Published online March 14, 2005

Nucleic Acids Research, 2005, Vol. 33, No. 5 ***e49***
doi:10.1093/nar/gni049

A dual-fluorescence reporter system for high-throughput clone characterization and selection by cell sorting

Juno Choe[1,2], Haiwei H. Guo[3] and Ger van den Engh[1,*]

[1]Institute for Systems Biology, 1441 North 34th Street, Seattle, WA 98103-8904, USA, [2]Department of Genome Sciences and [3]Department of Pathology, University of Washington, Seattle, WA 98195, USA

Received November 3, 2004; Revised February 14, 2005; Accepted February 22, 2005

ABSTRACT

Molecular biology critically depends upon the isolation of desired DNA sequences. Flow cytometry, with its capacity to interrogate and sort more than 50 000 cells/s, shows great potential to expedite clone characterization and isolation. Intrinsic heterogeneity of protein expression levels in cells limits the utility of single fluorescent reporters for cell-sorting. Here, we report a novel dual-fluorescence strategy that overcomes the inherent limitations of single reporter systems by controlling for expression variability. We demonstrate a dual-reporter system using the green fluorescent protein (*GFP*) gene fused to the Discosoma red fluorescent protein (*DsRed*) gene. The system reports the successful insertion of foreign DNA with the loss of DsRed fluorescence and the maintenance of GFP fluorescence. Single cells containing inserts are readily recognized by their altered ratios of green to red fluorescence and separated using a high-speed cell-sorter for further processing. This novel reporter system and vector were successfully validated by shotgun library construction, cloned sequence isolation, PCR amplification and DNA sequencing of cloned inserts from bacteria after cell-sorting. This simple, robust system can also be adapted for diverse biosensor assays and is amenable to miniaturization. We demonstrated that dual-fluorescence reporting coupled with high-speed cell-sorting provides a more efficient alternative to traditional methods of clone isolation.

INTRODUCTION

Many investigations in modern molecular biology require the construction of large DNA libraries and the subsequent isolation of clones of interest. The sequencing of the human genome necessitated the evaluation of a large number (in the order of 10^8) of clones (1). As many other genomes are sequenced, it is expected that the demand for high-throughput cloning capabilities will increase. Additionally, many gene regulatory networks and protein–protein interactions are being elucidated, often by means requiring the handling of thousands of individual clones. Within directed molecular evolution, improved gene products for medical and industrial purposes are being generated from large diverse mutagenesis libraries (2–4). All of these processes depend critically upon the creation of random libraries as well as rapid and reliable screening methods.

Library generation followed by clone selection is effective if it can be conducted inexpensively and on a large scale. In a traditional experiment, libraries of DNA fragments are cloned into bacterial expression vectors, transformed into bacteria and grown on solid media. A phenotypic marker identifies colonies that are derived from bacterial cells with desirable inserts. The commonly used lacZ system uses blue or white color to distinguish colonies of interest in the presence of X-gal (5). For high-throughput applications, clone picking robots are used to mechanically transfer individual colonies into multi-well plates.

Over the past 30 years, fluorescence activated cell-sorting (FACS) machines have evolved to become powerful tools for analyzing and isolating single cells at very high rates (6). If the properties of recombinant DNA can be expressed as fluorescent tags in carrier cells, cell-sorters can isolate desirable cells at very high rates. Cell-sorters are also adept at depositing small volumes of liquid precisely within individual wells for culturing or DNA amplification (7,8). FACS of reporter cell populations also constitutes a powerful quantitative analytical system. Each individual cell is, in effect, a single microplate well (6). Modern high-speed cell-sorters are currently capable of quantifying multicolor fluorescence and making sort decisions at the rates of up to 50 000 cells/s (6). Therefore, extending the analogy, a single cell-sorter can effectively examine and select from ~15 million 96-well microtiter plates per work-shift. Cell-sorting could become an extremely

---

*To whom correspondence should be addressed. Tel: +1 206 364 3400; Fax: +1 206 364 346; Email: engh@systemsbiology.org

© The Author 2005. Published by Oxford University Press. All rights reserved.

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oupjournals.org

efficient tool for screening large DNA libraries and reporter cell populations.

In order to use cell-sorting for library screening, fluorescent marker reporting strategies need to be further developed. Traditional sequencing vectors based on lacZ activity are incompatible with cell-sorters. Several groups have previously created cloning vectors containing fluorescent reporters. Inouye *et al*. (9) developed a plasmid vector that directs the production of green fluorescent protein (GFP) under a constitutive promoter. When an insert is present, the successful translation of GFP is prevented, and colonies lack fluorescence. Similarly, Roessner and Scott (10) developed a plasmid vector containing the uroporphyrinogen III methyltransferase (*cobA*) marker gene that produces a fluorescent product within *Escherichia coli* when an insert is not present. These strategies work well for colonies grown on solid media, but negative screening (the absence of fluorescence) turns out to be a poor criterion for selection using cell-sorters. Clonal *E.coli* populations exhibit substantial stochastic variations in phenotype (11). The degree of cell-to-cell variability in protein expression can make it impossible to distinguish between weak fluorescent cells and non-fluorescent cells. Background particles inherent to culture media are also weakly fluorescent and are difficult to segregate from non-fluorescent bacteria. For sorting to be a viable approach, desired bacterial cells must have a positive fluorescent signal.

In this study, we demonstrate a dual-fluorescence system that controls the intrinsic variations in reporter protein expression. The novel vector pGRFP was developed to report the integration of cloned inserts by shifting the ratio of green to red fluorescence. In the native vector, a fusion protein composed of GFP and rapidly maturing DsRed-T3 fluorescent protein (12) is expressed. Cloning of DNA within the linker region between GFP and DsRed abrogates DsRed translation and results in enhanced green to red fluorescence ratio within transformed *E.coli*. In this paper, this approach is successfully validated in real-world applications using DNA from the sea urchin *Strongylocentrotus purpuratus*. The described dual-fluorescence method allows for accurate and efficient clone characterization using flow cytometry. The adaptability of this method to a diverse range of biological assays could expand its role to an even greater number of biological investigations.

MATERIALS AND METHODS

Vector construction

The pGRFP vector is shown schematically in Figure 1A. A 1990 bp pUC backbone containing the origin of replication, an AmpR marker and a lac promoter was PCR amplified. A consensus *E.coli* Shine–Dalgarno sequence (13–15) and a 7 bp spacer were introduced using a $5'$ tail on the reverse primer. All vector construction PCRs were performed using *Pfu* Turbo (Stratagene). The PCR product was cleaned, cut with AatII (NEB) and treated with calf intestinal alkaline phosphatase (CIAP; NEB). The *GFPmut3.1* gene (16) was PCR amplified from pGFPmut3.1 (Clontech) excluding the stop codon. The reverse primer introduced an additional 56 bp tail containing a six amino acid linker sequence, M13 forward priming site, stop codon, BsmI site and AatII site. The 868 bp phosphorylated PCR product was cut with AatII.

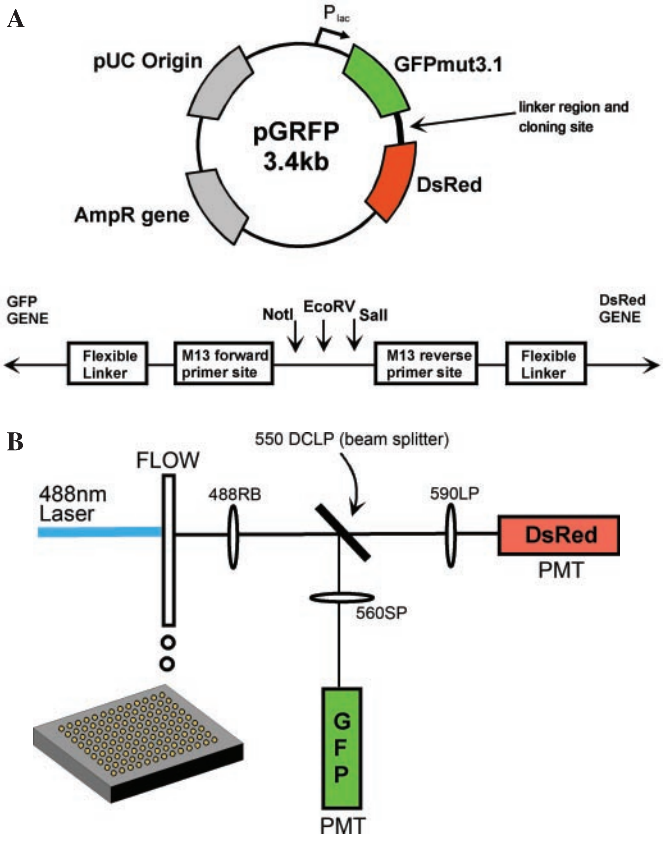

Figure 1. (**A**) The components of the pGRFP plasmid vector. The main components are the pUC origin of replication, ampicillin resistance marker and a fused GFP-DsRed gene separated by a linker. The linker region is shown with six amino acid linkers (SGSGSG and GSGSGS) on either side, M13 forward and reverse priming sites, and EcoRV, NotI and SalI sites. (**B**) Flow-cytometry configuration for dual-fluorescence quantification and sorting. A 488 nm laser excites the fluorescent proteins in individual *E.coli* suspended in the flow stream. The flow cytometer is configured to trigger either on forward scatter or GFP fluorescence. The fluorescence is split using a 550 nm dichroic long pass beam splitter. The green fluorescence is filtered through a 560 nm short pass filter before detection. The red fluorescence passes through a 590 nm long pass filter before detection.

Both DNA fragments were gel purified and ligated together with T4 DNA Ligase (Invitrogen) to generate a precursor vector. The ligation products were electrotransformed into Electromax DH10B cells (Invitrogen) and plated onto Luria–Bertani (LB)–ampicillin. Colonies expressing GFP were picked, miniprepped (Qiagen) and partially sequenced. A clone with expected sequence was digested with BsmI (NEB) to destroy the GFP stop codon, treated with Mung Bean nuclease (NEB) and $5'$ dephosphorylated.

To construct the DsRed portion, the rapidly maturing *DsRed-T3* mutant (12,17) gene was PCR amplified from vector obtained from Bevis and Glick (12). The upstream primer introduced NotI, EcoRV and SalI sites as well as an M13 reverse priming site and an additional six amino acid linker. The downstream primer contained a StuI site after the DsRed stop codon. The PCR product was cleaned and treated with DpnI (NEB). The product was further gel purified and blunt ligated to the precursor vector. Transformed colonies expressing both GFP and DsRed were selected by fluorescence microscopy after 24 h growth. The pGRFP vector was further sequence confirmed, and the sequence has been deposited in GenBank (accession no. AY916793).

Library construction

The SU66E20 bacterial artificial chromosome (BAC) from the sea urchin *S.purpuratus* was prepared using the Large Construct Plasmid Isolation kit (Qiagen). The template DNA was sonicated and end repaired with T4 DNA polymerase (Fermentas), Klenow fragment (Fermentas) and T4 polynucleotide kinase (Fermentas). Fragments of approximately 1.5–3 or 2–4 kb were isolated by gel purification. The pGRFP plasmid was cut with EcoRV (NEB), $5'$ dephosphorylated with CIAP, and gel purified. The linearized vector was blunt-end ligated to the SU66E20 BAC library inserts using T4 DNA ligase.

Preparation of cells

Ligation products were electrotransformed into Electromax DH10B cells and incubated in 1 ml SOC for 1 h. An aliquot of 5 ml of fresh LB–carbenicillin (50 μg/ml) was inoculated with 100 μl of the SOC mixture. Cells were grown at 30°C with shaking for 24 h. Approximately 10–50 μl of culture was added to 4 ml of 0.9% NaCl (saline) for flow-cytometric analysis.

Fluorescence microscopy

A culture transformed with pGRFP harboring ligated sea urchin DNA was harvested for cells that were fixed on a coverslip with 10% formaldehyde. Cells were photographed using a Zeiss Axiophot fluorescence microscope with attached Nikon digital camera. Illumination was by arc lamp light filtered to the blue range (450–490 nm).

Flow cytometry

Flow cytometry was performed using an inFlux flow cytometer (Cytopeia). Samples were injected into a sterile saline flow stream. The sample cells were analyzed for green and red fluorescence according to the set-up shown in Figure 1B. Excitation was by 5 W Innova argon laser (Coherent) at 488 nm emission with ~400 mW power. The flow cytometer was configured to trigger on the forward scatter channel. Emission fluorescence was passed through a 488 nm rejection band filter and split using a 550 nm dichroic long pass beam splitter. The green fluorescence was filtered through a 560 nm short pass filter before measurement by using a Hamamatsu PMT. The red fluorescence was passed through a 590 nm long pass filter before measurement. Red signal intensity was plotted versus green for each cell on linear bivariate dot plots. Cells displaying low red-to-green fluorescence ratio were selected for sorting. The flow cytometer was programmed to deposit a single cell into each well of a 96-well plate containing 200 μl LB–carbenicillin (50 μg/ml).

To establish enrichment of insert-bearing clones after cell-sorting, individual cells were sorted onto LB–carbenicillin plates in a grid pattern. Cells were either selected GFP+/DsRed− cells or randomly selected cells from the original culture. After 24 h growth at 37°C, plates were counted to determine the number of DsRed+ and DsRed− colonies.

PCR amplification

Shotgun library inserts from SU66E20 in the 1.5–3 kb size range were cloned into pGRFP. From the single sorted cells 46 cultures were grown. An aliquot of 5 μl from each culture was diluted into 50 μl ddH_2O and heated to 95°C for 5 min. As a template for subsequent PCRs 1 μl of lysate was used. *Taq* polymerase (Promega) was used according to the manufacturer's recommended protocol with forward (TGTAAAACGACGGCCAGT) and reverse M13 primers (CAGGAAACAGCTATGACC) flanking the linker region of pGRFP. Thermocycle conditions were as follows: 95°C for 5 min; and 35 cycles of 95°C for 45 s, 51°C for 45 s, 72°C for 3 min 30 s; and a final extension step of 72°C for 5 min. PCR products were photographed after gel electrophoresis.

BAC sequencing

Shotgun library inserts from SU66E20 in the 2–4 kb size range were cloned into pGRFP. Single cells were sorted into 96-well plates as described above. After overnight growth, 1 μl of cultures were used as templates for rolling circle amplification (RCA) using TempliPhi DNA Sequencing Template Amplification kits (Amersham) according to the manufacturer's recommended protocol. RCA reactions were performed in 10 μl volume at 30°C for 18 h and then stopped by heating to 65°C for 10 min. Products were dephosphorylated using 1 U shrimp alkaline phosphatase at 37°C for 45 min with heat inactivation at 80°C for 10 min. An aliquot of 1 μl of the product was used as template for Big Dye Terminator v3.0 (Perkin-Elmer) sequence reactions with 4 pmol of either forward or reverse M13 primers. Sequences were analyzed and assembled using Phred/Phrap (18) and the assembly was viewed using Consed (19).

RESULTS

Single reporter

Initial experiments using single fluorescent proteins were limited by large inter-bacterial variations in fluorescence intensity. Even when picked from a single colony on a plate, individual bacteria showed large variations in GFP concentration and occasionally bacterial cell size. Figure 3A shows a typical flow cytometry dot plot of *E.coli* expressing GFP from a constitutive promoter. The spread of fluorescence intensities reflects stochastic cell-to-cell variations, despite the cells' clonal nature. When using only a single fluorescent protein, large ambiguities exist owing to the large degree of inherent variation in expression. These observations are consistent with earlier studies in clonal *E.coli* populations (11).

Identification of insert bearing clones using dual-fluorescence reporting system

The problems associated with single fluorochrome markers can be circumvented with a dual-reporter system. We constructed the pGRFP vector that expresses two translationally fused fluorescent proteins (Figure 1A). The linker region between the two proteins contains several cloning sites. Without an insert, the uninterrupted protein linker will lead to a hybrid protein consisting of GFP and DsRed fluorochrome groups. On the other hand, insertion of a DNA fragment into the linker region will introduce stop codons so that these clones will express only GFP. Excitation then produces an intense green fluorescence signal.

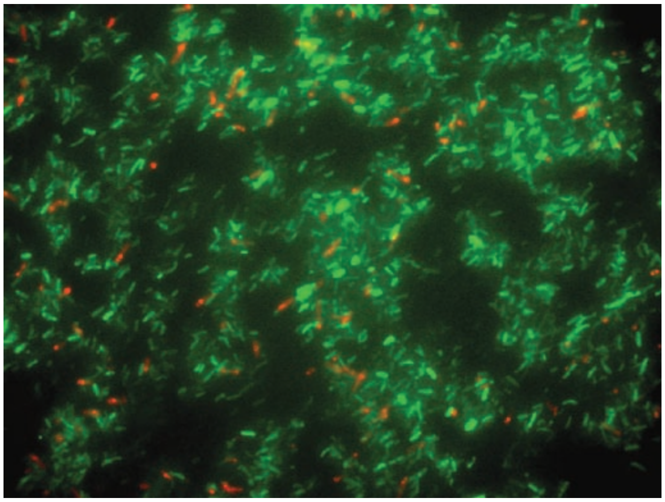

Figure 2. Fluorescence microscopy image of *E.coli* containing pGRFP with cloned BAC library. The green cells are GFP+/DsRed− and contain plasmids with successfully inserted fragments from the library. The orange colored cells are GFP+/DsRed+ and contain native pGRFP vector without inserts.

This approach was validated by subcloning a sea urchin BAC library into pGRFP. Figure 2 shows an image of fluorescence from transformed *E.coli* excited by 450–490 nm light. The cells containing native pGRFP vector appear orange, because they express both GFP and DsRed. The insert-containing cells are readily recognized by their green color. There is little ambiguity about the presence or absence of inserts in each cell in the image.

Cells were further analyzed using flow cytometry. Figure 3B and C show bivariate dot plots of DsRed versus GFP fluorescence for each analyzed sample. Native pGRFP vector yields cells with a high DsRed to GFP ratio (Figure 3B). When foreign DNA is cloned into pGRFP, a second population of cells appear that does not have red fluorescence (Figure 3C). This population is equivalent to the green cells shown in Figure 2 and represents clones with successfully integrated inserts. A large separation in the two populations is observed by the flow cytometer according to the DsRed/GFP fluorescence ratio resulting in a high degree of certainty when discriminating between them.

Reliability of sorting assessed by PCR amplification

Cloned DNA inserts were PCR amplified from 46 single sorted cells (see Figure 4). All the 46 PCR products indicate the presence of cloned inserts. A total of 39 clones had inserts in the 1.5–3 kb size range of the original library, while 7 clones had inserts in the 400–1500 bp range. This experiment demonstrated that the GFP+/DsRed− phenotype has a 100% predictive value in detecting the presence of inserts.

We estimated the initial fraction of insert containing clones in culture by analyzing colonies grown from single sorted cells. In plates containing random cells deposited from the initial culture, a total of 588 colonies with 536 DsRed− colonies and 52 DsRed+ colonies were observed. In plates containing cells sorted for the GFP+/DsRed− phenotype, 301 DsRed− colonies and one DsRed+ colony were observed. This indicates a sorting enrichment factor of ∼29.2 for DsRed− cells. Given this enrichment ratio, it is estimated that sorting from a culture with 90.0% GFP+/DsRed− cells will yield nearly 100% cells with the correct phenotype. If sorting from a culture containing only 10.0% GFP+/DsRed− cells, possibly as a result of inefficient cloning, it is expected that 96.8% of sorted cells will have the desired phenotype.

BAC sequencing

We demonstrated the performance of the pGRFP dual-fluorescence reporter system in addressing a biological question of interest. We sequenced a small portion of the genome of the sea urchin *S.purpuratus* as represented in BAC SU66E20. We subcloned this BAC, selected for insert-bearing clones by cell sorter, amplified plasmid DNA and sequenced the inserts. The average Phred20 score sequence read-length was ∼350 bases. Given the number of sequences used, coverage of ∼4.3X was achieved. Thirteen primary contigs were assembled that spanned 55 kb of sequence out of a total BAC length of 59 kb. Furthermore, sequenced clones were evenly spread throughout the SU66E20 BAC sequence. This level of successful assembly is expected, given the level of coverage and average sequence read-lengths. The results are similar to what would be achieved using traditional clone isolation and sequencing protocols.

We analyzed the final assembly data to evaluate the insert lengths of our flow-sorted clones. The locations of the forward and reverse reads within the final BAC assembly reveal the distance between these reads. We analyzed 113 clones that had both forward and reverse reads residing on either of the two largest contigs (13.0 and 7.3 kb in length). The mean insert length was 2.95 kb with a standard deviation of 860 bp. These insert sizes are well within the 2–4 kb range that was isolated during library construction. Six percent of the inserts were smaller than 2 kb in size. The insert sizes in the final assembly approximate the insert sizes of the original library.

DISCUSSION

In this study, we report the development of the dual-fluorescence approach to clone characterization and isolation. The flow-cytometry data support our proposed mechanism of action of two linked fluorescent proteins. Bacteria containing the pGRFP vector exhibit both green and red fluorescence when excited with 488 nm light (Figure 3B). Integration of foreign DNA into cloning sites between GFP and DsRed results in premature termination before DsRed is translated. This termination results in bacteria with primarily green fluorescence, represented as a second population of cells (Figure 3C). Despite the fact that bacteria have large cell-to-cell variations in fluorescence, GFP acts as an internal control to interpret the level of DsRed fluorescence. In this manner, two distinct populations can be easily distinguished.

The pGRFP vector performed very well in clone selection experiments. All the isolated GFP+/DsRed− bacteria (100%) had inserts amplifiable using PCR. Furthermore, analysis of colonies grown from single sorted cells indicated that cell-sorters can enrich for GFP+/DsRed− cells by at least a factor of 29.2. Given this enrichment factor, we estimate that cell-sorters could isolate cells containing inserts 96.8–100% of the time from cultures originally containing 10.0–90.0% insert bearing cells. The use of dual-fluorescence reporting with commercially available cell-sorters is capable of accurately isolating large numbers of insert containing clones from complex libraries.

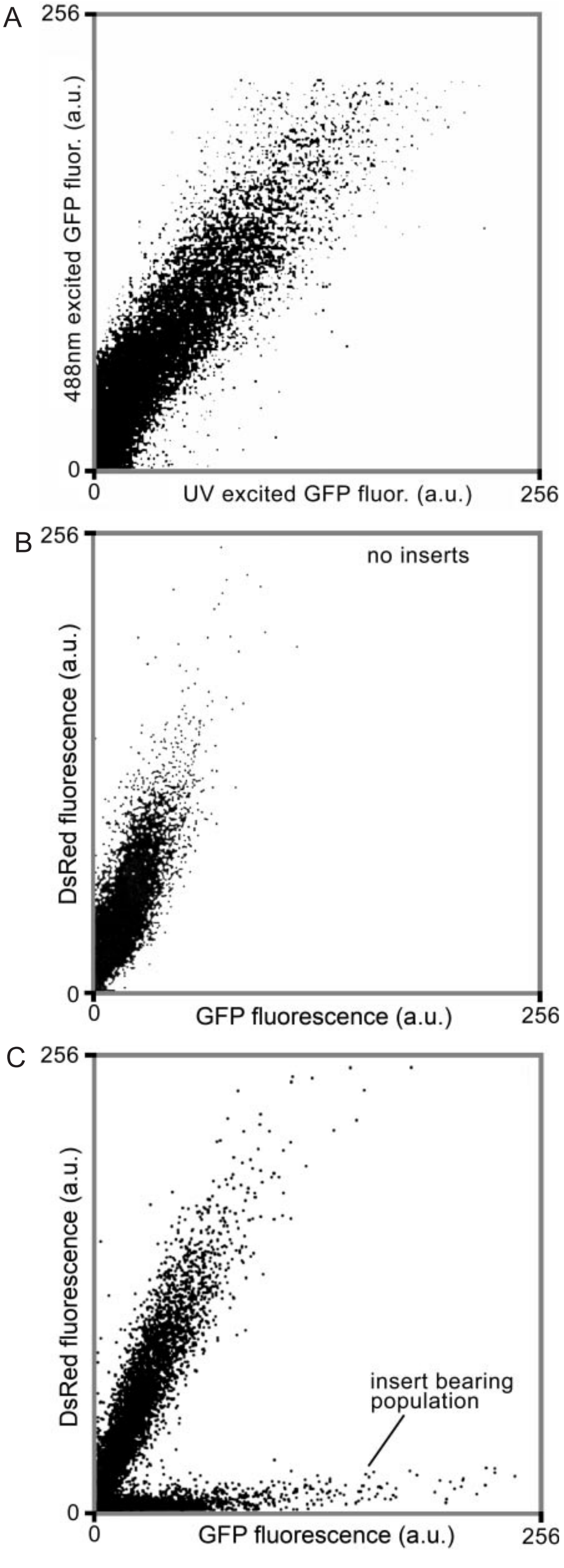

Figure 3. (**A**) Flow-cytometry bivariate dot plot of *E.coli* expressing GFPmut3.1 from a constitutive promoter. Cells were picked from a single colony and cultured overnight. Because of the large cell-to-cell variation of fluorescence from a single fluorophore, it is difficult to use a single fluorophore as an indication of the presence or absence of a cloned insert. (**B**) Dot plot shows a population of *E.coli* containing native pGRFP. All cells are observed expressing both GFP and DsRed. (**C**) Dot plot showing a *S.purpuratus* BAC library cloned into the pGRFP vector. A second population of cells that are GFP+/DsRed− appears owing to loss of function of the DsRed half of the fusion protein. This population contains cells with successfully incorporated inserts. This reporter system clearly distinguishes between insert containing and non-insert containing *E.coli*.

The shotgun sequencing project demonstrates the usefulness of the dual-fluorescence reporter system in real-world applications. Partial assembly of the SU66E20 BAC was achieved after sequencing the BAC to 4.3X coverage. Furthermore, insert size was not biased during the culturing and cell-sorting processes. The demonstration of sequencing capability serves as strong validation of the robustness of this method for selecting clones that are carrying inserts. This reliability is an important factor during any large-scale biological investigation.

Modern genomic studies require the processing of large numbers of samples. To date, advantages of scale have been obtained through the increasing use of automation. The development of clone picking robots guided by video recognition technology was essential to the completion of the Human Genome Project. Here, we demonstrate a faster alternative. Clone isolation by cell-sorting is a unique technology that is only limited by the speed at which wells can be transported under the stream of sorted droplets. Assuming a conservative rate of 2 wells/s, it is possible to isolate 7200 clones/h using one cell-sorter; this is faster than the latest generation of colony picking robots. Using one cell-sorting apparatus, 172 800 clones could be isolated in each 24-h period. Assuming 500 bp forward and reverse sequence reads, upwards of 173 Mb of raw sequence data could be generated per day. At this rate, the human genome could be sequenced to greater than 5X coverage in less than 90 days from clones isolated by one cell-sorter. Even further rate increases could be achieved with the use of multiple sort streams or by increasing the well transport rate.

With further downstream instrumentation, clone selection by sorting could be a powerful strategy for high-throughput applications. After single cells are cultured, sample handling robots could transfer small culture volumes for further amplification. Use of rolling circle amplification with sub-microliter volumes of culture can quickly generate sequencing templates in thousands of wells simultaneously. In addition, advances in single cell amplification technologies could obviate the need for culturing altogether. Groups have demonstrated amplification from single DNA molecules using RCA techniques similar to those utilized here (20). Another method of RCA developed by Tabor and co-workers (21) is nearly able to amplify DNA from single cells (personal communication). The integration of cell sorting, amplification technologies and instrumentation represents an efficient and practical method for sequencing large number of clones.

Looking further, the rapid handling of samples could be achieved on microfabricated DNA processing chips. Flow cytometers could deposit cells into a landing zone on a chip, and sorted cells are then transported into individual

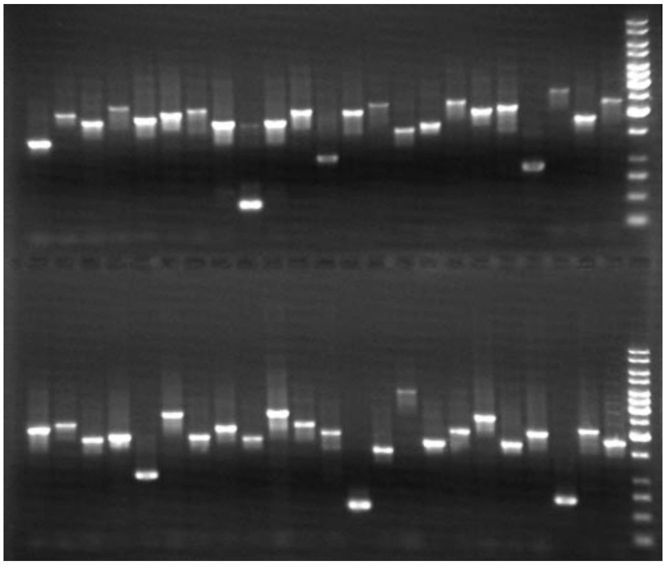

Figure 4. Gel image showing PCR amplifications of inserts from 46 cultures grown from single sorted GFP+/DsRed− cells containing members of a *S.purpuratus* shotgun library. All 46 cultures yielded a PCR product consistent with a cloned insert. The majority of these inserts were in the 1.5–3 kb size range selected during initial library construction.

reaction chambers. This transport has been achieved previously (22) and could occur at very fast rates. Cell lysis, DNA amplification and even DNA sequencing reactions can all be performed in individual chambers. Microfabricated chips have been previously constructed with the ability to perform thermocycling and sequencing using micro-capillary electrophoresis (23).

Beyond sequencing, the advantages offered by dual-fluorescence reporter systems make them attractive platforms for diverse biosensor based assays. For example, motifs responsive to *trans*-acting regulatory proteins can be inserted into the linker region between GFP and DsRed to serve as detectors for specific protein–nucleic acid interactions. Dual fluorescence can also be used to measure the frequency of mutational events. For example, repetitive DNA elements can be cloned between the fluorescent reporters to study DNA polymerase replication fidelity. After *in vitro* or *in vivo* replication of the reporter, the frequency of slippage events can be quantified by either loss or gain of fluorescence signal.

The pGRFP vectors can be reconfigured for other applications by putting the fluorescence proteins under the control of separate response elements and promoters that respond to environmental cues. For example, Axtell and Beattie (24) linked GFP under the control of the *E.coli proU* promoter, which is sensitive to solute concentration in environment surrounding the cell. Modified bacteria then provided rapid detection of water availability on plant leaves. Using in-field flow cytometers, dual or multiple fluorescent proteins could serve as efficient reporters of environmental conditions.

The dual-fluorescence approach can be further enhanced by the ability of flow cytometers to isolate rare sub-populations from a vast background of negatives, making cell-sorting a powerful technique for isolating rare events. For example, Koo *et al.* (25) were able to encode daa mRNA processing activity in *E.coli* as a loss of GFP signal. High-speed cell-sorters were used to examine 500 million cells before successfully isolating one rare *E.coli* mutant that had reduced daa mRNA processing ability.

Dual-fluorescence reporting lays the foundation for a large number of novel high-throughput processes based on flow cytometers. Hundreds of thousands of clones can be isolated every day from a diverse mixture of desirable clones, or a few exceedingly rare clones can be isolated from among millions of reporter cells. These methods create new opportunities for cell-sorting to address the genomics problems of today and tomorrow.

ACKNOWLEDGEMENTS

The authors wish to thank Edward Ramos (Fred Hutchinson Cancer Research Center), and Sherrif Ibrahim and Kathleen Kennedy (Institute for Systems Biology) for their invaluable help throughout this research effort and for many insightful discussions. This research was supported by Department of Energy Grant no. DE-FG03-00ER63051. Funding to pay the Open Access publication charges for this article was provided by the Institute for Systems Biology.

REFERENCES

1. Istrail,S., Sutton,G.G., Florea,L., Halpern,A.L., Mobarry,C.M., Lippert,R., Walenz,B., Shatkay,H., Dew,I., Miller,J.R. *et al.* (2004) Whole-genome shotgun assembly and comparison of human genome assemblies. *Proc. Natl Acad. Sci. USA*, **101**, 1916–1921.
2. Arnold,F.H. (2001) Combinatorial and computational challenges for biocatalyst design. *Nature*, **409**, 253–257.
3. Daugherty,P.S., Chen,G., Iverson,B.L. and Georgiou,G. (2000) Quantitative analysis of the effect of the mutation frequency on the affinity maturation of single chain Fv antibodies. *Proc. Natl Acad. Sci. USA*, **97**, 2029–2034.
4. Encell,L.P., Landis,D.M. and Loeb,L.A. (1999) Improving enzymes for cancer gene therapy. *Nat. Biotechnol.*, **17**, 143–147.
5. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
6. Wittrup,K.D. (2000) The single cell as a microplate well. *Nat. Biotechnol.*, **18**, 1039–1040.
7. Ibrahim,S.F. and van den Engh,G. (2003) High-speed cell sorting: fundamentals and recent advances. *Curr. Opin. Biotechnol.*, **14**, 5–12.
8. Durack,G. and Robinson,J.P. (2000) *Emerging Tools for Single-cell Analysis: Advances in Optical Measurement Technologies*. Wiley-Liss, NY.
9. Inouye,S., Ogawa,H., Yasuda,K., Umesono,K. and Tsuji,F.I. (1997) A bacterial cloning vector using a mutated Aequorea green fluorescent protein as an indicator. *Gene*, **189**, 159–162.
10. Roessner,C.A. and Scott,A.I. (1995) Fluorescence-based method for selection of recombinant plasmids. *Biotechniques*, **19**, 760–764.
11. Elowitz,M.B., Levine,A.J., Siggia,E.D. and Swain,P.S. (2002) Stochastic gene expression in a single cell. *Science*, **297**, 1183–1186.
12. Bevis,B.J. and Glick,B.S. (2002) Rapidly maturing variants of the Discosoma red fluorescent protein (DsRed). *Nat. Biotechnol.*, **20**, 83–87.
13. Chen,H., Bjerknes,M., Kumar,R. and Jay,E. (1994) Determination of the optimal aligned spacing between the Shine–Dalgarno sequence and the translation initiation codon of *Escherichia coli* mRNAs. *Nucleic Acids Res.*, **22**, 4953–4957.
14. Vellanoweth,R.L. and Rabinowitz,J.C. (1992) The influence of ribosome-binding-site elements on translational efficiency in *Bacillus subtilis* and *Escherichia coli in vivo*. *Mol. Microbiol.*, **6**, 1105–1114.
15. Tang,G.L., Wang,Y.F., Bao,J.S. and Chen,H.B. (1999) Overexpression in *Escherichia coli* and characterization of the chloroplast triosephosphate isomerase from spinach. *Protein Exp. Purif.*, **16**, 432–439.

16. Cormack,B.P., Valdivia,R.H. and Falkow,S. (1996) FACS-optimized mutants of the green fluorescent protein (GFP). *Gene*, **173**, 33–38.
17. Matz,M.V., Fradkov,A.F., Labas,Y.A., Savitsky,A.P., Zaraisky,A.G., Markelov,M.L. and Lukyanov,S.A. (1999) Fluorescent proteins from nonbioluminescent Anthozoa species. *Nat. Biotechnol.*, **17**, 969–973.
18. Ewing,B., Hillier,L., Wendl,M.C. and Green,P. (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.*, **8**, 175–185.
19. Gordon,D., Abajian,C. and Green,P. (1998) Consed: a graphical tool for sequence finishing. *Genome Res.*, **8**, 195–202.
20. Lizardi,P.M., Huang,X., Zhu,Z., Bray-Ward,P., Thomas,D.C. and Ward,D.C. (1998) Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nature Genet.*, **19**, 225–232.
21. Kato,M., Frick,D.N., Lee,J., Tabor,S., Richardson,C.C. and Ellenberger,T. (2001) A complex of the bacteriophage T7 primase-helicase and DNA polymerase directs primer utilization. *J. Biol. Chem.*, **276**, 21809–21820.
22. Thorsen,T., Maerkl,S.J. and Quake,S.R. (2002) Microfluidic large-scale integration. *Science*, **298**, 580–584.
23. Paegel,B.M., Blazej,R.G. and Mathies,R.A. (2003) Microfluidic devices for DNA sequencing: sample preparation and electrophoretic analysis. *Curr. Opin. Biotechnol.*, **14**, 42–50.
24. Axtell,C.A. and Beattie,G.A. (2002) Construction and characterization of a proU-GFP transcriptional fusion that measures water availability in a microbial habitat. *Appl. Environ. Microbiol.*, **68**, 4604–4612.
25. Koo,J.T., Choe,J. and Moseley,S.L. (2004) HrpA, a DEAH-box RNA helicase, is involved in mRNA processing of a fimbrial operon in *Escherichia coli*. *Mol. Microbiol.*, **52**, 1813–1826.