2 of Lynch *et al.* (1995), selected as follows: from panel 1, 25 members derived from all three F_1 clones (note that the members of panel 1 were initially obtained for a different purpose; only those having an odd number of genetic crossovers between the *mat* and *PMR1* loci were kept); from panel 2, 22 members derived from SB983, 41 members derived from SB990 and 9 members derived from SB1804, for a total of 72 members altogether; finally, from panel 3, 100 additional members derived from SB990.

RAPD PCR: RAPD polymerase chain reaction (PCR) reactions were performed as originally described (Williams *et al.* 1990), except that two primers were used instead of one (see Lynch *et al.* 1995). In a few cases (RAPD polymorphisms PM8, EO1, and BR4), PCR was done at a final Mg^{2+} concentration of 5 mm, because the banding pattern was clearer than at the standard 2.5 mm concentration. Gel electrophoresis was essentially as described (Brickner *et al.* 1996). After electrophoresis, gels were stained in 2 μg/ml ethidium bromide for 10–15 min, followed by destaining for 15–30 min in deionized water.

Linkage analysis and map construction: Genetic data were analyzed and maps were constructed using MAPMAKER/EXP 3.0 (Lander *et al.* 1987), with "error detection" (Lincoln and Lander 1992) enabled (see below). Map distances were calculated using the Haldane function, $\omega = -\frac{1}{2}\ln(1 - 2\theta)$, where ω represents genetic distance in Morgans and θ is the recombination fraction ($0 \leq \theta \leq \frac{1}{2}$). A marker order was considered solid if it had a LOD (Log of the ODds by maximum likelihood) score >3.0 relative to the next best marker order (that is, the best order is 1000 times more likely to have generated the observed data than the next best order).

Maps were constructed by first defining a "framework" of markers having a well-defined order (LOD > 3). The selection of these framework markers was somewhat arbitrary, but an effort was made to choose the "best" set of markers that spanned the linkage group for maximum coverage at roughly uniform spacing, while including the largest possible number of markers that had a well defined order, as defined above. Because the choice of best framework markers was not unique, many alternative sets were examined to verify that the choice did not significantly affect mapping results, and the final selection was partly based on heuristic criteria (primarily the perceived quality of the RAPD banding patterns and number of informative data points).

To determine LOD scores for framework marker orders, it was impractical in terms of computation time to perform full multipoint maximum-likelihood calculations on all possible marker orders (*e.g.*, for 14 markers, there are 14!/2, or $>4 \times 10^{10}$, possible orders). Consequently, a sliding "window" at least six (and sometimes as many as eight) markers wide was used for comparison of the likelihoods for alternative orders of adjacent markers. Markers outside the window were considered fixed in terms of relative order (but not in distance), and maximum-likelihood values were calculated for maps with all possible orders of markers within the window. To be considered solid, a marker order had to have a LOD score of at least 3.0 relative to the next best order, where likelihoods were calculated using the full set of framework markers, not just those within the window. In principle, using a window size less than the complete framework (*i.e.*, not including all markers) could lead to errors in ordering, but in practice this was not thought to be a problem because the window was advanced one marker at a time, and markers far outside the window are only weakly linked to those inside. In addition, many trials with different starting configurations were done to make sure that results did not depend on initial marker choices.

Next, all remaining markers were placed into intervals relative to the framework map. For this step, a given marker was placed in turn into each possible interval in the framework (including positions off of each end, and far away, or unlinked). Maximum-likelihood scores were then calculated for each position (this procedure is automated by the MAPMAKER "try" command). If a marker placed at LOD > 3 around a framework marker (that is, into the dual interval composed of the *two* intervals flanking the framework marker), then it was considered to place at LOD 3 and indicated on the map in its maximum-likelihood position. Note that this placement criterion for nonframework markers is different from that used for the framework markers, and we are therefore always careful to differentiate clearly between the two classes of markers in our primary mapping figures. Next a window, as above, in which all possible marker orders were considered, was placed around the new marker, and full maximum-likelihood calculations were done for all possible orders. Square brackets, indicating uncertainty in relative order, were placed around all markers that varied in relative order up to a threshold of LOD 3.

Odds against order reversal for adjacent markers were calculated as differences in maximum-likelihood scores for maps using the indicated orders. For the framework, only framework markers were used, but for all pairs involving nonframework markers as well, full maps containing all markers were used.

Maximum-likelihood estimate of the total mapped region: One method we used to estimate the total mapped region for our data (the total genetic length of the region on chromosome 1L that was accessible to our random screen for genetic markers) was the maximum-likelihood method of Chakravarti *et al.* (1991). The method and its rationale are described

in detail in the reference above, and only details of its application to our data set are presented here. The relevant ln-likelihood expression for the total data set is

$$\sum_{i \neq j} \ln\left\{\int_0^{\theta_{\max}} \binom{n_{ij}}{r_{ij}} \theta^{r_{ij}} (1-\theta)^{n_{ij}-r_{ij}} f(\theta) d\theta\right\},$$

where the sum is over all unique marker pairs (i, j); n_{ij} is the number of informative meioses for marker pair (i, j); r_{ij} is the number of recombinants for marker pair (i, j); L is the total mapped region; θ = recombination fraction; $\theta_{\max}$ $\frac{1}{2}(1 - e^{-2L})$ (note the dependence on L); and $f(\theta)$ is the theoretical probability density function of θ for a marker pair.

In our case, because all loci considered are syntenic, $f(\theta)$ reduces to $f_s(\theta)$ as given in Chakravarti *et al.* (1991):

$$f_s(\theta) = \begin{cases} \dfrac{2L + \ln(1-2\theta)}{L^2 (1-2\theta)}, & 0 \leq \theta < \theta_{\max} < \frac{1}{2} \\ 0, & \theta_{\max} \leq \theta < \frac{1}{2} \end{cases}.$$

Numerical calculations were implemented in the C programming language and run under the Linux operating system.

Estimate of map coverage fraction by Monte Carlo method: We estimated the map coverage fraction by means of a Monte Carlo method using maps containing 20, 40, or 80 markers. For each choice of number of markers to use, 1×10^6 independent random maps were generated by placing all markers on a map of unit length according to a uniform random distribution. For each map, the coverage fraction, defined here as the fraction of the map contained between the two most distal markers at the ends of the map, was recorded. The Monte Carlo simulation was implemented in the C programming language and run under the Linux operating system.

Scoring errors: We used the incomplete penetrance error detection mechanism (Lincoln and Lander 1992) of MAPMAKER to flag potential scoring errors in the data set. The error detection scheme treats all experimentally measured genotypes as "phenotypes" of the true underlying genotype, which is considered to be partially penetrant, and likelihood calculations are performed under this assumption. The method computes a LOD_{error} score for individual data points, which is the log of the odds ratio of the probability that the entire data set would arise if the genotype for the data point is scored in error, divided by the probability that the data set would arise if the data point is scored correctly. Because accurate flagging of potential errors is highly dependent on correct marker order, we waited until most of the segregants had been scored before examining putative errors. Then, we retested all scores having a $LOD_{error} > 1.5$ as determined by MAPMAKER. When an individual genotyping was rechecked, the RAPD PCR reaction was repeated at least in duplicate, with double strain B and C3 controls, and these were run side by side on an agarose gel as described above. In some cases of apparent high probability errors, we confirmed that the scores were indeed errors and corrected the data. Some of the apparent errors, even those having high LOD_{error} scores, were found *not* to be errors. In a few cases where the results were still ambiguous, the corresponding data point was left unscored. The error detection feature also facilitated the detection—and exclusion from the data set—of four meiotic segregants that appeared to be heterozygous for at least a segment of chromosome 1L.

Error rate estimate: Even after screening as above, some scoring errors almost certainly remain. An estimate of the percentage of scoring errors in the data set was obtained by the "drop one" error analysis technique (see Buetow 1991; Weeks *et al.* 1995; see results for a discussion of the rationale). Briefly, one nonterminal marker was dropped from the map and the total map length was calculated in its absence. This was then repeated for each nonterminal marker in turn, and the average drop one map length was computed. The error rate was estimated as one-half the difference between the full map length and the average drop one length, in units of Morgans (see Weeks *et al.* 1995).

Generation of pseudorandom numbers: Pseudorandom number generation was required for estimating the variance associated with the scoring error rate (to introduce random errors into the data set) and for the Monte Carlo method of estimating the map coverage fraction, used for determination of the total mapped region (see above). In both cases, functions from the standard C library were used.

Unfortunately, the applicable American National Standards Institute standards do not specify how the functions are to be implemented (the algorithm is not specified, and the minimum required precision is unacceptably low), and many implementations are seriously flawed (see chapter 7 of Press *et al.* 1992). For introduction of random errors into MAPMAKER files, the function drand48() was used (in SunOS 4.1.3, under which MAPMAKER was run). On this system, drand48() is considered superior to rand() or random(). For the map coverage fraction Monte Carlo (and all other calculations not directly involving MAPMAKER), the function rand() of the GNU (http://www.gnu.ai.mit.edu) standard C library (glibc-2.0.7) was used under Linux (Red Hat 5.1). This function is considered superior in recent releases of this library, and drand48(), although still available, has been declared obsolete.

Expected frequency of meiotic segregants that are nonrecombinant types over the entire linkage group: We used the map and its intermarker distances to calculate the expected probability of observing an individual segregant (either B or C3) with no crossovers over the full map, $P = 1/2 \prod_i (1 - \theta_i)$, where θ_i is the recombination fraction for the ith intermarker interval, and the product is over all intervals in the map for this region. The recombination fraction, θ, was calculated for each interval from the centimorgan distance between markers by the Haldane function, $\theta = \frac{1}{2}(1 - e^{-2\omega})$, where ω is the distance in centimorgans for the maximum-likelihood map. Multiplying this probability by the total number of segregants (197) gives the mean expected number of individuals (B or C3) that we expect to observe showing no recombination over the whole chromosome arm.

RESULTS

One major goal of our current studies was to investigate the correspondence between the micronuclear and macronuclear genomes of Tetrahymena. We chose to concentrate on the left arm of MIC chromosome 1, which was formerly known as the left arm of chromosome 2 (V. Merriam, P. Bruns and D. Cassidy-Hanley, personal communication), because it had been mapped in more detail than any other region. We have previously reported preliminary micronuclear maps (Lynch *et al.* 1995; Brickner *et al.* 1996; Longcor *et al.* 1996) that include all or portions of the left arm of chromosome 1. However, these maps either contained too few markers or were at too low a resolution to allow effective comparison of micronuclear and macronuclear genomes. Therefore, we set out to map this region in

TABLE 1

RAPD loci mapped

Locus	Primers		Size (kb)	Locus	Primers		Size (kb)
1AS2[a]	A12	C5	1.0	1JP11[a]	A2	B10	1.2
1AT3/R	B20	C5	0.55/0.42	1JP33	A4	B1	0.9
1BB1R[a]	E10	E12	1.05	1JP34[a]	A4	B5	0.95
1BD6	F1	F20	0.4	1KF2	A2	C6	0.6
1BD8R[a]	F2	F11	0.4	1KN3	A2	A17	0.3
1BD11	F3	F16	0.33	1LS15	A5	E4	1.1
1BR4[a]	C5	C6	0.65	1MJ10aR	G16	G16	0.95
1CH1	A6	B12	1.3	1PM8	B17	B20	0.5
1EM10	D6	E11	2.0	1RT1	F9	H15	0.75
1EO1[a]	A5	A6	1.3	1SN7a	D2	D18	0.4
1EO3R	A19	B4	0.4	1SN9	D3	D11	0.18
1GM9	E18	F2	0.8	1SP9	E6	F8	1.1
1JB3	A1	A9	1.0	1SP11R[a]	E1	F5	1.4
1JB10R	A2	A9	1.1	1XS10[a]	A8	B8	0.2
1JO7	A2	D8	1.4	1XS19[a]	C2	D8	0.7
1JO9	A11	D11	0.2	1XS24	C4	D2	0.45
1JO13R	A14	D15	0.95	1XS35[a]	C6	D20	0.75
1JO16	A5	D20	0.8	1XS36	C7	D8	1.55
1JO39R[a]	A14	D10	0.75	1YD19	E15	E18	0.95

[a] MIC-limited RAPD.

greater detail, which primarily involved scoring a larger number of independent meiotic segregants.

Micronuclear map of chromosome 1L: The micronuclear map of the left arm of chromosome 1, generated by this work, is based on conventional meiotic recombination and consists largely of RAPD polymorphisms identified between inbred Tetrahymena strains B and C3 by random screening (see materials and methods). Table 1 lists these RAPD markers along with their associated primers and band sizes. In addition to the RAPD markers, they include *mat*, which is the mating type determination locus, and *PMR1*, which confers resistance to the drug paromomycin. To construct the map, whole-genome homozygotes made from independent meiotic products of B/C3 heterozygotes were scored for each locus, and a maximum-likelihood genetic map was constructed using MAPMAKER/EXP 3.0 (see materials and methods). The raw segregation data are available at the Tetrahymena genome web site (http://lifesci.ucsb.edu/~genome/Tetrahymena).

Figure 1 shows details of the map and marker placements and their associated statistical confidence levels. As described in detail in *Linkage analysis and map construction* (see also discussion), there are two classes of markers represented here, each having different statistical criteria for placement. The first is the set of 14 "framework" markers, for which the unique marker order shown has a very high degree of statistical confidence, which is defined globally for the entire framework. The second class includes all nonframework markers, whose placement criteria are defined in terms of the framework. After construction of the framework, many other markers mapped close to a framework marker, with high and nearly equal LOD scores for placement into the two intervals flanking the framework marker and low LOD scores for placement in all other intervals. This pattern, coupled with maximum-likelihood marker positions close to the associated framework marker for both flanking intervals, primarily represents uncertainty in marker *relative order* in a small local region, but not in the overall *location* of the marker on the map.

Therefore, in Figure 1, we show all nonframework markers that placed into a unique (framework-marker-containing) interval at LOD 3 or better on the map in their maximum-likelihood positions, with square brackets indicating uncertainty in marker relative order at LOD 3. Of the 40 markers, only three (GM9, JO13R, and JO16) could not be placed in this way into a unique interval at LOD $>$ 3 (for details of their placement, see the Figure 1 legend).

The left side of Figure 1 illustrates the statistical confidence of the framework in terms of the odds against reversing adjacent framework marker pairs. In a similar fashion, Figure 2 shows the associated confidence levels for the relative order of markers within a typical set of square brackets on the map. Note that in this case, markers XS36 and BD11 are so close that their relative order cannot be resolved at all (1:1 odds of reversal), while the odds against reversal of XS36 and JB3 are 6.6:1. As expected, markers located farther away from each other generally showed higher odds against rever-

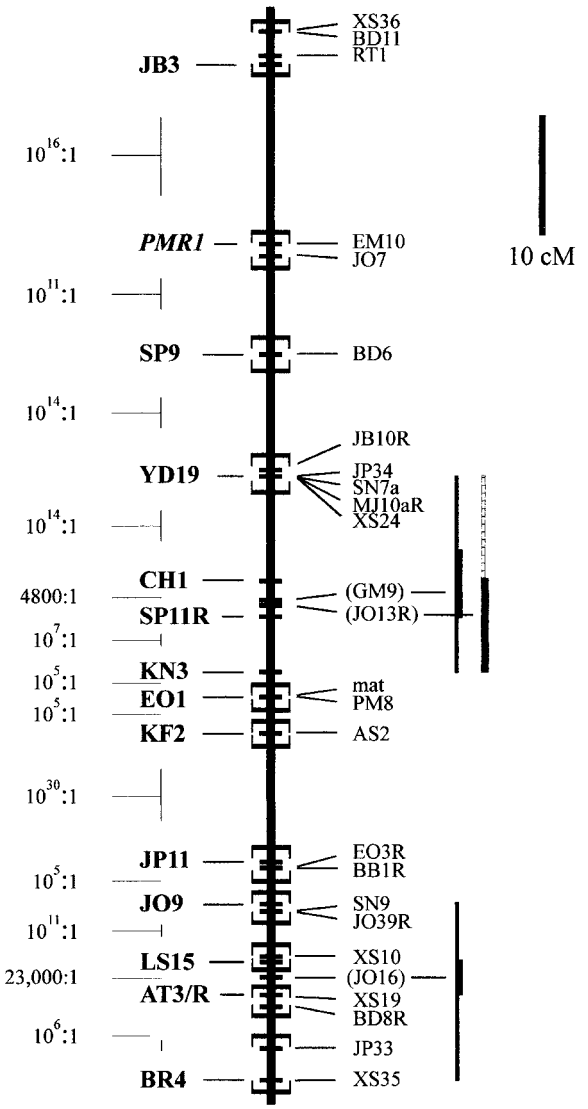

Figure 1.—Micronuclear map of chromosome 1L—details of marker placements and associated statistical confidence levels. The thick vertical line at center represents genetic distance in the micronucleus (10-cM scale bar shown at top right). Framework markers (see materials and methods) 1JB3–1BR4 (note: leading 1's have been removed from all marker names for clarity) are shown in boldface to the left of the thick line, with odds against reversal of adjacent framework markers shown to the left of each pair. Placement of all other markers relative to the framework is shown to the right of the thick line. All markers except 1GM9, 1JO13R, and 1JO16 (indicated by parentheses) placed into a unique interval spanning a framework marker (see materials and methods) at LOD 3.0 or better. Square brackets indicate marker relative orders that cannot be resolved at LOD 3. Placements for 1GM9, 1JO13R, and 1JO16 are indicated to the far right, with details as follows: The maximum-likelihood position for each marker is as shown on the map, with confidence intervals of placement shown to the right. Bars indicate odds of placement as follows: thick bars, <10:1 against; thin bars, 10–100:1 against; hatched bars, 100–1000:1 against. All other intervals have odds of placement ⩾1000:1 against in all cases.

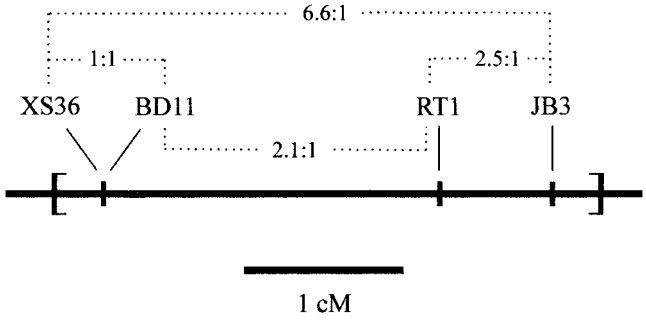

Figure 2.—Example of statistical confidence of marker order assignment within a cluster around a framework marker. A typical set of markers (1XS36, 1BD11, and 1RT1) that place at LOD > 3 around a framework marker (1JB3) is illustrated here (see Figure 1, top). The odds by maximum likelihood against reversing the relative order of various marker pairs in the map (indicated by dotted lines) are as shown. As in Figure 1, leading 1's have been removed from marker names for clarity. Square brackets have the same meaning as in Figure 1.

sal. Because the odds against reversal for each marker pair within the square bracket clusters are all well below 1000:1 (other data not shown), which is a commonly used threshold for mapping, we show only this one representative example.

The map contains 40 markers altogether, at an average spacing of 2.3 cM in the Haldane function. The largest interval is 15.1 cM (between JB3 and *PMR1*), and the map has a total length of 88.6 cM. The map reported here supercedes previously reported maps on this chromosome arm and represents the first mapping of any large region of the Tetrahymena genome that has been done at this level of detail.

Error rate: For dense maps, the rate of scoring errors in the data set is an important consideration, because these errors have a major impact on apparent map length and marker order (Lincoln and Lander 1992; Buetow 1991). We used the error detection feature of MAPMAKER (Lincoln and Lander 1992) to flag high-likelihood error candidates in our data set and to check and correct them if necessary as described in materials and methods. However, even after rechecking the data in this way, some scoring errors almost certainly remain in a data set of this size.

We estimated the percentage of scoring errors remaining using the drop one technique (see Weeks *et al.* 1995). The rationale is that most scoring errors introduce spurious double crossovers, thereby inflating the apparent map length (in a nonuniform way that is dependent on marker spacing). Therefore, if a marker is dropped from the analysis, there should be a decrease in apparent map length, due (mostly) to the removal of spurious double crossovers associated with that marker. In contrast, for reasonably dense maps, removal from the analysis of a marker that has no scoring errors is expected to have a negligible effect on map length on average. From the average decrease of the drop one

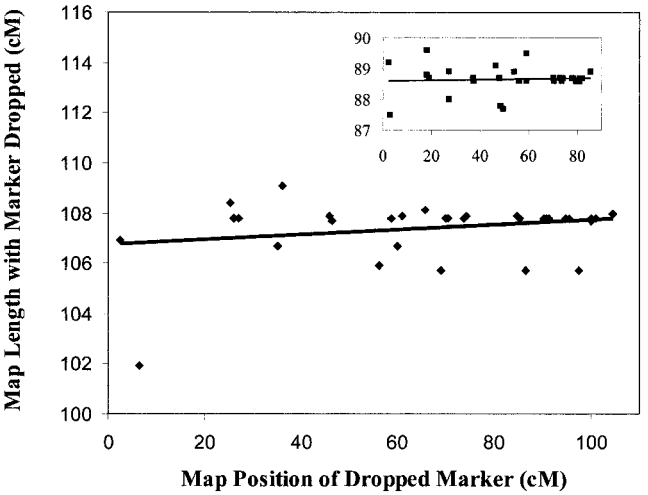

Figure 3.—"Drop one" micronuclear map lengths. For each nonterminal marker on the map (diamonds), the total map length with that marker removed is plotted as a function of marker position. MAPMAKER's "error detection" was disabled. Solid line, least-squares linear fit to the data points. Inset (squares), the same analysis but with "error detection" enabled; axis labels are the same.

lengths relative to the full map length with all markers included, we can estimate the overall error rate (see materials and methods).

The results of drop one analysis are shown in Figure 3. The error rate appears to be mostly uniform, with no obvious trend over the length of the map, as shown by the essentially flat trendline. In this analysis, MAPMAKER's error detection was disabled to allow errors to have their full effect on map length. With error detection enabled (Figure 3, inset), the effective error rate drops to zero. This is mostly a reflection of the effectiveness of the error detection scheme and is shown only for reference. Nevertheless, it suggests that any remaining errors in our data set have not greatly influenced map calculations.

The drop one lengths in Figure 3 suggest an error rate of ~0.2% (see materials and methods for details). To get a more accurate estimate, and to determine its precision, we introduced random errors into the data set at a rate of 0.20% and observed their effects. See the discussion for more information on the rationale behind this approach. Briefly, we assumed that the variance in error rate estimation associated with actual errors in the data set would be the same as the variance caused by errors artificially introduced in a truly random fashion. We generated 100 independent data sets in which random errors had been introduced into the actual data set and subjected them to a full drop one analysis of map lengths. The results are shown in Figure 4. For the 100 data sets, the calculated error rate, ε', was $\varepsilon' = (0.36 \pm 0.03)\%$ (mean ± SD). Because 0.20% random errors were artificially introduced, we conclude that the actual error rate, ε, for our data set is $\varepsilon = \varepsilon' - 0.20\% = (0.16 \pm 0.03)\%$.

One major consequence of errors in scoring is anomalous map expansion. We checked the consistency of our error rate estimate by examining map expansion with our data. From the work of Buetow (1991), a map of length L expands under the addition of scoring errors (rate ε) to a length L', given approximately by $L' \approx [(2)(100)\varepsilon + 1]L$. For our data set, L' (the full map length with MAPMAKER's error detection mechanism disabled) is 107.8 cM. From this, using the above relation, we obtain a value $L \approx L'/(200\varepsilon + 1) \approx 82 \pm 4$ cM for the estimated "zero error" map length. This value is in reasonable agreement with the 88.6-cM total map length reported by MAPMAKER with error detection enabled. As noted previously, MAPMAKER's error mechanism is quite effective in mitigating the effect of scoring errors on map length, so we expect the reported length with error detection enabled to be a close approximation to the zero error length, although some small expansion effects may not be completely nullified.

Total mapped region: A fundamental parameter in any genomic mapping project is the size in centimorgans of the entire region mapped, which in this case corresponds to the total genetic length of the region on chromosome 1L that was accessible to our random screen for genetic markers. We shall refer to this number as the total mapped "region," primarily to distinguish it from "map length," by which we mean the distance between the most distal markers at each end of the map. The distance between most distal markers is a crude estimate of the total mapped region, but underestimates it, because these distal markers, if distributed in a uniform random fashion over the mapped region, are expected to fall somewhat short of the region's endpoints. We used two different methods to estimate the total mapped region, and they are in approximate agreement.

The first is a maximum-likelihood method described by Chakravarti *et al.* (1991). It assumes that markers are distributed randomly throughout the mapped region and uses the theoretical distribution of distances between marker pairs to define an expression for the ln-likelihood of the total data set (see materials and methods). Because this expression depends on the length of the total mapped region, L, as a parameter, maximizing the likelihood with respect to L yields an estimate of the total mapped region. An additional advantage of this method is that it provides statistical confidence limits on the estimate. The experimental inputs to the calculation are a set of two numbers for each marker pair: (1) the number of recombinants and (2) the number of informative meioses for the pair. There are $n(n-1)/2$ unique marker pairs for n markers, which for our data set of 40 markers correspond to 780 marker pairs. See the discussion for a more thorough treatment of the assumptions underlying the model.

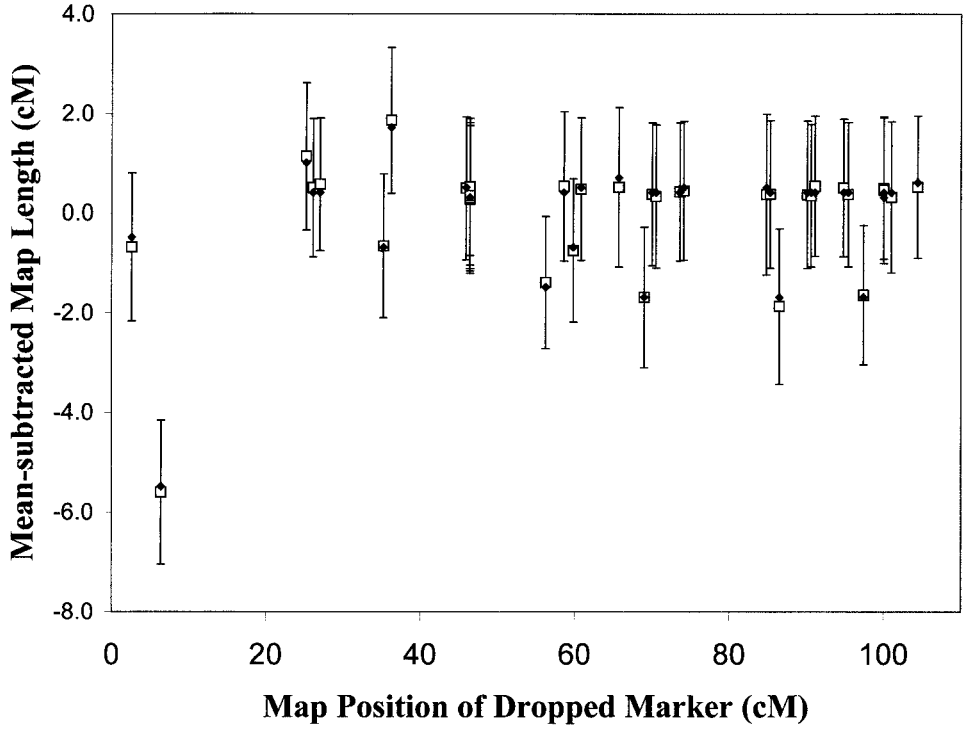

Figure 4.—Comparison of mean-subtracted drop one map lengths. Open squares, mean and standard deviation (error bars) of drop one lengths for 100 independent data sets with 0.20% random errors introduced into the actual data set. Solid diamonds, drop one lengths for the actual data set, mean subtracted and replotted from Figure 3 for comparison. For both sets of lengths, the mean of the corresponding set of drop one lengths was subtracted to nullify map expansion caused by introduction of the random errors to better show the correspondence between the two sets of drop one lengths.

Because the validity of the model is strongly contingent upon the conformance of the experimental intermarker distance distribution to the theoretically expected one, we first tested our data against this criterion. The theoretical cumulative intermarker distance distribution is given by $F(\omega) = (2L\omega - \omega^2)/L^2$ (see Chakravarti *et al.* 1991), where ω represents map distance in Morgans or centimorgans and L is the total mapped region in the same units. At this stage of the analysis, we did not know the length of L, so we used the best estimate that was then available, the total map length (88.6 cM), *i.e.*, the distance between the most distal markers. This approximation underestimates L (further analysis suggested that the underestimation is ~5–10%; see below), but it was close enough to check whether our data fit the model. Figure 5 plots the experimentally observed cumulative intermarker distance distribution for the data set (histogram) and compares it with the expected theoretical distribution (solid curve). We concluded that our data fit the theoretical expectations sufficiently well to proceed with the analysis.

The results of the maximum-likelihood calculation (see materials and methods) are shown in Figure 6. (In all cases, numerical values quoted below were derived from high-precision application of the calculations over relevant ranges, but this level of precision is not represented in the figure for the sake of clarity and computation time). The likelihood is maximized at an L value of 100.3 cM (arrow in Figure 6). In addition, (asymmetrical) confidence limits on the total mapped region can be read directly from the ln-likelihood plot. The two points where the ln-likelihood falls from its maximum by two units represent the approximate edges of a 95% confidence limit interval for the value of L. These limits are shown by the vertical lines in Figure 6 and occur at 95.3 and 105.7 cM. The inset shows the ln-likelihood over a wider range than the main figure so that its asymmetrical form may be more clearly seen.

The second method we used to estimate the length

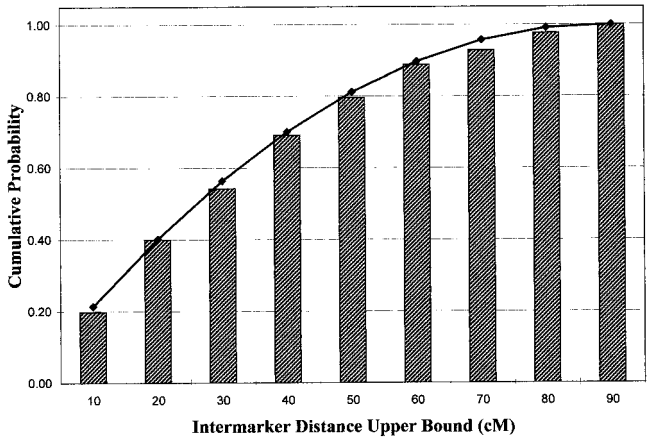

Figure 5.—Comparison of theoretical and observed cumulative probability distributions of intermarker distance between marker pairs. Hatched bars, histogram showing fraction of all intermarker pairs in the actual data set having an intermarker distance less than the indicated upper bound. For our data set, there were 780 such unique marker pairs (see text). Solid curve, theoretical cumulative probability distribution, as given in the text, using for L the map length, *i.e.*, the distance between the two most distal markers and evaluated at the values indicated by the histogram bin labels.

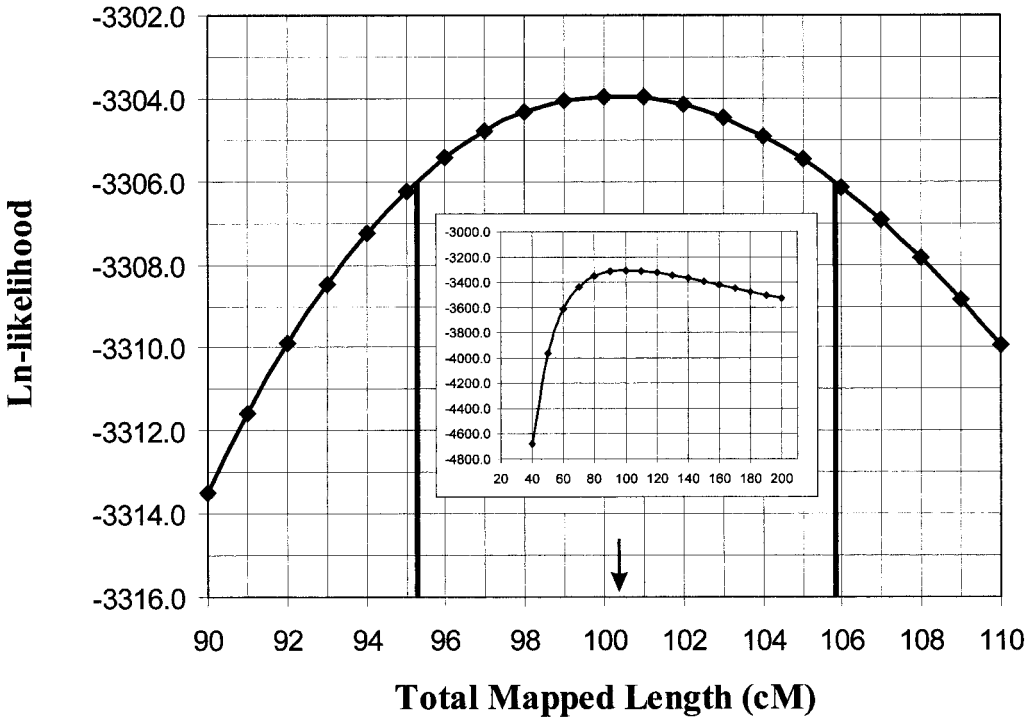

Figure 6.—ln-likelihood for the entire data set as a function of L, the total length of the region mapped. The arrow marks the length at which the ln-likelihood is maximum. Vertical lines, points at which the ln-likelihood falls by two units from its maximum value. Inset, same plot over a wider range of L values; axis labels are the same.

of the total mapped region was an evaluation of its degree of marker coverage. This method has the advantage of requiring very few assumptions—only that the markers be distributed randomly. If markers are distributed in a uniform random fashion over a region of length L, then the map length, as represented by the distance between the two most distal markers at the ends of the map, is expected to be somewhat less than L, because the most distal markers will not fall exactly at the ends of the region. The total of the "uncovered" areas at the ends of L represents the difference between what we have called the "map length" and the "total mapped region," and we can quantitatively determine its value in a statistical sense. We used this to estimate the total mapped region from the map length.

To accomplish this, we constructed Monte Carlo-generated maps of markers distributed randomly over a region of unit length. For each map, we recorded the fractional marker coverage, defined in this case as the fraction of the total mapped region that is contained between the most distal markers. The resulting probability density for maps containing 20, 40, or 80 markers is plotted in Figure 7. As expected, the probability density peaks higher and more narrowly, and at a higher coverage fraction, when more markers are used.

To calculate the total mapped region for our map, we focused on the coverage fraction probability density for the case of 40 markers, as in our data set. The mean (expectation value) coverage fraction for this distribution is given by $(n-1)/(n+1)$, where n is the number of markers (see David 1970). For $n = 40$, the mean coverage fraction is $39/41 \approx 0.9512$ (direct numerical integration using the Monte Carlo-generated distribution yielded a mean that agreed with this value to within 0.02%). For our map length (distance between most distal markers) of 88.6 cM, this corresponds to a total mapped region of $(88.6 \text{ cM})/0.9512 \approx 93.1$ cM.

The coverage fraction distribution also provides a confidence interval on the total mapped region. The cumulative probability for this case (the integral of the solid curve in Figure 7) is plotted in Figure 8. For a 95% confidence interval, we used the region of the

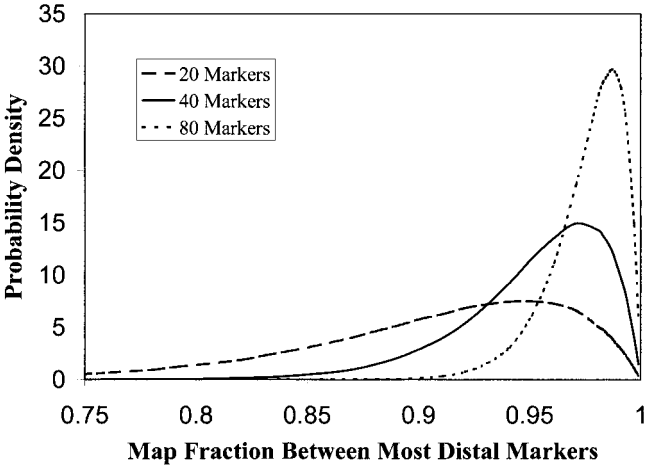

Figure 7.—Probability density of map coverage fraction for markers placed according to a uniform random distribution. The coverage fraction is here defined as the fraction of the map that is contained between the two most distal markers at the ends of the map. Maps contain 20, 40, or 80 markers as indicated.

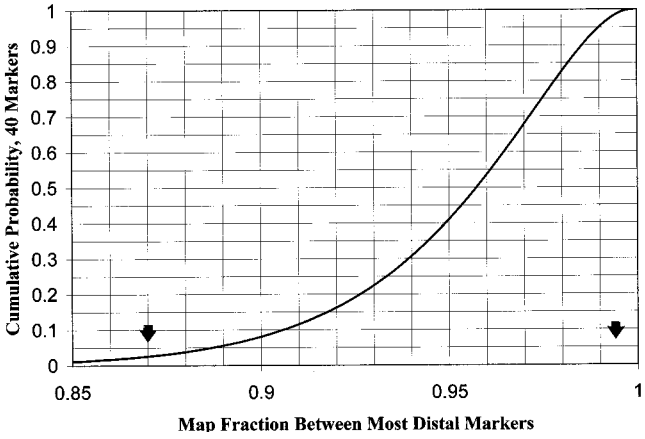

Figure 8.—Cumulative probability distribution of map coverage fraction for the case of 40 markers (the integral of the solid curve in Figure 7). Arrows indicate map fractions corresponding to probabilities of 2.5 and 97.5%; *i.e.*, map coverage fraction is between these values with 95% probability.

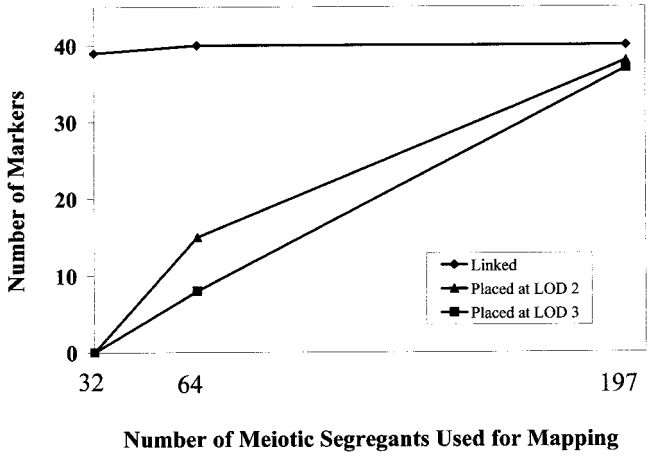

Figure 9.—Micronuclear mapping precision on the left arm of chromosome 1 as a function of the number of meiotic segregants. Of the 40 markers in this region, the number that can be (◆) linked at LOD 3, (■) placed at LOD 3, and (▲) placed at LOD 2 is shown for each case.

cumulative probability distribution between 2.5 and 97.5%. As can be seen from the figure, the coverage fraction under these conditions is between 0.870 and 0.993 with 95% confidence. This translates to a total mapped region between 89.2 and 101.8 cM, which is in reasonable agreement with that from the other method (see discussion).

Estimate of kilobase pairs/centimorgans for chromosome 1L: If we assume that the map in Figure 1 covers most of the left arm of chromosome 1 (and other assumptions; see discussion), we can make a crude estimate of the kilobase per centimorgan value for this region. For the total map length in centimorgans, we used the average of the two estimates above for the total mapped region: 96.7 cM. Tetrahymena has a haploid genome physical length of $\sim 2.2 \times 10^8$ bp (reviewed in Prescott 1994), or 220,000 kb. There are five MIC chromosomes, with chromosomes 1 and 2 being the largest: two large metacentrics that are roughly indistinguishable in size (Bruns and Brussard 1981). To estimate the fraction of the genome contained in 1L, we used the photometric work of Seyfert (1979). We averaged the genomic DNA fractions for the two largest Tetrahymena chromosomes in Table 5 of Seyfert (they had not yet been assigned numeric designations) to obtain an estimate of 15.1% for the genome fraction of chromosome 1 (note that because of the way they are calculated in Table 5 of Seyfert, the chromosome fractions do not sum to unity; we therefore renormalized them). Because chromosome 1L is metacentric, we estimated its physical length as (1/2)(220,000 kb)(15.1%) ≈ 17,000 kb. This leads to a kilobase per centimorgan value for 1L of (17,000 kb)/(96.7 cM) ≈ 200 kb/cM. This value is higher than our previous estimates (reviewed in Orias 1998), but is based on more precise mapping.

Comparison of mapping with 32, 64, and 197 meiotic segregants: This detailed study of one chromosome arm provided an opportunity to determine how mapping resolution depends on the number of meiotic segregants used in our analysis to guide our efforts in mapping the rest of the genome. We therefore examined mapping results using panel sizes of 32, 64, and 197 independent meiotic segregants at each locus. The results are presented in Figure 9. As can be seen, 32 individuals were sufficient to place all of the markers into one linkage group at LOD > 3, but were generally insufficient for establishing relative orders and distances. Different unique subsets of 32 individuals from the full data set were tried, and no significant differences in results were noted (data not shown). Using 64 segregants, correct maps (as seen with 197 segregants) were reproduced for the most part, but with lower LOD scores (~ 2).

B allele segregation bias in panels of meiotic segregants: While checking the data for the expected 1:1 allele segregation, we observed a B allele bias in the meiotic segregant panels (see Table 2). The overall B:C3 allele ratio was 1.29:1 (probability of $\chi^2 \approx 10^{-25}$ against 1:1 segregation). A bias in the segregation of *mat* and *PMR1* was already evident when the first two meiotic segregant panels were first obtained (Bleyman *et al.* 1992). Curiously, the bias occurs preferentially in the JB3 half of the map (Table 2). The bias is even more striking in the numbers of B and C3 segregants that show no recombination at all over the entire length of the map. Based on the calculated map intermarker distances, we expected to observe 43 such nonrecombinants of each type (see materials and methods). However, the actual numbers of such nonrecombinants seen were 50 of the B type and 27 of the C3 type, a ratio of 1.85:1 (probability of $\chi^2 \approx 0.009$ against a 1:1 segregation pat-

TABLE 2

Allele ratios in meiotic segregation

Locus	No. tested	Total B	B fraction	Locus	No. tested	Total B	B fraction	Locus	No. tested	Total B	B fraction
1XS36	189	112	0.59*	1XS24	179	102	0.57	1BB1R	183	94	0.51
1BD11	183	107	0.58*	1CH1	189	115	0.61**	1JO9	188	100	0.53
1RT1	190	116	0.61**	1GM9	90	59	0.66**	1SN9	155	78	0.50
1JB3	192	121	0.63**	1JO13R	179	106	0.59*	1JO39R	188	94	0.50
1EM10	94	55	0.59	1SP11R	184	106	0.58*	1XS10	175	99	0.57
PMR1	197	112	0.57	1KN3	195	112	0.57*	1LS15	188	101	0.54
1JO7	89	56	0.63*	mat	196	107	0.55	1JO16	138	71	0.51
1BD6	95	59	0.62*	1PM8	193	106	0.55	1AT3/R	195	103	0.53
1SP9	180	104	0.58*	1EO1	194	107	0.55	1XS19	117	67	0.57
1JB10R	187	108	0.58*	1AS2	192	105	0.55	1BD8R	182	91	0.50
1JP34	180	106	0.59*	1KF2	195	104	0.53	1JP33	190	101	0.53
1YD19	192	109	0.57	1JP11	194	101	0.52	1BR4	193	105	0.54
1SN7a	179	101	0.56	1EO3R	188	97	0.52	1XS35	186	101	0.54
1MJ10aR	185	105	0.57								
Average over all loci											0.56**
* and ** loci excluded											0.54**

The loci are listed in map order (Figure 1).
*Difference from 1:1 ratio is statistically significant ($P\ \chi^2 < 0.05$).
**Difference from 1:1 ratio is statistically significant ($P\ \chi^2 < 0.01$).

tern). The observed allele segregation bias was complex and nonuniform, and we have found no single satisfactory explanation for it and the observations above, despite considerable effort. Its possible distorting effects on map calculations are expected to be correspondingly complex and nonuniform.

DISCUSSION

We have mapped a significant portion of the Tetrahymena MIC genome (chromosome 1L) at a higher resolution than has been available previously. This study represents the first mapping of any large region of the Tetrahymena genome that has been done at this level of detail. The resulting map, consisting of 40 markers with an average spacing of 2.3 cM in the Haldane function and a total genetic length (between the most distal markers) of 88.6 cM, has already proven invaluable for investigating the relationship between Tetrahymena's MIC and MAC genomes (Wickert *et al.* 2000) and should be useful in many other contexts as well.

Map length: The preceding map length (distance between most distal markers) is likely to be a slight underestimate because we know that our data set contains some apparent double crossovers in short adjacent intervals that are not scoring errors. Such apparent double crossovers are essentially ignored by MAPMAKER in its calculations of map distance with error detection on. If they were to be counted as true double crossovers, we estimate that they would add roughly another 5 cM to the map (data not shown). This expansion would be reduced if any of these represent single recombinational events, *i.e.*, gene conversions.

Statistical support for the micronuclear map: Because of the density of markers, it is important to interpret carefully the information on statistical confidence levels of marker placements and relative orders in the micronuclear map. The framework markers have a high confidence of placement. Relative to the framework, most other markers have placements localized to small regions of the map, but are often so near other markers that very little recombination was observed, and their relative order could not be resolved. This is unlikely to be a serious limitation to the usefulness of the map for classical linkage studies involving other experimentally important loci, or for map-based cloning, etc. In selecting markers to work with for such studies, it should be understood that the choice of framework markers was somewhat arbitrary (see materials and methods), so that the apparent dichotomy between framework markers and others is mostly artificial. One is therefore free to choose the most convenient markers among many alternative sets without appreciable loss of statistical confidence.

Marker clustering: It is not clear whether the fact that most other markers mapped close to a framework marker is due to definite clustering at these locations or is just a consequence of high marker density coupled with a framework chosen to span the map with even spacing. The end of the map near JB3 is suggestive of the former, while the end near BR4 perhaps suggests the latter. A statistical analysis of marker clustering was not conclusive on this point (data not shown).

There are several sets of markers in the map that appear to be coincident with one another, *i.e.*, show no recombination. In most cases, we do not know whether they are actually coincident or just located very close together. In one instance, however, we do have more information. We have shown physically that the coincident pair *PMR1* and EM10 is located on the same MAC ARP, the rDNA (see Wickert *et al.* 2000). The rDNA is one of the smallest ARPs, an ~21-kb palindrome derived from a 10.3-kb segment of MIC DNA. Assuming an approximate ratio of 200 kb/cM (see results), this would represent only at most ~0.05 cM on the map, so observing any recombinants with the number of segregants we have used would be highly unlikely.

The highest concentration of coincident markers is near the framework marker YD19, where five markers (YD19, JP34, SN7a, MJ10aR, and XS24) show no recombination at all. XS24 and SN7a have one primer in common (D2), and one of the YD19 primers (E15) also has an identical three-prime end (see Table 1). None of the others have any obvious similarities in primers, and all of the RAPD polymorphic bands of the YD19 group are of different sizes. Despite the fact that all of these markers are coincident on the map, XS24 is in a different macronuclear coassortment group from the others (see the companion article on macronuclear genetics; Wickert *et al.* 2000). In addition, JP34 is MIC-limited, whereas the others are MAC-destined (see the companion article).

The fact that all of these markers were identified independently in random screens raises the possibility that there may be unusual sequence structure in this region. A deletion, though it also could lead to the same result, seems unlikely to be the sole explanation, because the region includes RAPDs in which the polymorphic band is templated both by B and C3 DNA. The cluster of RAPDs having D2 (or E15) in common may be indicative of a cluster of repeated DNA sequences with similar but not identical copies. Alternatively, a local inversion between inbred strains B and C3 may be suppressing recombination. More work is required to determine the exact cause of the clustering at this site.

Error rate estimate: The drop one method is a straightforward and useful procedure for estimating the overall rate of scoring errors in an experimental data set, without requiring additional genotypings or type checks. Unfortunately, the basic drop one procedure does not provide a means of determining a precision for this estimate, so the number by itself, without any confidence limits, is of limited informativeness. One simple method of estimating the variance would be to just use the variance of the set of drop one lengths for the map to calculate the associated variance of the error rate estimate. However, this would be incorrect and results in gross overestimation. The reason is that the drop one lengths for the various nonterminal markers are not independent, but depend on the details of the map, particularly the local marker density.

An entirely satisfactory method for gauging the appropriate variance to use for the error rate estimate is difficult to find, but it is important to do so. Because we do not know the locations of actual errors, we cannot reconstruct a known "error free" data set to work with. We therefore adopted the strategy of adding random known errors to the actual data set and examining the effect of these artificially introduced errors on map calculations. If the introduced errors are truly random, this should allow an accurate determination of the variance of the error rate estimate. We have to assume, for this approach to be valid, that the magnitude of the effect of individual errors does not depend strongly on the overall error rate or at least does not change substantially over the range of ~0.2% to 0.4% error rate. If this assumption is correct, then the actual variance of the drop one error rate estimate for the real data set is expected to be the same as the variance in the estimates of this number that result from the introduction of truly random known errors. We have assumed that actual errors in the data set are random, but we of course have no way to verify this directly. Under the preceding assumptions, we arrived at an error rate estimate of $(0.16 \pm 0.03)\%$ for our data set, which seems quite reasonable.

Total mapped region: The maximum-likelihood method of Chakravarti *et al.* (1991) for estimating total genetic length is useful for maps that contain a sufficiently large number of markers. However, the method was originally proposed for calculating total genome length, using data simultaneously from all genomic chromosomes. The method can be adapted for strictly syntenic markers, but may suffer somewhat in accuracy.

The model is based on the assumption that all intermarker distances for locus pairs are statistically independent, which is clearly not the case for syntenic markers, as Chakravarti *et al.* readily acknowledge. However, they show that violation of this assumption should not lead to large errors if markers on several different chromosomes are considered simultaneously. The assumptions of the model are further strained by considering only syntenic markers, as we have done.

Estimation of the total mapped region by the expected map coverage fraction does not suffer from this problem. It has the advantage of requiring fewer assumptions, but this is also its biggest disadvantage, because it uses less information from the data set. In particular, it relies exclusively on the most distal markers at the ends of the map, and the distance between them, for making a prediction of the length of the total mapped region. In contrast, the method of Chakravarti *et al.* uses the entire distribution of intermarker distances for marker pairs, so it should be less sensitive to changes involving only the terminal markers.

Estimating kilobase per centimorgan on chromosome 1L: Our estimate of 200 kb/cM for chromosome 1L relies on some assumptions: first, that our map covers most of the chromosome arm. We cannot demonstrate this with certainty, but we have reasons to believe this may be the case. At the present stage of the genome project, 95% of the roughly 400 genomic markers identified to date, and all of those in 1L, fall into linkage groups (Orias 1998). Nevertheless, there is a formal possibility that some significant portion of chromosome 1L is inaccessible to our random search for polymorphic markers. We have no reason to suspect this and are not aware of its having been reported to be the case for RAPD polymorphisms in other organisms.

Second, we have assumed that the frequency of meiotic recombination is approximately uniform over the chromosome arm, with no significant "hotspots" or localized suppression of recombination. This is almost certainly not exactly correct, as suggested by the observed marker clustering of the YD19 group. Nevertheless, it should be a reasonable approximation at the current marker density of our maps, as suggested by the agreement between observed and predicted intermarker distance distributions (Figure 5) under the assumption of a uniform random distribution of marker locations. However, we noted a second possible anomaly in the form of an apparent discrepancy between the physical sizes and map lengths of the PM8 and KN3 ARPs [compare physical lengths of the ARPs from Longcor *et al.* (1996) to the map lengths of the same ARPs as seen in Wickert *et al.* (1999)].

Our previously reported estimates of micronuclear kilobase per centimorgan for Tetrahymena (see Orias 1998) were probably somewhat low. As we mapped more loci with more meiotic segregants, the estimated total length in centimorgans of our maps decreased as scoring errors became easier to detect and correct. Most errors and inaccuracies cause apparent map expansion, because a single scoring error often adds two spurious crossover events. Because the assumptions discussed above are untested, we caution that our current estimate of 200 kb/cM should still be considered subject to revision as more data become available.

Mapping resolution and uses of the map: Because a much higher micronuclear mapping resolution was required for this work than we had previously attained over any region of this size (a whole chromosome arm), we also had the opportunity to gauge mapping quality as a function of the number of meiotic segregants used. We concluded that 64 segregants represent a good tradeoff between labor required and mapping resolution for rapid construction of initial maps in most areas of the genome. Regions that need to be studied with greater resolution should be mapped subsequently with more segregants as required. The full set of 197 meiotic segregants, which is the largest number we have used, provides more resolution and statistical confidence than are often reported in genetic maps at this scale, but this was necessary in this study to allow investigation of the relationship between micronuclear and macronuclear maps (Wickert *et al.* 2000). In general, mapping quality degraded gracefully and predictably as the number of segregants was reduced.

The micronuclear map reported here is the first medium- to large-scale genomic mapping that has been done in Tetrahymena at this resolution and is probably near the limit of marker density that is currently useful for classical genetic maps. Beyond this limit, diminishing returns in map resolution accompany large increases in panel size and associated labor required. In addition, the accuracy of distances for a classical genetic map (compared to the actual physical map) is limited in reality, because real recombination frequency is nonuniform and does not exactly match any simple model. The map we have constructed seems to be sufficiently accurate to map coassortment groups (Wickert *et al.* 1999) and thus macronuclear pieces (L. Wong, L. Klionsky, S. Wickert, V. Merriam, E. Orias and E. Hamilton, unpublished results) to the micronuclear map and eventually to the genome sequence. It is this macronuclear mapping that in turn may well be the most useful for cloning mutant genes of interest.

We thank Laura Wong for maintenance of the PCR supplies, and Eileen Hamilton, John Cotton, Ruth Finkelstein, and Tim Lynch for valuable comments on the manuscript. The National Institutes of Health supported this work through grant RR 09231. The work reported here is being submitted by S.W. in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Molecular, Cellular, and Developmental Biology at the University of California, Santa Barbara.

LITERATURE CITED

Bleyman, L. K., M. P. Baum, P. J. Bruns and E. Orias, 1992 Mapping the mating type locus of Tetrahymena thermophila: meiotic linkage of mat to the ribosomal RNA gene. Dev. Genet. **13:** 34–40.

Brickner, J. H., T. J. Lynch, D. Zeilinger and E. Orias, 1996 Identification, mapping and linkage analysis of randomly amplified DNA polymorphisms in *Tetrahymena thermophila*. Genetics **143:** 811–821.

Bruns, P. J., 1986 Genetic organization of *Tetrahymena*, pp. 27–44 in *Molecular Biology of the Ciliated Protozoa*, edited by J. G. Gall. Academic Press, New York.

Bruns, P. J., and T. E. Brussard, 1981 Nullisomic *Tetrahymena*: eliminating germinal chromosomes. Science **213:** 549–551.

Buetow, K. H., 1991 Influence of aberrant observations on high-resolution linkage analysis outcomes. Am. J. Hum. Genet. **49:** 985–994.

Chakravarti, A., L. K. Lasher and J. E. Reefer, 1991 A maximum likelihood method for estimating genome length using genetic linkage data. Genetics **128:** 175–182.

David, H. A., 1970 *Order Statistics*, Chap. 6, pp. 93–136. John Wiley & Sons, New York.

Karrer, K. M., 1999 *Tetrahymena* genetics: two nuclei are better than one, chapter 3 in *Tetrahymena thermophila* (*Methods in Cell Biology*, Vol. 62), edited by D. J. Asai and J. D. Forney. Academic Press, New York.

Lander, E., P. Green, J. Abrahmson, A. Barlow, M. Daley *et al.*, 1987 MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics **1:** 174–181.

Lincoln, S., and E. Lander, 1992 Systematic detection of errors in genetic linkage data. Genomics **14:** 604–610.

Longcor, M. A., S. A. Wickert, M.-F. Chau and E. Orias, 1996 Coassortment of genetic loci during macronuclear division in *Tetrahymena thermophila*. Eur. J. Protistol. Suppl. **1:** 85–89.

Lynch, T. J., J. Brickner, K. J. Nakano and E. Orias, 1995 Genetic map of randomly amplified DNA polymorphisms closely linked to the mating type locus of *Tetrahymena thermophila*. Genetics **141:** 1315–1325.

Orias, E., 1998 Mapping the germ-line and somatic genomes of a ciliated protozoan, *Tetrahymena thermophila*. Genome Res. **8:** 91–99.

Prescott, D. M., 1994 The DNA of ciliated protozoa. Microbiol. Rev. **58:** 233–267.

Press, W. H., S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, 1992 Random numbers, pp. 274–277 in *Numerical Recipes in C: The Art of Scientific Computing*, Ed. 2. Cambridge University Press, Cambridge, United Kingdom.

Seyfert, H.-M., 1979 Evidence for chromosomal macronuclear substructures in *Tetrahymena*. J. Protozool. **26:** 66–74.

Weeks, D. E., T. G. Nygaard, M. Neystat, L. D. Harby and K. C. Wilhelmsen, 1995 A high-resolution genetic linkage map of the pericentromeric region of the human X chromosome. Genomics **26:** 39–46.

Wickert, S., L. Nangle, S. Shevel and E. Orias, 2000 Tetrahymena macronuclear genome mapping: colinearity of macronuclear coassortment groups and the micronuclear map on chromosome 1L. Genetics **154:** 1155–1167.

Williams, J. G. K., A. R. Kubelik, K. J. Livak, J. A. Rafalski and S. V. Tingey, 1990 DNA polymorphisms amplified by arbitrary primers are useful genetic markers. Nucleic Acids Res. **18:** 6531–6535.

Communicating editor: S. L. Allen