Journal of Structural and Functional Genomics **6**: 89–94, 2005.
DOI 10.1007/s10969-005-3084-1

Methods and results for semi-automated cloning using integrated robotics

Heath E. Klock, Aprilfawn White, Eric Koesema & Scott A. Lesley*
*Joint Center for Structural Genomics, Genomics Institute of the Novartis Research Foundation, 10675 John Jay Hopkins Drive, San Diego, CA 92121, USA; *Author for correspondence (e-mail: slesley@gnf.org; fax: +858-812-1746)*

Received 2 August 2004; accepted in revised form 15 February 2005

Key words: automation, cloning, high-throughput, JCSG, structural genomics

Abstract

The Joint Center for Structural Genomics (JCSG) has emphasized automation and parallel processing approaches. Here, we describe automated methods used across the cloning process with results from JCSG projects. The protocols for PCR, restriction digests and ligations, as well as for gel electrophoresis and microtiter plate assays have all been automated. The system has the capacity to routinely process 384 clones a week. This throughput can adequately supply our expression and purification pipeline with expression-ready clones, including novel targets and truncations. The utility of our system is demonstrated by our results from three diverse projects. In summary, 94% of the PCR amplicons generated to date have been successfully cloned and verified by sequencing (83% of the total attempted targets). Our results demonstrate the capabilities of this robotic platform to provide an avenue to high-throughput cloning which requires little manpower and is rapid and cost-effective while providing insights for method optimization.

Introduction

As the JCSG continues to expand its structural genomics pipeline, the need to provide expression-ready clones quickly grows. Recombinatorial cloning strategies, although fast and efficient, are not well suited for crystallography due to the additional encoded amino acids that recombination sites typically add to the target and to the supply of costly enzymes associated with recombination. Our efforts began by using traditional, but labor intensive, techniques to clone the 1877 predicted open-reading frames of *Thermotoga maritima* [1]. Here, we describe methods that semi-automated this scheme. Key steps of the cloning process were automated by using integrated, robotic liquid and plate handling workstations. These steps included PCRs, gel electrophoresis, restriction digests and ligations. Transformations and colony picking were performed manually in 96-well microtiter plates. Colonies were screened using diagnostic PCR (dPCR) followed by an automated, fluorescence plate assay that was developed to quickly identify insert-positive expression clones. Exonuclease I/Shrimp Alkaline Phosphatase (Exo/SAP) treatment prepared insert-positive dPCR products for sequencing [2]. The resulting sequence files were analyzed by a Perl script-mediated database search.

This automation is fully compatible with other recombinatorial cloning applications, so laboratories already vested in such a strategy could apply these methods for additional benefit [3, 4].

Materials

The principle expression vector (pMH4) used by the JCSG encodes a non-cleavable, amino-terminal, twelve-residue tag consisting of the first six residues of thioredoxin and six histidines (MGSDKIHHHHHH) and a 3-frame stop sequence at the 3′-end of the cloning site [5]. A

unique *Pml*I (CAC↓GTG) restriction site allows the first codon of a target reading frame to be ligated into the vector immediately after the 6th histidine codon of the tag. The cohesive end of the *Pac*I (TTAAT↓TAA) restriction site, located at the 3′-end of the cloning site, guides insert directionality into the vector. A *Pml*I/*Pac*I (New England Biolabs, Beverly, MA) digest followed by dephosphorylation using Calf Intestinal Phosphatase [CIP] (New England Biolabs) prepares the vector for ligation. Fusion proteins expressed from this vector compare favorably with larger tags, in terms of expression levels, while the small size typically eliminates the need for additional proteolytic cleavage and subsequent purification for crystallography.

The forward oligonucleotides (IDT, Coralville, IA) were designed such that the 5′-ends start with the first codon of each target reading frame. The reverse oligonucleotides were designed to include a *Pac*I restriction site, a 6-base insert-specific sequence and a stop codon along with complementary sequence to the 3′-end of the target reading frame (5′-CTC TTA ATT AAG TCG CG-Stop Codon-End of Reading Frame-3′). The 8-base recognition sequence of *Pac*I minimized the probability of cleavage inside the PCR product being processed for ligation. Oligonucleotides were phosphorylated with T4 Polynucleotide Kinase (Invitrogen, Carlsbad, CA) prior to PCR to facilitate subsequent ligation into CIP-treated pMH4.

Methods

Most of the following steps were performed using our integrated liquid and plate handling work environment (MWG Biotech, High Point, NC). PCRs, gel loading, digests, ligations, fluorescence assays and sample re-racking can all be done with this system (Figure 1a). Processing 96 targets from oligonucleotides and template to sequence submission takes 3 days (Figure 1b). Staggering batches of 96 targets allows up to 384 cloning attempts per week.

Open reading frames were amplified from either bacterial genomic DNA (ATCC, Manassas, VA) or mouse cDNA (MGC) templates using *PfuTurbo* Hotstart DNA Polymerase (Stratagene, La Jolla, CA). Cycling conditions were an initial 5 min denaturation at 95 °C followed by 30 cycles of 95 °C for 30 s, 55 °C for 45 s and 75 °C for 3 min. PCR success was determined using the E-gel 96 High-Throughput Agarose Electrophoresis System (Invitrogen) and manipulated into notebook-ready format using the manufacturer's included software application, E-Editor.

PCR products were purified to remove buffer, oligonucleotides, unincorporated dNTPs and polymerase using QIAquick 96 Purification Kits (QIAGEN, Valencia, CA) as recommended by the manufacturer. The 3′-ends were digested with *Pac*I. T4 DNA Ligase (Invitrogen) was used to ligate *Pac*I-digested inserts into the *Pml*I/*Pac*I/CIP-treated vector. The expression strain HK100 (JCSG), an enhanced, phage-resistant derivative of GeneHogs (Invitrogen), was transformed with the ligated DNA.

Four colonies were picked from each target transformation and cultured overnight. Colony dPCR used 3 μl of cells from the overnight cultures as templates for 50 μl reactions. A vector-specific, forward-directed oligonucleotide was paired with an insert-specific, reverse-directed oligonucleotide for primers in the dPCR. The nature of this oligonucleotide pairing was such that amplification only occurred when a target reading frame had correctly ligated into the vector.

A simple, microtiter plate, fluorescence assay exploited the conditional nature of the amplification in these dPCRs. Since the fluorescence of the dye SYBR Gold (Invitrogen) is quantifiably enhanced upon binding to dsDNA [6] or accumulated PCR product, this dye could be used to identify insert-containing plasmids in a microtiter plate format in lieu of traditional agarose gel analysis. First, the SYBR Gold dye was diluted 500-fold in 10 mM Tris, pH 7.5. Then, 5 μl samples from the 96-well plate of conditional dPCRs were mixed with 200 μl of the diluted SYBR Gold dye in the corresponding wells of a clear, flat-bottom, 96-well microtiter plate. Finally, the plate was assayed for fluorescence (Ex:485 nm/Em:525 nm).

The assay data was imported into a Microsoft Excel spreadsheet for analysis. Two assay-positive well locations were selected and sorted into re-rack files for each target automatically. Corresponding dPCR products were re-racked and treated with an Exo/SAP cocktail prior to submission for DNA sequencing. The Exo/SAP reaction involved mixing 5 μl (4 μl water, 0.5 μl Exo I, 0.5 μl SAP) of the Exo/SAP cocktail with 25 μl of dPCR product and incubating at 37 °C

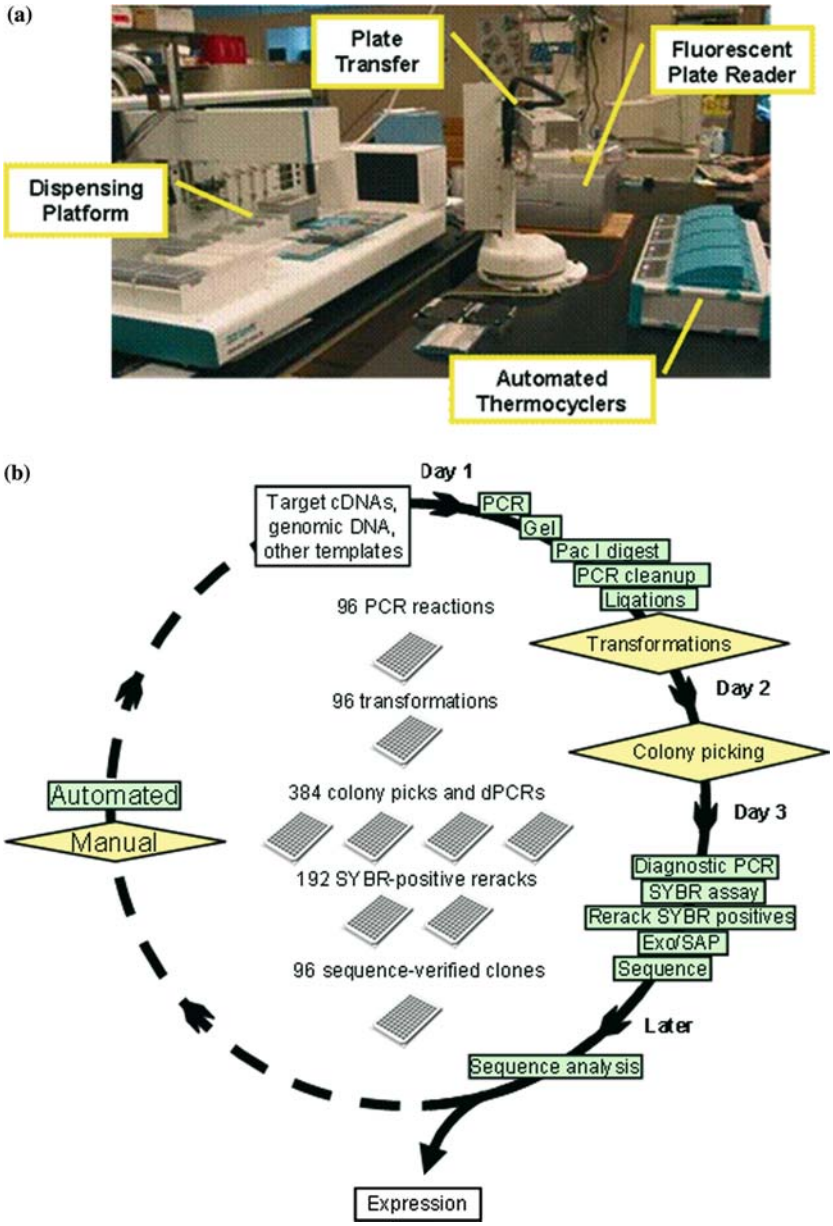

Figure 1. The semi-automated cloning environment. (a) Integrated MWG Biotech robotics. A R16 arm moves multi-well plates between liquid handling, plate reading and thermocycling workstations. (b) The JCSG semi-automated cloning process for each plate of 96 targets and the workload required at various steps. Automated (rectangle) and manual (diamond) steps are indicated.

for 30 min before heat-inactivating the enzymes at 75 °C for 15 min.

Clone analysis

Sequencing was performed using an ABI 3730 Sequencer (Applied Biosystems, Foster City, CA). Trace files were converted to FastA format by the application SeqManII (DNAStar, Madison, WI) and processed with CloneCompare (K. Ching, unpublished results). CloneCompare is a Perl script which searched for either the encoded 6thio/6his tag or the 3′ three-frame stop and *Pac*I site of the vector and then compared the adjacent, downstream sequence against a precompiled database of clone candidates via BLAST [7]. Correct clones are sorted from clones containing frameshifts and/or deletions.

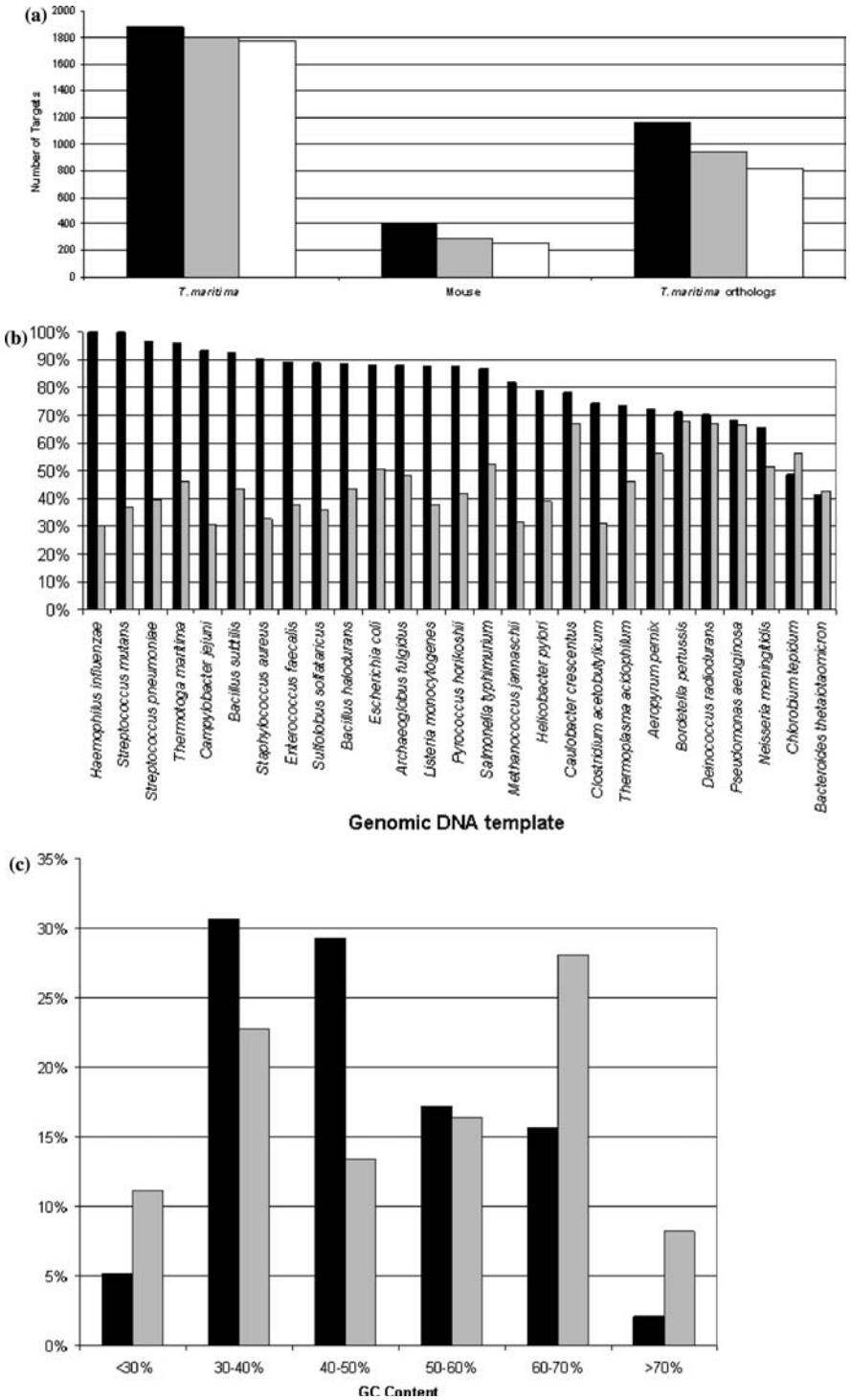

Figure 2. Cloning results and trends. (a) The number of total targets (black) is compared to the number of targets that were amplified by PCR (gray) and to the number of targets that became sequence-verified clones (white). (b) Comparison of the PCR amplification success rates using various bacterial genomic DNA templates (black) vs. the respective genomic %GC content of the DNA templates (gray). (c) Comparison of the %GC content distribution of all of the *T. maritima* ortholog targets where PCR amplification was attempted (black) vs. the %GC content distribution of only those targets that failed PCR amplification (gray).

Results

We were able to generate expression-ready clones for mouse and *T. maritima* targets as well as for orthologues of *T. maritima* genes from other bacterial genomes (Figure 2a). The use of automation and multi-well formats compelled us to use standardized amplification and cloning protocols. These protocols were optimized for the *T. maritima* project but were also used in subsequent projects. About 1800 of 1877 (96%) *T. maritima* targets amplified, while 1777 (99% of the amplified targets) of these PCR products were successfully cloned. Of 400 mouse targets attempted, 287 amplified successfully using mouse cDNA templates with 256 (89% of the amplified targets) resulting in clones. The *T. maritima* ortholog project required target amplification from 26 bacterial genomic DNA templates. The Bioinformatics Core of the JCSG selected these targets based on their homology to *T. maritima* targets where structure determination was incomplete. Here, 935 of the 1156 (81%) targets were amplified and 810 (87% of the amplified targets) were cloned correctly. Amplification success rates varied greatly across the various genomic DNA templates used, ranging from 41% to 100%. Cloning success rates of those targets which amplified from within a given organism ranged from 47% to 100%. However, cloning success rate did not correlate to the amplification success rate. In total, 2843 sequence-positive clones were obtained from 3022 PCR products amplified from 3433 targets. The robustness of the system was demonstrated by the level of success observed across this diverse set of projects.

We observed a 10:1 difference in fluorescence between amplified and non-amplified samples using the fluorescence assay for conditional dPCR analysis. Correlation of the fluorescence assay to standard gel analysis was 80–90% correct with no false-negatives observed across several comparison studies (data not shown). This method proved to be an efficient and reliable alternative to agarose gel electrophoresis.

Discussion

Using the automated system described, our clone-generating throughput was enhanced while tedious, labor-intensive manipulations were diminished. All PCRs were setup and processed robotically. Gel electrophoresis was performed on the initial PCRs using a robotically loaded, precast E-gel96. The process from loading the gel to obtaining the results took only 30 min. Colony screening was greatly facilitated by the fluorescence assay. Since our dPCRs worked only when an insert was present in the vector, the conditional enhancement of SYBR Gold fluorescence allowed dPCR success to be determined in a simple microtiter plate format instead of by gel electrophoresis and hand-scoring of gel images. Automated spreadsheet data processing of the insert-positive samples into re-rack files was of additional value.

As our data set of target successes and failures grew, we began to uncover the tendencies and biases of our approach. Because our strategy was optimized around the cloning of genes from *T. maritima*, we could expect to find continued success using other genomic DNA templates with similar characteristics. Indeed, if we sorted bacterial templates based on observed amplification success rates (Figure 2b), we found a general trend of greater success when using templates with GC-contents in the 30–50% range (GC-content of *T. maritima* is 46.1%). This trend was more apparent when analyzing the distribution of the GC-content of the specific target genes actually attempted within all of the bacterial templates (Figure 2c). We observed that the 60% of targets that fell within the 30–50% GC-content range accounted for only 36% of the amplification failures, while the 23% of targets below 30% or above 60% in GC-content claimed 47% of the failures. We also noted that the average attempted target size was 853 ± 456 base pairs while the average failed target size was 1052 ± 484 base pairs suggesting a slight bias towards smaller targets using our standard conditions.

These results will allow us to plan future cloning projects such that, where appropriate, either organisms that have shown greater amplification success or targets that are within the 30–50% GC-content range are emphasized in target selection and prioritized in the work flow. Additionally, we can devise alternate protocols that improve amplification success in organisms and targets that are difficult using our current

approach. Undoubtedly, we will uncover other trends that give insight into the causes of specific failures and use these trends to direct salvage pathways. Since cloning success was independent of amplification success and GC content, we will have the opportunity to improve parameters downstream of amplification as well.

These results illustrate the current capacity of this semi-automated approach to provide ample quantities of expression-ready clones to our structure determination pipeline. Since cloning efficiency seems to remain relatively stable across projects, we can expect proportionate output growth with throughput increase.

Acknowledgements

The authors thank Keith Ching, Tony Orth, Michael DiDonato, Daniel McMullan and Jaume Canaves for their various contributions. This work was supported by NIH grant PSI P50 GM62411.

References

1. Lesley, S., Kuhn, P., Godzik, A., Deacon, A., Mathews, I., Kreusch, A., Spraggon, G., Klock, H., McMullan, D., Shin, T., Vincent, J., Robb, A., Brinen, L., Miller, M., McPhillips, T., Miller, M., Scheibe, D., Canaves, J., Guda, C., Jaroszewski, L., Selby, T., Elsliger, M., Wooley, J., Taylor, S., Hodgson, K., Wilson, I., Schultz, P. and Stevens, R. (2002) *Proc. Natl. Acad. Sci.* **99**, 11664–11669.
2. Nordström, T., Nourizad, K., Ronaghi, M. and Nyrén, P. (2000) *Anal. Biochem.* **282**, 186–193.
3. Hartley, J., Temple, G. and Brasch, M. (2000) *Genome. Res.* **10**, 1788–1795.
4. Aslanidis, C. and de Jong, P. (1990) *Nucleic Acids Res.* **18**, 6069–6074.
5. DiDonato, M., Deacon, A., Klock, H., McMullan, D. and Lesley, S. (2004) *J. Struct. Funct. Genom.* **5**, 133–146.
6. Tuma, R., Beaudet, M., Jin, X., Jones, L., Cheung, C., Yue, S. and Singer, V. (1999) *Anal. Biochem.* **268**, 278–288.
7. Altschul, S., Madden, T., Schäffer, A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. (1997) *Nucleic Acids Res.* **25**, 3389–3402.