DOE/ER/60686-T1

# FINAL REPORT

DOE/ER/60686--T1

DE92 015220

**Department of Energy Grant No. DE-FG05-88ER60686**

**Project Period:** June 1, 1988 - November 30, 1990

## Overview

Large numbers of clones will be generated during the Human Genome Project. As each is characterized, subsets will be identified which are useful to the scientific community at large. These subsets are most readily distributed through public repositories. The ATCC is experienced in repository operation, but before this project had no history in managing clones and associated information in large batches instead of individually. This project permitted the ATCC to develop several procedures for automating and thus reducing the cost of characterizing, preserving, and maintaining information about clones (Maglott and Nierman, 1990, Appendix I).

The experimental clones were contributed by Maynard Olson, who has been generating a physical map of the genome of *S. cerevisiae* by determining the sizes of *Eco*RI/*Hin*dIII restriction fragments (Olson *et al.*, 1986). A subset of minimally overlapping clones was identified for deposit with the ATCC. As a result of this project, approximately 850 have now been recharacterized, preserved, and made available to the scientific community from the ATCC. In summary, the following goals necessary to operate a repository of clones on a large scale were accomplished:

- Clone transfer from depositor to ATCC in large batches
- Electronic information transfer between information resources and the ATCC, including updates
- Automation of clone structural verification by determining restriction fragment sizes using an ABI DNA sequencer
- Automation of clone structural verification by determining restriction fragment sizes using agarose gel electrophoresis and image analysis
- Preservation of clones in tubes compatible with microtiter plates and thus with laboratory robotic systems
- Initial preservation of clones in small batches to permit subsequent correlation of distribution stock with demand
- Design of a database to permit queries on the basis of such characteristics as chromosome position, fragment size, gene content, and ATCC number
- Development of software to incorporate values from the depositor's database to the ATCC database
- Development of software to generate composite chromosome diagrams (Appendix II)
- Distribution of clones to the scientific community beginning in May, 1990.

## Clone Transfer

Since our progress report of September, 1989, we received an additional 37 clones (15 bacteriophage, 1 plasmid, 21 cosmids). They were transferred to the ATCC in September, 1990 after gaps in the physical maps of the smaller chromosomes had been filled in the Olson laboratory. We currently have a total of 894 clones (872 bacteriophage, 1 plasmid, 21 cosmids)

which were received as follows:

| | |
|---|---|
| 336 | June 23, 1988 |
| 521 | February 1, 1989 |
| 37 | September 26, 1990 |

**Information Transfer**

Clone-associated information was also received in batches. A tape (TK50) of the complete Olson database was transferred with the first set of clones. A replacement version of the database was received in September, 1990, as well as a preprint of a paper describing the *Sfi*I and *Not*I maps as correlated to some of the clones at the ATCC. Two interim files identifying clones containing known genes were received as Paradox3 tables.

A electronic version of the genetic map (Version 10) was received from R. Mortimer, Yeast Genetic Stock Center, and used as a source of recombination distances, gene symbols and gene names. Files containing symbols and strain genotypes were also transferred to the ATCC from J. Kans, National Center for Biotech. · ·gy Information.

At the time the original information and clones were received, the clones were identified as belonging to 255 different contigs. This number has been reduced to 181, with the largest contig of about 563 kb corresponding to chromosome V.

**Preparation of Seed and Distribution Stocks**

Approximately 830 clones in bacteriophage lambda vectors have been amplified, dispensed into vials for archiving (seed, 4 vials) and distribution (5 vials), and preserved in liquid nitrogen according to the procedures detailed in our previous project reports. These procedures were developed to facilitate automated handling and to reduce storage costs. The small number of plasmid and cosmid clones were processed individually by standard ATCC procedures. There has not yet been sufficient distribution of any of these clones to warrant represervation by freeze-drying of larger amounts of distribution stock.

At the beginning of this project, a vial from each preservation was titered after freezing to assess potential loss of viability. Because all were adequately preserved (Nierman *et al.*, 1987), subsequent viability checks were limited to spotting lysate onto bacteria-containing plates and checking for plaques.

**Verification of Clone Structure**

The physical structure of all clones deposited to this pilot repository were verified by analysis of restriction fragment sizes. For the most part, fragment sizes were determined by methods described in our previous progress reports, namely labeling fragments with one of four fluorescent dyes and sizing them during electrophoretic separation on a 3% polyacrylamide gel using a DNA sequencer (Applied Biosystems, Inc., Model 370A) (Carrano *et al.*, 1989). This protocol resolved fragments only less than 1.5 kb, but this was sufficient for verification.

The latest set of clones provided to the repository were verified using horizontal agarose gel electrophoresis. Photographs of the ethidium bromide-stained gels were analyzed using an image analysis system (BioImage 60) and the restriction fragment sizes were computed using software provided by BioImage.

Software was written to display ATCC-determined fragment sizes on a screen with Olson-determined fragment sizes for each clone. Matching bands were identified interactively and recorded. A clone was considered verified when the fragments reported by the depositor within the experimental size range were identified by the ATCC.

The automated methods for determining fragment sizes and for comparing them to the expected patterns permitted us to establish the throughput necessary to complete this project.

## Information Management

To manage the clone-related information, the system utilizing a Sun 4/110 workstation, an AST 386 personal computer, a MicroVAXII computer, the Vectra associated with the ABI 370A sequencer, and a PostScript printer was developed as detailed in previous reports. Software included commercial packages (Sybase), contributions (from Dr. Carrano's group at the Livermore National Laboratory), and programs written at the ATCC. The latter software was developed to extract data from the Olson database to store in Sybase, to manage laboratory processing (notebook) information, to facilitate comparison of ATCC-generated to Olson-generated restriction fragment data, to generate clone-specific product sheets, to design a PC-based system to handle calls from investigators using our clones, and to produce the chromosome diagrams correlating the genetic and physical maps (ATCC Catalog of Yeasts, 1990 and Appendix II).

Having clone-related and genetic data supplied electronically facilitated the processes of both verification of clone structure and management of genetic and physical mapping data. Whereas the other recombinant collections require a staff of 3 to obtain, process, and develop software to manage information on about 500 strains a year, the software development and data management for these materials required only one staff member.

## Future Activities

Now that these clones have been recharacterized and preserved, it is possible to increase the usefulness of the *S. cerevisiae* genomic repository. Blots of DNA from all the clones of the overlapping set are being developed. These blots will serve as a useful mapping tool for determining the position of any *S. cerevisiae* DNA fragment within the clone set. A single hybridization reaction in a single bag may localize any yeast sequence to one or a few clones of the set and thus to a precise location on the physical map. Blots will be distributed with the request that the investigator report the hybridization results to the ATCC to receive the mapping information and clones of interest. Thus increased correlations between the genetic and physical maps will be generated.

The ATCC will continue to receive additional clones and information from the Olson laboratory as gaps in the physical map are closed.

## Projected Costs of Repositories supporting the Human Genome Project

Extrapolations from repositories managing clones individually generated estimates of $60 million (Roberts, 1987) or $15 million (Office of Technology Assessment, 1988) for maintaining a repository for the Human Genome Project for the first 5 years. We have recently reported our evaluation of costs for maintaining a genomic repository (Maglott and Nierman, 1990) based on our experience from this project. It is apparent that significant economies can be realized by automating laboratory data collection for restriction fragment analyses and by electronic data transfer from the depositor and other resources. Coupling this rate of productivity with a minimum rate (5%) of clone verification suggests that projections for the cost of initial preservation can be reduced to the order of $200,000 per 10,000 clones. Continuing costs can be recovered from fees assessed to the scientific community.

The initial costs must be compared to the benefits expected from a public repository. One benefit is that each clone developer is not required to become a storage and distribution center--with the inventory maintenance, information management, "customer service", and mailing costs required by such activities. This effort is not trivial and may easily exceed expectations. Another consideration is that clone repositories do meet a need in the scientific community. Although the *S. cerevisiae* clones have been made available from the ATCC too recently to generate significant usage data, in 1990 the ATCC/NICHD Repository distributed

11,548 clones and 615 libraries to more than 2200 different investigators worldwide. This argues that although storing information such as STS and ETS sequence data may be sufficient for some genomic analyses, there is significant need for well-characterized clones to serve as starting materials for a variety of applications such as transfection and site-directed mutagenesis. Although some of these sequences may be able to be generated by PCR, the cost of synthesizing the oligonucleotide primers themselves is greater than the cost to the investigator of ordering the required clone. The experience of the ATCC with clone repositories demonstrates that centralized resources are useful to the scientific community. The experience of this project documents that the cost of establishing repositories on a larger scale can be reduced significantly by automating clone processing and information management.

**Literature cited**


Carrano, A. V., Lamerdin, J., Ashworth, L. K., Watkins, B., Branscomb, E., Slezak, T., Raff, M., de Jong, P. J., Keith, D., McBride, L., Meister, S., Kronick, M. 1989. A high-resolution, fluorescence-based, semiautomated method for DNA fingerprinting. Genomics **4**:129-136.

Maglott, D. R. and Nierman, W. C. 1990. Clone and genomic repositories at the American Type Culture Collection. Genomics **8**: 601-605.

Nierman, W.C., Trypus, C., Deaven, L. L. 1987. Preservation and stability of bacteriophage lambda libraries by freezing in liquid nitrogen. Biotechniques **5**:724-728.

Office of Technology Assessment. 1988. Mapping our genes, genome projects: how big, how fast? Congress of the United States, Washington, D.C.

Olson, M.V., Dutchik, J.E., Graham, M.Y., Brodeur, G.M., Helms, C., Frank, M., MacCollin, M., Scheinman, R., Frank, T. 1986. Random-clone strategy for genomic restriction mapping in yeast. Proc Natl Acad Sci USA **83**:7826-7830.

Roberts, L. 1987. Human genomic questions of costs. Science **237**: 1411-1412.

## Appendix I

## Publications

Removed and
cycled separately -

# Appendix II

## *Saccharomyces cereviciae* AB 972
## Chromosome/Clone Maps

## OVERLAPPING GENOMIC CLONES FROM *SACCHAROMYCES CEREVISIAE*

The ATCC maintains and distributes a set of overlapping recombinant clones containing inserts from the genome of *Saccharomyces cerevisiae* AB972. These clones were constructed for a physical mapping project in the laboratory of Dr. Maynard Olson, Washington University (Proc. Natl. Acad. Sci USA *83*: 7826-7830, 1986). Sets of minimally overlapping clones groups in contigs (MERG and CONTG units named by Olson) were identified and deposited with the ATCC. The inserts are either in bacteriophage lambda vectors (lambdaMG3 or lambdaMG14, ATCC 70000 to 70872), in plasmid vectors (pNF400, ATCC 70873), or in cosmid vectors (pHC79, ATCC 70874 to 70894).

The following pages diagram the positions of each clone as related to the physical and genetic maps. Rather than repeat explanations for each figure, the diagram on the first page is to be used as the general legend.

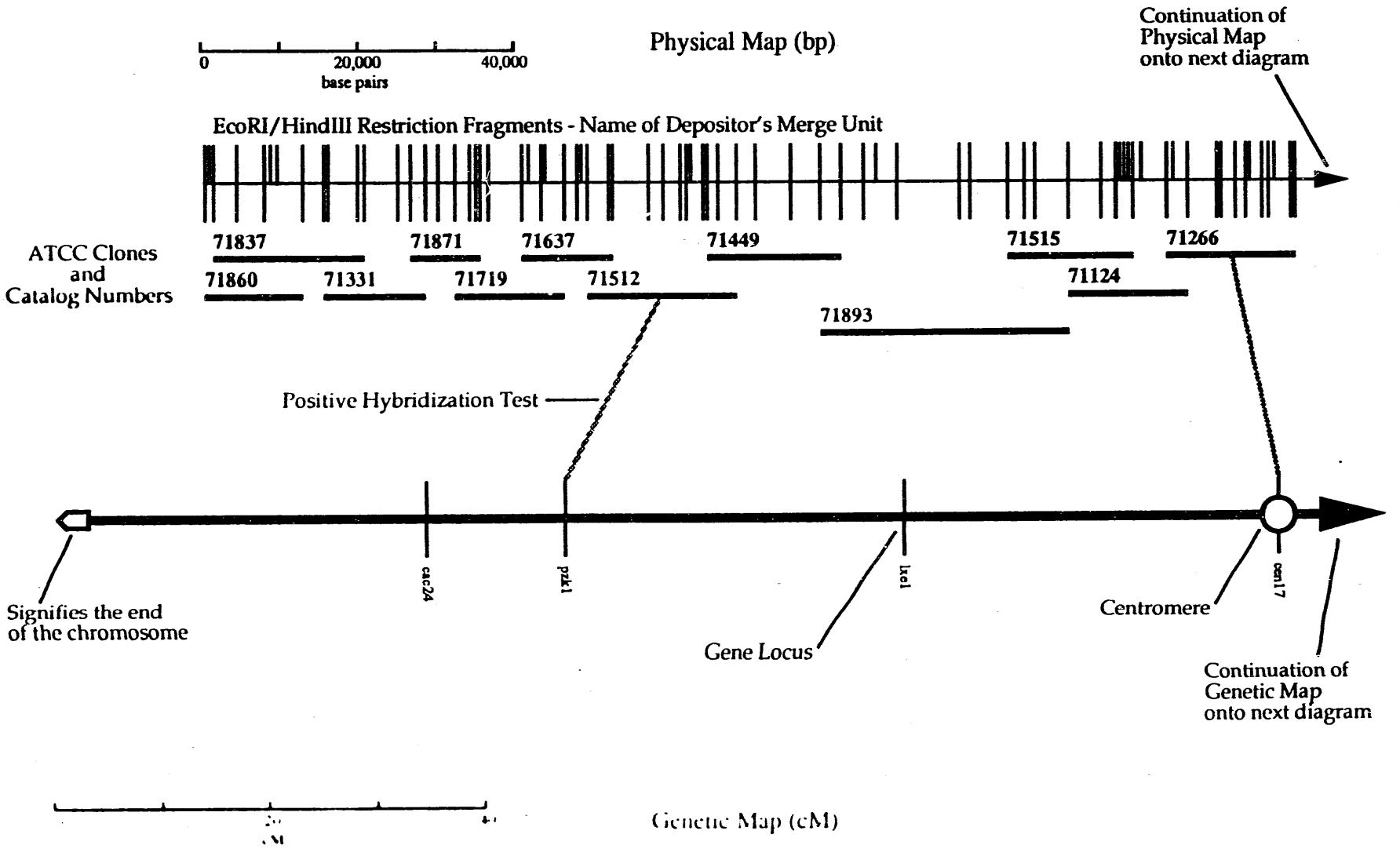
Physical mapping data were provided by M. Olson. When the order of *Eco*RI/*Hin*dIII fragments is unambiguous, the ends are represented by a vertical line intersecting the horizonal. Unordered fragments are depicted by the short vertical line in order of size. Please note that in some cases, such as when only one gene has been identified in a contig, the orientation of the physical map is undefined, so that the orientation displayed is arbitrary.

The genetic map (Edition 10) is based on data generously provided by Dr. R. Mortimer. If a gene has been identified in a clone by hybridization analysis, a shaded line has been drawn connecting the locus symbol on the genetic map to the middle of the clone diagram. This line does not indicate the position of the gene in the clone which, for the most part, is currently unknown.

Correlations between these maps and *Not*I and *Sfi*I maps are beginning to be established (A. Link and M. Olson, pers. comm.). The following clones in the collection are known to link *Sfi*I or *Not*I restriction fragments as follows:

| | *Sfi*I sizes (kb) | ATCC# | *Not*I sizes (kb) | ATCC# |
|---|---|---|---|---|
| Chromosome I | 120, 120 | 70893 | | |
| | 20, 100 | 70124 | | |
| Chromosome VI | 250, 20 | 70200 | | |
| Chromosome VIII | 225, 305 | 70715 | 250, 80 | 70537 |
| Chromosome IX | 35, 60 | 70769 | | |

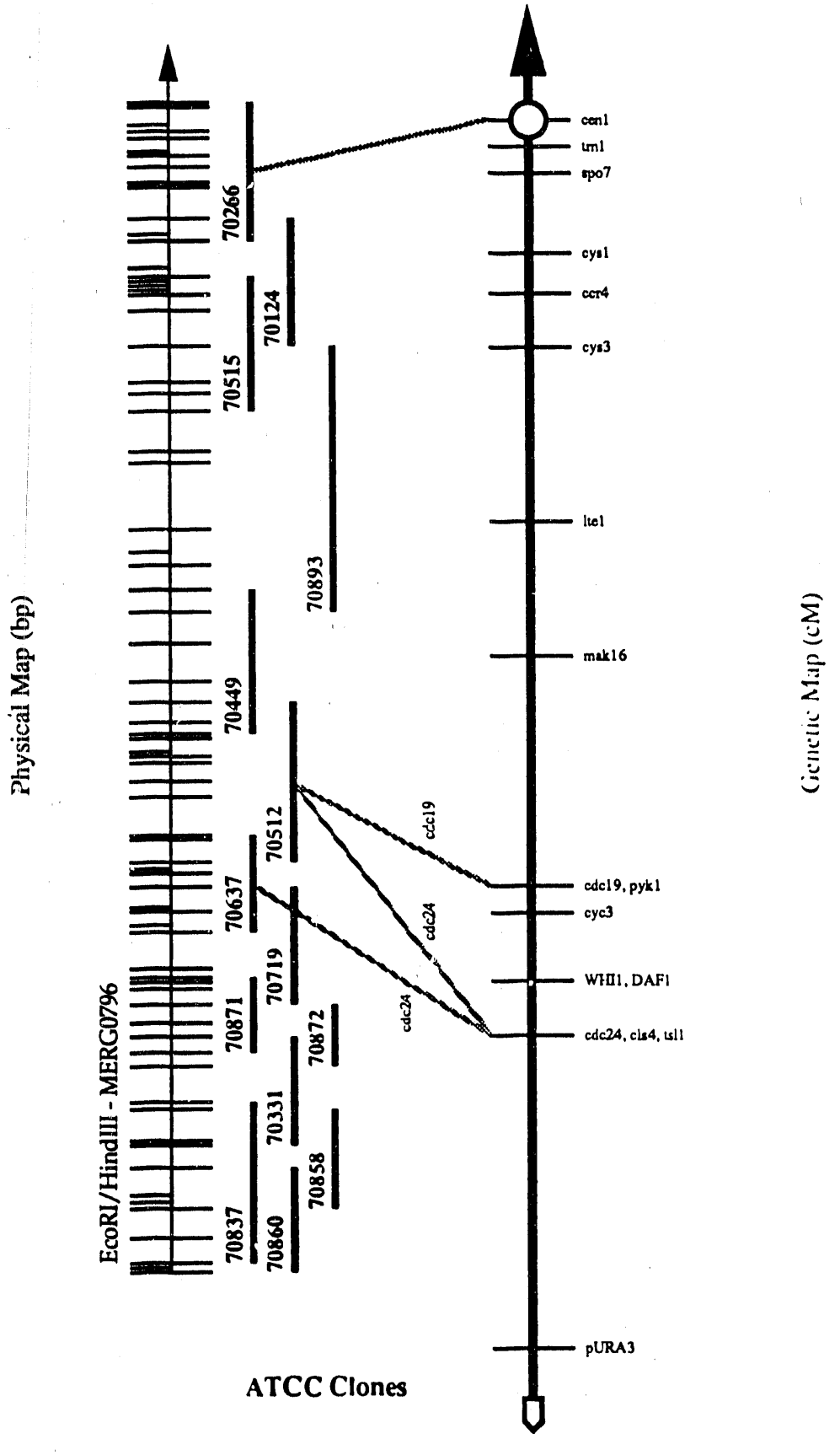
# Chromosome Diagram Legend




CLONES FROM THE *SACCHAROMYCES CEREVISIAE* GENOME

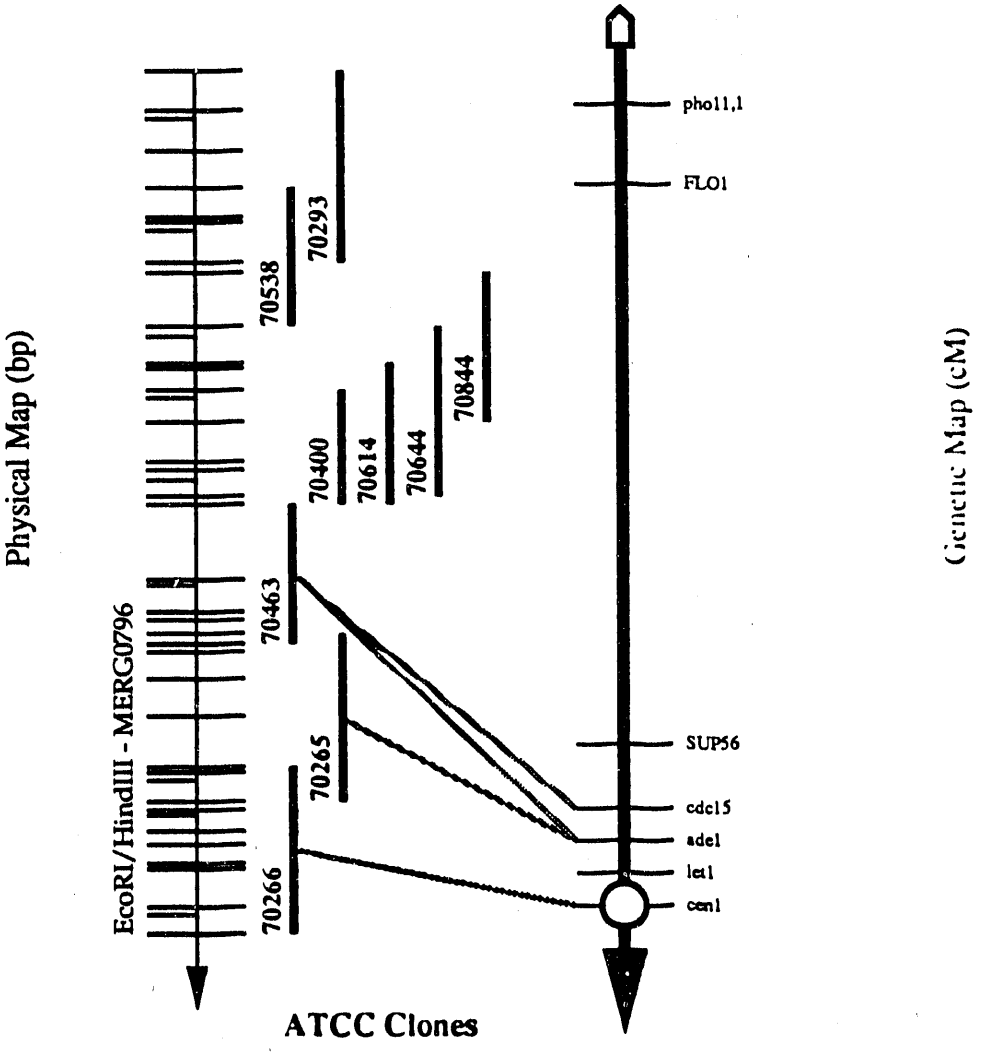
**CLONES FROM THE *SACCHAROMYCES CEREVISIAE* GENOME**

# Chromosome: I Diagram 1 of 2

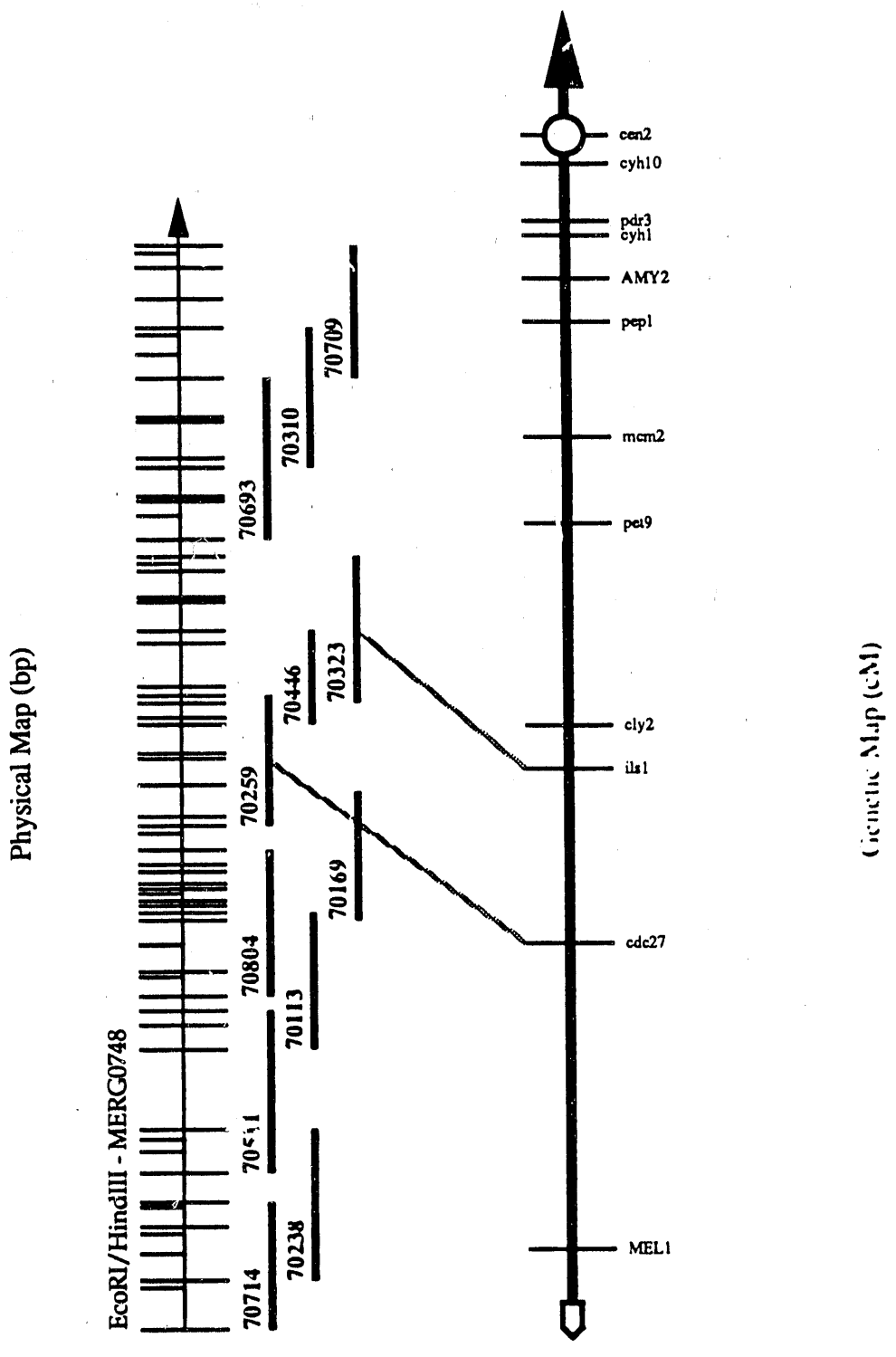


**CLONES FROM THE *SACCHAROMYCES CEREVISIAE* GENOME**

Chromosome: I Diagram 2 of 2



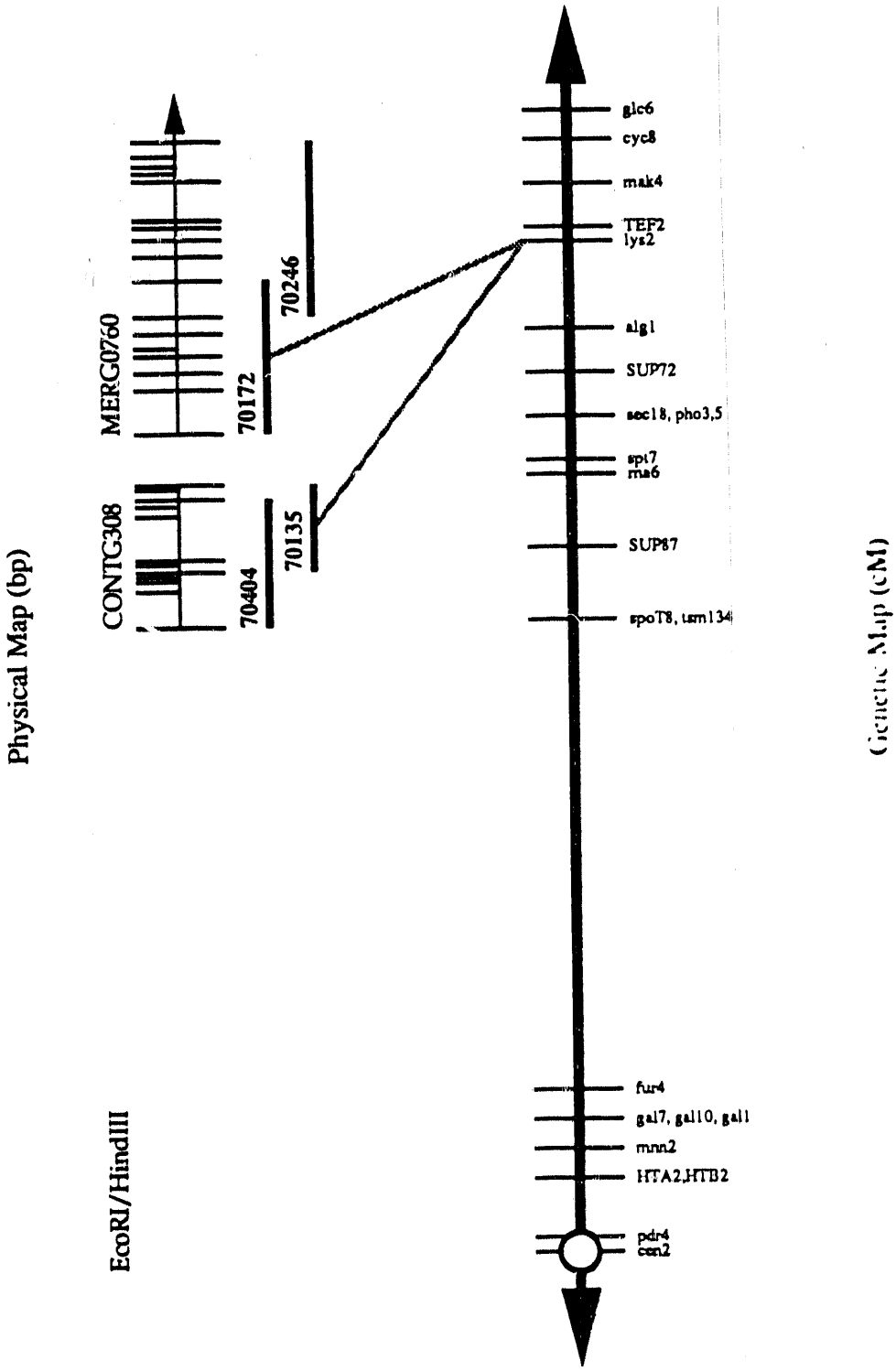
# Chromosome: II Diagram 1 of 4




ATCC Clones

# CLONES FROM THE *SACCHAROMYCES CEREVISIAE* GENOME

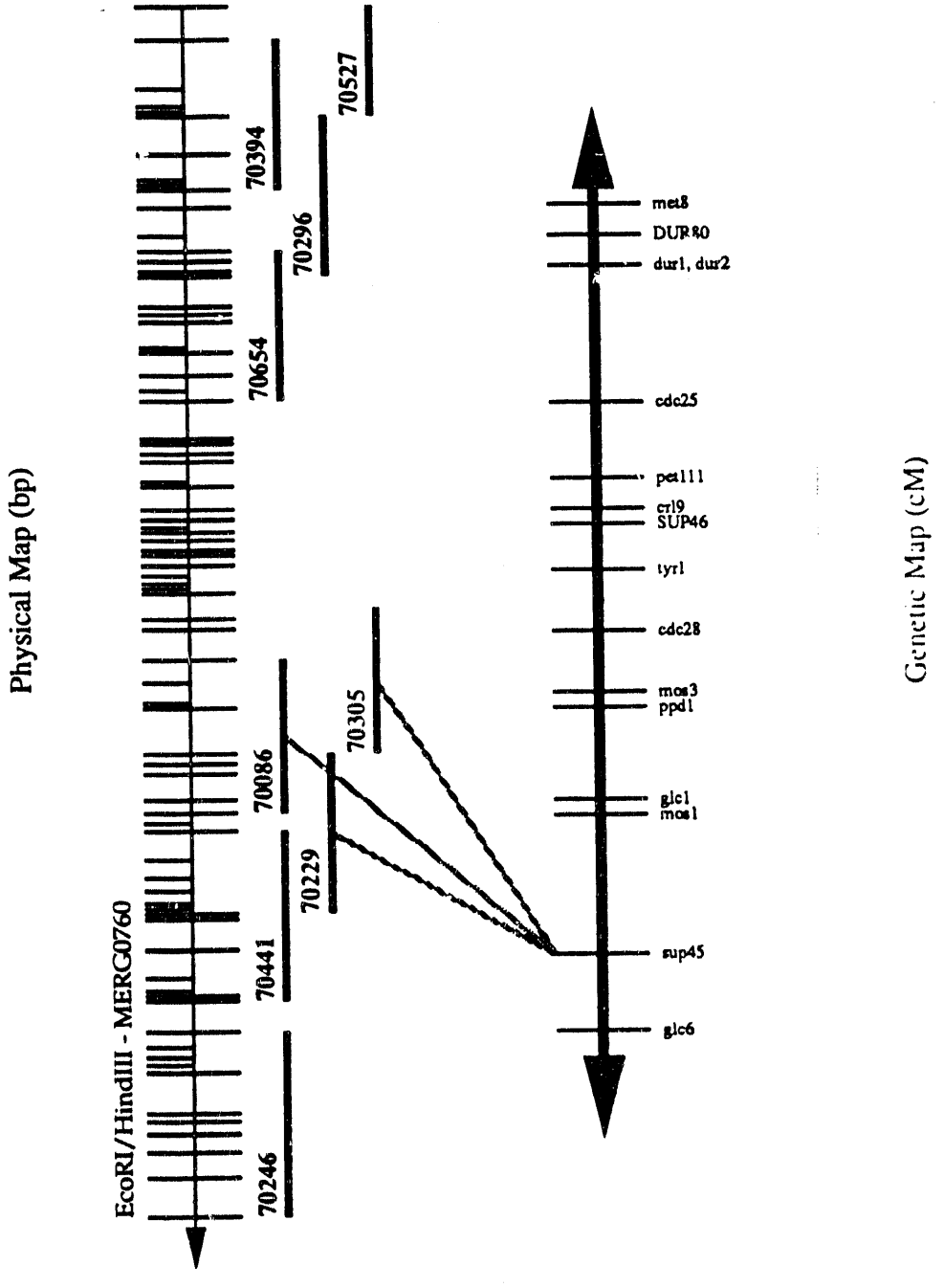
## Chromosome: II  Diagram 2 of 4




ATCC Clones

**CLONES FROM THE *SACCHAROMYCES CEREVISIAE* GENOME**

# Chromosome: II Diagram 3 of 4

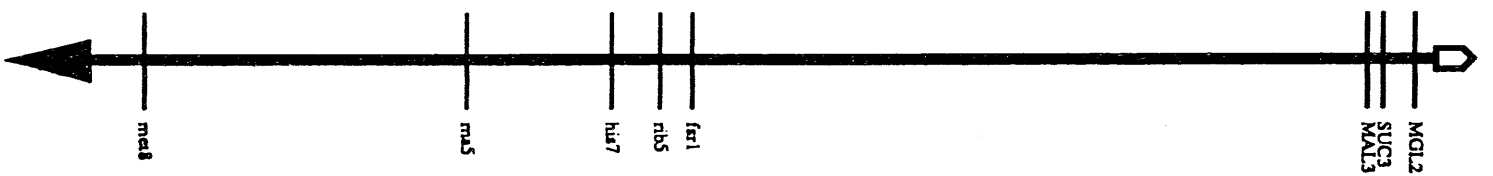



**ATCC Clones**

# Chromosome: II

# Diagram 4 of 4

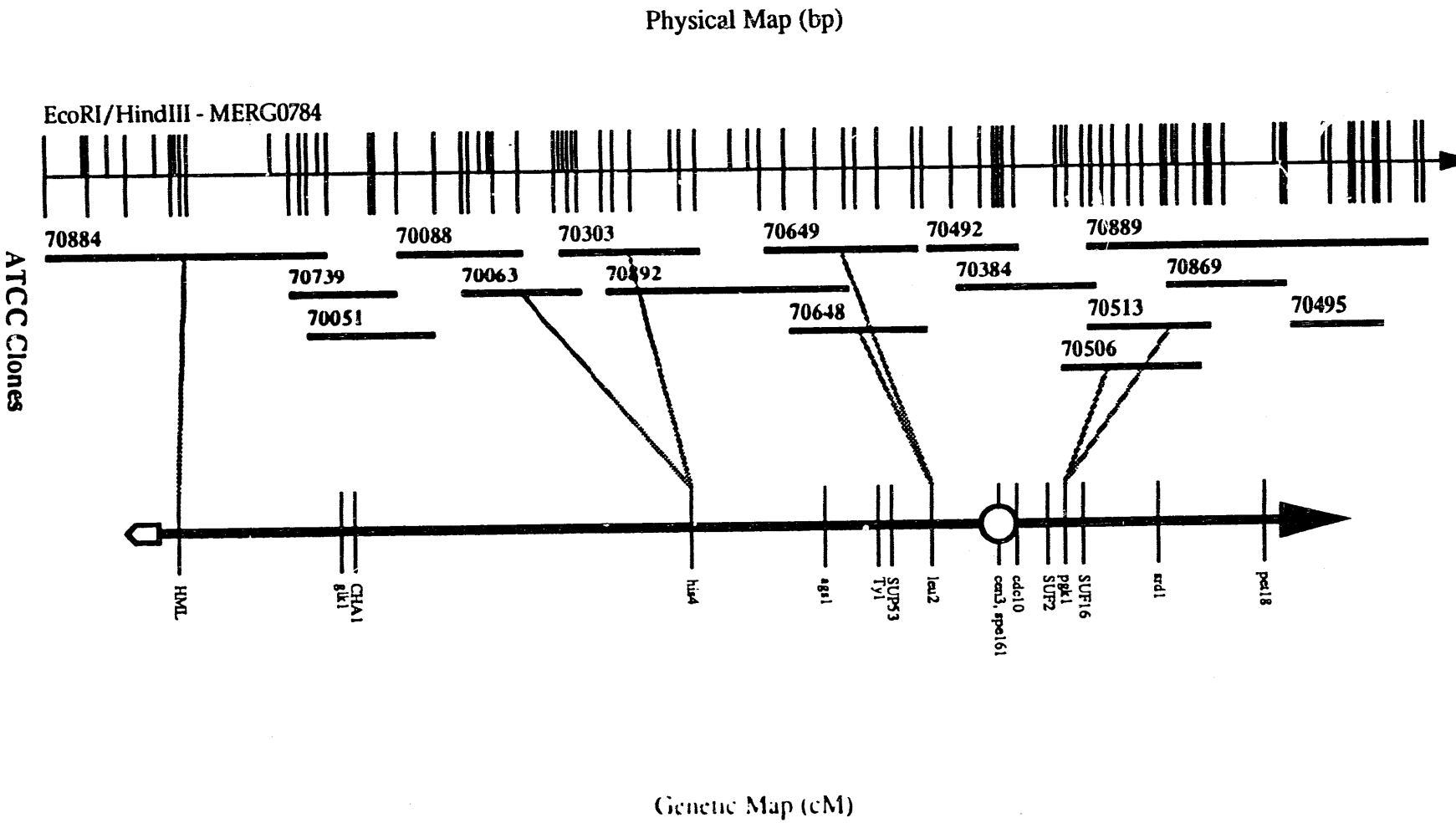
## No Physical Map Data Available



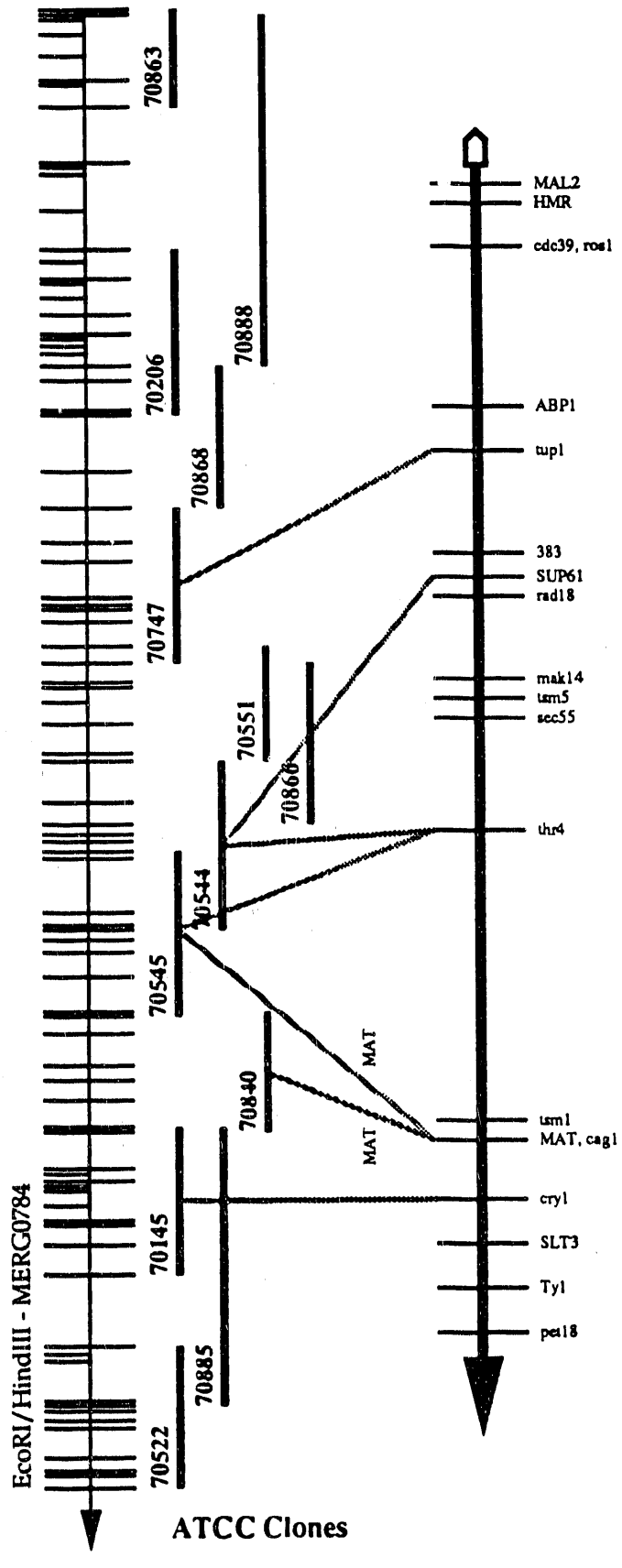

Genetic Map (cM)

# Chromosome: III Diagram 1 of 2

CLONES FROM THE *SACCHAROMYCES CEREVISIAE* GENOME



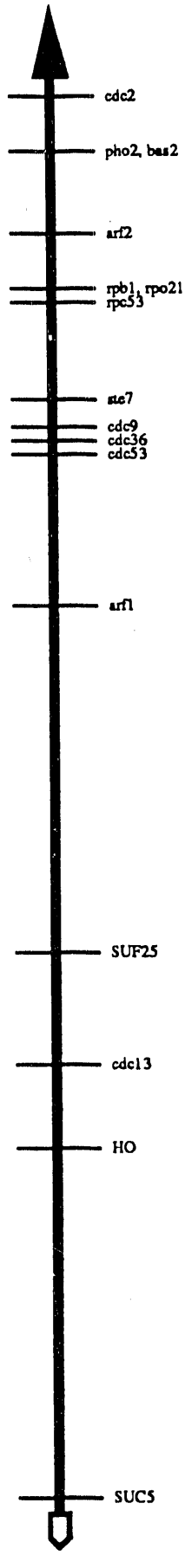
# Chromosome: III Diagram 2 of 2



**CLONES FROM THE *SACCHAROMYCES CEREVISIAE* GENOME**

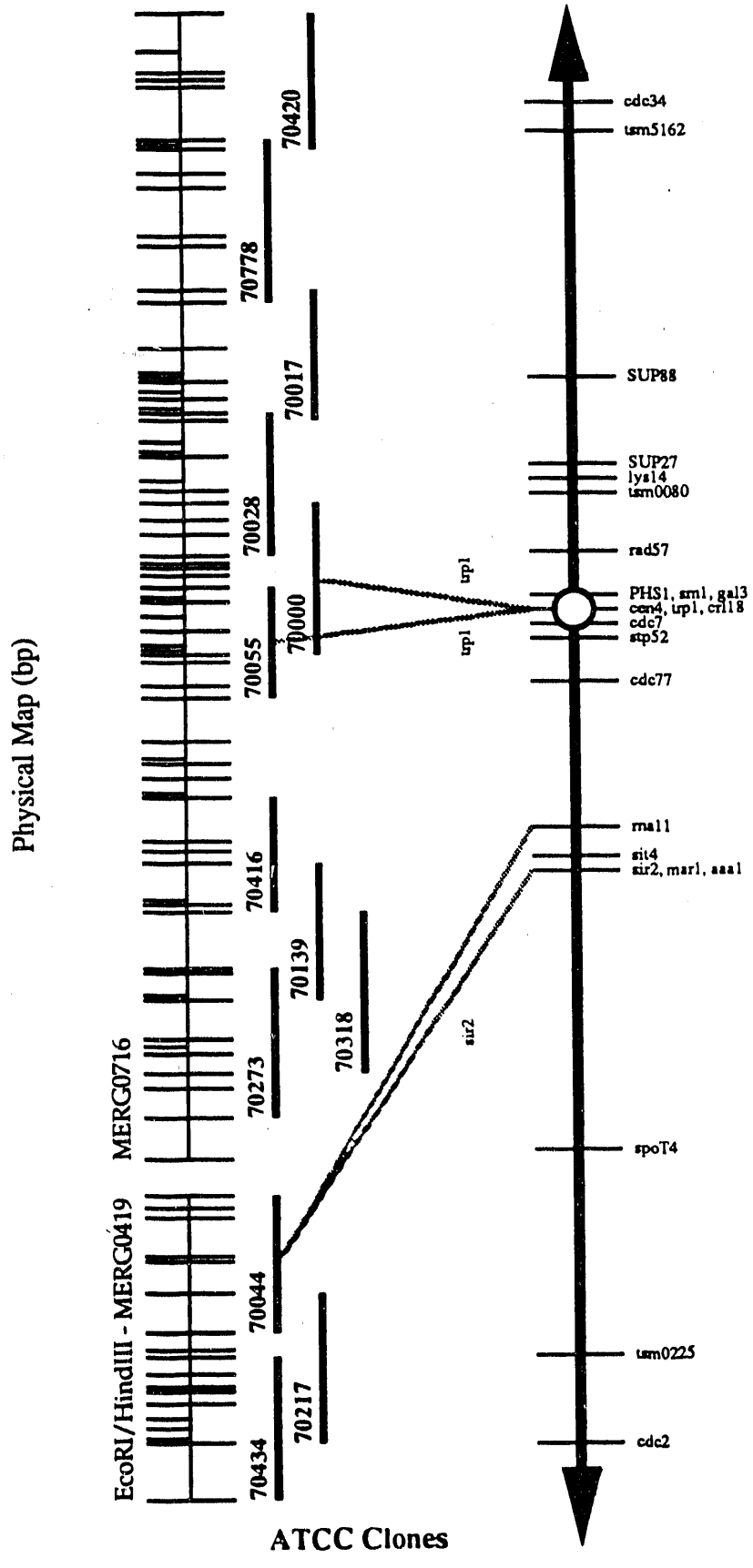
Chromosome: IV Diagram 1 of 6

No Physical Map Data Available



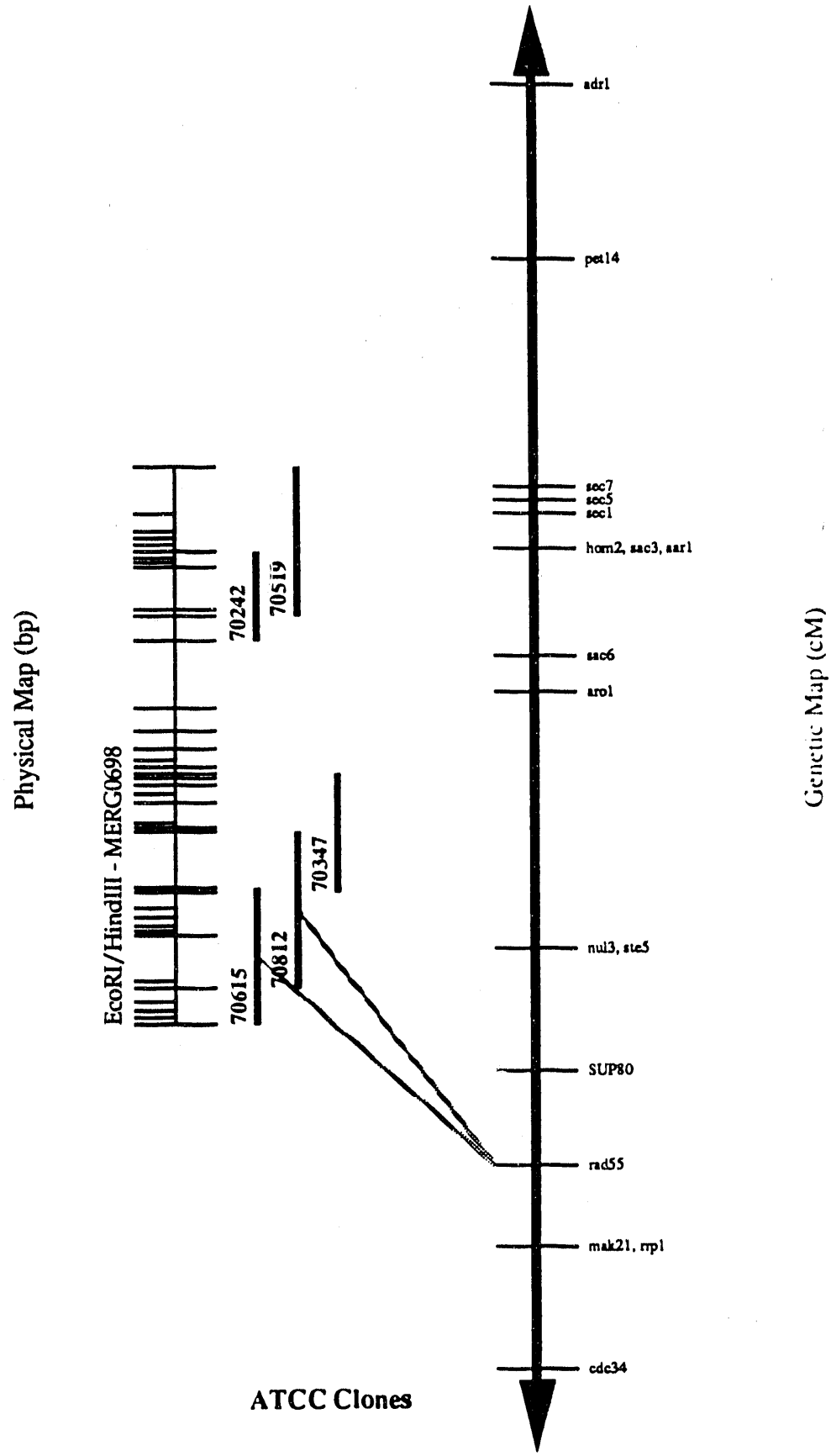

Genetic Map (cM)

# Chromosome: IV Diagram 2 of 6

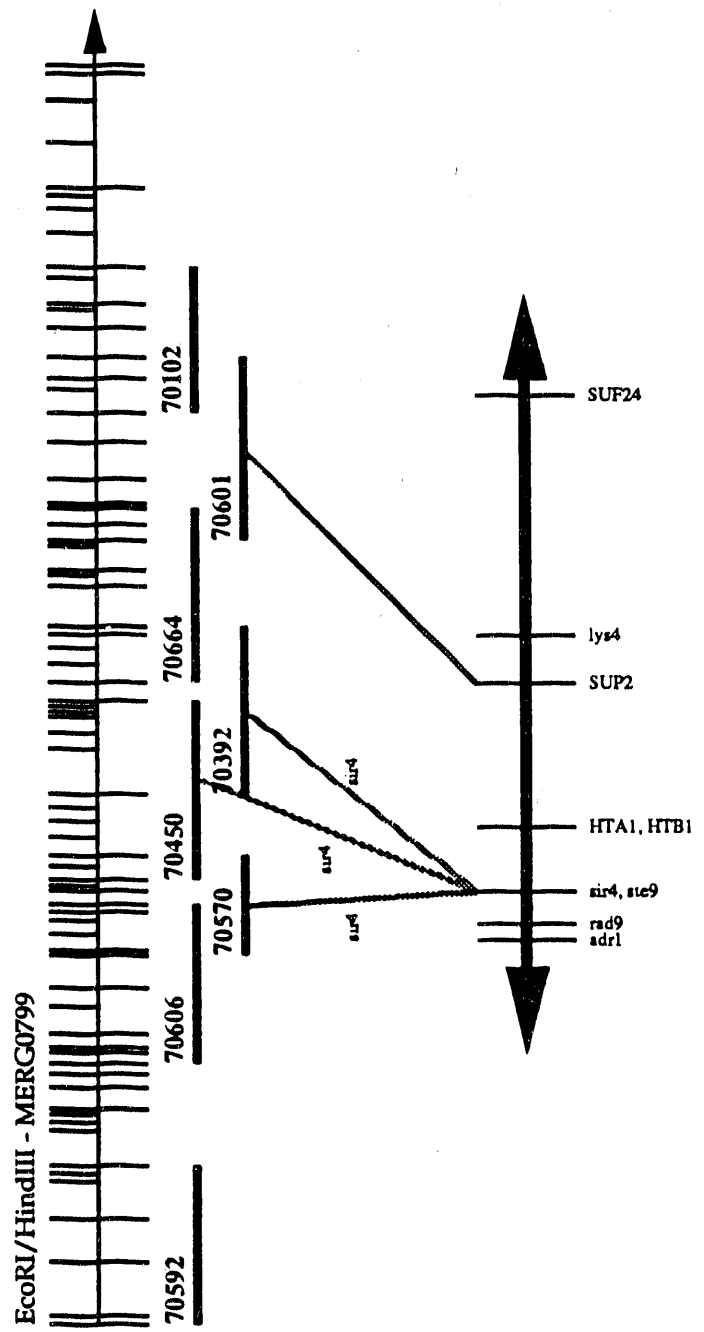


**CLONES FROM THE *SACCHAROMYCES CEREVISIAE* GENOME**

Chromosome: IV Diagram 3 of 6



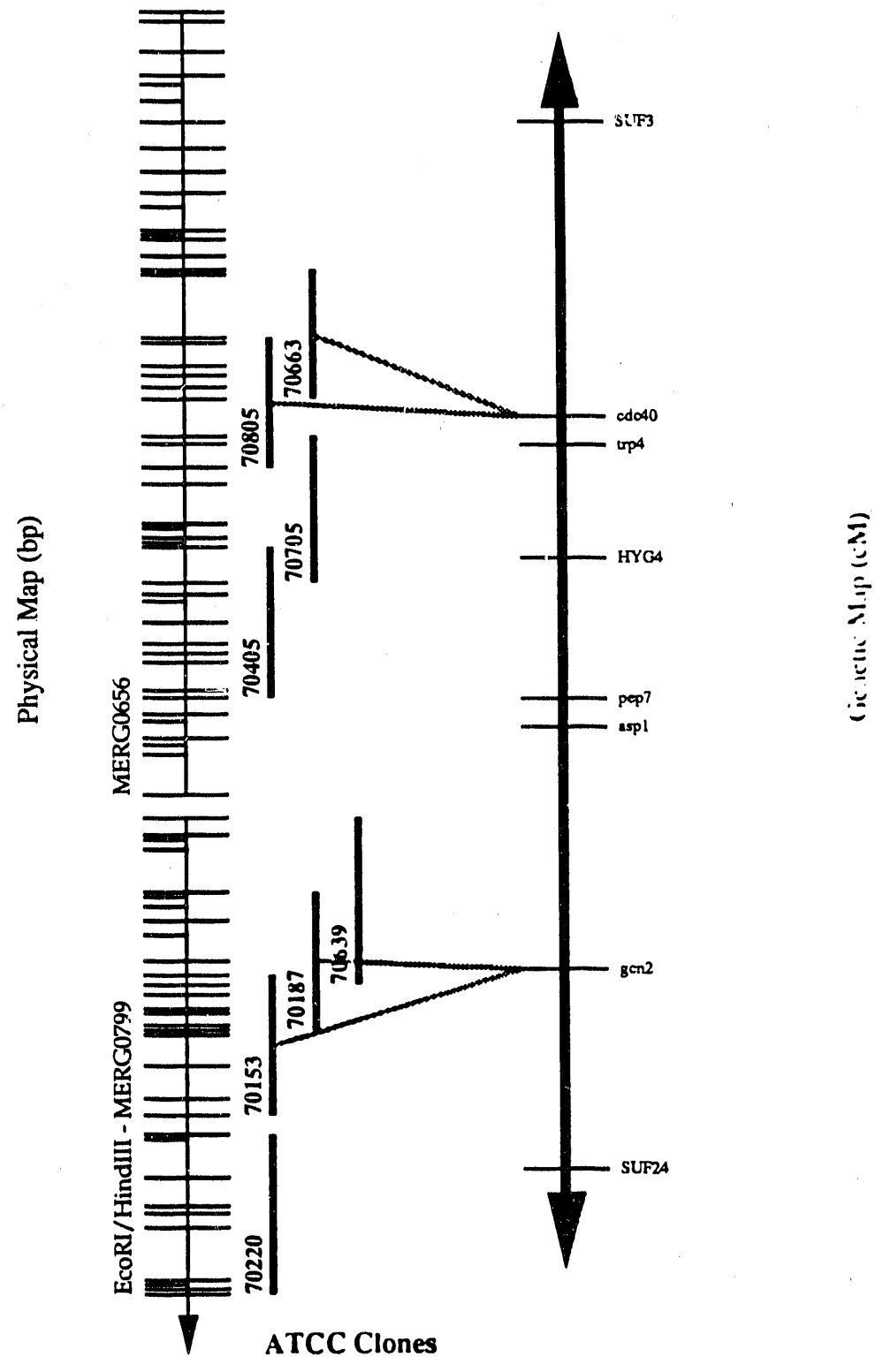
# Chromosome: IV Diagram 4 of 6




**ATCC Clones**

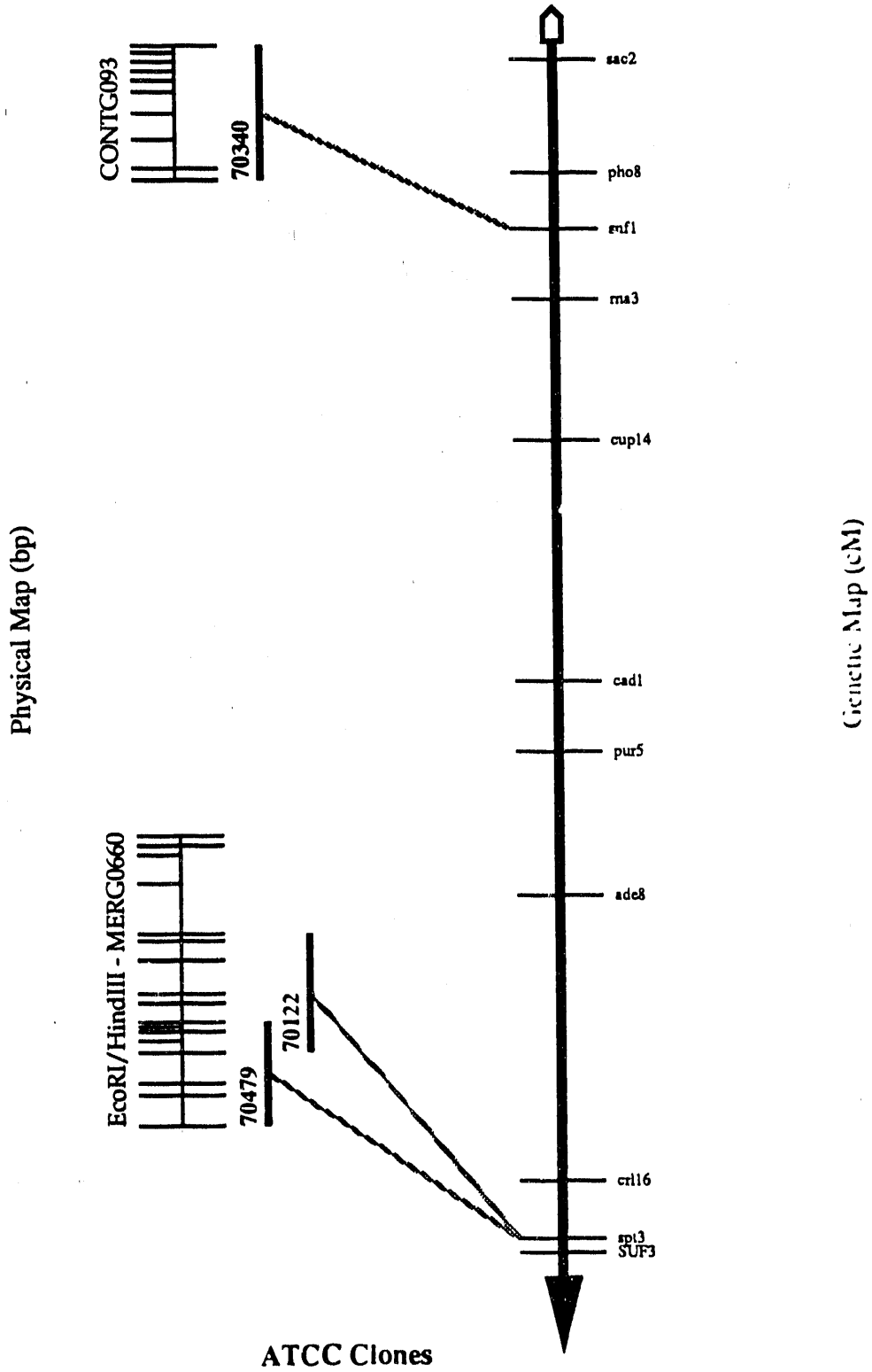
**CLONES FROM THE *SACCHAROMYCES CEREVISIAE* GENOME**

Chromosome: IV Diagram 5 of 6



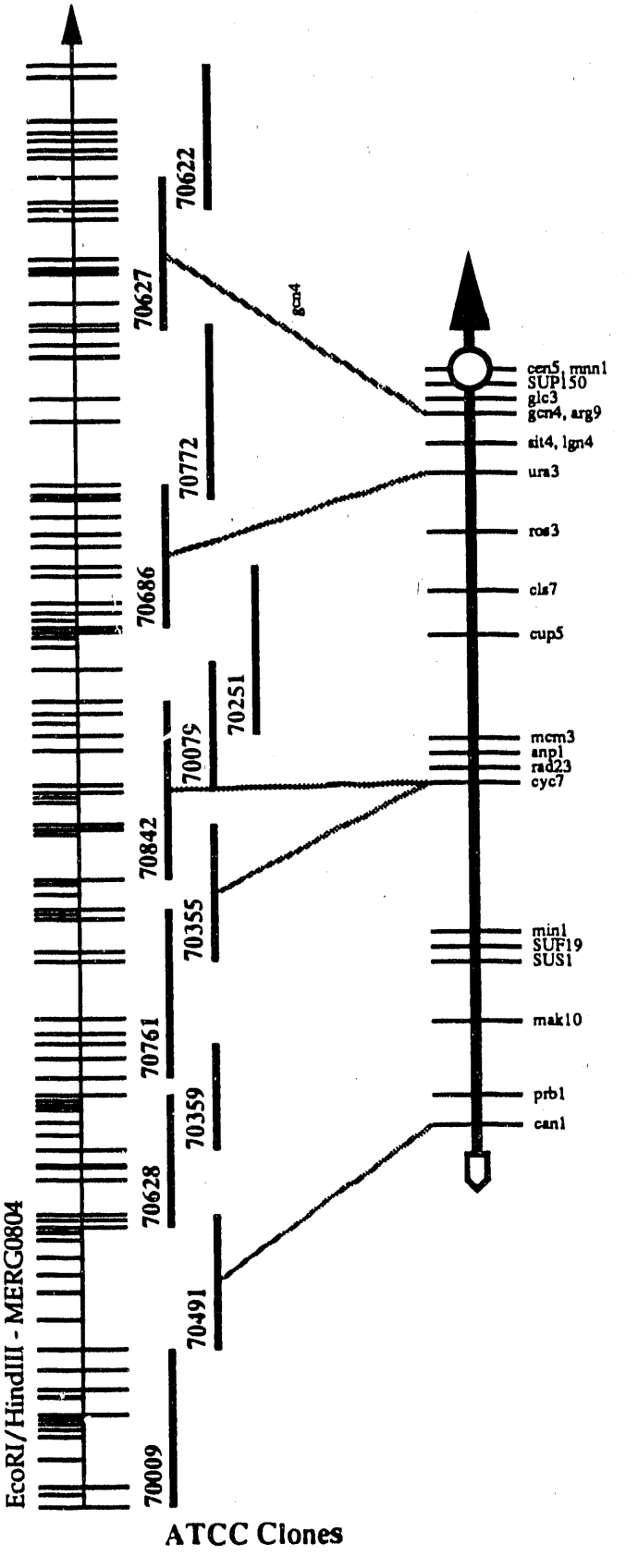
**CLONES FROM THE *SACCHAROMYCES CEREVISIAE* GENOME**

# Chromosome: IV Diagram 6 of 6



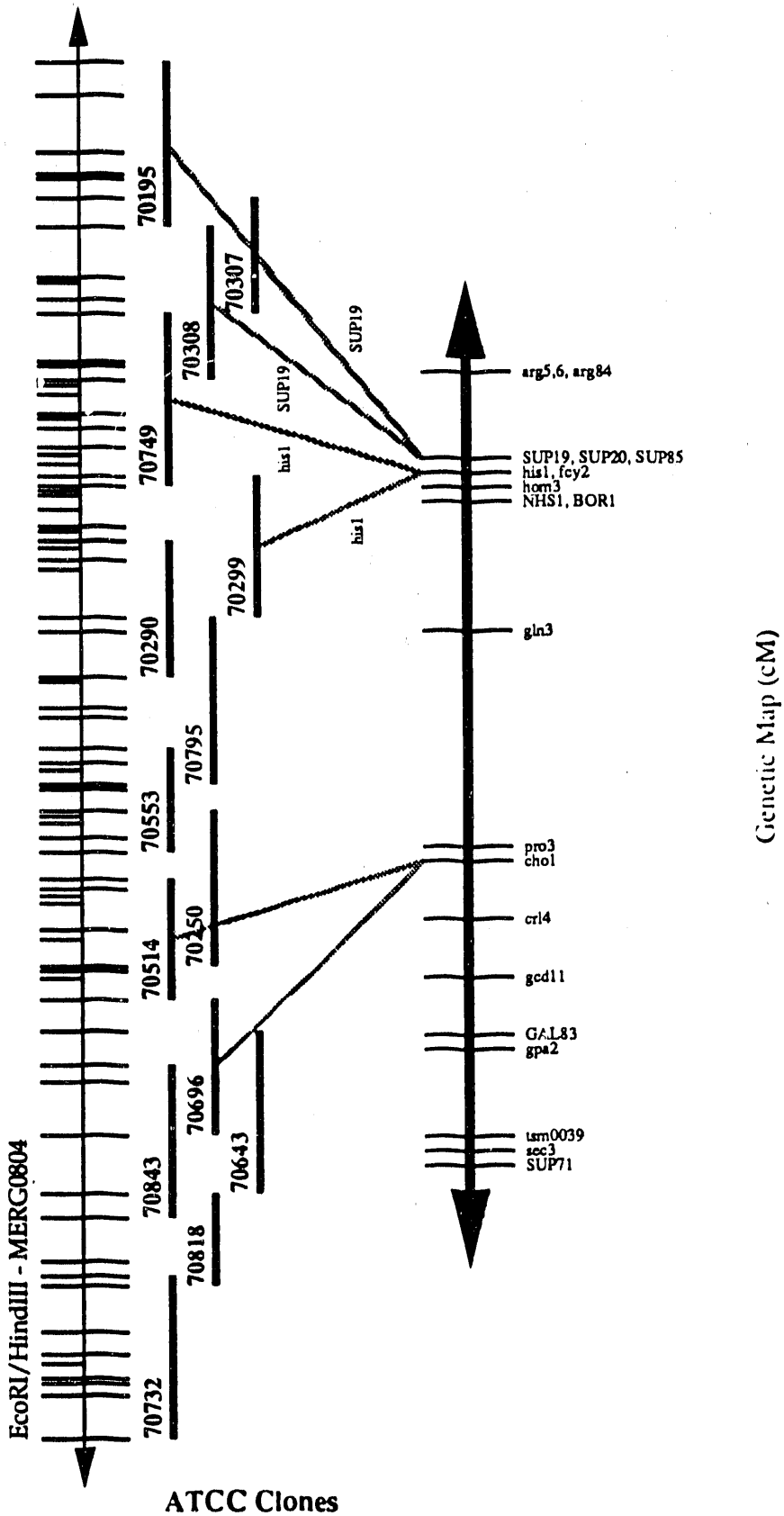
**CLONES FROM THE *SACCHAROMYCES CEREVISIAE* GENOME**

# Chromosome: V Diagram 1 of 4



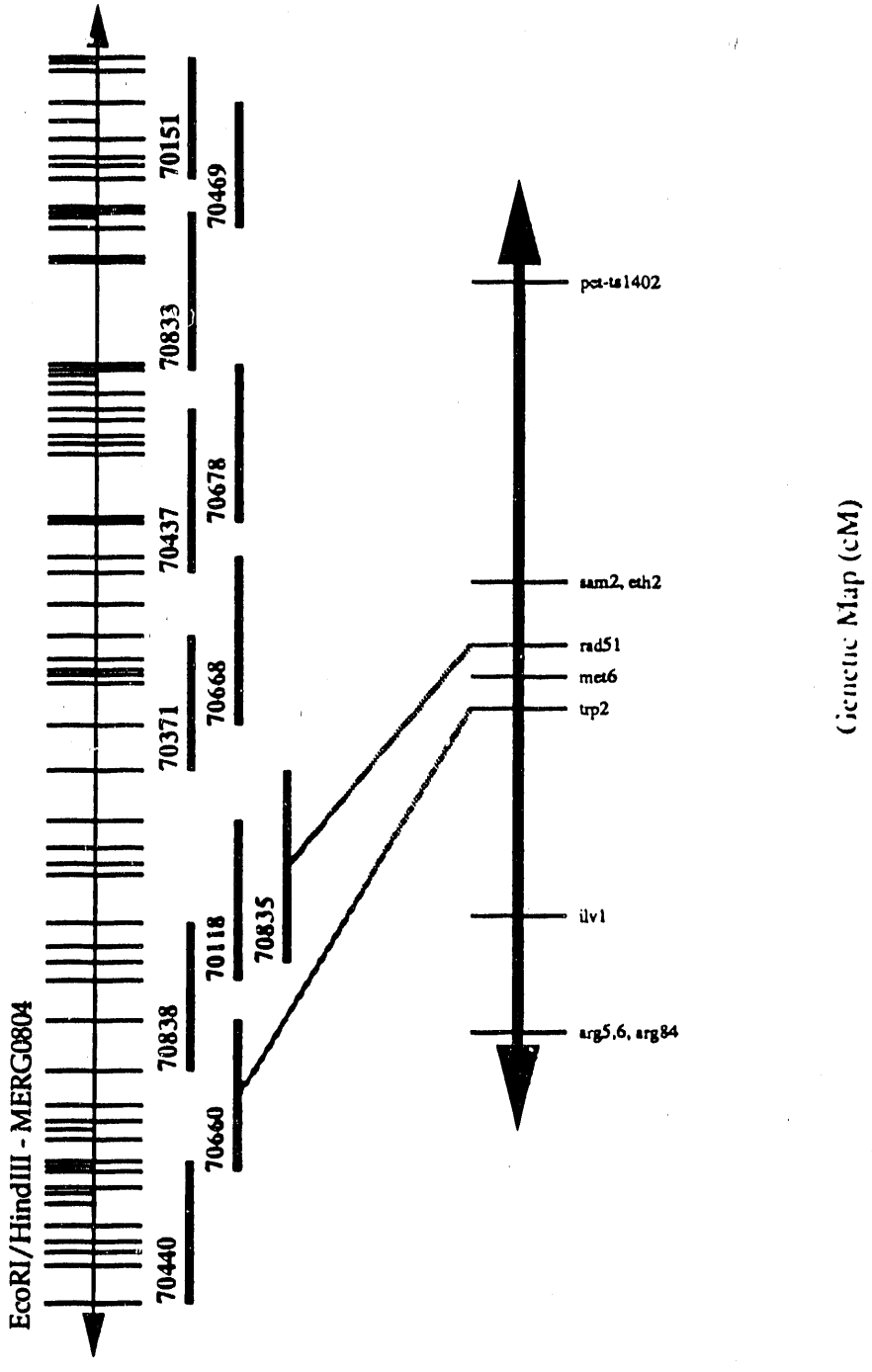
**CLONES FROM THE *SACCHAROMYCES CEREVISIAE* GENOME**

# Chromosome: V Diagram 2 of 4



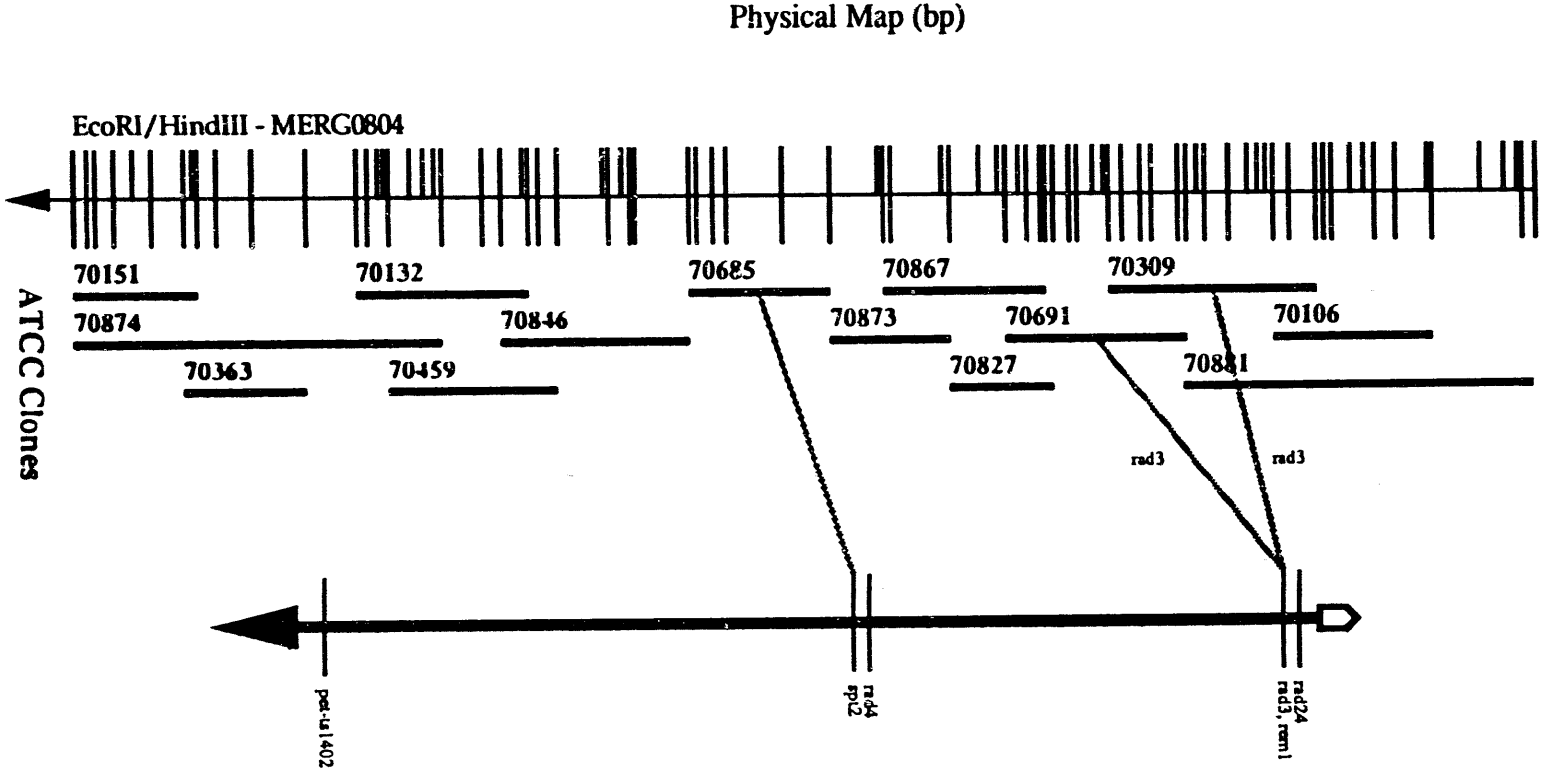
**CLONES FROM THE *SACCHAROMYCES CEREVISIAE* GENOME**

Chromosome: V Diagram 3 of 4



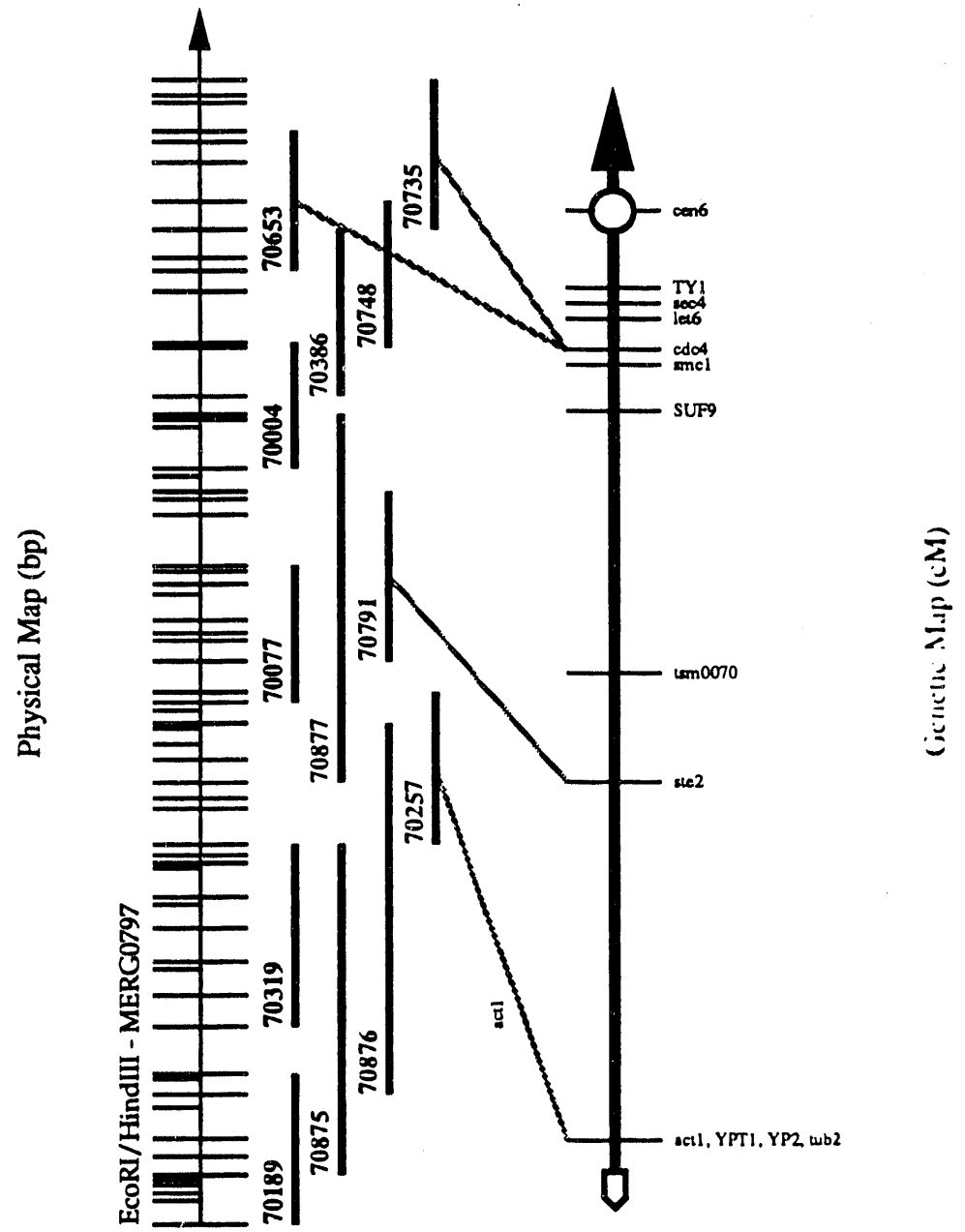

**ATCC Clones**

# Chromosome: V Diagram 4 of 4

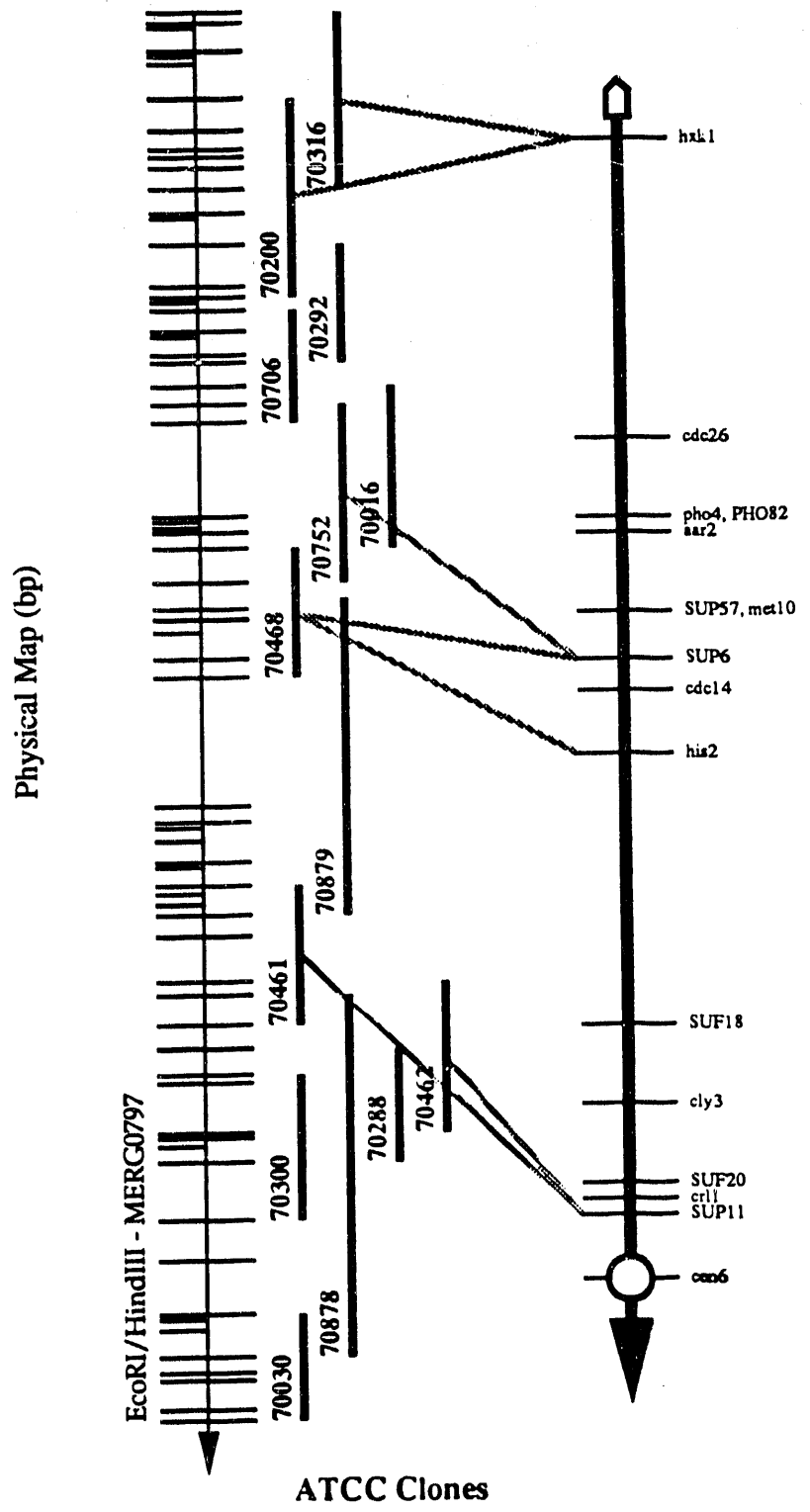


**CLONES FROM THE *SACCHAROMYCES CEREVISIAE* GENOME**

# Chromosome: VI Diagram 1 of 2

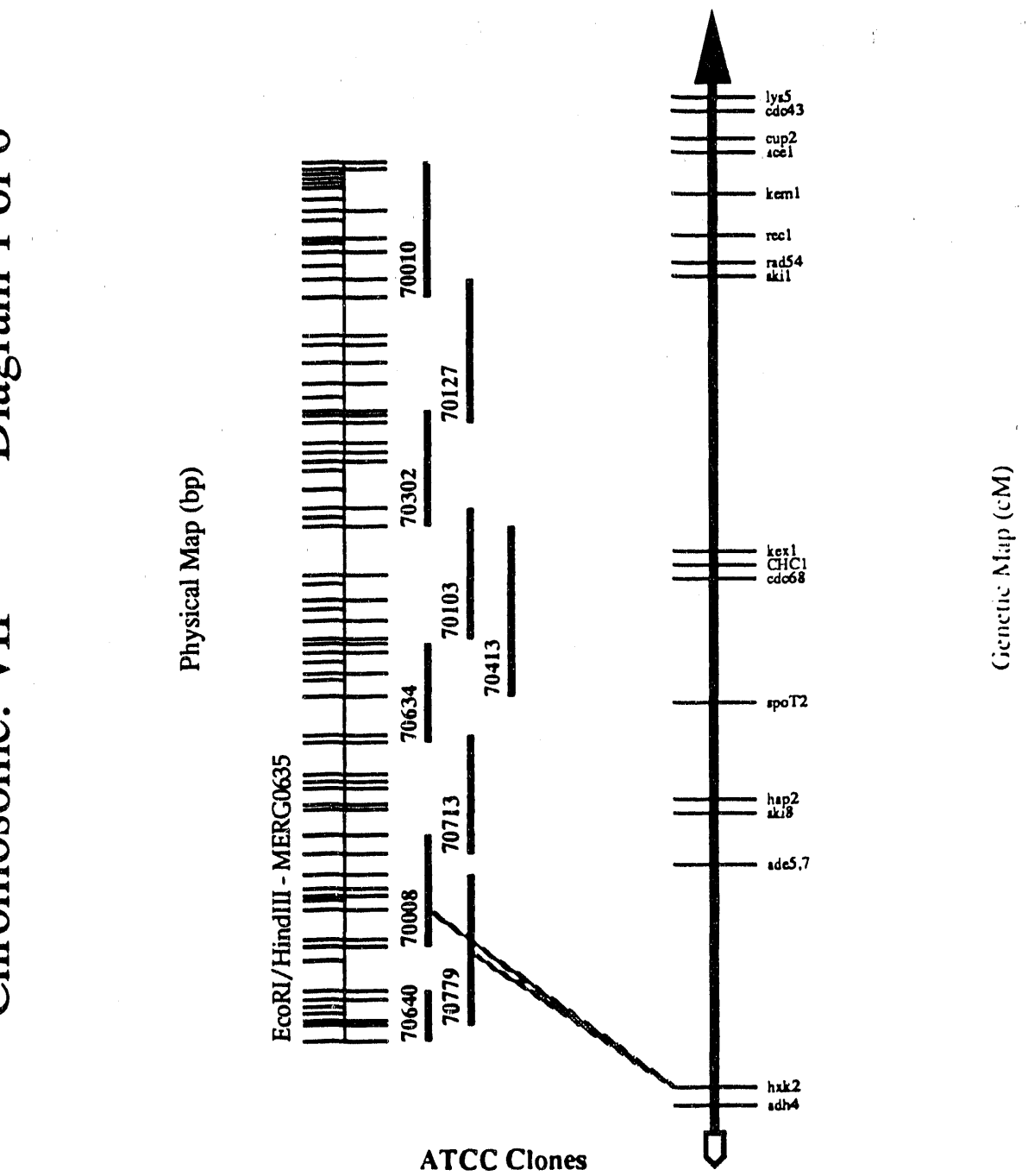


# Chromosome: VI Diagram 2 of 2

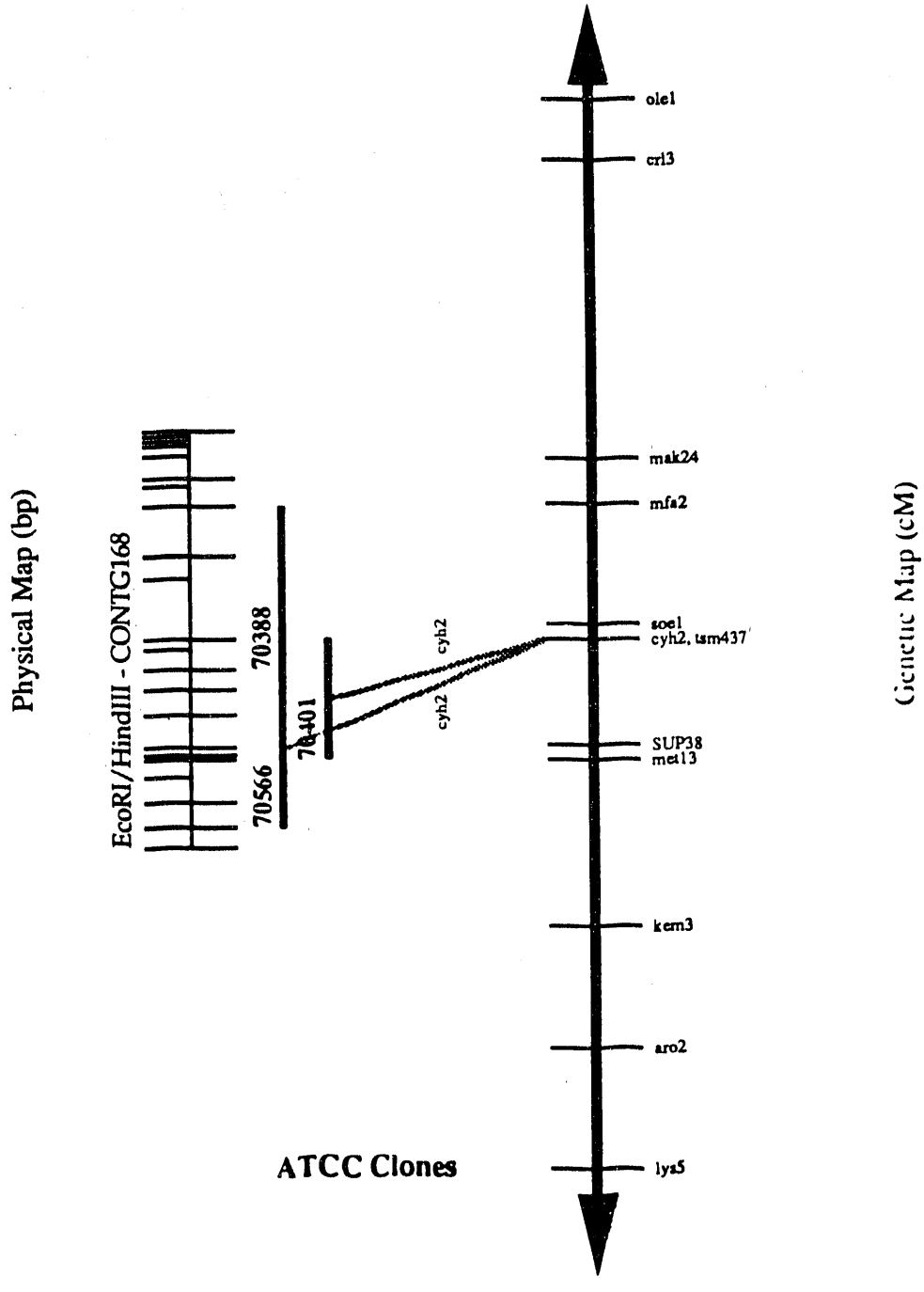


**CLONES FROM THE *SACCHAROMYCES CEREVISIAE* GENOME**

# Chromosome: VII Diagram 1 of 6

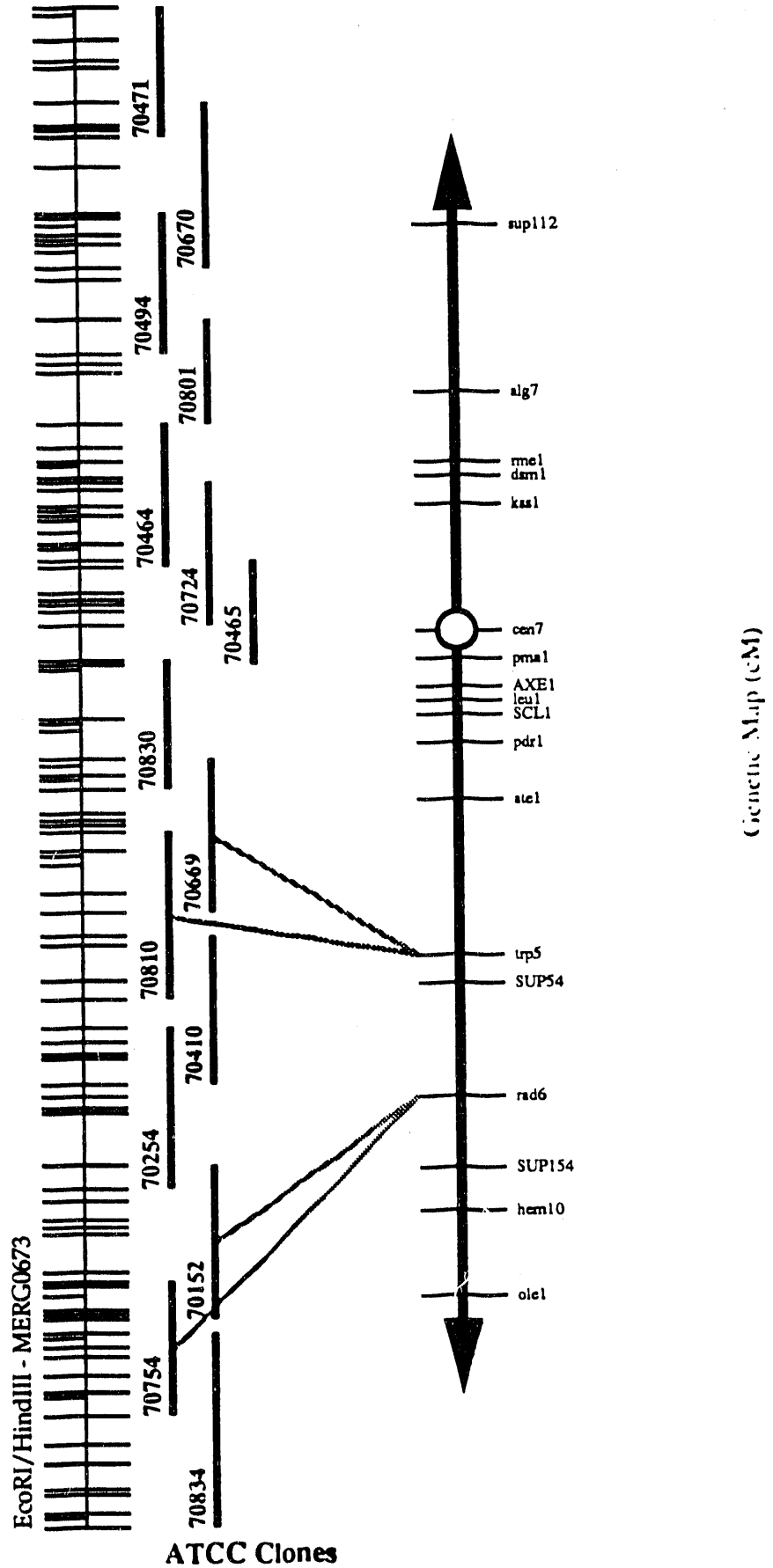


# Chromosome: VII Diagram 2 of 6



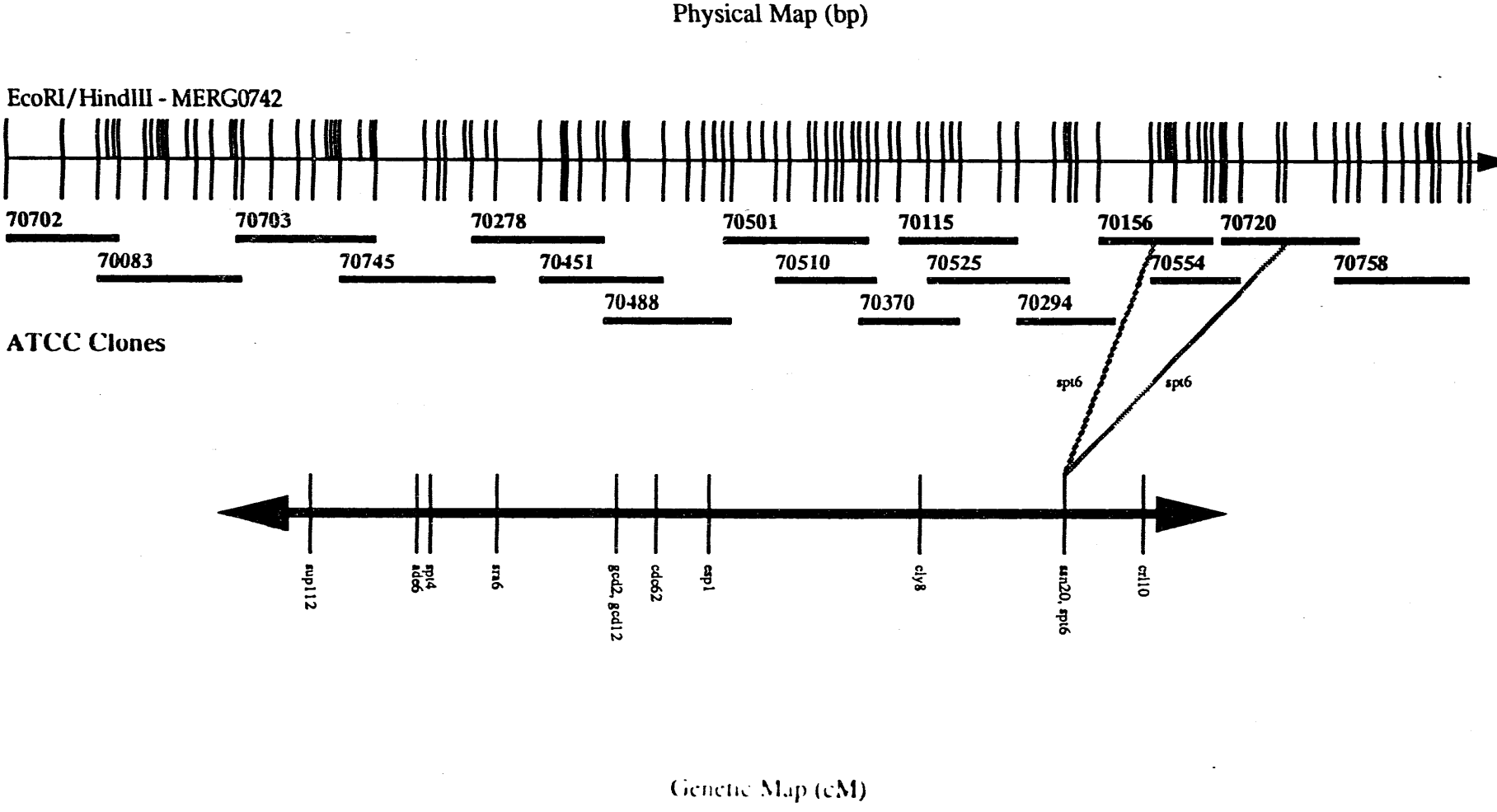
**CLONES FROM THE *SACCHAROMYCES CEREVISIAE* GENOME**

# Chromosome: VII Diagram 3 of 6

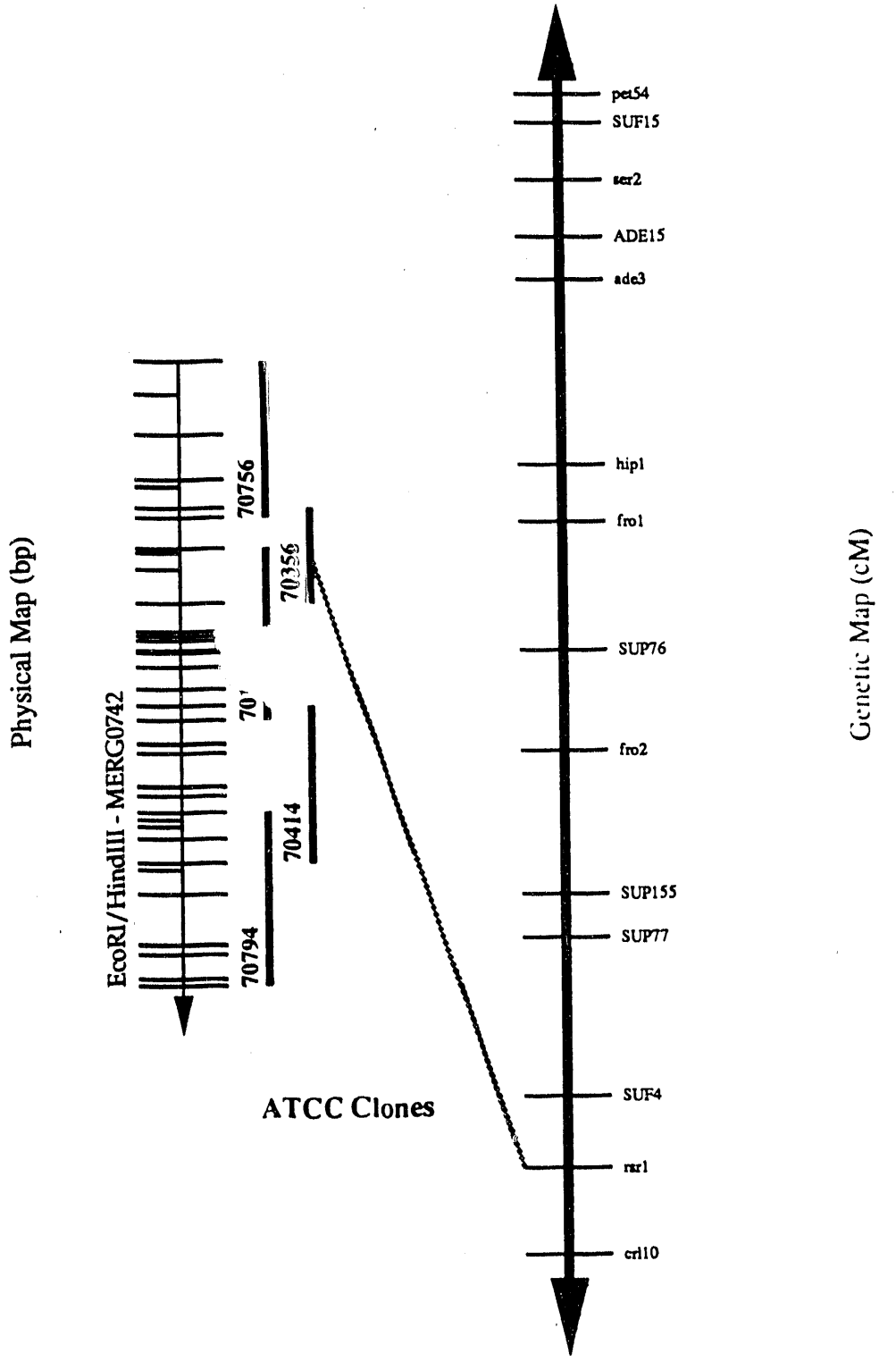


# Chromosome: VII Diagram 4 of 6



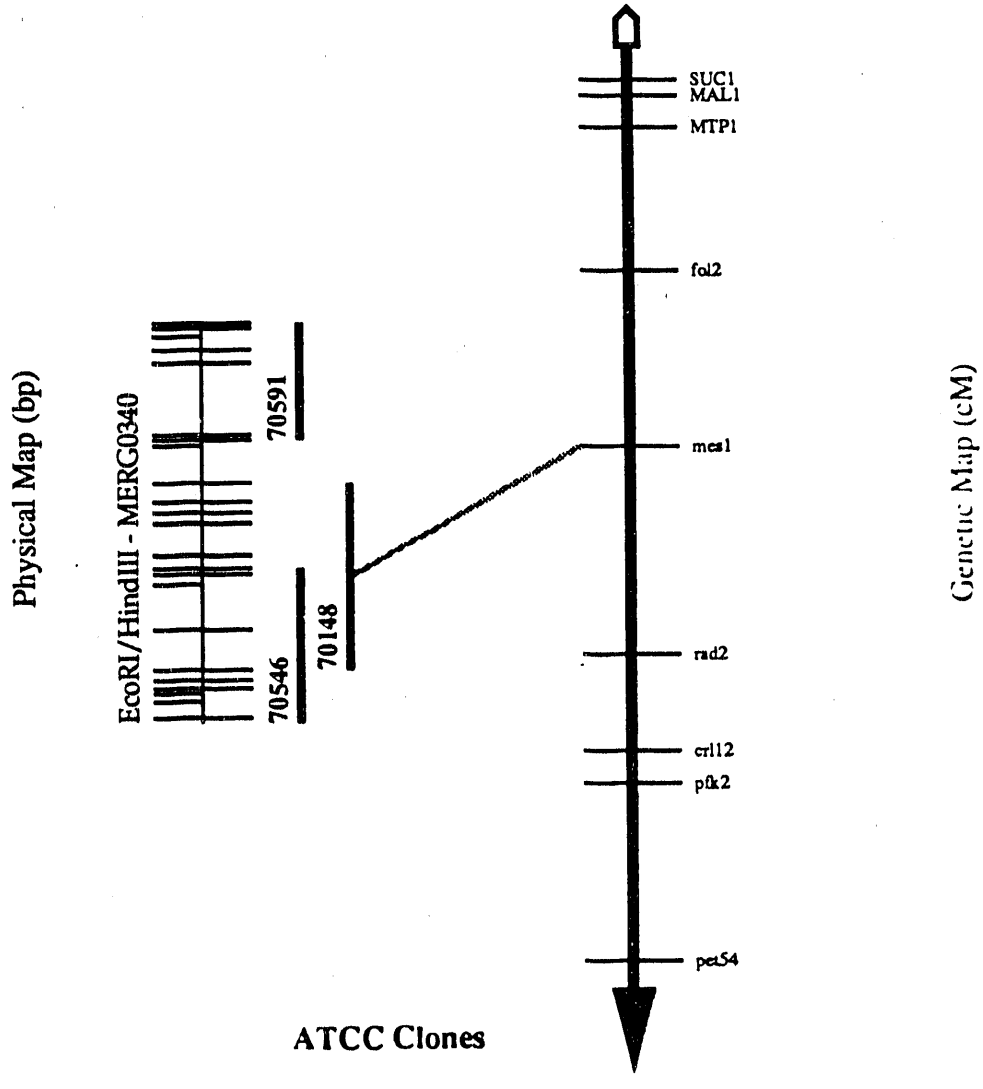


# Chromosome: VII Diagram 5 of 6



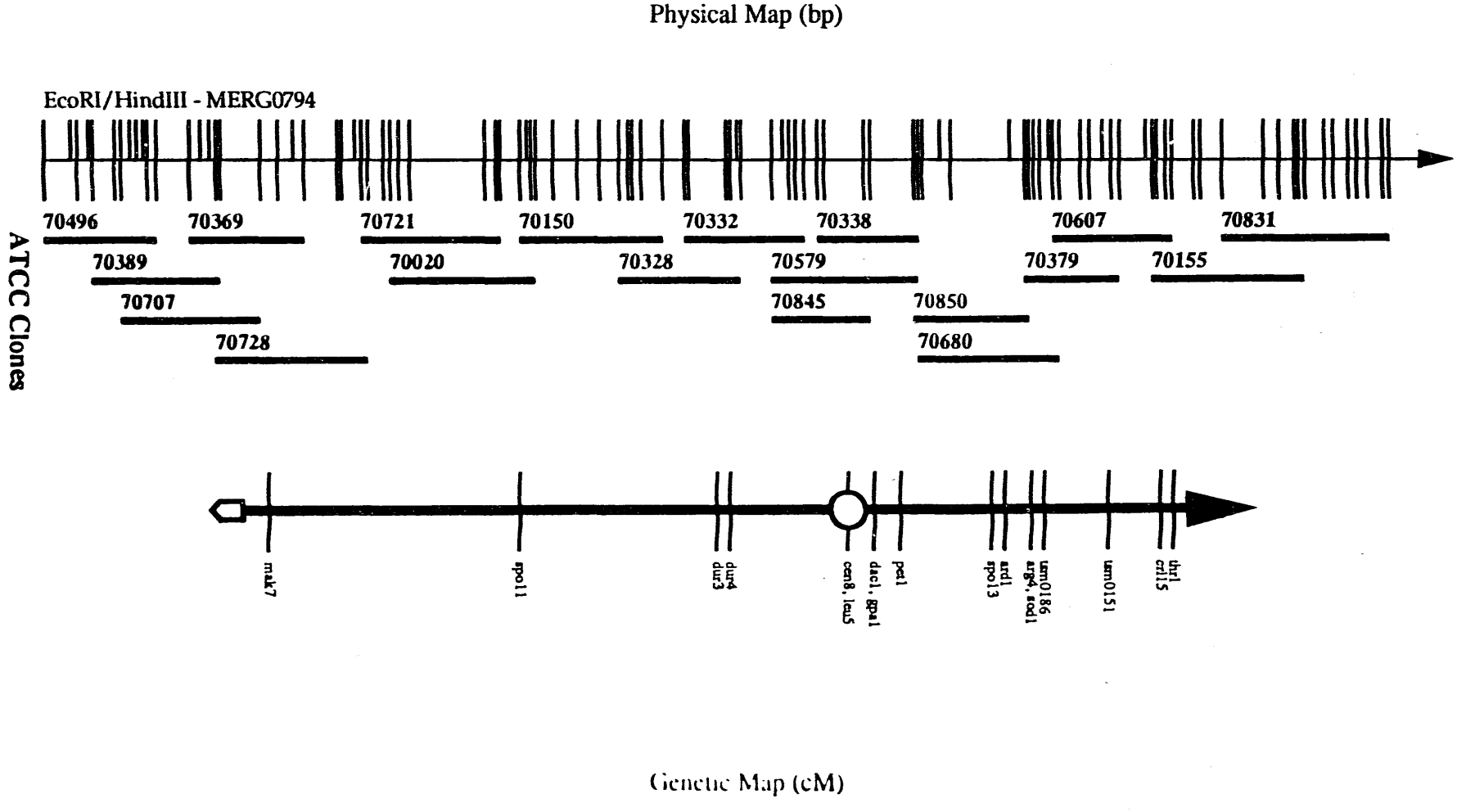
**CLONES FROM THE *SACCHAROMYCES CEREVISIAE* GENOME**

Chromosome: VII Diagram 6 of 6



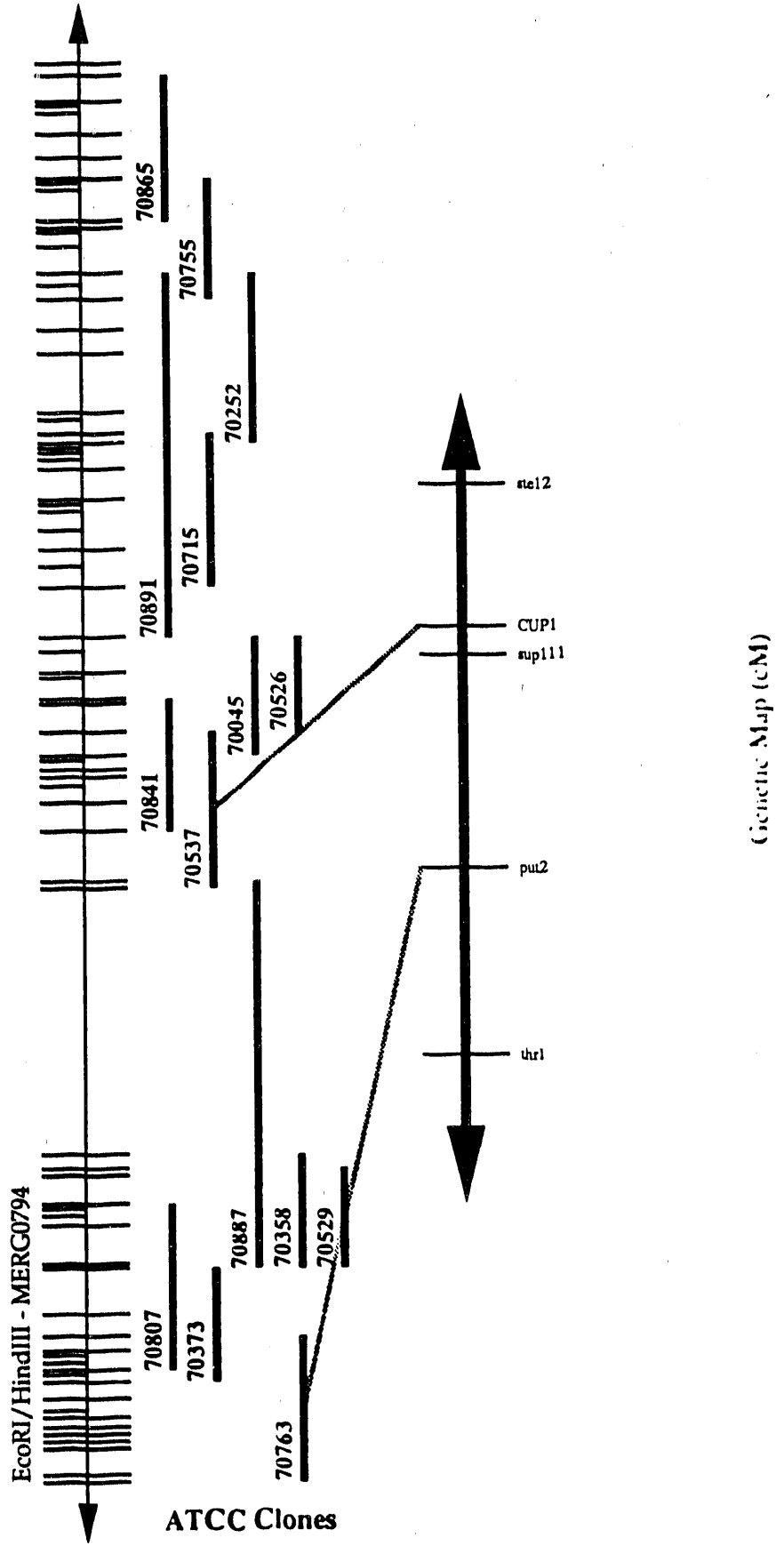
# Chromosome: VIII Diagram 1 of 3





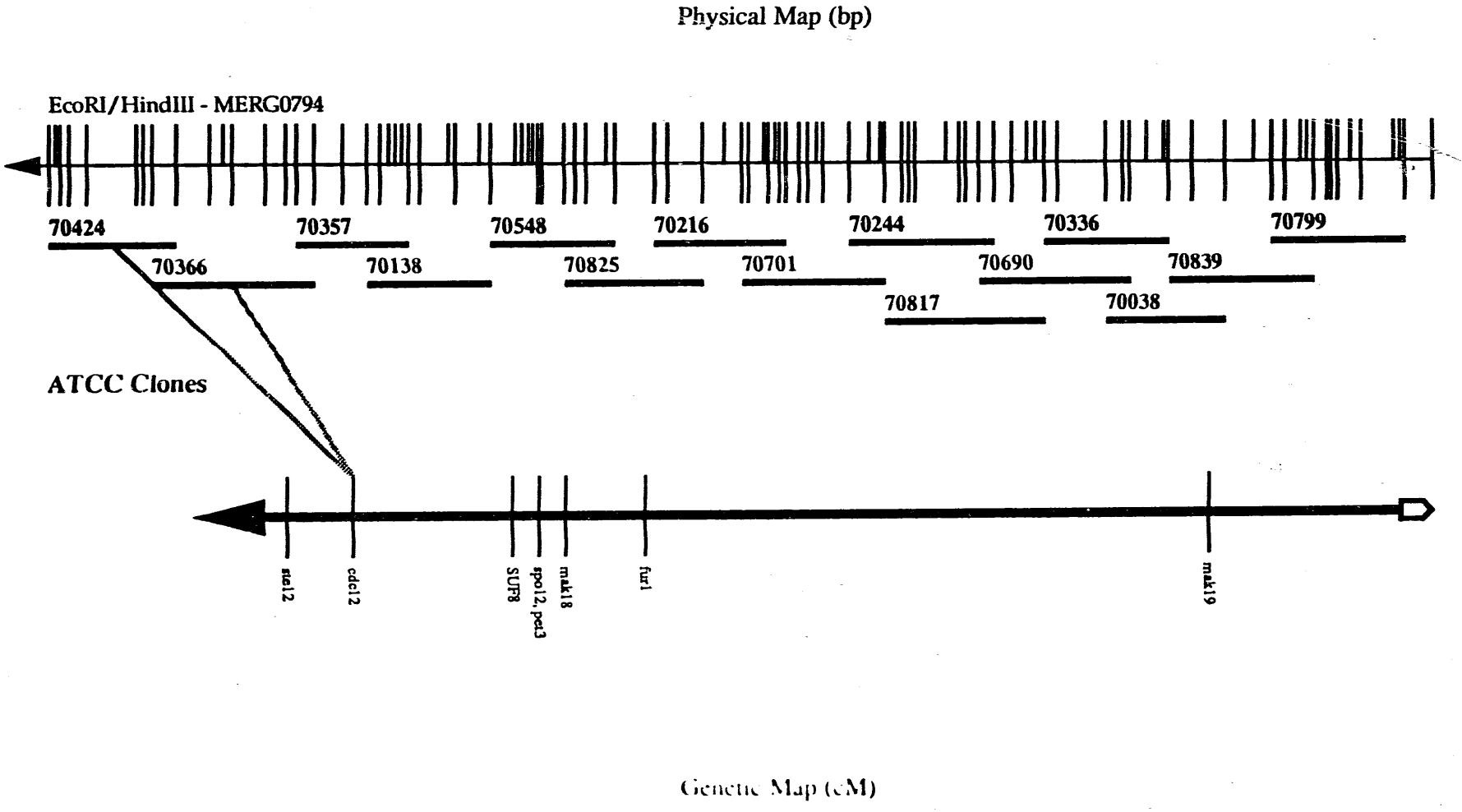
**CLONES FROM THE *SACCHAROMYCES CEREVISIAE* GENOME**

# Chromosome: VIII Diagram 2 of 3



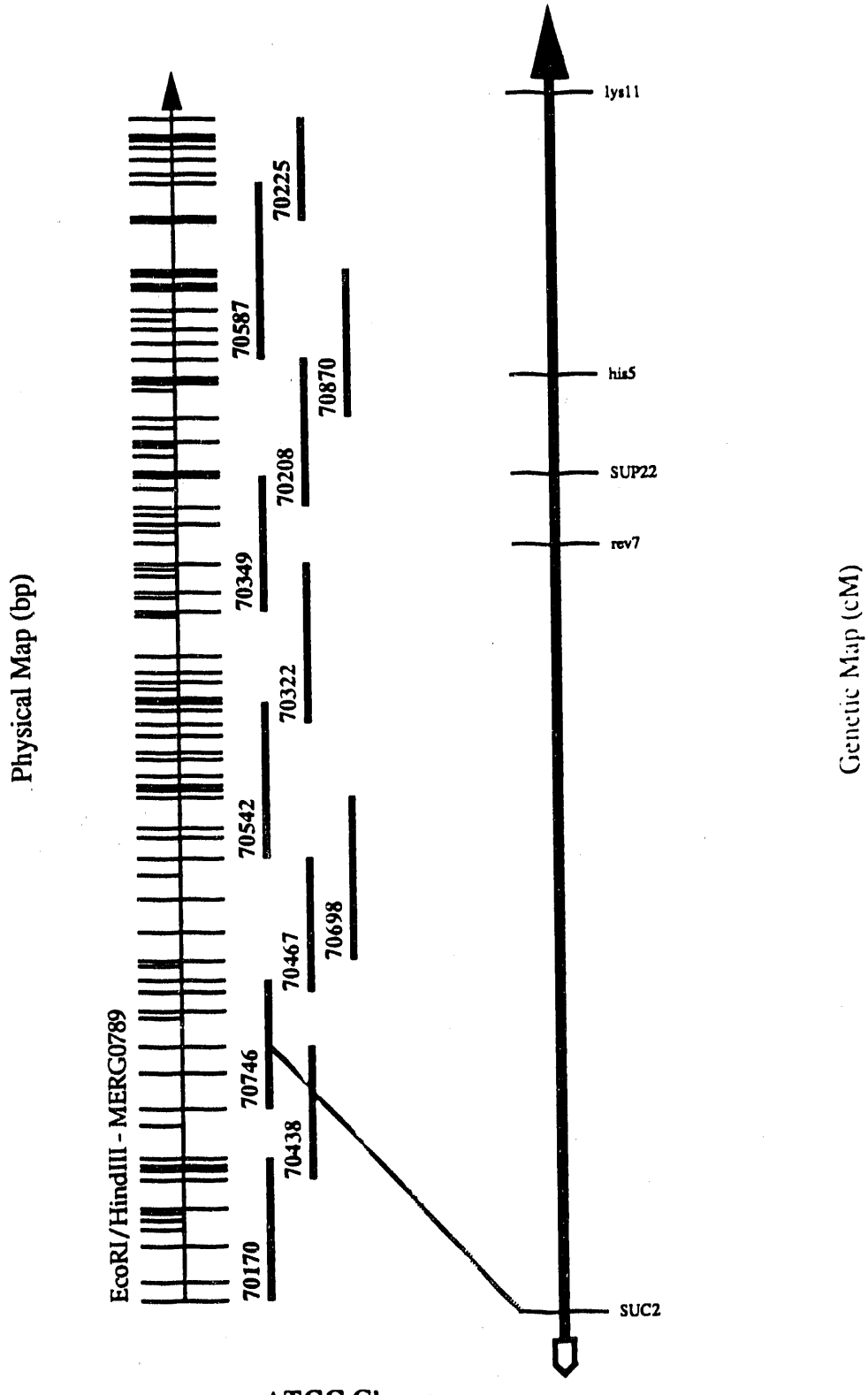
# Chromosome: VIII Diagram 3 of 3





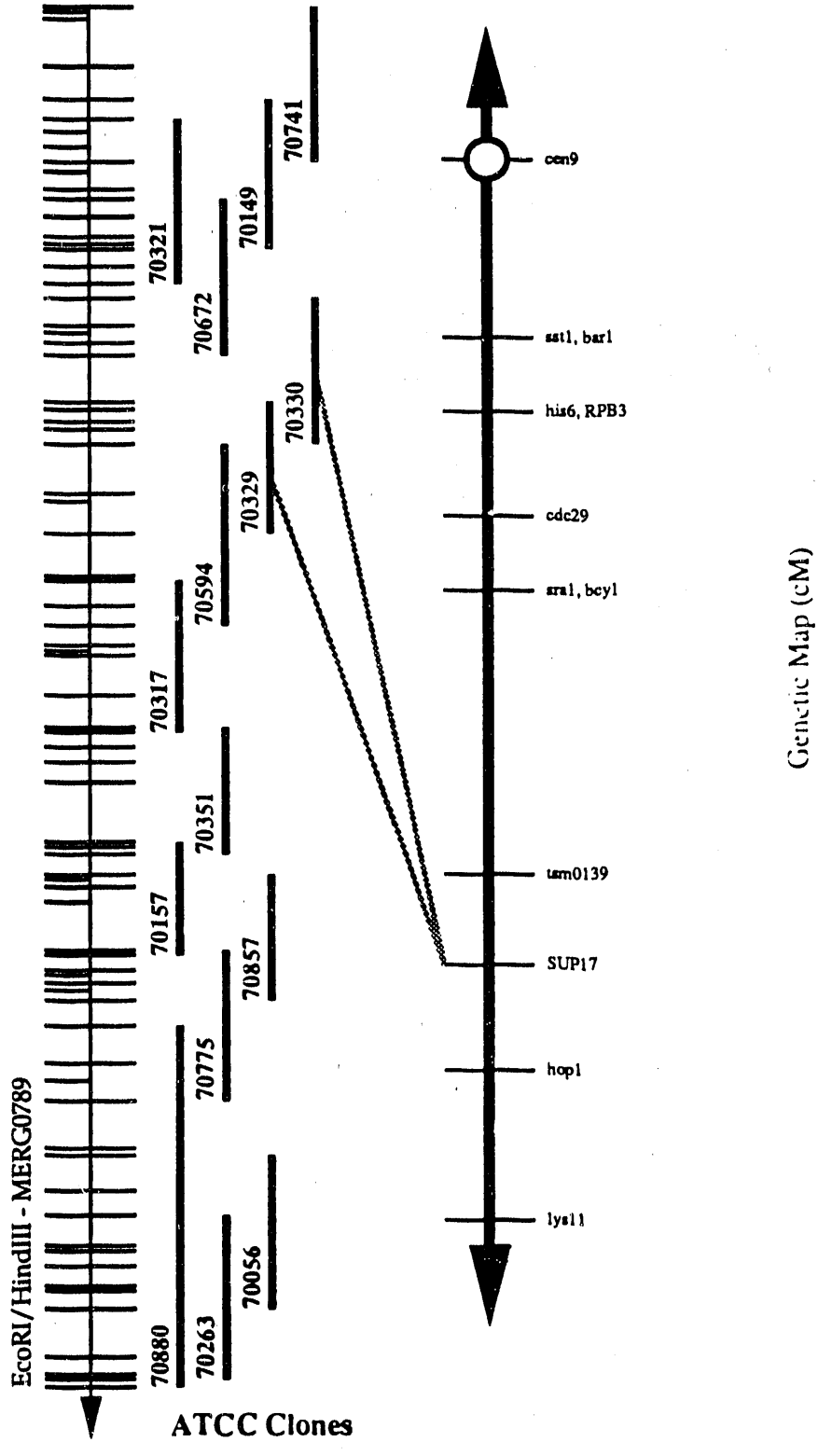
**CLONES FROM THE *SACCHAROMYCES CEREVISIAE* GENOME**

# Chromosome: IX Diagram 1 of 3



**CLONES FROM THE *SACCHAROMYCES CEREVISIAE* GENOME**

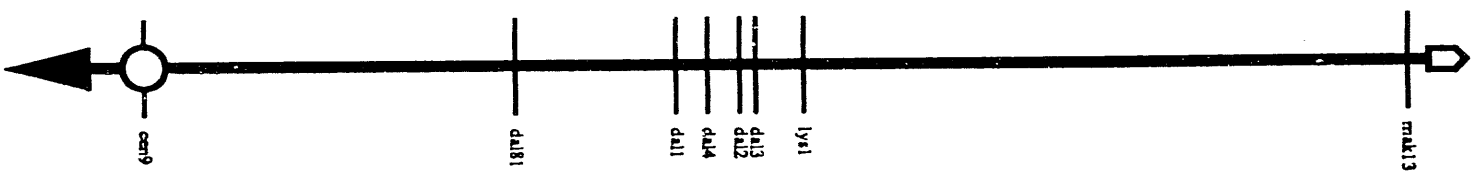
# Chromosome: IX Diagram 2 of 3



# Chromosome: IX Diagram 3 of 3



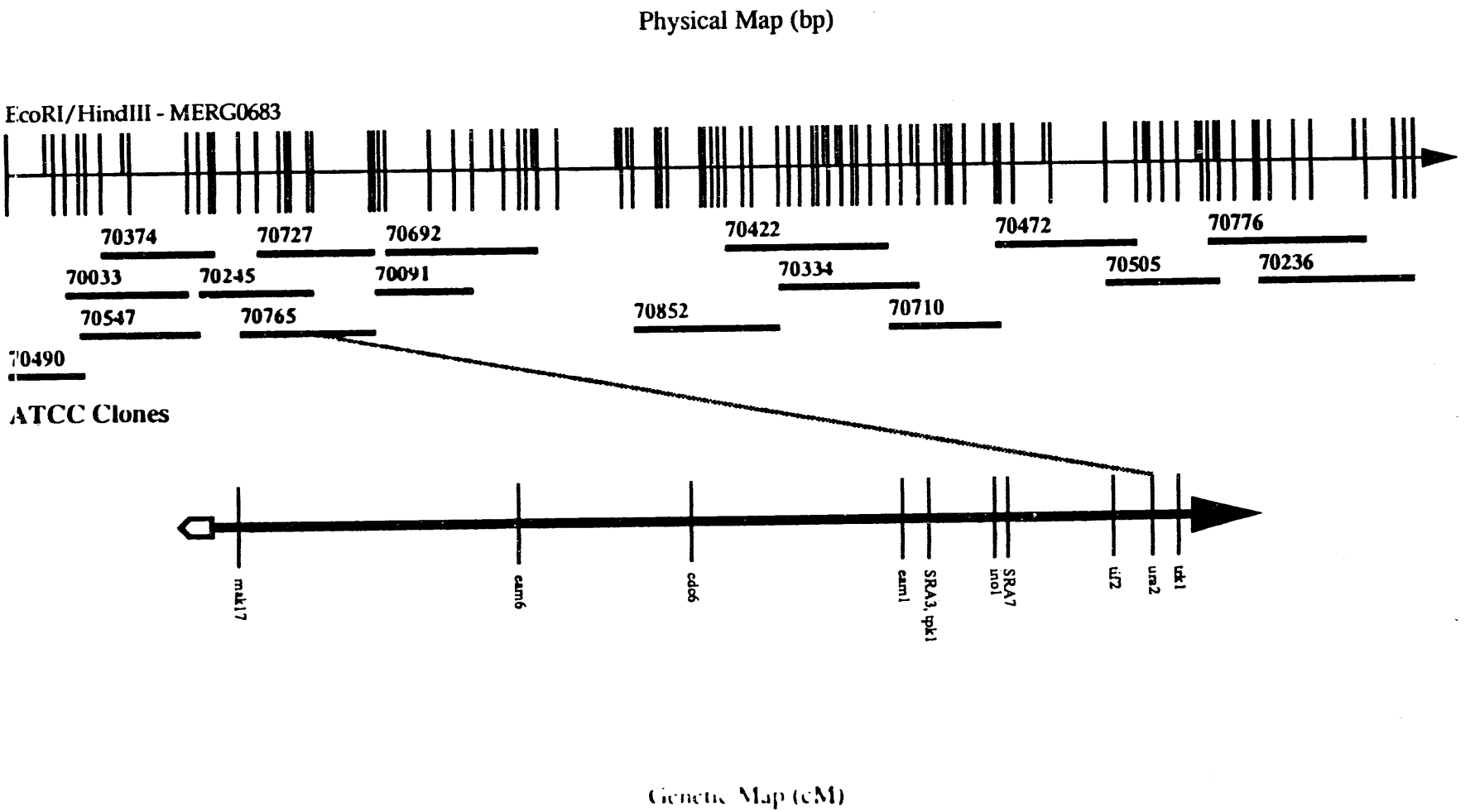
No Physical Map Data Available




Genetic Map (cM)

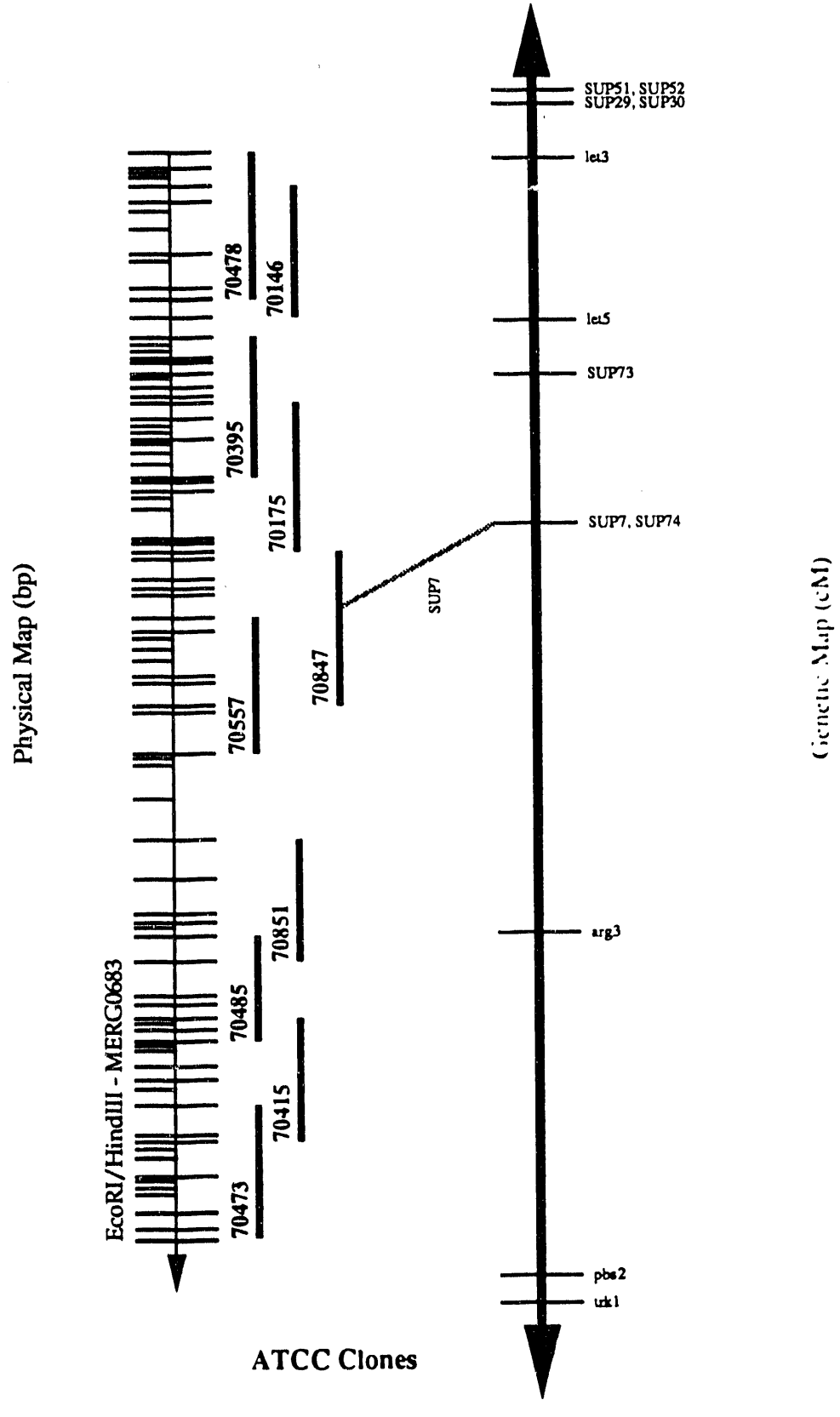
# Chromosome: X Diagram 1 of 4



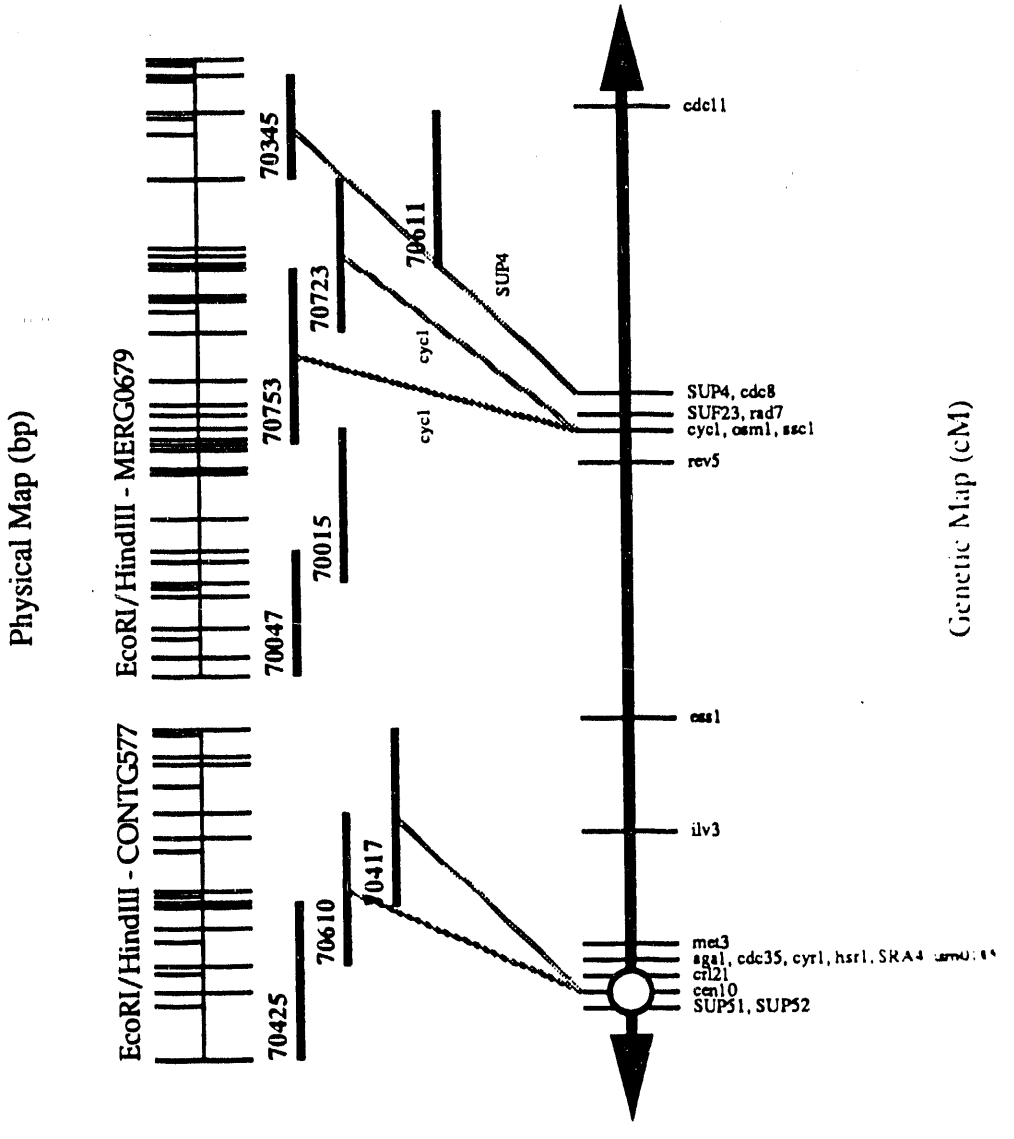


**CLONES FROM THE *SACCHAROMYCES CEREVISIAE* GENOME**

# Chromosome: X Diagram 2 of 4

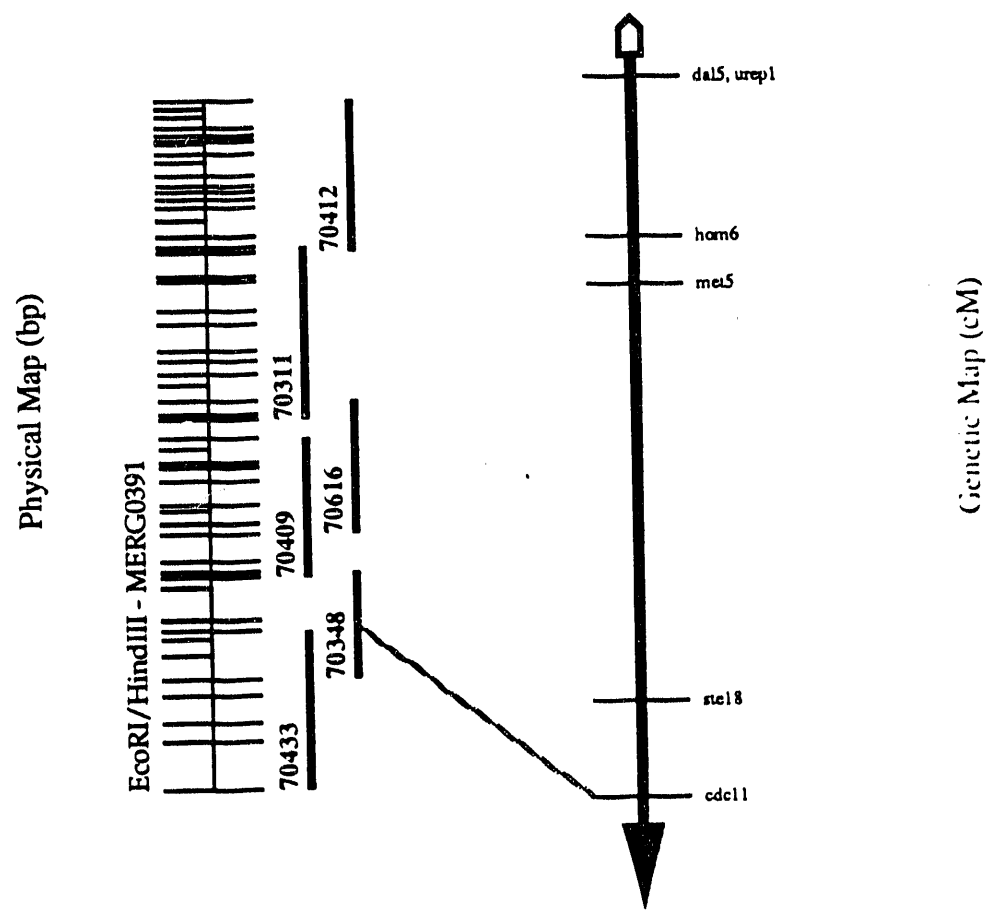


Chromosome: X Diagram 3 of 4



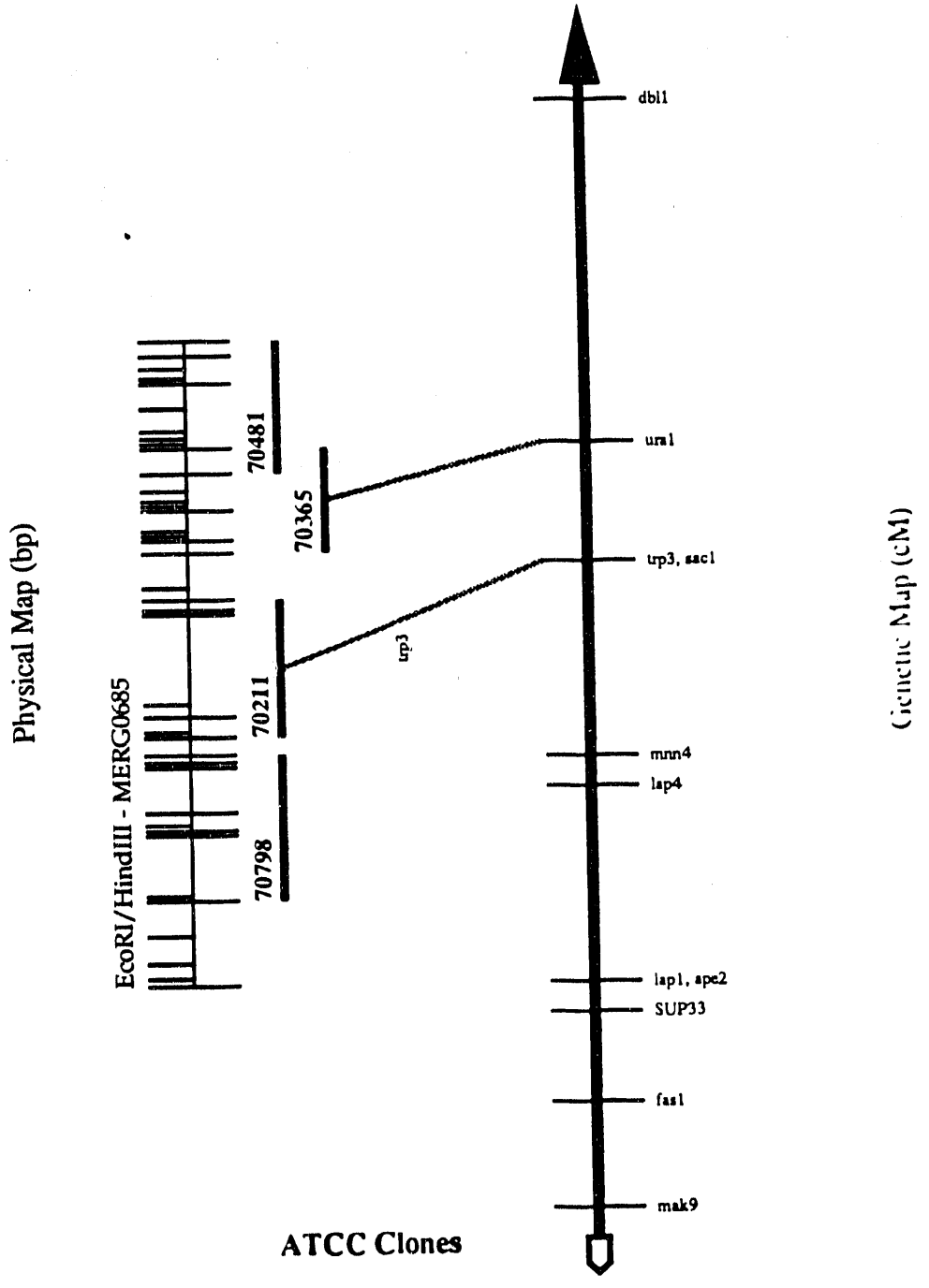
**CLONES FROM THE *SACCHAROMYCES CEREVISIAE* GENOME**

Chromosome: X Diagram 4 of 4




**ATCC Clones**

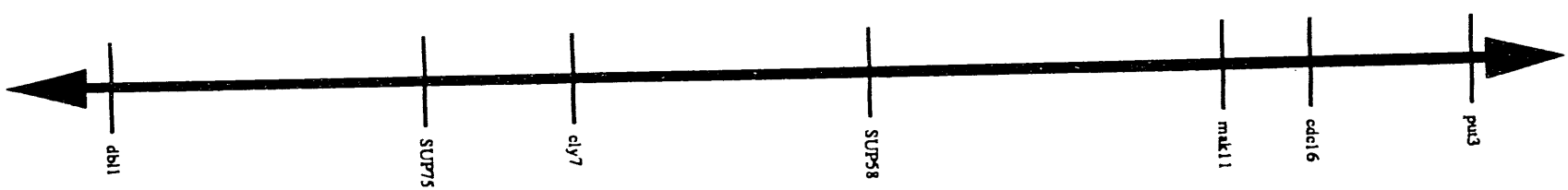
# Chromosome: XI Diagram 1 of 4



# Chromosome: XI

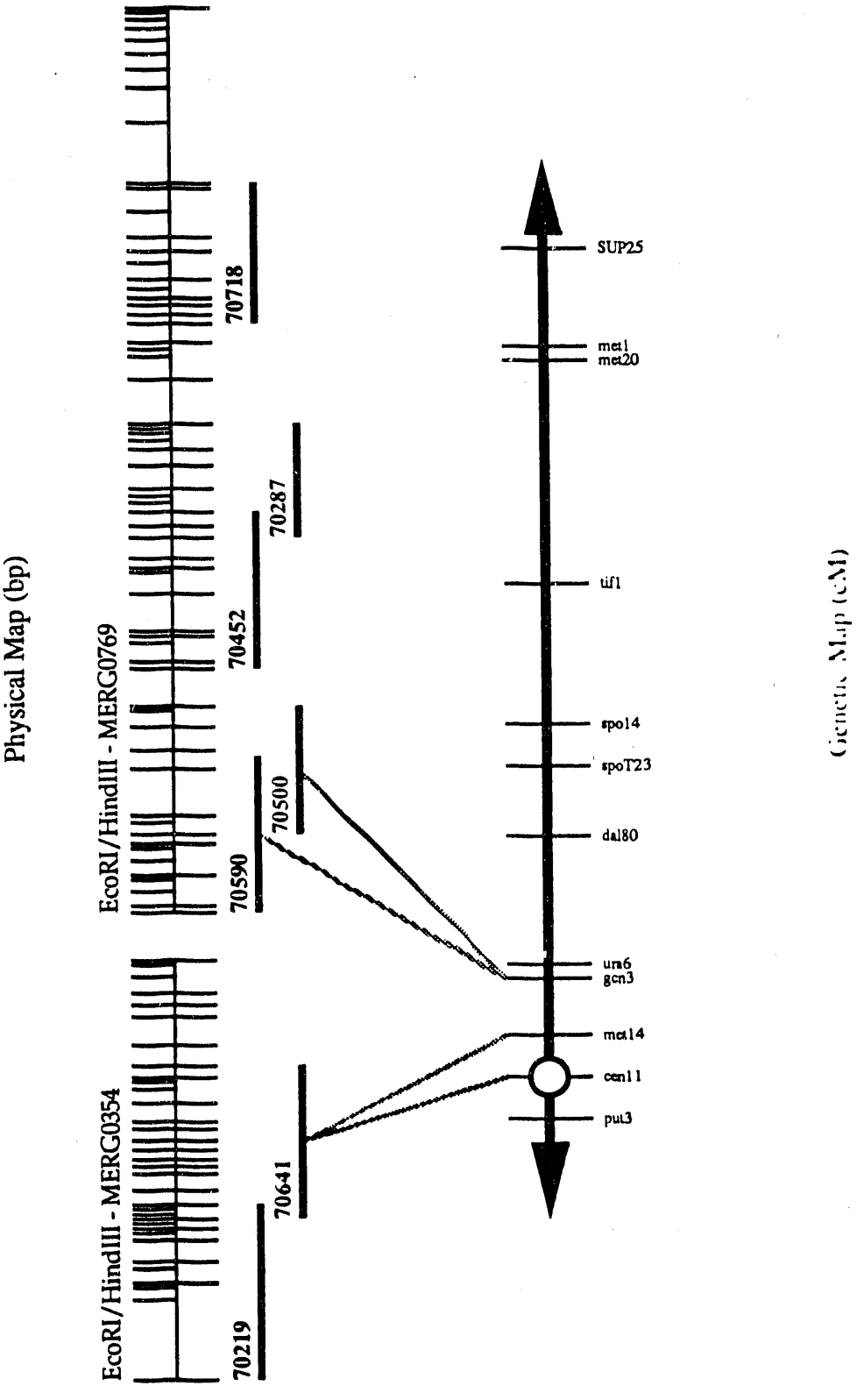
Diagram 2 of 4

No Physical Map Data Available



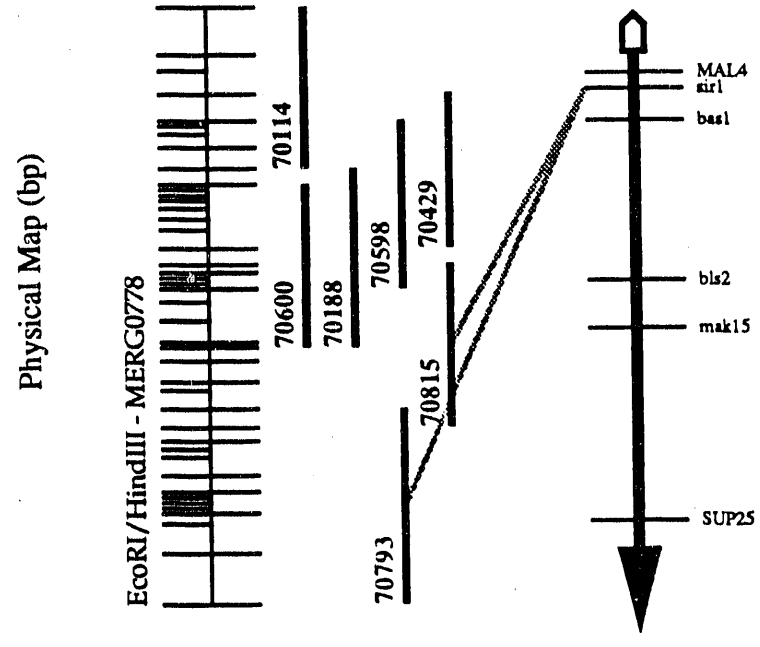

Genetic Map (cM)

# Chromosome: XI Diagram 3 of 4



**CLONES FROM THE *SACCHAROMYCES CEREVISIAE* GENOME**

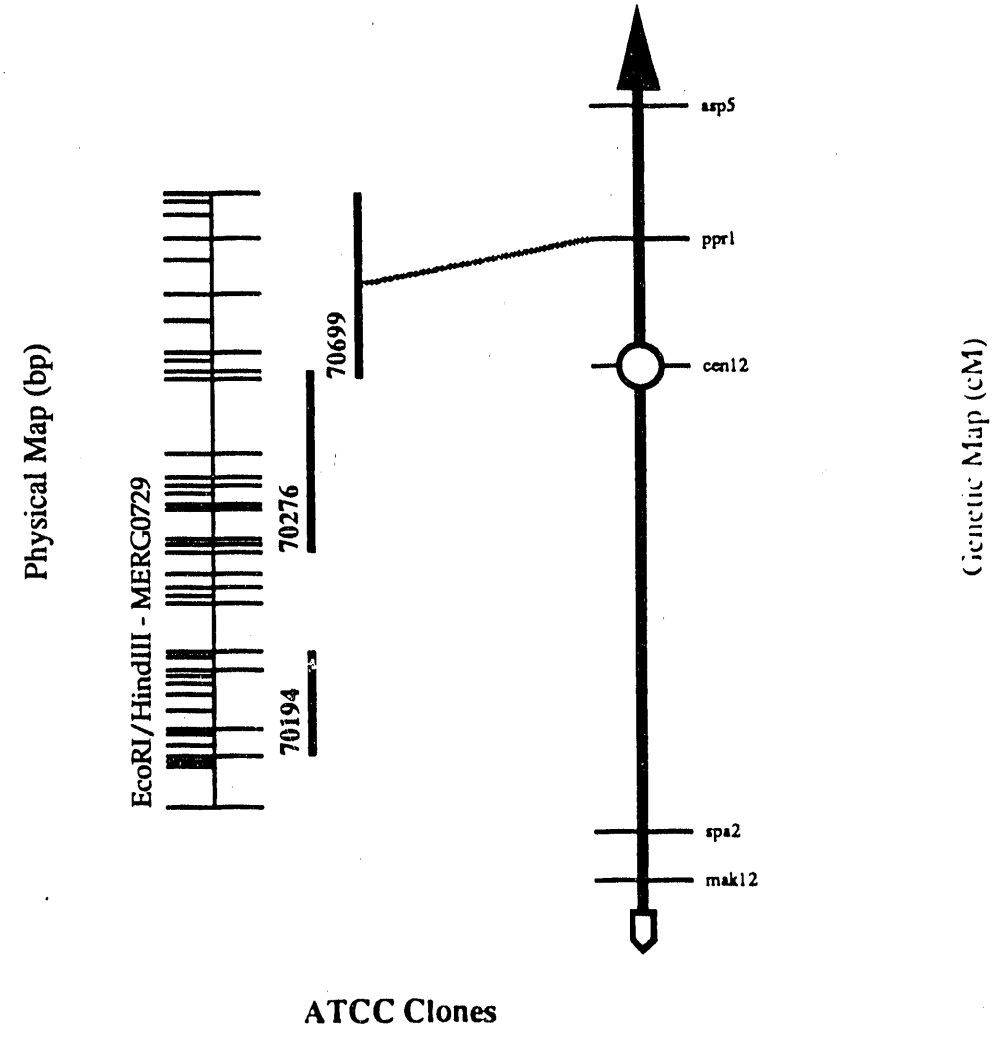
Chromosome: XI Diagram 4 of 4



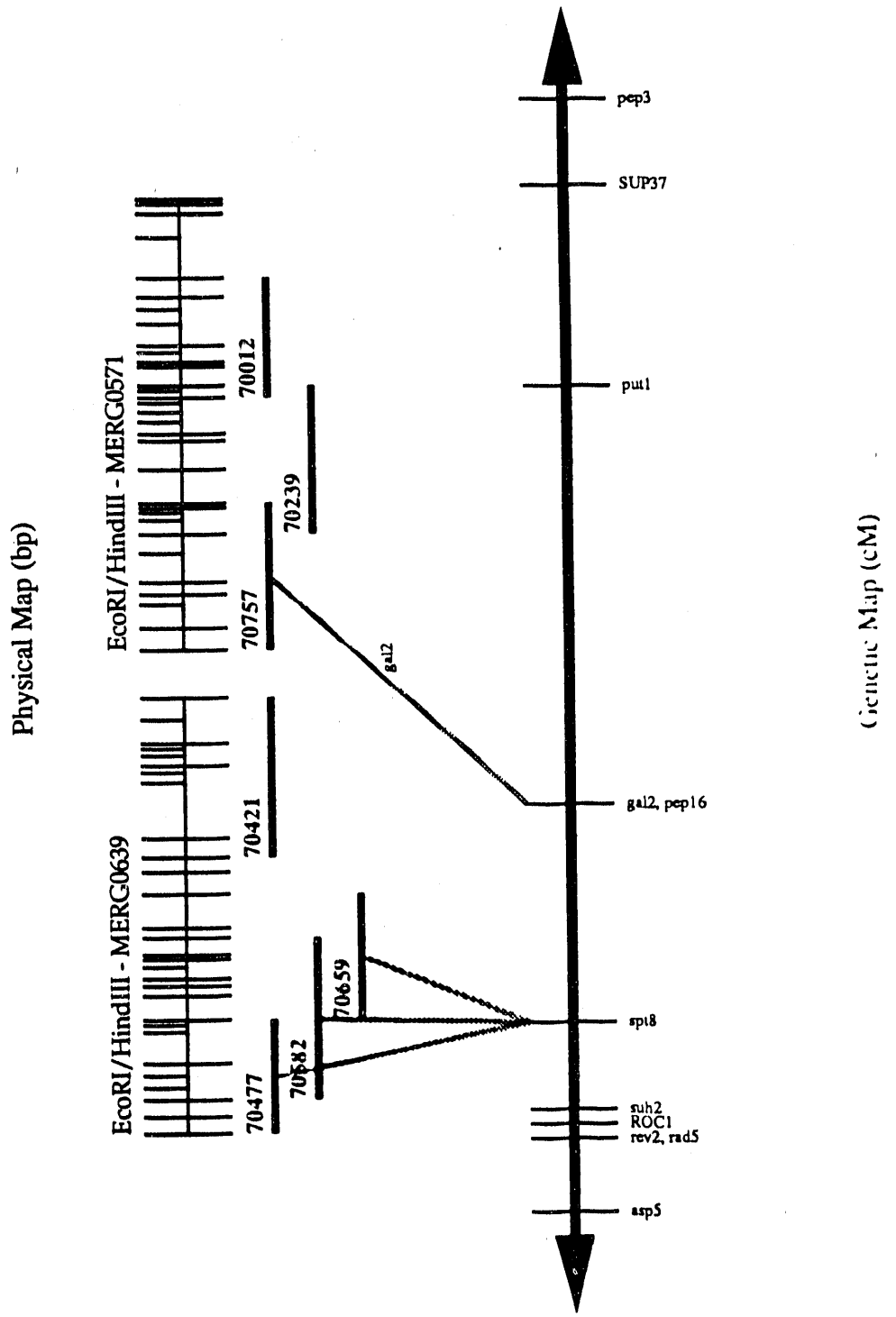

**ATCC Clones**

**CLONES FROM THE *SACCHAROMYCES CEREVISIAE* GENOME**

Chromosome: XII Diagram 1 of 5

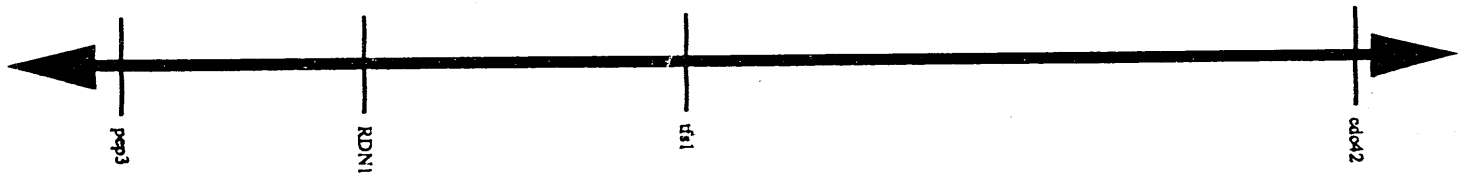


# Chromosome: XII Diagram 2 of 5



# Chromosome: XII Diagram 3 of 5

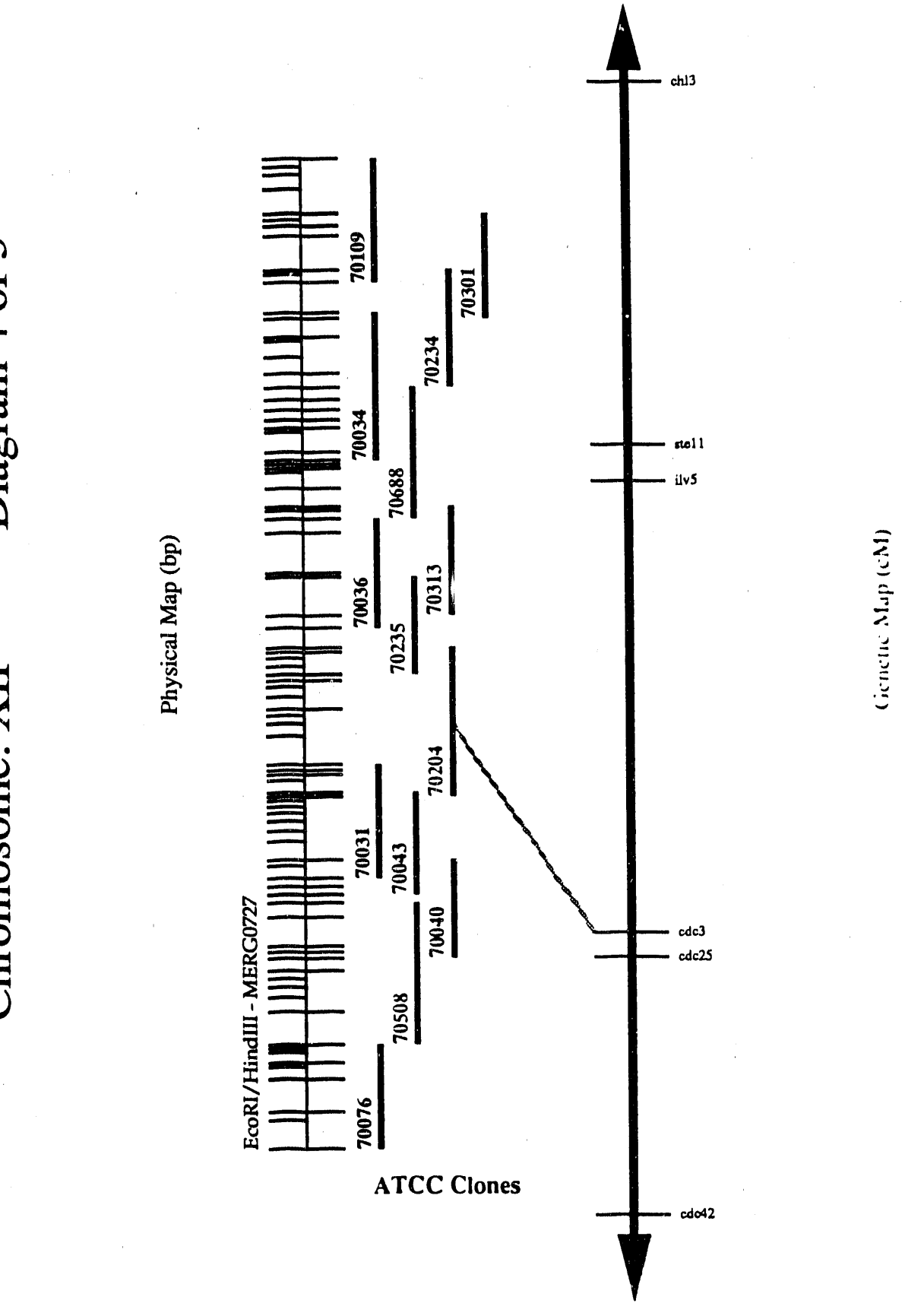
## No Physical Map Data Available




Genetic Map (cM)

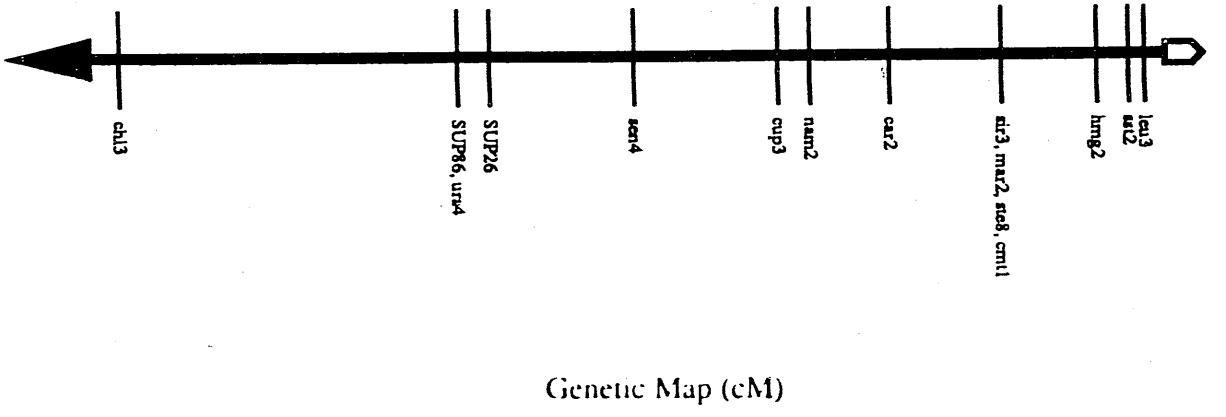
**CLONES FROM THE *SACCHAROMYCES CEREVISIAE* GENOME**

# Chromosome: XII Diagram 4 of 5



# Chromosome: XII Diagram 5 of 5

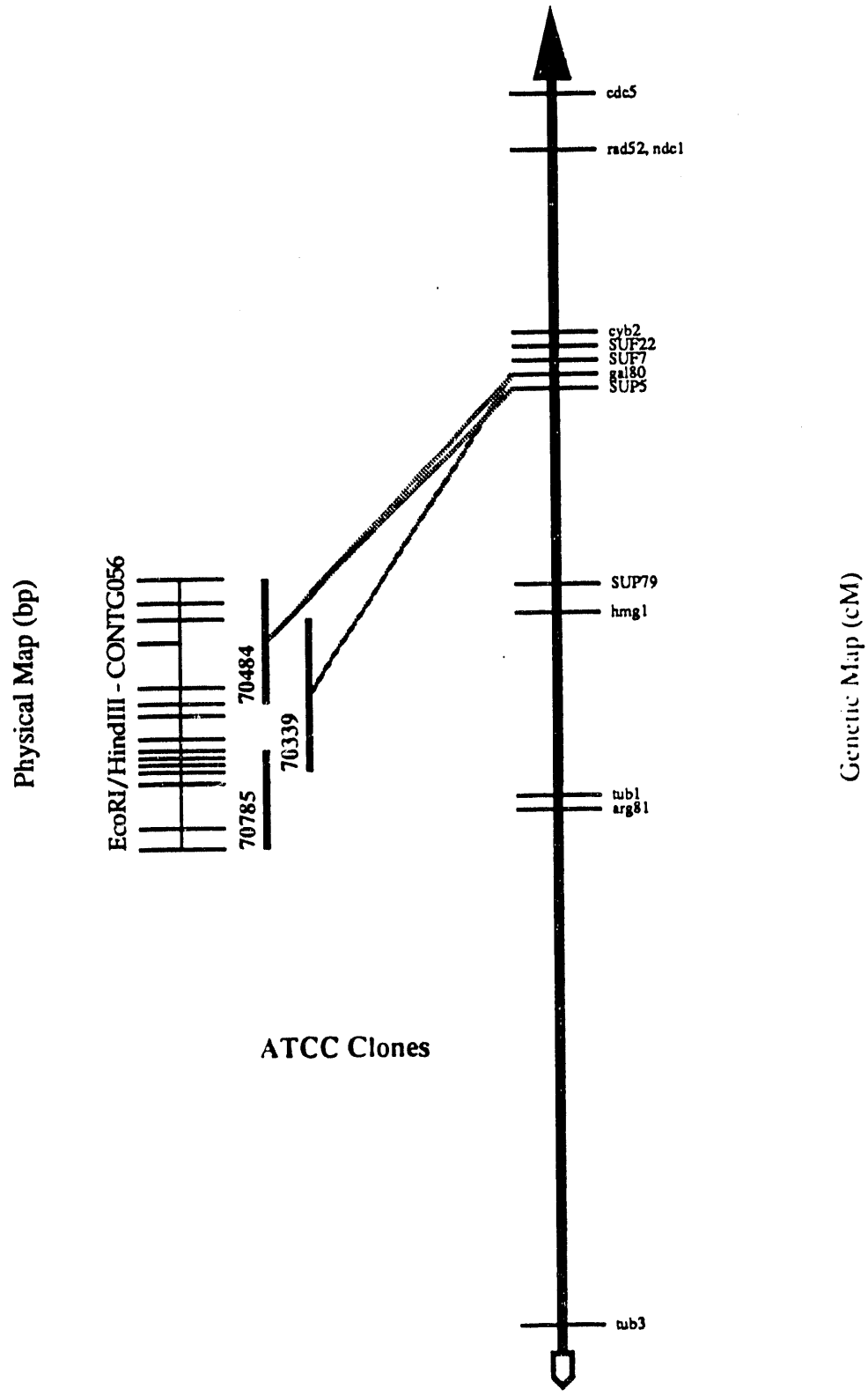
## No Physical Map Data Available




Genetic Map (cM)

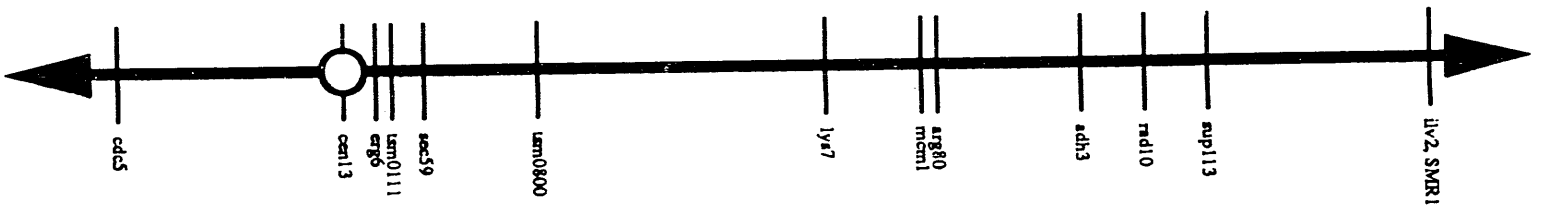
**CLONES FROM THE *SACCHAROMYCES CEREVISIAE* GENOME**

# Chromosome: XIII Diagram 1 of 4



# Chromosome: XIII Diagram 2 of 4

## No Physical Map Data Available

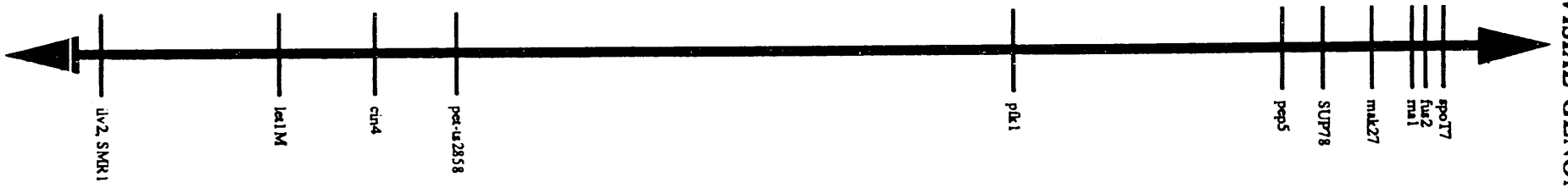



Genetic Map (cM)

# Chromosome: XIII Diagram 3 of 4

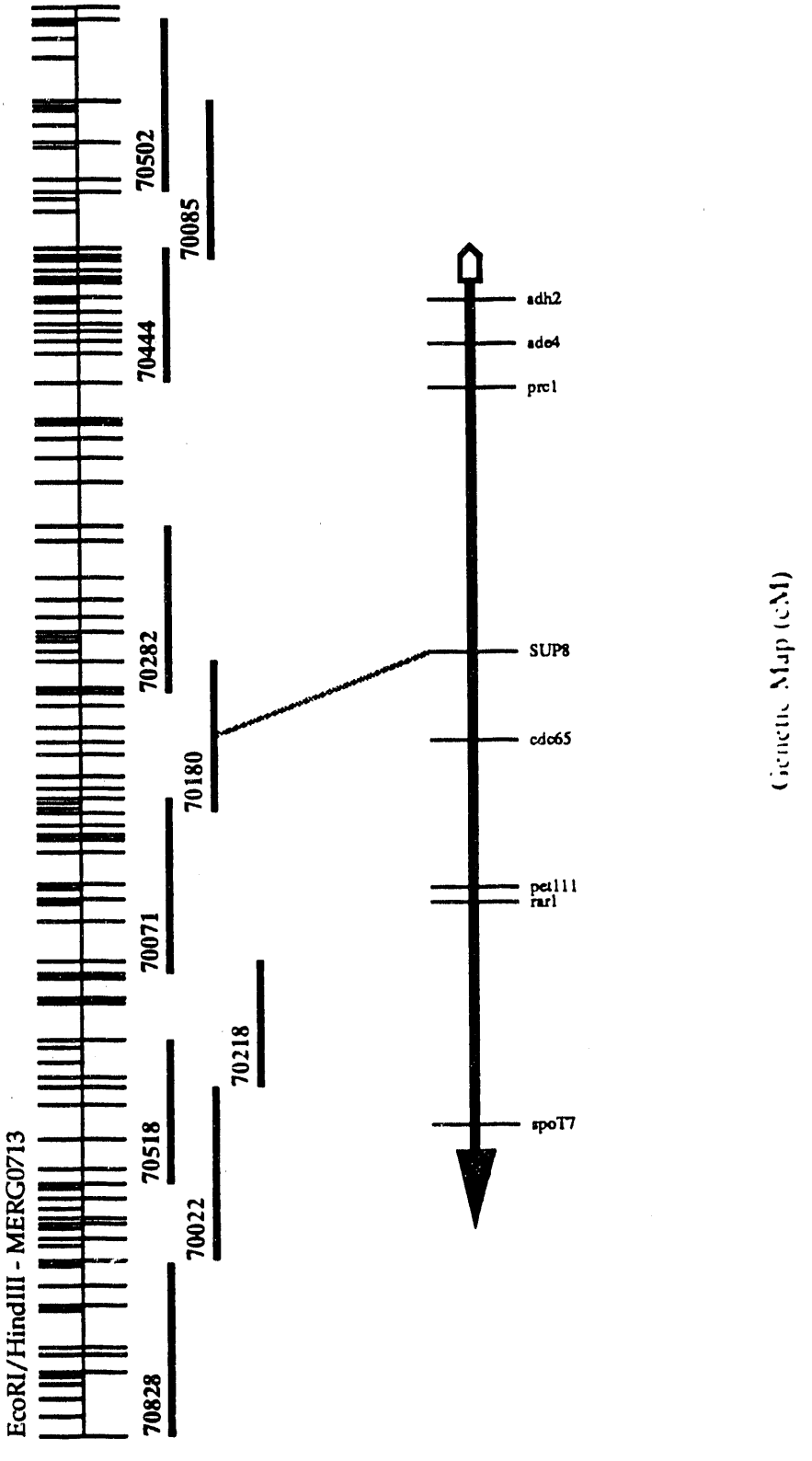
CLONES FROM THE *SACCHAROMYCES CEREVISIAE* GENOME

No Physical Map Data Available



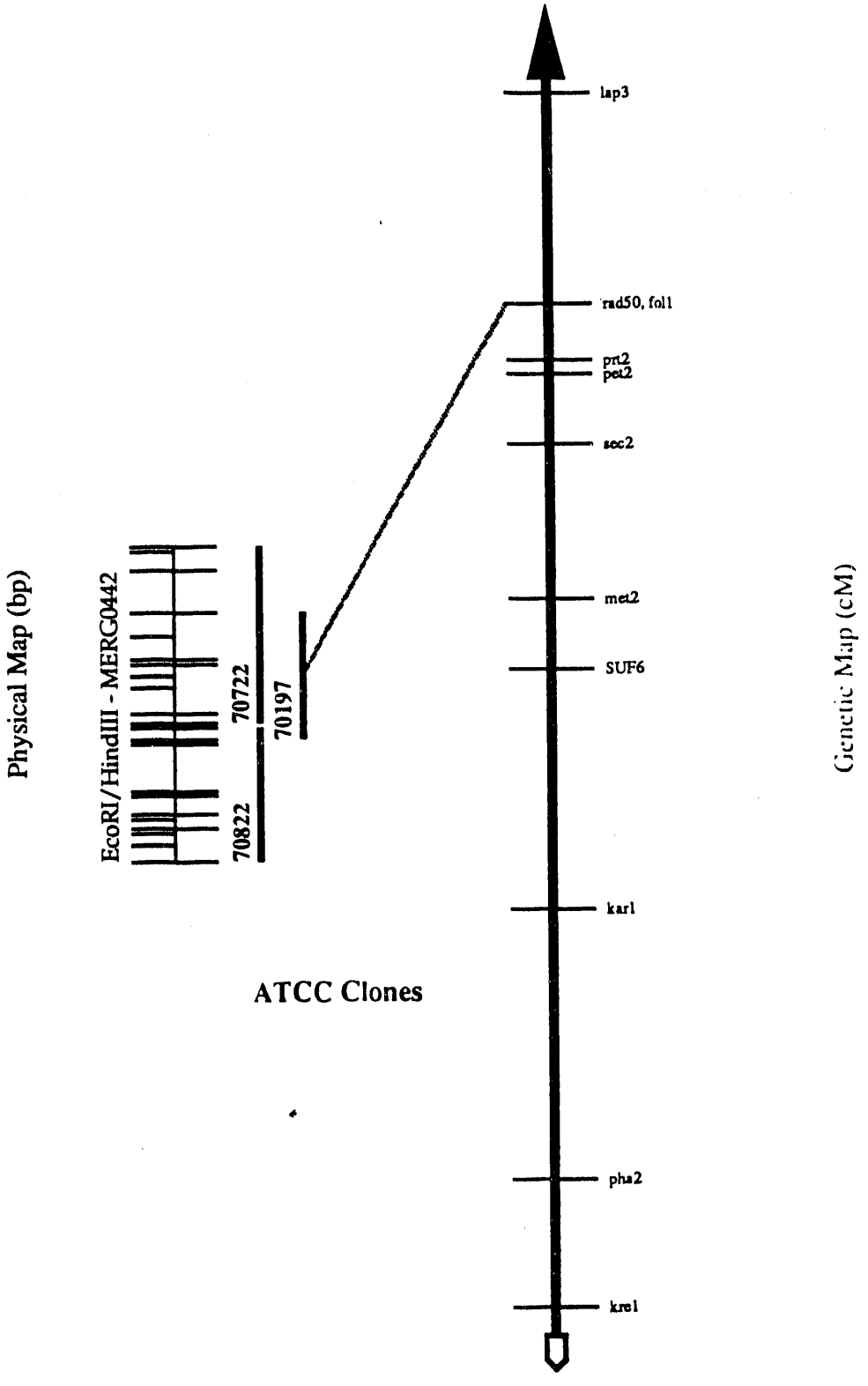

Genetic Map (cM)

# Chromosome: XIII Diagram 4 of 4

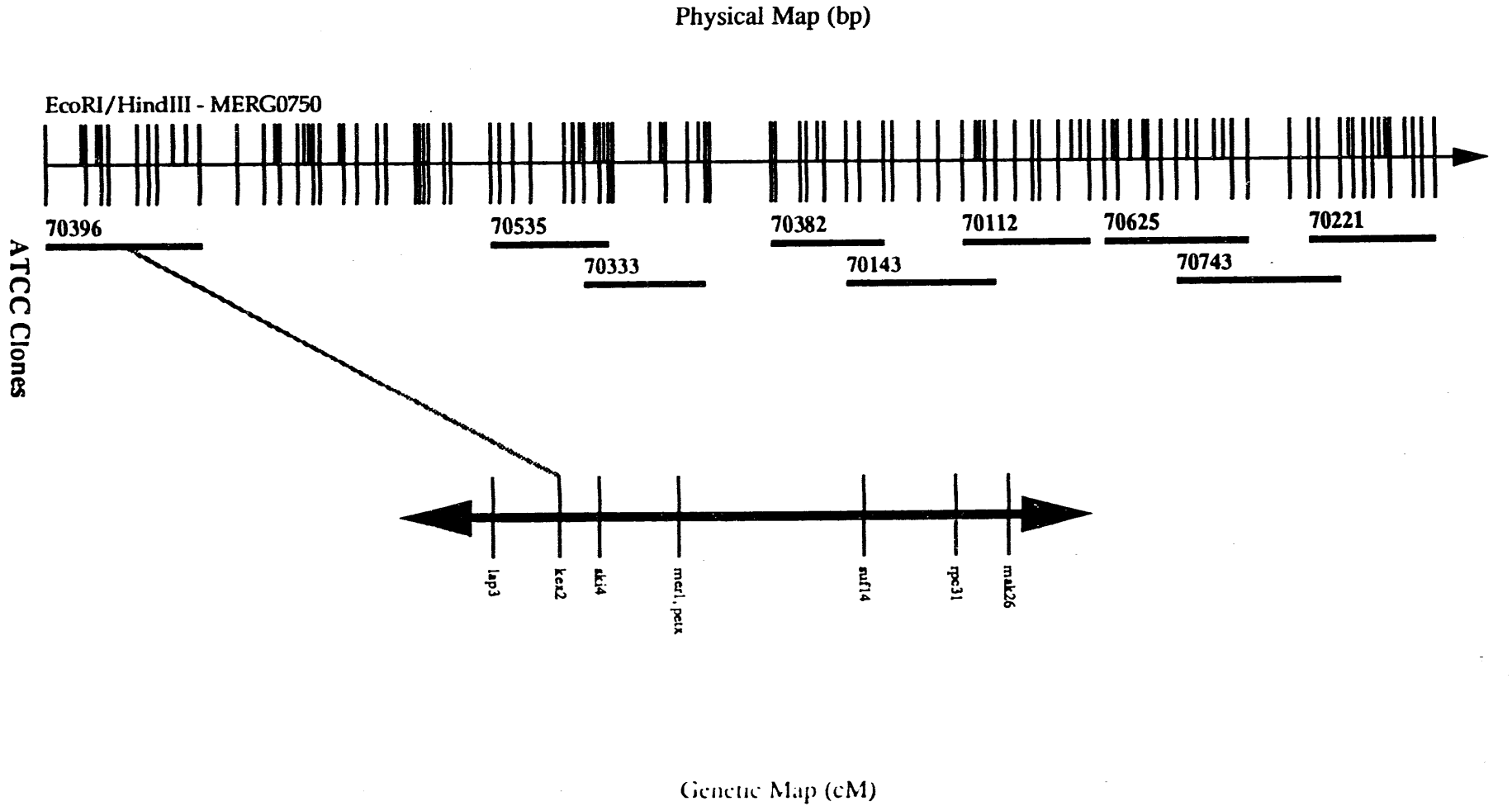


**CLONES FROM THE *SACCHAROMYCES CEREVISIAE* GENOME**

Chromosome: XIV Diagram 1 of 5

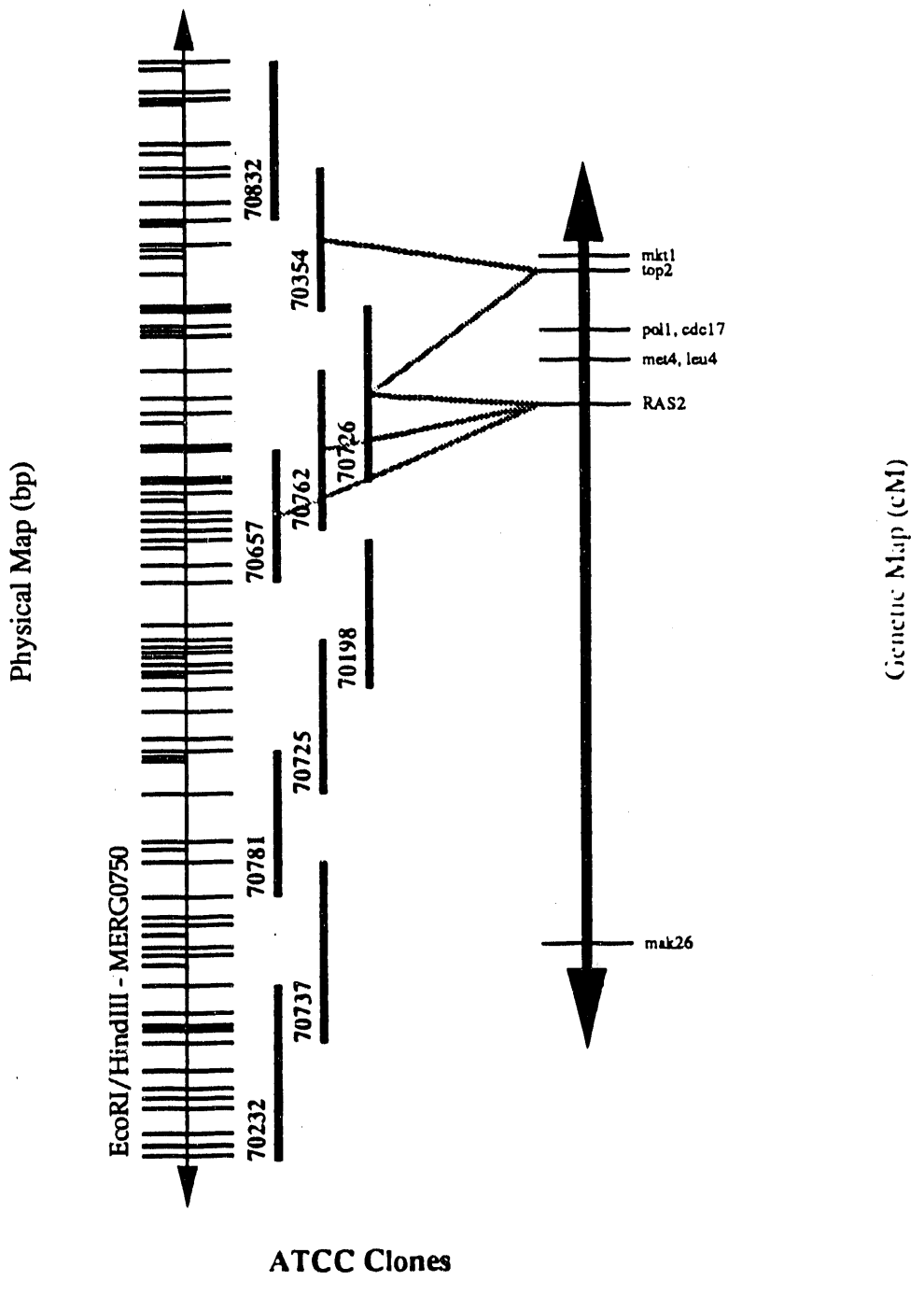


# Chromosome: XIV Diagram 2 of 5



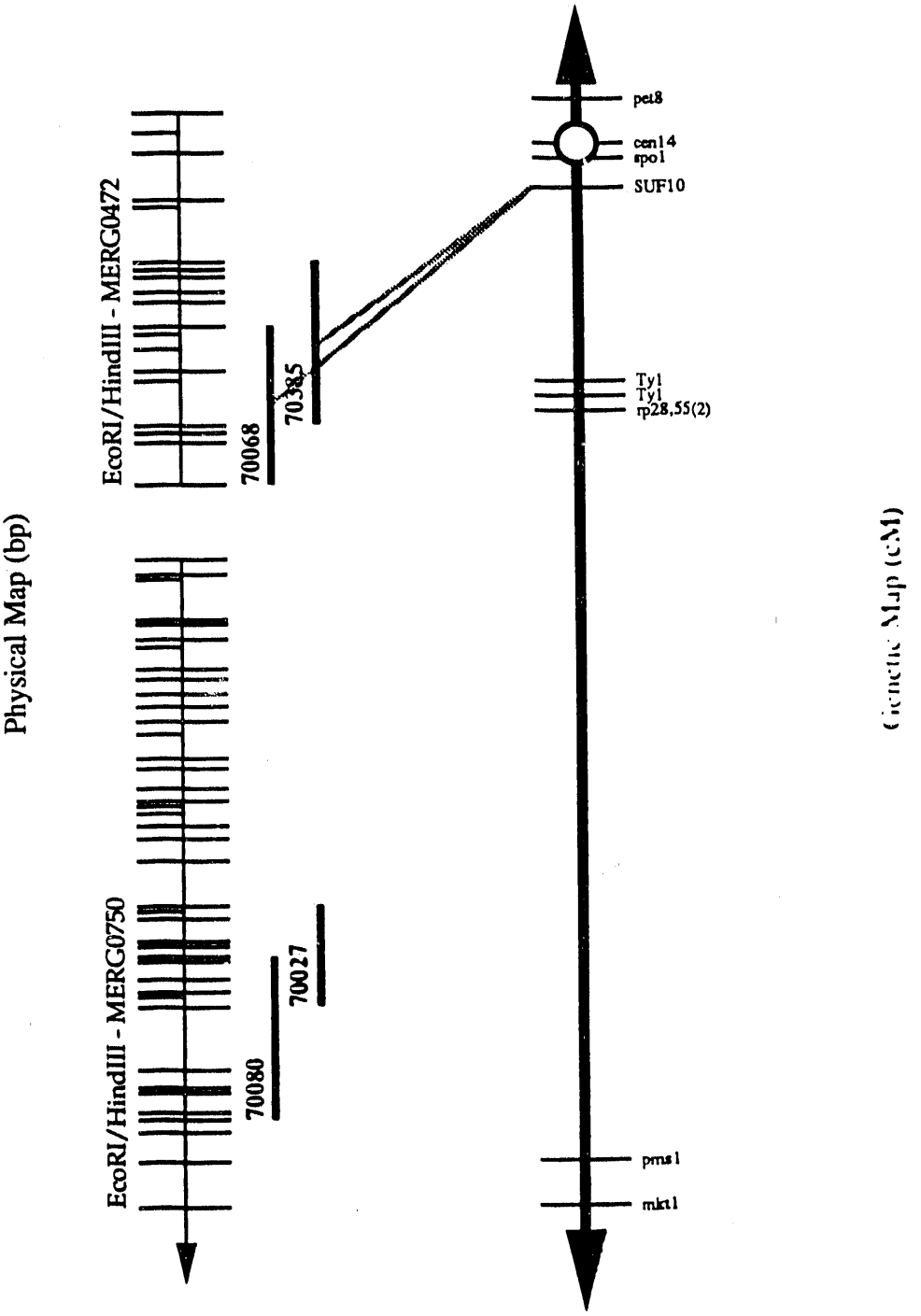

CLONES FROM THE *SACCHAROMYCES CEREVISIAE* GENOME

# Chromosome: XIV Diagram 3 of 5



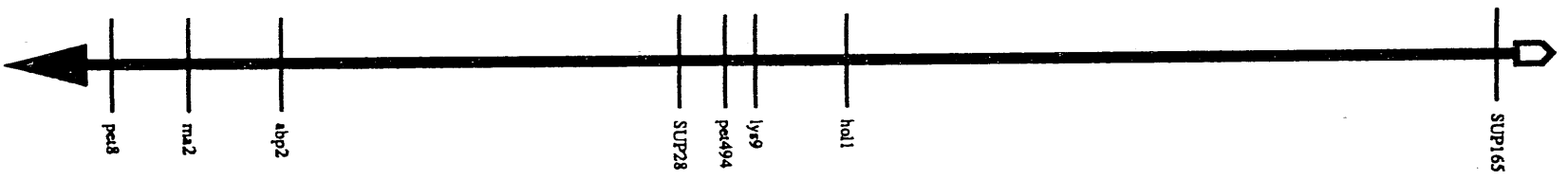
**CLONES FROM THE *SACCHAROMYCES CEREVISIAE* GENOME**

Chromosome: XIV Diagram 4 of 5



# Chromosome: XIV Diagram 5 of 5

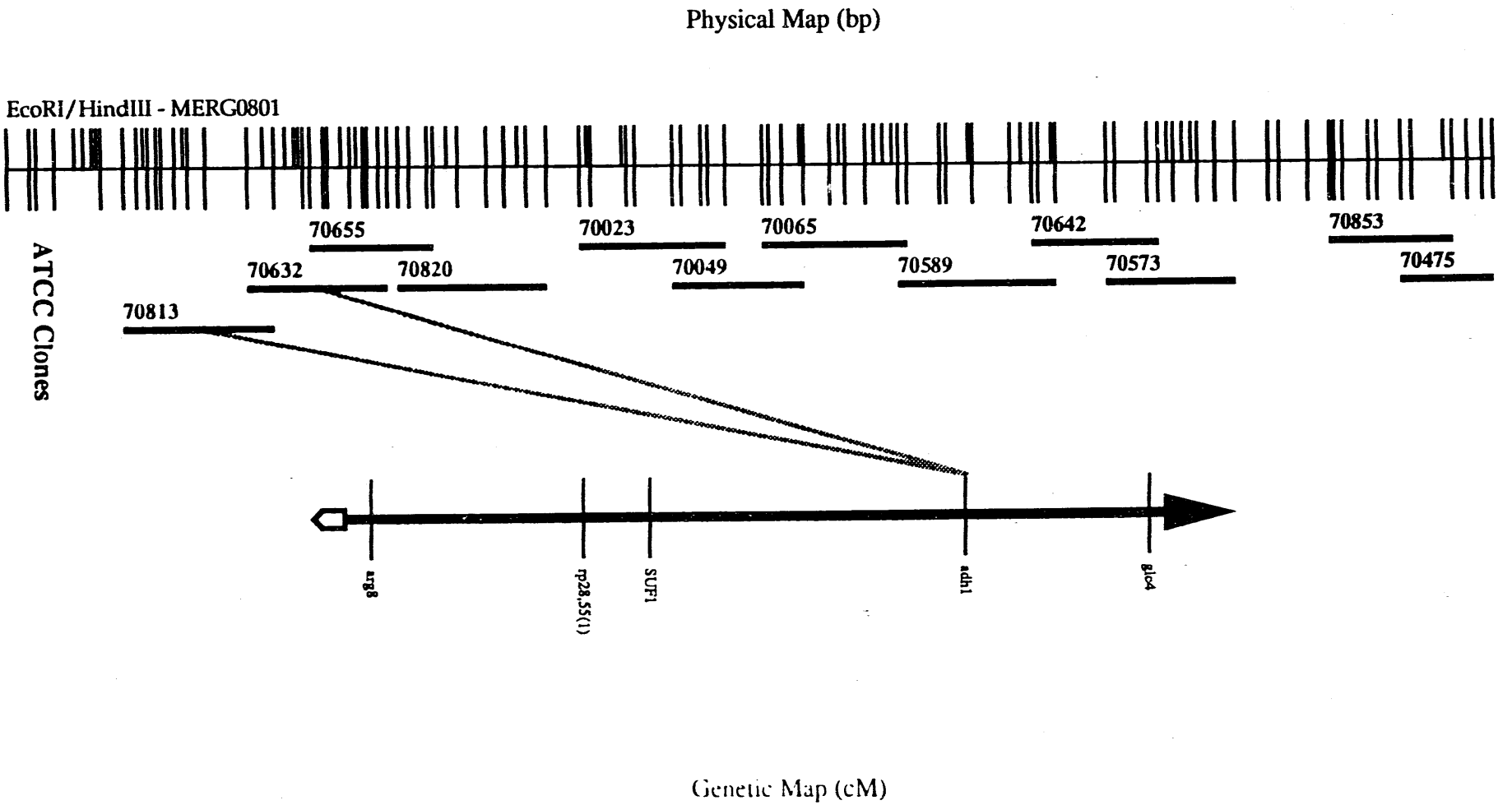
No Physical Map Data Available




Genetic Map (cM)

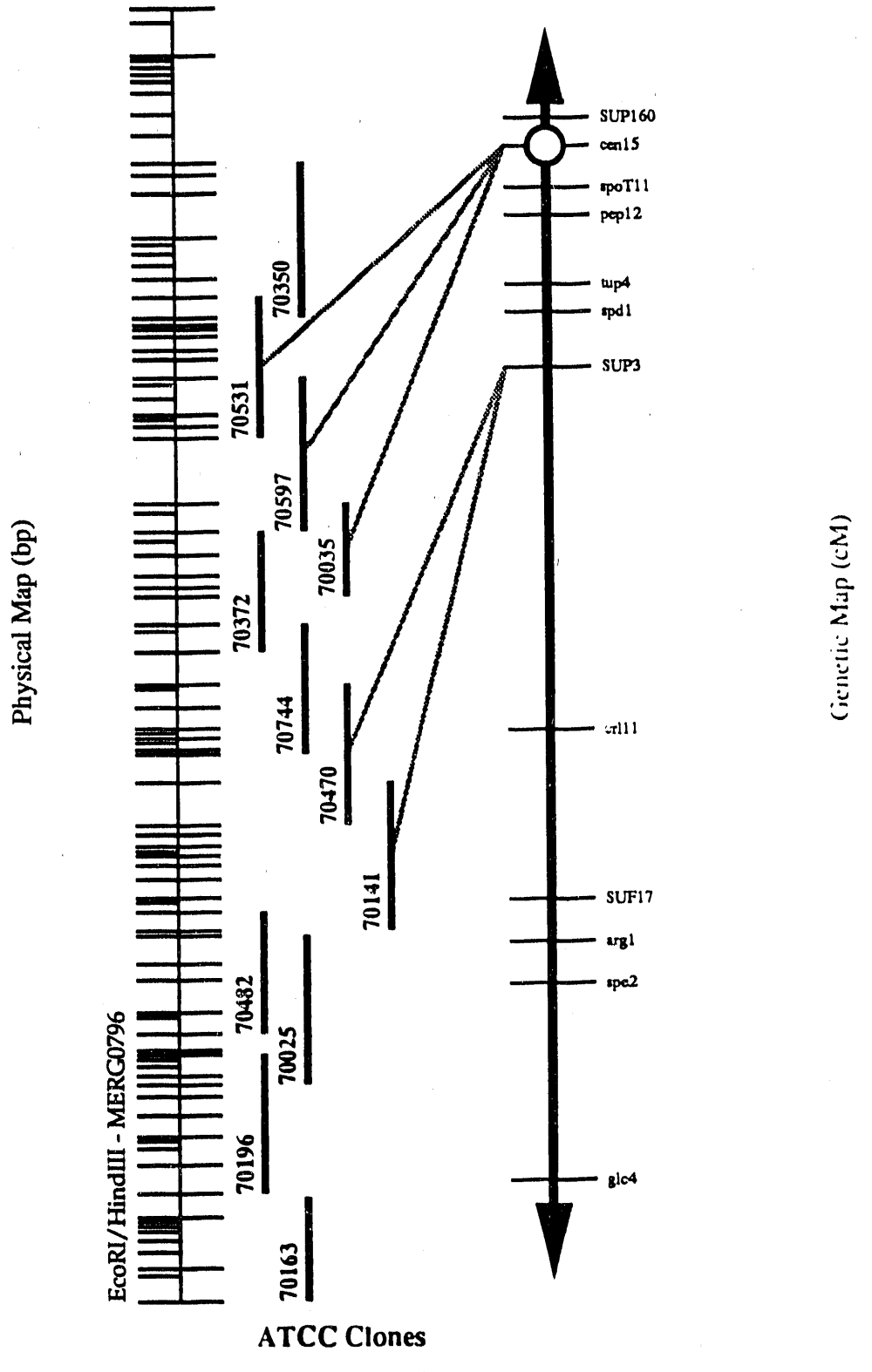
# Chromosome: XV Diagram 1 of 5





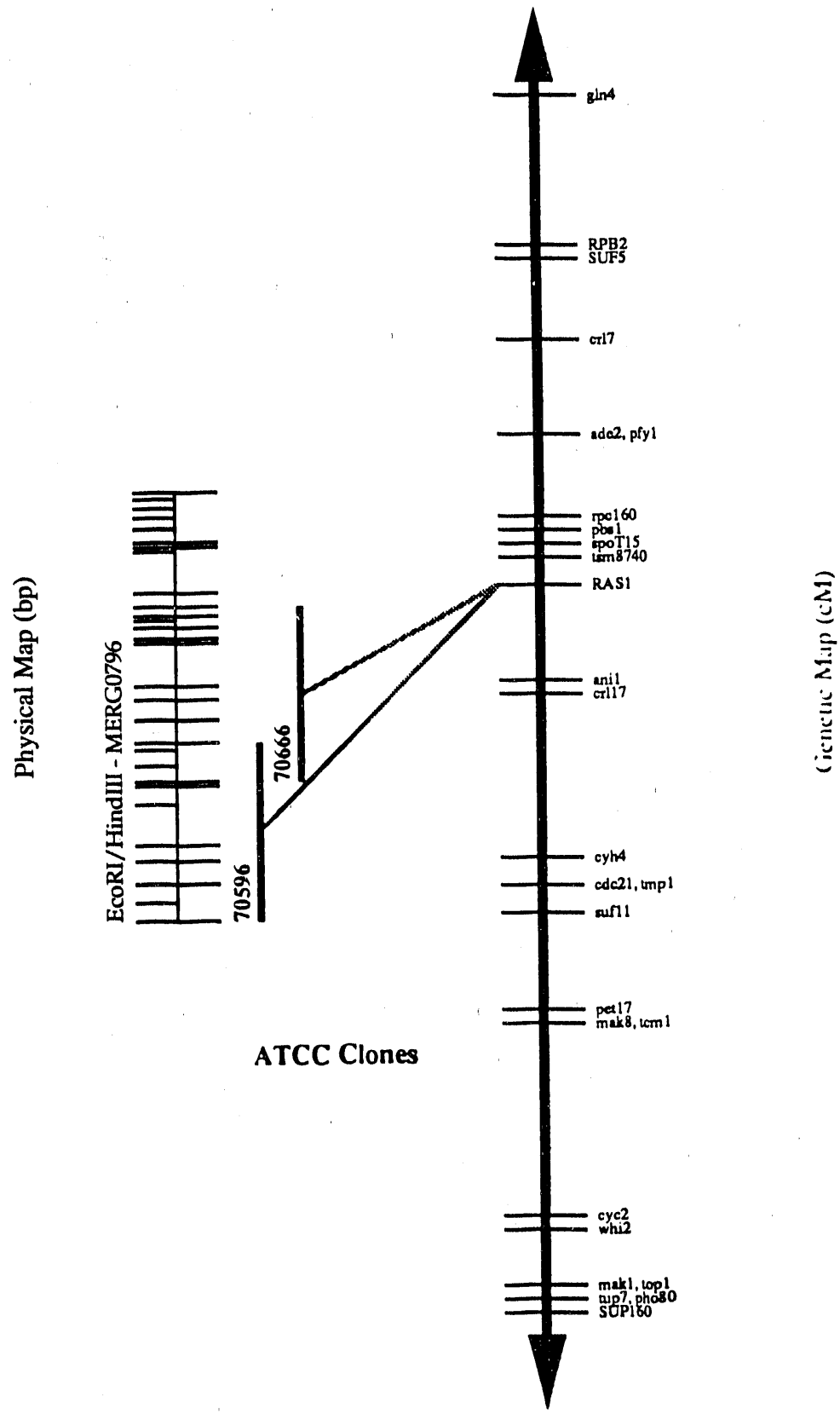
**CLONES FROM THE *SACCHAROMYCES CEREVISIAE* GENOME**

# Chromosome: XV Diagram 2 of 5

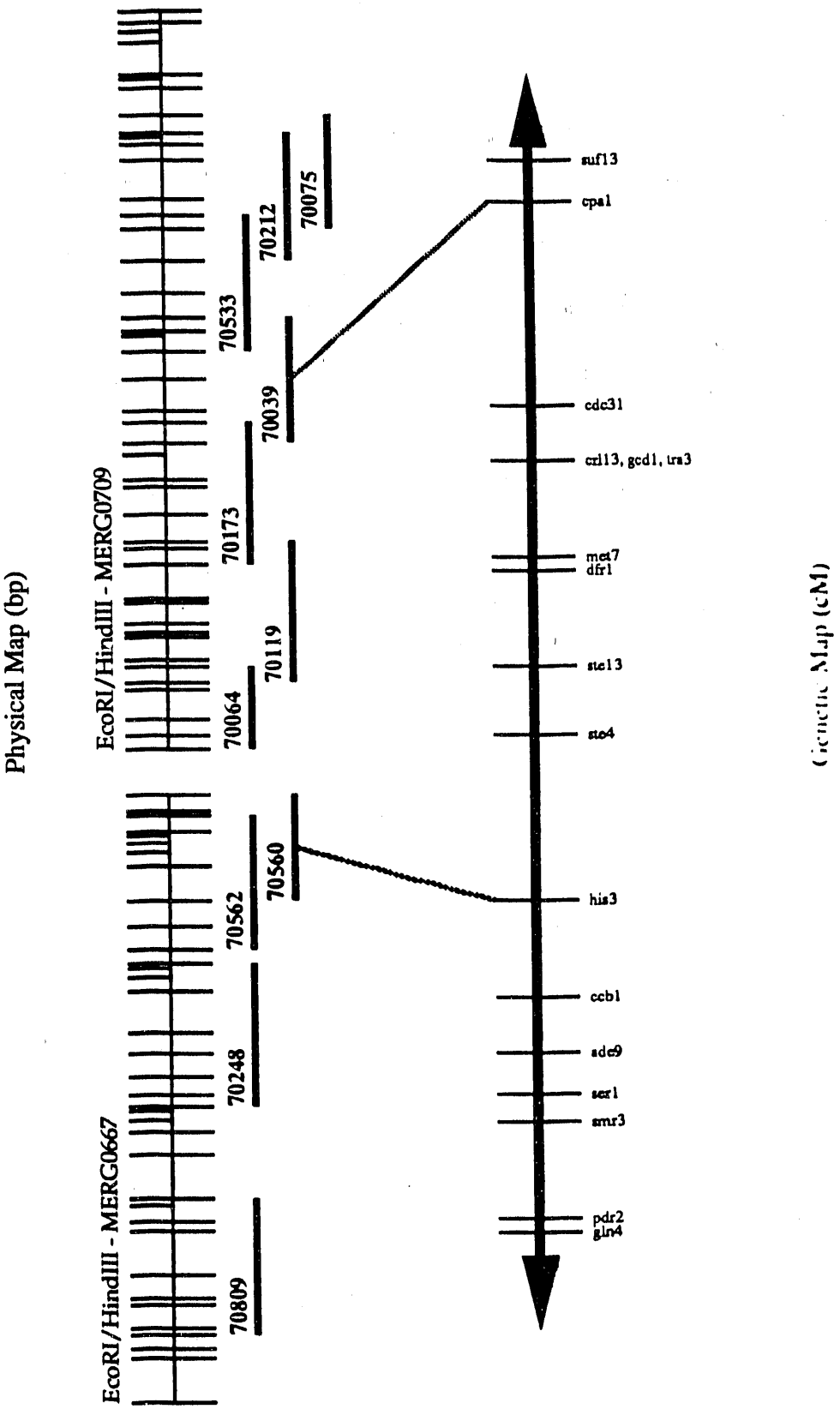


# CLONES FROM THE *SACCHAROMYCES CEREVISIAE* GENOME

Chromosome: XV Diagram 3 of 5

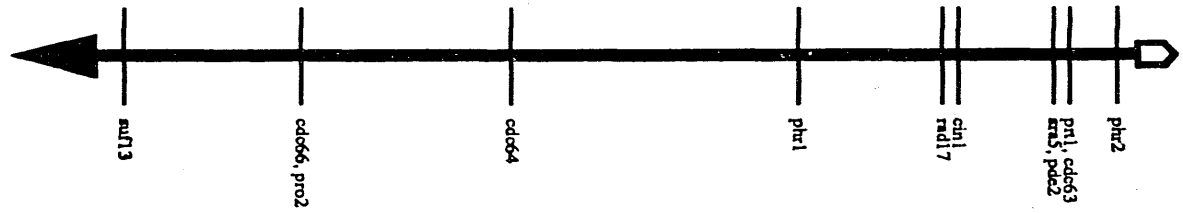


# Chromosome: XV Diagram 4 of 5



# Chromosome: XV Diagram 5 of 5

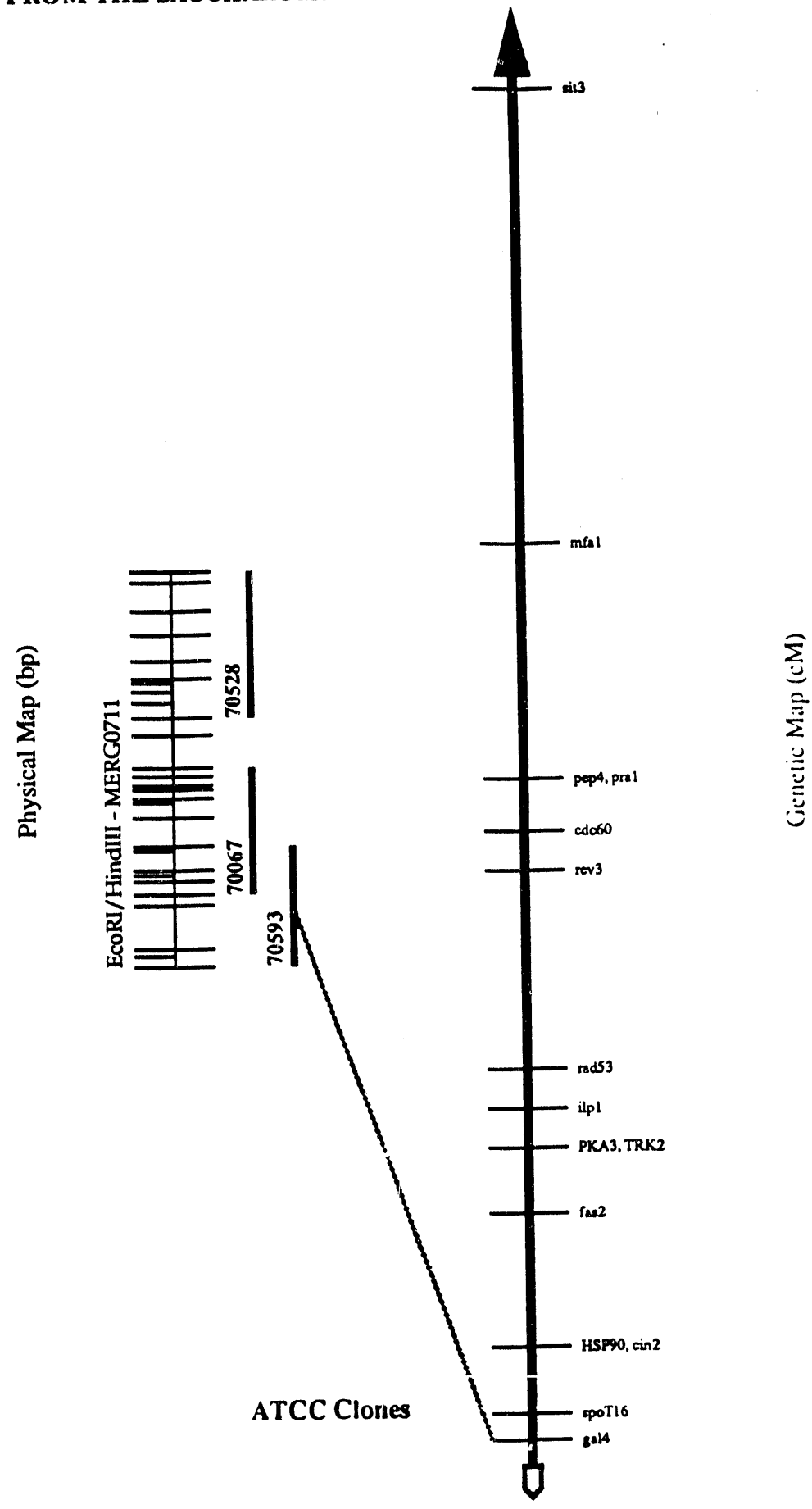
## No Physical Map Data Available



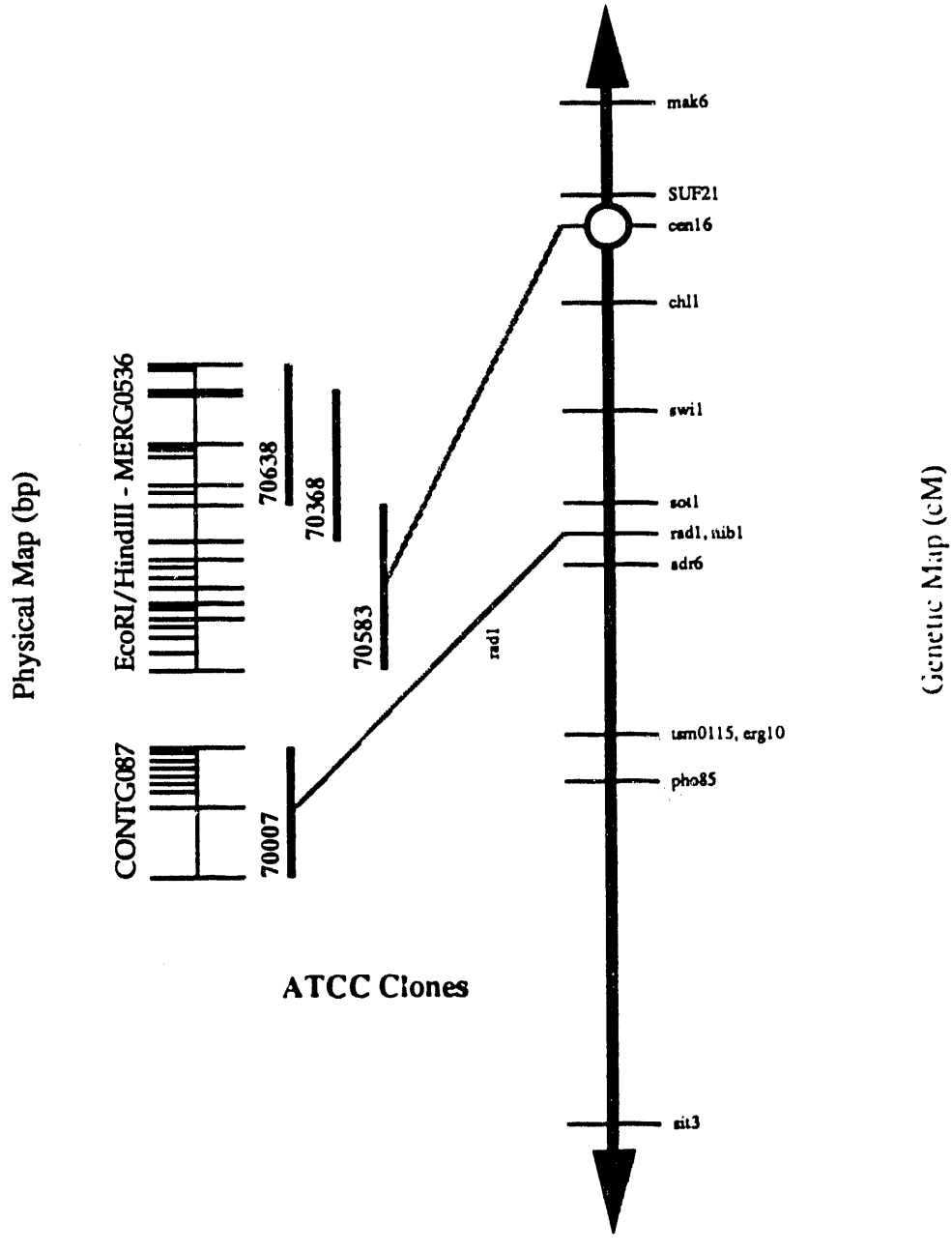

Genetic Map (cM)

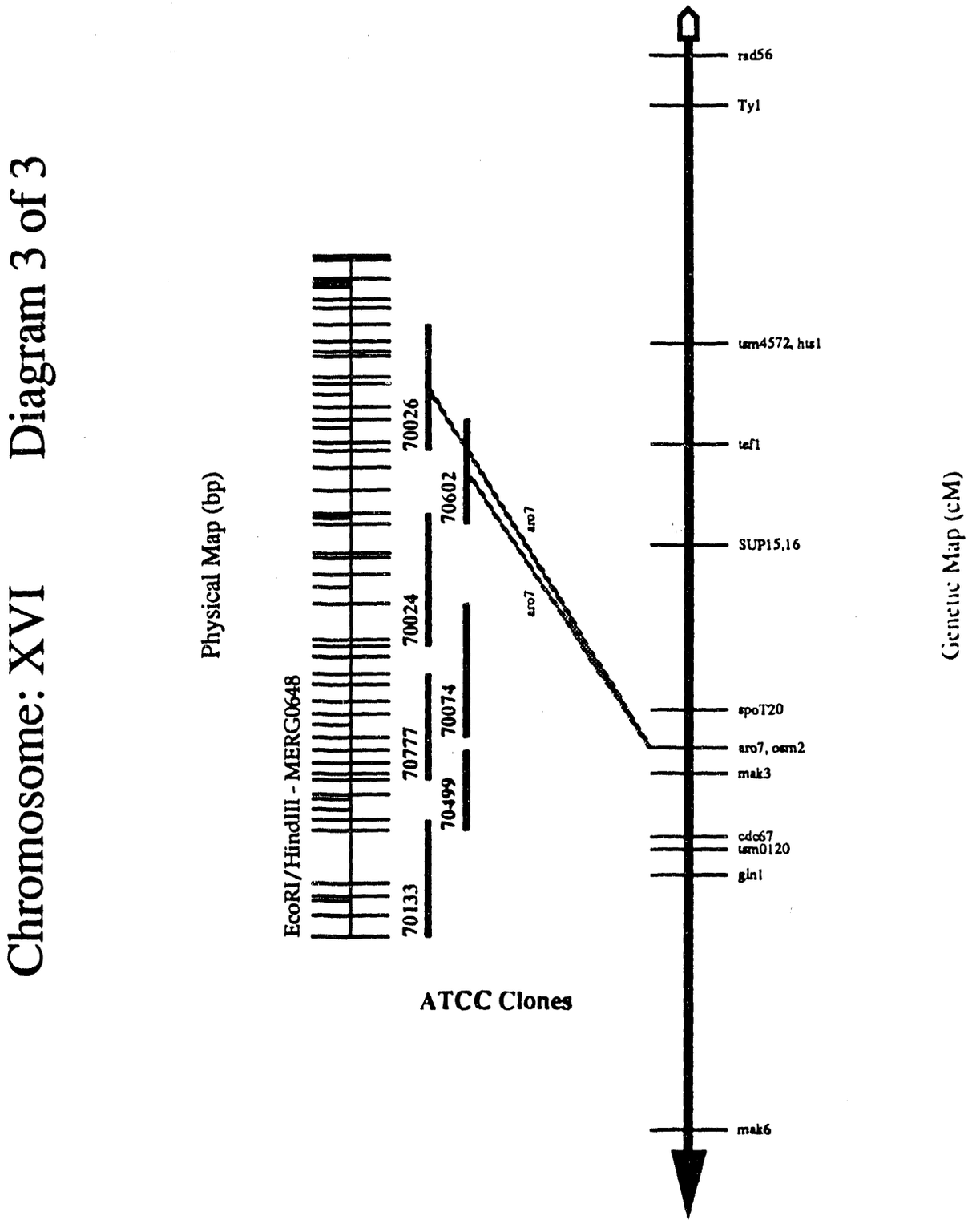
**CLONES FROM THE *SACCHAROMYCES CEREVISIAE* GENOME**

# Chromosome: XVI Diagram 1 of 3



# Chromosome: XVI Diagram 2 of 3




Chromosome: XVI
Diagram 3 of 3
Physical Map (bp)
EcoRI/HindIII - MERG0648
70133
70499
70777
70074
70024
70602
70026
ATCC Clones
aro7
aro7
rad56
Ty1
tsm4572, hus1
tef1
SUP15,16
spoT20
aro7, osm2
mak3
cdc67
tsm0120
gln1
mak6
Genetic Map (cM)

# END

DATE
FILMED
8107192