THE FINISHING OF HUMAN CHROMOSOMES 5 AND 19

Stanford Human Genome Center

JGI / SHGC Collaboration

- Finish human chromosome 5 and 19 clones from the shotgun data produced by the Joint Genome Institute
- Obtain long range contiguity for chromosomes 19 and 5
- Complete human chromosome 16 clones and mouse clones syntenic to human chromosomes 19, 5 and 16

JGI / SHGC Collaboration

JGI

>6x assembly of plasmid shotgun reads

Plasmid subclone libraries

Glycerol stock of large insert clone

SHGC

Finish clones at a rate of 8 Mb per month

What is Finishing?

- The process of turning a rough draft assembly composed of shotgun sequencing reads into a highly accurate finished DNA sequence with a defined maximum error rate
- Sequence must be completely contiguous, be confirmed by at least 2 templates, have no gaps and have a final estimated error rate of less than 1/10,000 bases

Finishing Steps (Prefin)

- Resequencing Reads (universal primers) ABI3700
  Read into gaps
  Raise low quality areas

- Custom Primer Round 1 (BDT) ABI377
  Close gaps
  Raise internal contig quality

- Custom Primer Round 2 (dGTP) ABI377
  Walk further into large gaps
  Read through sequencing stops

Finishing Steps (Human)

- Finisher Ordered Reads (dGTP) ABI 377

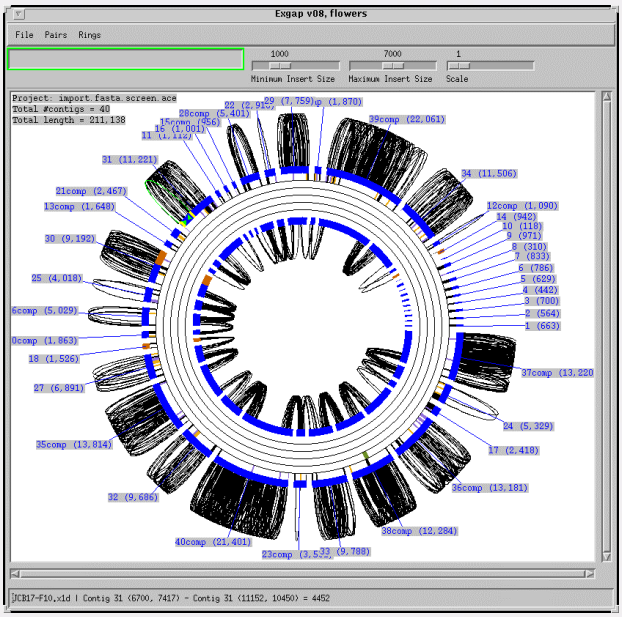
  Custom / Universal primer on subclone templates
  Custom / Universal on BAC template

  'Enhanced' custom / universal reads

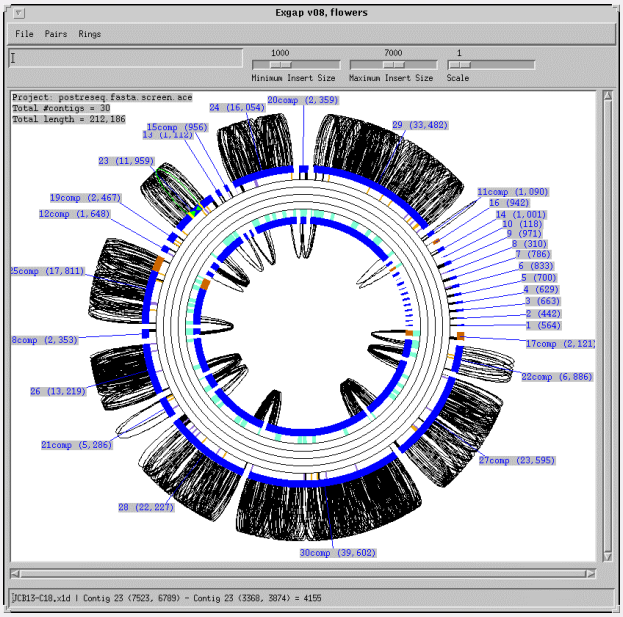
- Special reads ABI 3700

  Transposons
  Shatter libraries

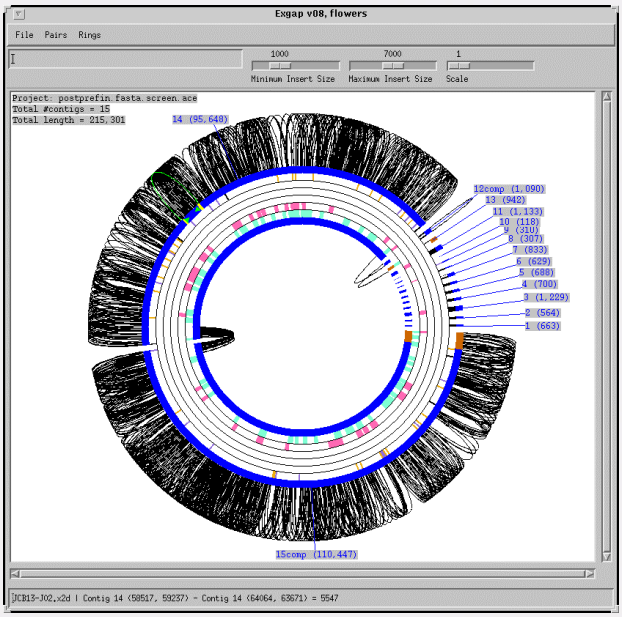
On Import from the JGI



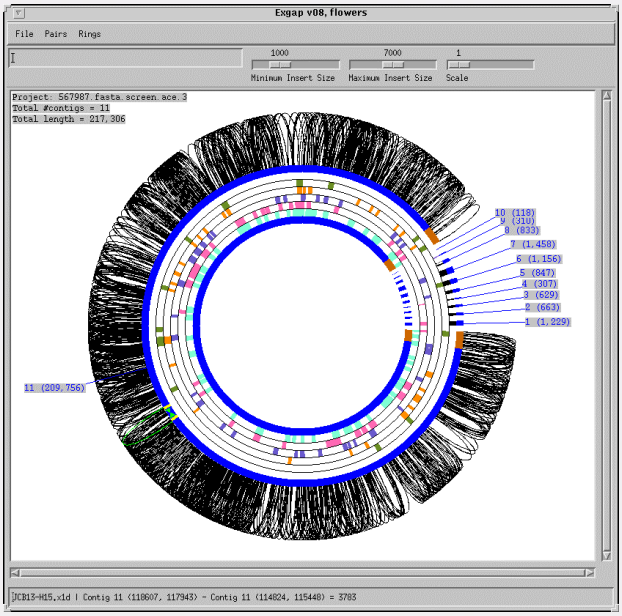
After Resequencing



After 2 automated custom primer rounds



After ‘human’ finishing



Finishing reads by type over a 12 month period (11/1/00 through 11/1/01)

- Prefinishing (automated)

Universal Resequencing	51,072
Primer Walking subclone (BDT)	29,952
Primer Walking subclone (dGTP)	15,360

- Finishing (human selected)

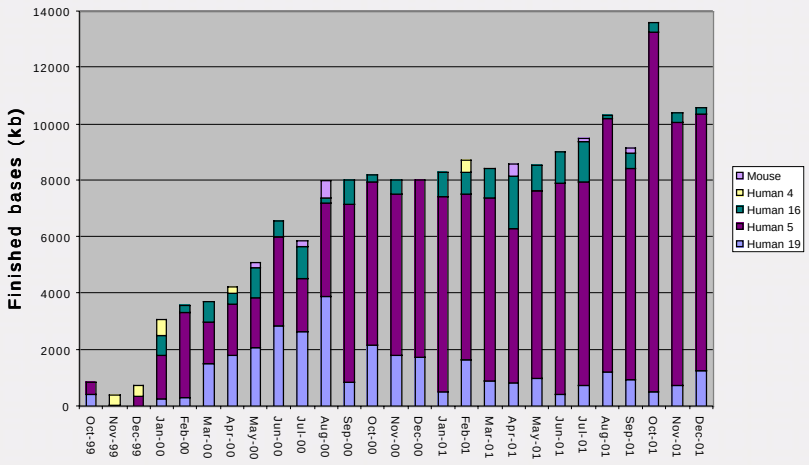
Universal resequencing	32,736
Primer walking subclone	25,920
Primer walking BAC	13,679
BAC end sequencing	1,655
Enhancer reads	2,331

- Custom library reads 6,336
- Transposons 688

Average Finishing Reads Per 100kb (11/1/00 - 11/1/01)

- Subclone universal 67.3
- Subclone custom primer 49.9
- BAC custom primer 7.2
- BAC universal 1.0

Finished sequence by month



Finished human sequence by Genome Center

Sanger	762 Mb
Washington University	461 Mb
SHGC	199 Mb
Whitehead	170 Mb
Baylor	143 Mb

Human Genome Project

The goal of the human genome project is to produce a sequence of human DNA that is of sufficiently high accuracy and long-range contiguity to allow a full interpretation of all information encoded in the human genome

Target date for completion of all human chromosomes

April 2003

Human Genome Project

1/24/02

Finished Human Sequence

From NCBI:

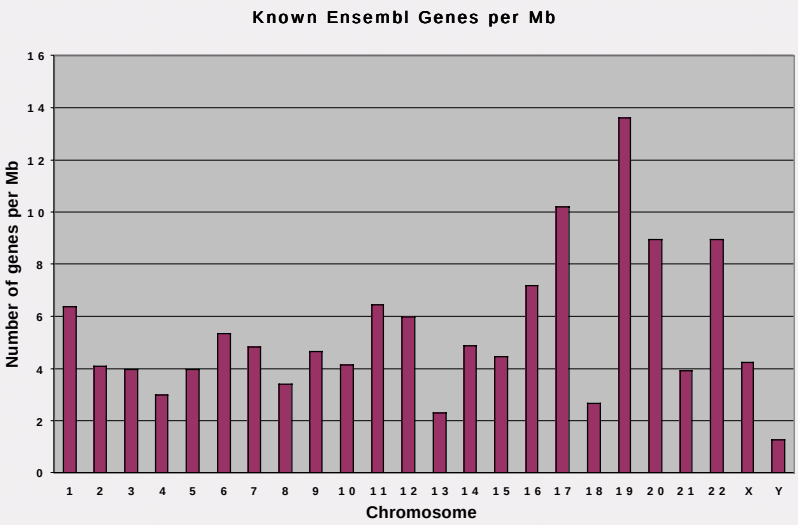
2,073,539 bp

Or about 64.8% of the human genome

Chromosome 19

Target completion date

July 2002



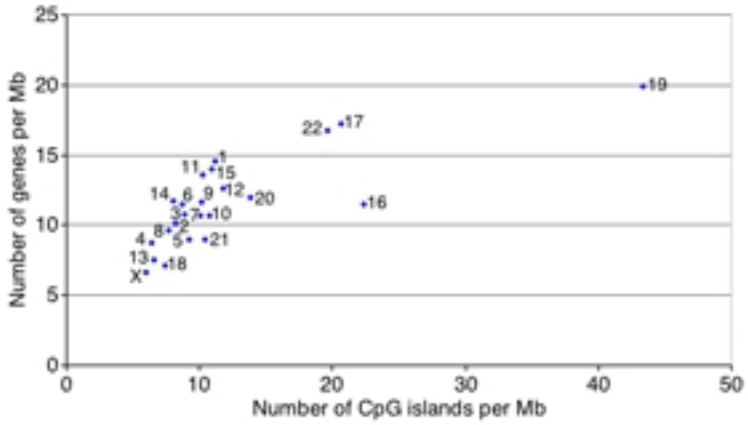


Figure 14 Number of CpG islands per Mb for each chromosome, plotted against the number of genes per Mb (the number of genes was taken from GeneMap98 (ref. 100)). Chromosomes 16, 17, 22 and particularly 19 are clear outliers, with a density of CpG islands that is even greater than would be expected from the high gene counts for these four chromosomes.

NATURE | VOL 409 | 15 FEBRUARY 2001 | www.nature.com

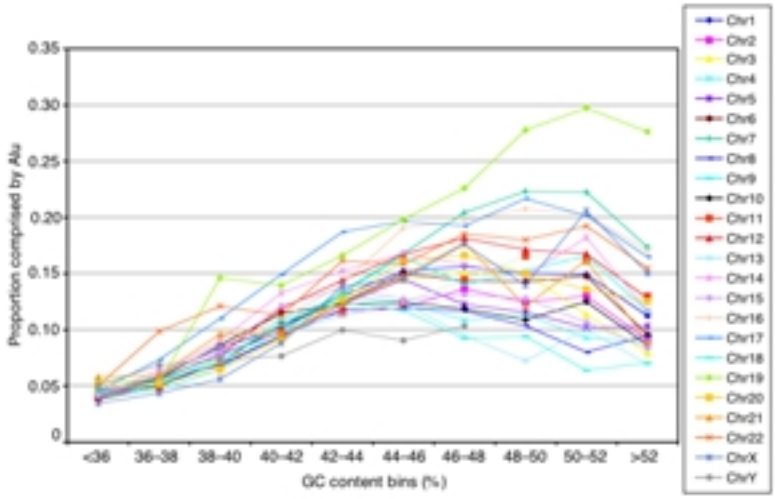


Figure 26 Comparison of the Alu density of each chromosome as a function of local GC content. At higher GC levels, the Alu density varies widely between chromosomes, with chromosome 19 being a particular outlier. In contrast, the LINE1 density pattern is quite uniform for most chromosomes, with the exception of a 1.5 to 2-fold over-representation in AT-rich regions of the X and Y chromosomes (not shown).

NATURE | VOL 409 | 15 FEBRUARY 2001 | www.nature.com

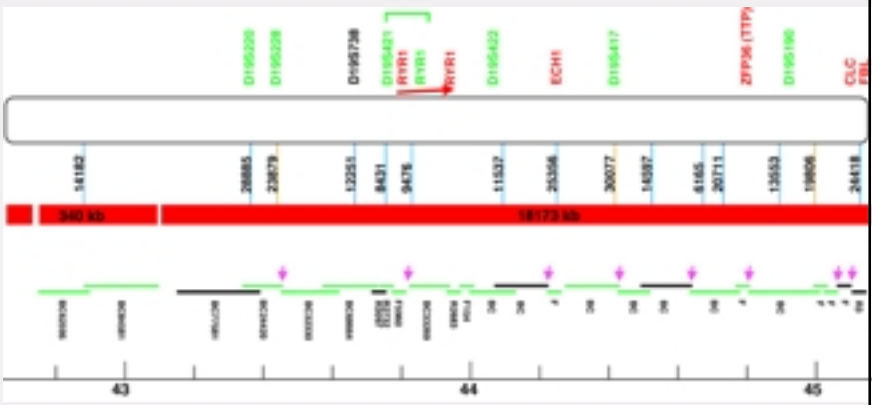
Chromosome 19 tiling path

851 clones on the tiling path

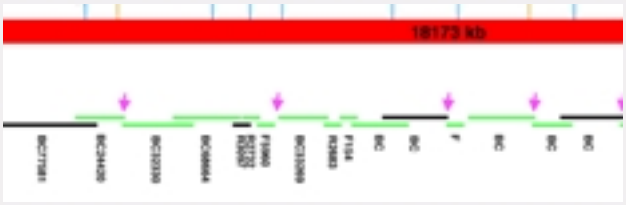
- 327 BAC
- 513 cosmids
- 9 fosmids
- 2 P1 clones

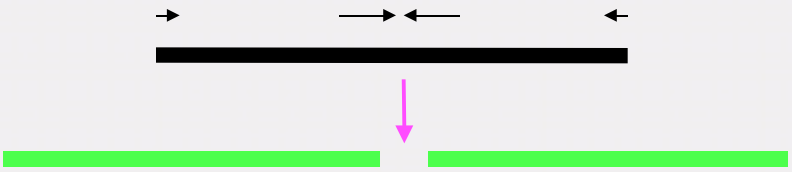
Euchromatin estimate 58 Mb

Chromosome 19 Map and Sequence (Section)



Small sequence gaps (114)





Directed BAC walks to close the small sequence gaps

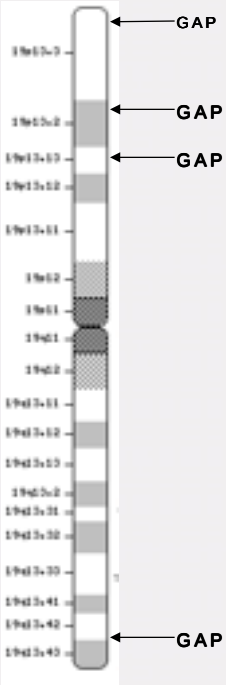
Current status of Chromosome 19

- 43.56 Mb finished (63.1%)
- 154 finished contigs

- 662/851 clones finished (77.7%)
- 189 clones in finishing (14 Mb)

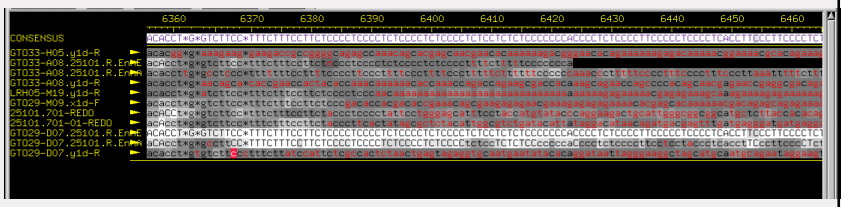
- 4 clone gaps

Chromosome 19



Challenges to finishing Chromosome 19

- Many strings of simple sequence repeats
- Small repetitive areas
- Large repetitive areas




```
>Contig12 357334.fasta.screen.ace.5 from 42310 to 44540
CTACACTGGGAATCACCCGGAAGGAGGCCAAGGGTGTGTAATTCTGGGGAAGTTTGGCGGCAGCCTAGGGC
CTCCTCAATCTGAAAGCTCTTGTTTCGCTTTGTAAAGCGTTTGCTTGTTTCTTTTTCTTTCTCTCTTTTTT
TCTTTCTTTCTTTTTCTTTCGCCTGCCTCCCTCCCTCCTTCCCTCCCTCCCTTCCCTCCTCTCTCCGTCCC
TCCTTCTCTCCCTCTCCTGCCTCTCTCCCTCTCTCCCCCCCTTTCTCTTTCCTTCCTCCCTCCCTCCCTTAC
TTCCTCCCTCCCTCCTCTCCCTCTCCTGCCTCTCTCCCTCTCTCCCCCCTTTCTCTTTCCTTCCTCCCTCC
CTCCCTTACTTCCTCCCTCCCTCCTCTCCCTCTCCTGCCTCTCTCCCTCTCTCCCCCCTTTCTCTTTCCTT
CCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTCCCTCCCTCCCTTCCTCCCTCCCTCCCTTCCTCCCTCCCT
CCCTTACTTCCTCCCTCCCTCCTCTCCCTCTCCTGCCTCTCTCCCTCTCTCCCCCTTTCTCTCTCCTTCCT
TCCTTTCTTCCCTTCCTCCCTCCCTCCCTCCTTCTCTCCCTCTCCTGCCTCTCTCCCTCTCTTCCCCTTTT
CTCTCCTTCCTTCTTTTCCTTCCTTCCTTCCTTTCTTTCCTTTCTTTCTTTCTTTCGCCTTCCTCC
CTCCCTCCTTCCCTCCCTCCCTTCCTCCCTCCCTCCCTCTCCTGCATCTCTCCCTCTCTCTCCCCTTTCTC
TTTCCTCCCTTCCTCCCTCCCTTCCTCCCTCCCTTCCTTCCTCCCTCCCTCCTTCTCTCCCTCTCCTGCCT
CTCTCCCTCTCTCCCCCCTTTTCTCCCCTTCCTTCCTTCCTTTCTTTTCCTTCCTTCCTTCCTTCTTCCTT
TTCTTTCTCTCTTTCTCTTTCTTTCCTTTGCCTTCCTCCCTCCCTCCCTCCTTCCCTCCCTCCGTCCTTCC
CACCCTCCCTCCTTCCCTCCCTCCCTTCCTTCCTCCCTCTCTCCTTCTCTCCCTCTCCTGCATCTCTCCCT
CTCTCTCCCCTTTCTCTTTCCTTCCTTCCTTCCTCTCTCCCTTCTTTCCTCCCTCCCTCCTTCTCTCCCTC
TCCTGCCTCTCTCCCTCTCTCCCCGCTTTTCTCTCCTTCCTTCTCCTTCCTTTCTTTTCCTTCCTTCCTTCTT
CCTTTTCTTTCTTTTCTTTTTCTCTTTCTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT
TTCTTTCTCTCTTTCTTTTTCTTTCTTTCTTTCTCTTTTCTGTCTCTCTCTCTCTCTCTCCCCCACCTTTT
TTTTTCAGACAGGGTCTTGCTAAGTTGCCCAGGCTGGTCTCAAACTCCTGGCCTCAAGTGATCCTCCTGCC
TTGGCCTCCCAAAGGAGGAATTATTATCACTATTTTGCCTTAGATGTGATAATGGT
```

Small repeat regions on a chromosome 19 BAC clone

GCCTGGACTCCTGGGTCTGAGGGAGGAGGGGCTGGGGGCCTGGGCTCCAGGGTCG.....CCAGGGTCTGAGGGAGGAGGGGCTGGGGGCCTGGACTCCTGGGTCTGAGGGAG

94.8% 304bp

94.8%

90.6% 219p

90.6%

96.5% 171bp

96.5%

96.5%

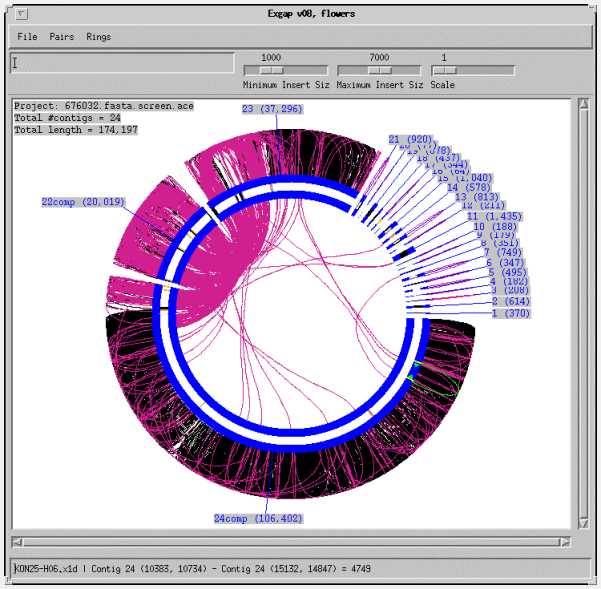
100% 54bp

100% 54bp

100% 54bp

100% 54bp

Large repeat regions on a chromosome 19 BAC clone



Current status of Chromosome 5

- Euchromatin estimate 186 Mb
- 1828 Clones on the tiling path
- 105.4 Mb of finished sequence (54%)

- 177 Type 1 and 2 gaps (clone selected)
- 44 Type 3 gaps (clone gap)

Current status of Chromosome 5

- SHGC has now finished 100.2 Mb of unique chromosome 5 sequence
- Current finishing ramp will increase to finish majority of the chromosome by October 2002
- Target completion date Dec 2002

Acknowledgements

Richard Myers
Jeremy Schmutz
Mark Dickson
Sequencing group at SHGC

Trevor Hawkins
Joe Monforte
Anne Olsen
Susan Lucas
Kristen Kadner
Joint Genome Institute

Marvin Frazier and the DOE

Stanford Human Genome Center