
Preface

Several developmental and historical threads are woven and displayed in these two volumes of *Bacterial Artificial Chromosomes*, the first on *Library Construction, Physical Mapping, and Sequencing*, and the second on *Functional Studies*. The use of large-insert clone libraries is the unifying feature, with many diverse contributions. The editors have had quite distinct roles. Shaying Zhao has managed several BAC end-sequencing projects. Marvin Stodolsky during 1970–1980 contributed to the elucidation of the natural bacteriophage/prophage P1 vector system. Later, he became a member of the Genome Task Group of the Department of Energy (DOE), through which support flowed for most clone library resources of the Human Genome Program (HGP). Some important historical contributions are not represented in this volume. This preface in part serves to mention these contributions and also briefly surveys historical developments.

Leon Rosner (deceased) contributed substantially in developing a PAC library for drosophila that utilized a P1 virion-based encapsidation and transfection process. This library served prominently in the Drosophila Genome Project collaboration. PACs proved easy to purify so that they substantially replaced the YACs used earlier. Much of the early automation for massive clone picking and processing was developed at the collaborating Lawrence Berkeley National Laboratory. However, the P1 virion encapsidation system itself was too fastidious, and P1 virion-based methods did not gain popularity in other genome projects.

Improving clone libraries was an early core constituent of the DOE genome efforts. Cosmid-based libraries with progressively larger inserts were developed within the DOE National Laboratories Gene Library Program. But quality control tests by P. Youdarian indicated that perhaps 25% of human insert cosmids had some instability, possible owing to the multicopy property of the system. Both for this reason and to provide for larger inserts of cloned DNAs, DOE supported the investigation of several new cloning systems. Of the eukaryotic host systems, the Epstein-Barr virus-based system from Jean-M. Vos (deceased) was quite successful. But the added costs and care needed for use of eukaryotic cells precluded its wide adoption in HGP production efforts.

Among the bacterial host systems, two developed in the lab of Melvin Simon provided pivotal service. Ung-Jin Kim developed fosmids. They are maintained as single copy replicons and utilize the reliable encapsidation pro-

cesses developed for cosmids. Fosmids proved to be highly stable. BACs were developed by Hiroaki Shizuya. They were introduced into *E. coli* by electroporation and stability was generally good, though there is an unstable BAC minority (1). This BAC resource emerged after the chimeric properties of the large YACs was recognized. BACs were thus initially viewed with appropriate suspicion. But at the nearby Cedar-Sinai Medical Center, J. R. Korenberg and X.-N. Chen implemented a very efficient FISH analysis. They found that chimerism in any of the BACs was at worst around 5% and the BACs were well distributed across all the chromosomes. Overall human genome coverage was estimated in the 98–99% range, with even centromeric and near telomeric regions represented.

Two examples of this good coverage soon emerged. Isolation of the BRAC1 breast cancer gene had failed with all other clone resources. But when Simon's group was provided with a short cDNA probe, they soon returned a BAC clone carrying an intact BRAC1 gene. Pieter de Jong had acquired the technology of cloning long DNA inserts from the Simon lab, initially using a PAC vector and electroporation. After a first successful library, DOE advised de Jong to broadly distribute this new PAC resource. Shortly thereafter, he assembled a 900 kb contig for the candidate region of the BRAC2 gene. The subsequent DNA sequence generated at the Washington University then revealed the BRAC2 gene. These striking easy successes stimulated broad usage of the BAC and PAC resources.

End sequences of clonal inserts have been used to facilitate contig building since the 1980s in small-scale mapping and sequencing projects. Glen Evans for example was piloting with DOE support a "mapping plus sequencing" strategy on chromosome 11, before the BAC resources were available. Once a covering set of cloned DNAs with sequenced ends is generated, clones to efficiently extend existing sequence contigs can be chosen. As the need for high throughput genome sequencing to meet HGP timelines became imminent, only a few human chromosomes had adequate contig coverage. L. Hood, H. Smith, and C. Venter proposed a Sequence Tag Connector (STC) strategy to alleviate this bottleneck. With application to the entire human genome, concurrent BAC contig building and sequencing would be implemented.

The DOE instituted a fast track review of two STC applications in the spring of 1996 (2). One was from a team comprised of L. Hood, H. Smith, and C. Venter, and the second from a team comprised of G. Evans, P. de Jong, and J. R. Korenberg. A panel with broad international representation reviewed applications from two teams. Interested colleagues from the NIH and NSF were observers. Although the overall STC concept was reviewed favorably,

initial pilot implementations to better define the economics were recommended. A year later, progress was reviewed and a DOE commitment to a full scale implementation was made. At the request of the NIH, the DOE later increased support to accelerate a 20-fold coverage of the genome.

The STC data set has had multiple beneficial roles. Sequence Tag Sites (STSs) were defined within the STC sequences and used to enrich the Radiation Hybrid (RH) maps of the genome, thus providing for an early correspondence of the RH maps and the maturing contig maps. Validity constraints on sequence contigs were provided by the spanning BACs. Most broadly, the STC resource had an indispensable role for both the strategies of Celera Genomics Inc., and the international public sector collaboration, in the rapid generation of draft sequences of the human genome. The STC strategy is now implemented in many current genomic projects, including the NIH sponsored mouse and rat genome programs.

Bacterial Artificial Chromosomes in its two volumes provides a comprehensive collection of the protocols and resources developed for BACs in recent years. These two volumes collectively cover four topics about BACs: (1) library construction, (2) physical mapping, (3) sequencing, and (4) functional studies. The laboratory protocols follow the successful *Methods in Molecular Biology*TM series format by containing a clear sequence of steps followed by extensive troubleshooting notes. The protocols cover simple techniques such as BAC DNA purification to such complex procedures as BAC transgenic mouse generation. Both routine and novel methodologies are presented. Besides protocols, chapter topics include scientific reviews, software tools, database resources, genome sequencing strategies, and case studies. The books should be useful to those with a wide range of expertise from starting graduate students to senior investigators. We hope our books will provide useful protocols and resources to a wide variety of researchers, including genome sequencers, geneticists, molecular biologists, and biochemists studying the structure and function of the genomes or specific genes.

We would like to thank all those involved in the preparation of this volume, our colleagues, and friends for helpful suggestions, and Professor John Walker, the series editor, for his advice, help and encouragement.

Shaying Zhao
Marvin Stodolsky

References

1. <http://www.ornl.gov/meetings/ecr2/index.html>
2. <http://www.ornl.gov/meetings/bacpac/body.html>

Bacterial Artificial Chromosomes

Volume 1: Library Construction, Physical Mapping, and
Sequencing

Zhao, S.; Stodolsky, M. (Eds.)

2004, XX, 360 p. 71 illus., Hardcover

ISBN: 978-0-89603-988-9

A product of Humana Press