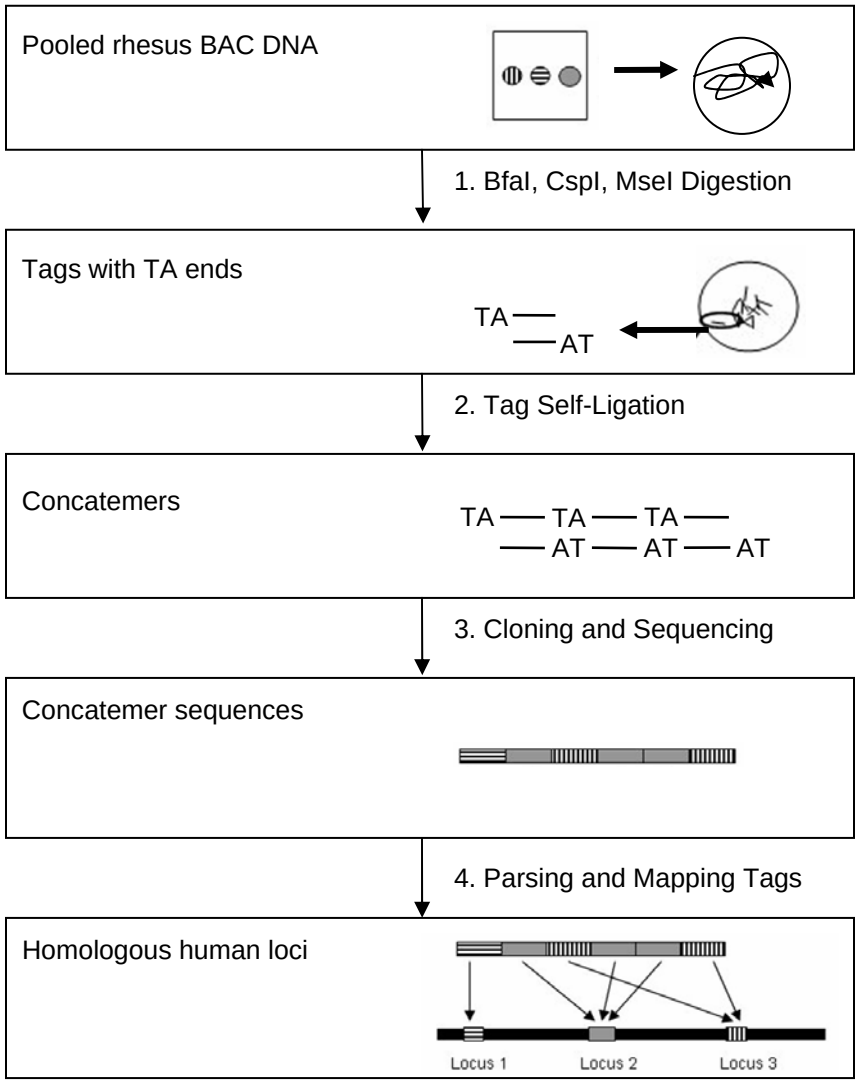
Mapping pooled BACs using the short-tag method

ST-PGI samples pooled rhesus BAC DNA by cutting it into fragments ("short tags"). As indicated in the **diagram below,** the tags are ligated together to form concatemers. The concatemers are then cloned and sequenced. The sequencing reads are compared to the human genome using BLAT (3), and the comparison results are examined by the Tagamizer program. The Tagamizer program uses comparative information to find the boundaries between tags and map the rhesus tags onto the human sequence.



In the following we describe the four steps of the short-tag method, as illustrated in the figure above.

Step 1. BfaI, CspI, MseI Digestion

Restriction enzymes BfaI, CspI, and MseI (all recognizing different 4bp sites and generating compatible “TA” overhangs) are used. The compatible “TA” overhangs are used in the subsequent ligation step. The enzymes work under common buffer conditions enabling single-step complete digestion and are heat-inactivated, further simplifying the protocol.

BAC DNA is prepared and pooled as previously described (1). Ten micrograms of pooled macaque BAC DNA is digested at 37 C overnight with 25 units of Bfa I and 50 units of Mse I (New England Biolabs, Beverly, MA) with 1X NEB #4 buffer and 0.5 µg of BSA in a 250 microliter volume as directed by the manufacturer. Digestions are purified using the Qiagen (Valencia, CA) PCR purification columns as directed by the manufacturer and re-digested overnight at 37 C with 10 units of Csp6 I (Fermentas, Hanover, MD) in a 200 microliter volume. The final digest is phenol/chloroform extracted, ethanol precipitated and resuspended in 40 microliters of TE. One twentieth of the reaction is analyzed by 1.5% agarose gel electrophoresis for completeness of the digestion.

Step 2. Tag Self-Ligation

The digested DNA is concatenated at 16 C overnight in a reaction containing 50 mM ATP, 20 units of T4 DNA ligase (Amersham Biosciences, Piscataway, NJ) and 1 X blunt ligation buffer as supplied by the manufacturer. The concatenation reaction is purified by Qiagen PCR purification columns as previously described and analyzed for size on 1.5% agarose gel electrophoresis.

Step 3. Cloning and Sequencing

DNA fragments are blunt-ended and ligated to oligonucleotide adaptors as previously described (1). The DNA from 1.5 to 4 kb is purified from adaptor oligonucleotides and other DNA fragments using Clontech (Palo Alto, CA) 1000s spin columns or a low-melting 0.8% agarose gel in 1 X TEA buffer. The DNA fragments are purified from the agarose using a previously described freeze-thaw method (2). PEKII+8 vector is prepared and annealed to the insert as previously described (1). One tenth of the annealing reaction is transformed into 110 µl of XL10 Gold cells (Stratagene, La Jolla, CA) using the heat shock method as described by the manufacturer. One hundred transformants are selected on carbenicillin plates and picked into 400 µl of selective media. Plasmid DNA is purified using the Life Technologies, Inc. plasmid kit and cycle sequenced on Applied Biosystems 3700 or 3730 DNA analyzers.

Step 4. Parsing and Mapping Tags

Parsing and Mapping of Tags is performed using the Tagamizer program. Tagamizer is available free of charge for academic use. Program download and commercial licensing information are available at
http://www.brl.bcm.tmc.edu/castl/tagamizerDownload.html.

Tagamizer uses BLAT Client/Server version 23, which is available from its authors at http://www.soe.ucsc.edu/~kent/src/blatSrc23.zip.
Tagamizer exclusively relies on comparative information, and does not require any delimiting sequence between tags. Reads are first compared against the reference genome using the BLAT program (3). The Tagamizer program identifies those BLAT hits that correspond to correct tag mappings

Acknowledgements

We are grateful to many individuals the Baylor Human Genome Sequencing Center for feedback about this work and for the use of computing resources, and to Elke Norwig-Eastaugh for administrative support. KJK was partially supported by a training fellowship from the W.M. Keck Foundation to the Gulf Coast Consortia through the Keck Center for Computational and Structural Biology. This research was funded by the NIH NCRR grant U01 RR18464-01 to AM.

References

1. Andersson, B., Wentland, M.A., Ricafrente, J.Y., Liu, W. and Gibbs, R.A. (1996) *Anal Biochem*, **236,** 107-113.
2. Qian, L. and Wilkinson, M. (1991) *Biotechniques*, **10,** 736, 738.
3. Kent, W.J. (2002) *Genome Research*, **12,** 656-664.