

The use of multiple displacement amplification to amplify complex DNA libraries

Melissa J. Fullwood, Jack J. S. Tan, Patrick W. P. Ng, Kuo Ping Chiu, Jun Liu, Chia Lin Wei and Yijun Ruan*

Genome Institute of Singapore, Agency for Science, Technology and Research (A*STAR), 60 Biopolis Street, Genome #02-01, Singapore 138672

Received December 14, 2007; Revised January 17, 2008; Accepted February 5, 2008

ABSTRACT

Complex libraries for genomic DNA and cDNA sequencing analyses are typically amplified using bacterial propagation. To reduce biases, large numbers of colonies are plated and scraped from solid-surface agar. This process is time consuming, tedious and limits scaling up. At the same time, multiple displacement amplification (MDA) has been recently developed as a method for *in vitro* amplification of DNA. However, MDA has no selection function for the removal of ligation multimers. We developed a novel method of briefly introducing ligation reactions into bacteria to select single insert DNA clones followed by MDA to amplify. We applied these methods to a Gene Identification Signatures with Paired-End diTags (GIS-PET) library, which is a complex transcriptome library created by pairing short tags from the 5' and 3' ends of cDNA fragments together, and demonstrated that this selection and amplification strategy is unbiased and efficient.

INTRODUCTION

A mainstay of genomic technologies to interrogate genomes and functional genomic elements is the generation of complex cloning-based DNA libraries. Examples of such libraries include genomic DNA libraries used in the sequencing of the human genome (1) as well as other genomes (2); full-length cDNA (fcdNA) libraries (3) and Gene Identification Signatures with Paired-End diTags (GIS-PET) libraries used for elucidating the transcriptome (4); as well as Chromatin Immunoprecipitation with

Paired-End diTags (ChIP-PET) libraries used for elucidating transcription factor binding sites (5).

In constructing such libraries, the starting DNA samples are often limited, and therefore DNA amplification is often necessary. The method of choice has been bacterial propagation of DNA fragments in plasmid vectors. To ensure accurate representation, the bacteria must not be allowed to compete with each other for nutrients. Therefore, growth and scraping from solid-surface agar is commonly used because colonies are spread out on solid-surface agar such that they will not encounter each other and compete. As the libraries are complex and contain many different DNA molecules, a large number of colonies must be scraped from the agar to ensure that the resulting library contains sufficient coverage of the different DNA molecules present in the original pool. Plating and scraping large numbers of solid-surface agar bacteria clones then results in methods that are tedious, time consuming and difficult to scale up.

Multiple displacement amplification (MDA) has been recently developed as a method for *in vitro* amplification of DNA. MDA is a method for amplifying plasmids and long strands of DNA in a cell-free system using phi29 polymerase, a newly discovered polymerase enzyme that has very high fidelity (6), proof-reading activity (7) and processivity (8). Such a system would be ideal for replacing the tedious solid-phase agar scraping steps used for the amplification of complex cloning-based libraries. The use of MDA would remove this bottleneck, as MDA is able to amplify complex mixtures with high accuracy and efficiency.

However, one obstacle to the use of MDA for the amplification of complex cloning-based libraries is the fact that cloning ligation reactions into vectors typically results in multimers of plasmid vectors and DNA fragments. Bacterial propagation can remove multimers because replication constructs that contain multiple origins of replication will not survive during bacterial replication,

*To whom correspondence should be addressed. Tel: (65) 6478 8073; Fax: (65) 6478 9059; Email: ruanyj@gis.a-star.edu.sg

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

© 2008 The Author(s)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.0/uk/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

while MDA alone is not capable of such selection to eliminate multimers during amplification.

To overcome this problem, we developed a method, called Selection-MDA, which combines the selection capability of bacterial replication for single vector/insert constructs with the efficiency and convenience of MDA. In this method, we first transfer the vector/insert ligation into electrocompetent *E. coli* for a short period of replication and selection in liquid media. Because the bacteria are harvested after a short period of growth in liquid media, the bacteria would not have multiplied to such an extent that they begin to compete for nutrients, yet plasmids with multiple origins of replication would be selected out. The multimer-free pool of plasmids is then purified from liquid media and used for MDA, which amplifies large quantities of multimer-free DNA, thus eliminating tedious and time-consuming plating and scraping of solid-surface agar. As such, the selective advantage of bacterial propagation can be combined with the efficiency convenience of the MDA method without the disadvantages of sample bias or chimeras. The end result is an MDA-amplified library of the same quality as a similar library prepared by bacterial propagation.

To validate the Selection-MDA method in a complex library, we prepared a GIS-PET library (4) with the Selection-MDA method, and compared it with the same library prepared by conventional bacterial amplification on solid surface agar (9). Short Paired-End diTag (PET) libraries, including GIS-PET, were conceived of in order to improve sequencing efficiency. In GIS-PET, the 5' and 3' signatures of each full-length cDNA are covalently linked into structures in which the 5' and 3' tags were paired together, and then sequenced, allowing a 20- to 30-fold increase in efficiency compared with bidirectional sequencing of DNA (10). The paired-end nature of the method also allows the use of GIS-PET to study unconventional fusion transcripts (11). The same concept has also been applied to ChIP DNA characterization (ChIP-PET) (5). The PET analysis method involves the construction of two libraries: the original DNA insert library (fcdDNA library for GIS-PET), and the single PET library, which is derived from the original DNA insert library. The amplification of the libraries using bacteria propagation is time consuming and labor intensive. To further improve PET analysis, we applied the Selection-MDA method to replace the single PET library amplification step.

MATERIALS AND METHODS

HES3 human embryonic stem (ES) cells were grown and prepared as described (9). Briefly, cells were obtained from ES Cell International, and cultured in a feeder-free medium. Flow cytometry analysis was used to ensure that cells were human ES cells.

A fcdDNA library was constructed from the human embryonic stem cells and PETs were prepared for sequencing as described in the classic bacterial propagation protocol (12). Briefly, RNA was isolated from HES3 cells (Figure 1A), and poly A+RNA was isolated from

RNA using the μ MACS mRNA isolation kit (Figure 1B). The poly A+RNA was converted into cDNA by oligo-dT-primed reverse transcription. RNA ends were biotinylated. Cap-trapper selection was performed to select full-length first strand cDNA. 5' adapters were added to prime for second strand cDNA synthesis, and the material was then digested to give rise to sticky ends for cloning. The fcdDNA was then ligated with pGIS4b vector cut with NotI (NEB) and GsuI (Fermentas). The fcdDNA library was amplified by bacterial amplification at 37°C on solid-surface agar Q-trays (Figure 1C) followed by scraping and plasmid extraction by Maxiprep (Qiagen).

An aliquot of the Maxiprep was used to prepare a GIS-PET library by the classic bacterial propagation GIS-PET protocol (12). Briefly, MmeI digestion was performed, and the single-PET plasmids were end-polished with T4 polymerase (Promega). The single-PET plasmids were then self-ligated and amplified by bacterial amplification at 37°C on solid-surface agar Q-trays (Figure 1C) followed by scraping and plasmid extraction by Maxiprep (Qiagen). Single PETs were released with BseRI, purified and concatenated. The concatemers were then blunted by T4 DNA polymerase (Promega), cloned into EcoRV-cut pZErO-1 vectors (Invitrogen) (Figure 1D), and 300 384-well plates were sequenced with Sanger capillary sequencing. This library was called SHE001. The library was analyzed, and the results were reported separately (9).

To construct the MDA-amplified library using the new Selection-MDA protocol (Figure 2), we took an aliquot of 8 ng of maxiprep from the GIS-PET full-length cDNA library and added it to 50 μ l of Templphi 500 sample buffer (GE Healthcare). The sample was denatured at 95°C for 3 min, and then cooled to 4°C. 2 μ l of Templphi 500 enzyme mix (GE Healthcare) was added to 50 μ l Templphi sample buffer on ice, and the mixture was then added to the 50 μ l sample buffer with denatured template. The reaction was incubated at 30°C for 18 h, and then heat inactivated at 65°C for 10 min. The material was quantitated with Picogreen Fluorimetry (Invitrogen), and an MmeI (New England Biolabs) digestion was performed following the Single PET construction method as described (12). 800 ng of self-ligation reaction was purified to remove salts before electroporation by phenol/chloroform isopropanol precipitation as described (12). The pellet was resuspended in 5 μ l of Elution Buffer (Qiagen). The entire ligation mix was transformed into 50 μ l of Top10 *E. coli* electrocompetent cells (Invitrogen) and recovered in 1 ml of Lucigen Recovery Medium (Lucigen) with shaking at 37°C for 4 h. Because recovery was for only 4 h, the bacteria would not have multiplied sufficiently so as to compete with each other; hence the library should contain no size bias. To monitor bacterial growth, the optical density at 600 nm (OD₆₀₀) of aliquots were taken at various time points by an ND-1000 spectrophotometer (Nanodrop). Cells were spun down at 10 000 g for 5 min and washed twice with 750 μ l of Lucigen Recovery Medium to remove free-floating DNA that was not introduced into the cells. Next, plasmids were extracted by performing Miniprep (Qiagen). 40 μ l of elution buffer was used for the elution, and the DNA

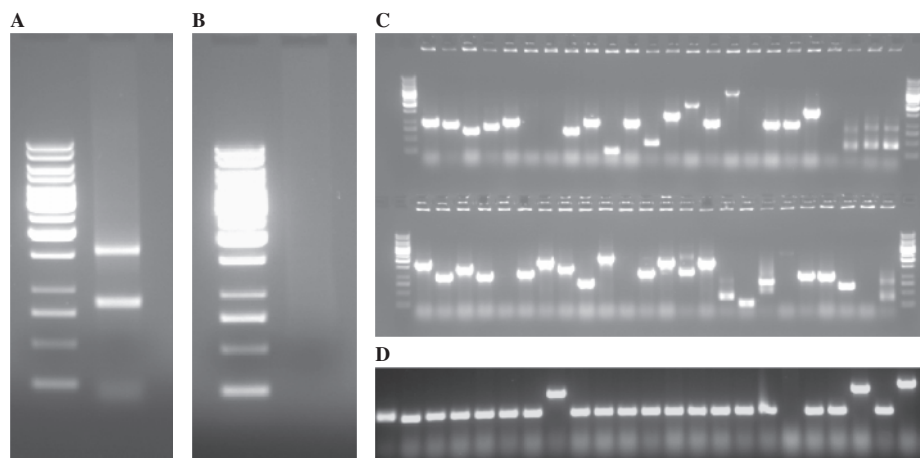


Figure 1. Library quality controls. (A) HES3 Human embryonic stem cells were grown and prepared as described (9). Total RNA was prepared by the Trizol isolation method. A smear of RNA with two bright bands corresponding to the 28S and 18S rRNA was obtained. The ladder used in all panels is Generuler 1 Kb (Fermentas) (<http://www.fermentas.com/catalog/electrophoresis/images/generuler031123.jpg>). (B) The mRNA prepared by the use of the μ MACS mRNA isolation kit on total RNA showed no bright bands corresponding to the rRNA. (C) A f1cDNA library was prepared by the Captrapper method, which had a titer of 4.6×10^6 cfu. Colony PCR quality control of the library was performed. An empty vector will produce a PCR product of size 260 bp (corresponds to the first band of the ladder); insert sizes were therefore calculated by subtracting off the size of the empty vector. Colony PCR therefore showed a range of insert sizes from 250 to 2000 bp (corresponds to the second to seventh bands of the ladder). This is expected, as a f1cDNA library is expected to give a range of different-sized inserts, with no single dominant size. Given that the library was of good quality, as can be seen from the colony PCR, the library was used to prepare two libraries: A single-PET library by the classic method, and a single-PET library by the Selection-MDA method. (D) A single-PET library was prepared from the full-length library as per the classic bacterial propagation method. Colony PCR quality control of this library showed a single predominant fixed size of 300 bp in many colonies, which is expected, as single-PET plasmids all have a fixed size of 2800 bp, and hence upon PCR, will give a band of 300 bp. Certain clones do not show this fixed size, which could be the result of the incorporation of foreign DNA, or other factors. Colony PCR quality control showed an insert ratio of 75% based on the number of wells that had PCR products of the correct size (300 bp).

was quantitated with Picogreen fluorimetry. 1 μ l was run on a PAGE gel to check that plasmids were prepared correctly (Figure 2B, 'purified plasmids'). Plasmid-safe DNase (Epicenter) treatment was then performed to remove any linear species, such as bacterial genomic DNA, that might be present. Phenol/chloroform ethanol precipitation was then performed and pellets were resuspended in 20 μ l of Elution Buffer (Qiagen). MDA was performed on aliquots of 8 ng of material as described earlier. The material was quantitated with Picogreen Fluorimetry, and digested with BamHI (New England Biolabs) according to the manufacturer's protocols. The PETs were PAGE gel-purified (Figure 2B, '50 bp ditags obtained after BamHI digest'), then cloned, concatenated (Figure 2B, 'concatenated BamHI-cut PETs'), partially digested with BamHI, cloned into BamHI-cut pZER0-1 vectors (Invitrogen), and prepared for sequencing as described (12). Ten plates of 384 colonies consisting of concatenated PETs were sequenced as a GIS-PET library, SHE002. A more detailed protocol is provided in the Supplementary Data.

Data analysis was performed using PET-Tool for PET extraction and genome mapping (13), followed by visualization in the T2G browser, a specially designed visualization system for PETs mapped to genome assemblies (4). Calculations were performed with Microsoft Excel. Categories of the genes were identified using RefSeq (14), UCSC Known Genes (15), Genbank mRNA (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=Nucleotide>), MGC (16), Ensembl (17), ESTs (18), Twinscan (19), SGPGene (20,21) and Genescan (22) databases.

RESULTS

The starting point for this analysis was HES3 human ES cell RNA, from which we generated a f1cDNA library (Figure 1A, B and C). We then generated two libraries: (1) a GIS-PET library by the standard approach, called SHE001 (Figure 1D), which comprised 613 905 unique PETs that were collapsed into 25 845 transcriptional units; and (2) a GIS-PET library prepared by the Selection-MDA approach, called SHE002 (Figure 2), which comprised 12 888 unique PETs which were collapsed into 3584 transcriptional units. To construct the MDA-amplified library (schematic in Figure 2B), a single-PET ligation mixture was generated from the maxiprep of the f1cDNA library, transformed into bacteria, and recovered for 4 h in the 'Selection' part of the procedure. The short 4 h growth in liquid media, allows for the selection of single insert clones because multiple insert clones have multiple origins of replication and cannot survive. However, the time is not long enough to result in crowding of bacteria in liquid media, such that size bias is minimized. To investigate whether the bacteria would have multiplied such that they crowd, we analyzed the optical density of the liquid media at 0, 1, 2 and 4 h. The optical density absorbance at 600 nm (OD_{600}) of the media increased from 0.728 at 0 h to 0.897 over 4 h. Using the approximation that 1 OD_{600} is $\sim 1 \times 10^9$ cells/ml (23), our bacteria increased from 7.3×10^8 to 9.0×10^8 cells over 4 h. Hence, our bacteria are still in log growth and not yet saturated (23), thus the increase in cell number should not be sufficient to cause crowding. At the end of 4 h, the bacteria were washed well and harvested.

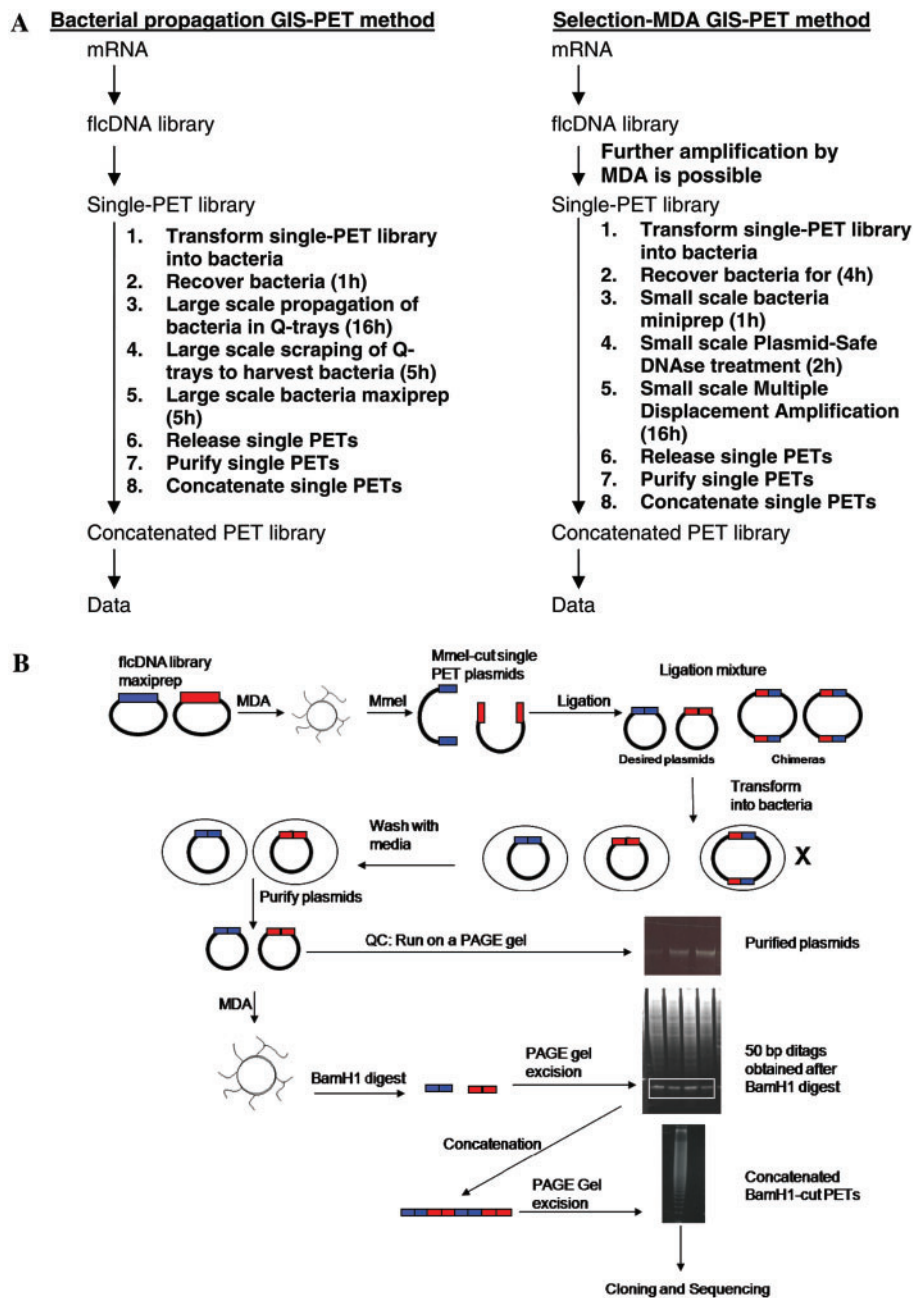


Figure 2. Schematic of the Selection-MDA approach. (A) Schematic showing the differences between the GIS-PET approach using bacterial propagation, and the GIS-PET approach using Selection-MDA. The Selection-MDA version allows for further amplification of the f1cDNA library maxiprep by MDA, as well as amplification of the single-PET library solely by Selection-MDA without the need for tedious plating and scraping of large numbers of bacterial colonies from solid-surface agar. Approximate times required for steps that are different between different protocols are given in brackets. Comparing the steps between Selection-MDA and the bacterial propagation method, it is clear that Selection-MDA requires much less hands-on labor and time, and also, in terms of absolute time, is at least 4 h shorter. (B) Detailed schematic of the MDA protocol. F1cDNA maxiprep was cut with MmeI, self-ligated, and transformed into bacteria, which were recovered for 4 h. After this, cells were washed with media, plasmids were extracted. MDA was then performed, followed by enzymatic digestion, concatenation and then cloning and sequencing. We ran quality control aliquots of the reactions on PAGE gels after the plasmid purification. Clean plasmids of the correct size, 2800 bp, were obtained. After BamHI digestion, 50 bp PETs were successfully recovered, as may be seen from the PAGE gel which shows a band of 50 bp (marked by a white box) separated from a high molecular weight smear from the plasmid backbone. PETs were successfully excised and concatenated, as may be seen from the smear from the concatemers, which was seen on a third PAGE gel. The concatemers were excised from the PAGE gel and prepared for subsequent cloning and sequencing.

Plasmids were prepared by miniprep and DNase cleanup. A quality control check showed that clean plasmids (Figure 2B) were obtained. PETs were then released by BamHI digestion (Figure 2B). Released PETs were concatenated for Sanger sequencing (Figure 2B). These

quality controls indicate that the Selection-MDA procedures were successful in producing PETs for sequencing.

We analyzed the library of PET sequences derived from the MDA approach using standard GIS-PET quality control measures (4), to investigate whether libraries

Table 1. Analysis of GIS-PET library quality control measures

Category		SHE002 (Selection-MDA)	SHE004 (Classic)	SHE005 (Classic)	SHE006 (Classic)
PET sequences					
Total number of unique PETs		12 888	13 196	12 988	13 102
PET matches to the genome	0 matches	2953 (22.9%)	2903 (22.0%)	2895 (22.3%)	2925 (22.3%)
	1 match	9641 (74.8%)	8266 (62.5%)	9851 (75.8%)	9936 (75.8%)
	>1 match	294 (2.3%)	2027 (15.4%)	242 (1.9%)	241 (1.8%)
Mapping accuracy	All PETs	88.4%	89.1%	88.2%	88.4%
	PETs from the top 20 transcriptional units	98.5%	97.9%	98.5%	99.2%
GC percentage		49.7%	48.9%	48.2%	48.3%
Categories of PETs with 1 match to the genome	Known	5697 (59.1%)	5253 (63.6%)	6080 (61.7%)	6083 (61.2%)
	ESTs	3512 (36.4%)	2678 (32.4%)	3291 (33.4%)	3385 (34.1%)
	Gene predictions	380 (3.9%)	303 (3.7%)	431 (4.4%)	420 (4.2%)
	Novel	52 (0.5%)	31 (0.4%)	48 (0.5%)	48 (0.5%)
Transcriptional units					
Total number of transcriptional units		3584	3362	3780	3776
Categories of Transcriptional units	Known	2278 (63.6%)	2309 (68.7%)	2490 (65.9%)	2506 (66.4%)
	ESTs	997 (27.8%)	817 (24.3%)	965 (25.5%)	956 (25.3%)
	Gene predictions	265 (7.4%)	209 (6.2%)	287 (7.6%)	280 (7.4%)
	Novel	44 (1.2%)	27 (0.8%)	38 (1.0%)	34 (0.9%)

prepared by the MDA approach are of good quality. Of a total 12 888 unique PETs sequenced, the number of PETs that could not be mapped to the human genome was 22.9%. This number is comparable to the percentage of unmappable PETs (26%) shown in a mouse embryonic stem cell library (4), and indicates that the MDA approach has a low percentage of chimeras due to multimers as well as high accuracy amplification, which allows the amplified sequences to map well to the genome. In addition, the mapping accuracy (percentage within ± 100 bp of the transcription start site or polyadenylation site) for all known PETs in SHE002 was 92.5% for 5' tags and 91.9% for 3' tags, comparable to the mouse ES cell GIS-PET (4), which showed results of 90.7% for 5' tags and 86.9% for 3' tags. Overall, the percentage of PETs with both 5' and 3' tags that map accurately is 88.4% for the entire library. While high, this measure includes mRNAs that have alternative splicing and alternative transcription start sites and hence represents a lower bound. The 12 888 unique PETs were collapsed into 3584 transcriptional units. To more accurately measure the mapping accuracy of the library, we examined PET sequences from the top 20 most abundant transcriptional units, which are well-annotated. The overall mapping accuracy is 98.5% for the top 20 transcriptional units of SHE002. This high level of mapping accuracy indicates that Selection-MDA method can accurately capture gene identification signatures.

In order to directly compare the performance of the Selection-MDA protocol with the standard protocol, we wanted to compare the quality control measures of the MDA-prepared GIS-PET library with those of a GIS-PET library (SHE001) prepared by conventional bacterial amplification. As the size of the data sampled from library SHE001 (the total number of PETs is 613 905) is almost 50-fold larger than the size sampled from library SHE002 (the total number of PETs is 12 888), a direct comparison of these two libraries will not be meaningful. Therefore, in

order to compare the two libraries at the same number of PETs, we created three smaller virtual libraries, SHE004, SHE005 and SHE006 (Table 1), by random selection of data from bacterial propagation library SHE001, such that the virtual libraries had the same approximate size as that of the MDA-prepared SHE002. Differences within the set of these three virtual libraries would reflect sampling variation. Hence, if the differences between the MDA approach and the conventional approach are significant, then the differences between SHE002, and SHE004, SHE005 and SHE006 should be much larger than the differences between SHE004, SHE005 and SHE006. The percentages of PET matches to the genome, numbers of transcriptional units, as well as mapping accuracies of SHE004, SHE005 and SHE006 are comparable to that of SHE002, indicating that the MDA-prepared library is of similar quality as that of the conventionally-prepared library constructed from the same starting material (Table 1).

Next, we checked whether the MDA procedure caused any biases in the sample. Because MDA is a different amplification method from bacterial amplification, we wished to investigate if there was any base bias. Base bias was measured by calculating the GC percentage of the library. There is minimal base bias between the MDA method and the conventional method (Table 1).

Again because MDA is a different amplification method, we investigated whether there is any bias towards any category of genes, such as novel genes. We grouped the PETs and transcriptional units into 'known genes', 'gene predictions', 'ESTs' and 'novel genes'. All libraries showed similar distributions, indicating minimal category bias (Table 1).

The Selection-MDA step could not have introduced a length bias in this particular library, because Selection-MDA was performed on single PET clones, which are all of a fixed size. Therefore, we could not test whether

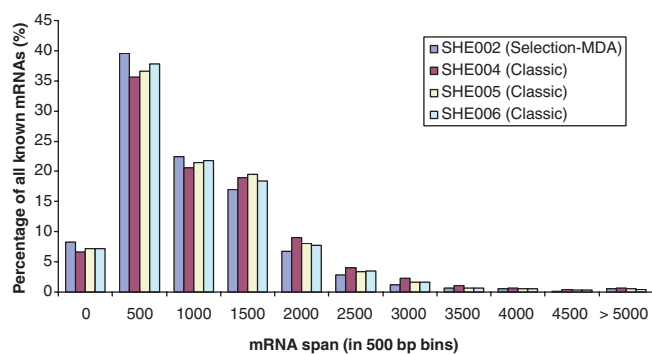


Figure 3. Analysis of length bias between the MDA approach and the bacterial amplification approach. We tested for the presence of length bias by classifying the mRNA lengths of the best-matching Known Genes, ESTs, or Gene Predictions from each library into 500-bp bins, which were then plotted on a graph. There is a small length bias. Because the length bias is small, it is possible that the apparent bias is due to sampling variation.

Selection-MDA would result in length biases or not. However, given that MDA was performed on the full-length cDNA library maxiprep to obtain more material for the construction of the single-PET library in the MDA procedure, we reasoned that this step might have introduced a length bias, and hence investigated whether there was a length bias. We tested for the presence of length bias by investigating the mRNA lengths of the best-matching known genes, ESTs or gene predictions, and found there was a length bias towards shorter mRNAs on the part of Selection-MDA, but the bias is small (Figure 3). Given that the bias is small, it is possible that the apparent bias could still be the result of sampling variation.

Next, we reasoned that the contents of the SHE002, SHE004, SHE005 and SHE006 libraries should be similar, because the same starting full-length cDNA library was used for the preparation of the two libraries. Hence, we compared the top 20 most abundant transcriptional units of each library with each other. The average number of transcriptional units that are the same between SHE002 (the MDA-prepared library) and any randomly selected library from a bacterial propagation library is 13. The average number of transcriptional units that are the same between the bacterial propagation libraries is 14, suggesting that the agreement between the MDA method and the bacterial amplification method is similar to the agreement between randomly selected libraries chosen from the same bacterial propagation library (Table 2). This analysis thus indicates that the contents of the MDA-prepared library show a good match to those of the conventionally prepared library.

DISCUSSION

Taken together, we have shown the method of inserting plasmids into bacteria for a short selection interval followed by MDA is a feasible method for the construction of a complex library. We have successfully applied Selection-MDA to the construction of a complex

Table 2. Identities of Top 20 most abundant transcriptional units (genes)

Rank	SHE002 (Selection-MDA)	SHE004 (Classic)	SHE005 (Classic)	SHE006 (Classic)
1	FTL	FTL	FTL	FTL
2	GAPDH	MIF	ENO1	MIF
3	MIF	ENO1	MIF	ENO1
4	TPI1	PRDX1	RPL13	LOC388817
5	ENO1	IFITM1	RPS2	RPS2
6	LOC388817	C14orf172	TPI1	RPL13
7	RPL13	K-ALPHA-1	LOC388817	H3F3A
8	OAZ1	PGK1	RPL9	PRDX1
9	FTH1	RPL13	K-ALPHA-1	TMSL3
10	TMSL3	PFN1	FTH1	TPI1
11	H3F3A	LOC388817	MDK	H2AFZ
12	IFITM1	RPL18	H2AFZ	K-ALPHA-1
13	H2AFZ	IFITM3	H3F3A	FTH1
14	PRDX1	ACTG1	PGK1	PRDX4
15	C14orf172	PRDX4	RPL18	PFN1
16	PFN1	MDK	TMSL3	C14orf172
17	RPL15	OAZ1	IFITM1	IFITM1
18	TPT1	RPL8	PRDX1	RPL9
19	RPL9	RPLP0	C14orf172	RPL18
20	RPL10	STOML2	OAZ1	PGK1

GIS-PET library and found that the Selection-MDA method results in a library with similar content and quality control statistics as compared with a library constructed from the same starting material that was amplified with bacteria and harvested through scraping bacterial colonies from solid surface agar.

Comparing the steps between the MDA version and the bacterial propagation method, it is clear that the MDA version requires much less hands-on labor. In terms of the physical handling, the MDA version uses small scale 1.5 ml tubes of material whereas the bacterial propagation method uses 10 large Q-trays and many maxiprep columns. The approximate times for each step that differed between the two protocols was estimated (Figure 2A). Comparing the absolute times required, the MDA method requires 4h less time than the bacterial propagation method. Considering the fact that many of the time-consuming steps in MDA do not require hands-on activities and hence allows other projects to be carried out in parallel, the time requirement of the MDA method is much less than the bacterial propagation method. With recent improvements in the MDA method (for example, the Illustra Genomiphi V2 DNA Amplification kit from GE Healthcare), further time savings could be possible.

The concept of performing bacterial selection followed by MDA (Selection-MDA) may be used to replace amplification steps in complex libraries, and represents a substantial improvement to existing cloning-based protocols. The Selection-MDA method is an effective and simple method for the unbiased amplification of a pool of complex clones, which allows scale-up and elimination of tedious scraping steps in library-preparation protocols. The method may be readily integrated and applied to current cloning-based protocols.

In conclusion, Selection-MDA is a novel method for the amplification of cloned libraries consisting of complex DNA. We applied Selection-MDA to a GIS-PET

library, an example of a cloned, complex DNA library, to illustrate the benefits of Selection-MDA. Library preparation was made simpler, and differences between the MDA-prepared library and a library prepared by the classic protocol were minimal. Hence, Selection-MDA is an effective and useful improvement to current cloning-based protocols.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Mr H. Thoreau and the Genome Technology & Biology Group at the Genome Institute of Singapore for high-throughput sequencing support. National Institutes of Health (1R01HG003521-01 to C.L.W. and Y.J.R.); Agency for Science, Technology and Research (A*STAR grants to C.L.W. and Y.J.R.; A*STAR National Science Scholarship to M.J.F.). Funding to pay the Open Access publication charges for this article was provided by the Agency for Science, Technology and Research.

Conflict of interest statement. None declared.

REFERENCES

- Lander, E.S., Linton, L.M., Birren, B., Nussbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
- Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P. *et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature*, **420**, 520–562.
- Strausberg, R.L., Feingold, E.A., Klausner, R.D. and Collins, F.S. (1999) The mammalian gene collection. *Science*, **286**, 455–457.
- Ng, P., Wei, C.L., Sung, W.K., Chiu, K.P., Lipovich, L., Ang, C.C., Gupta, S., Shahab, A., Ridwan, A., Wong, C.H. *et al.* (2005) Gene identification signature (GIS) analysis for transcriptome characterization and genome annotation. *Nat. Methods*, **2**, 105–111.
- Wei, C.L., Wu, Q., Vega, V.B., Chiu, K.P., Ng, P., Zhang, T., Shahab, A., Yong, H.C., Fu, Y., Weng, Z. *et al.* (2006) A global map of p53 transcription-factor binding sites in the human genome. *Cell*, **124**, 207–219.
- Esteban, J.A., Salas, M. and Blanco, L. (1993) Fidelity of phi 29 DNA polymerase. Comparison between protein-primed initiation and DNA polymerization. *J. Biol. Chem.*, **268**, 2719–2726.
- Garmendia, C., Bernad, A., Esteban, J.A., Blanco, L. and Salas, M. (1992) The bacteriophage phi 29 DNA polymerase, a proofreading enzyme. *J. Biol. Chem.*, **267**, 2594–2599.
- Blanco, L., Bernad, A., Lazaro, J.M., Martin, G., Garmendia, C. and Salas, M. (1989) Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication. *J. Biol. Chem.*, **264**, 8935–8940.
- Zhao, X.D., Xu, H., Chew, J.L., Liu, J., Chiu, K.P., Choo, A., Orlov, Y.L., Sung, K.W., Shahab, A., Kuznetsov, V.A. *et al.* (2007) Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. *Cell Stem Cell*, **1**, 286–298.
- Ng, P., Tan, J.J., Ooi, H.S., Lee, Y.L., Chiu, K.P., Fullwood, M.J., Srinivasan, K.G., Perbost, C., Du, L., Sung, W.K. *et al.* (2006) Multiplex sequencing of paired-end ditags (MS-PET): a strategy for the ultra-high-throughput analysis of transcriptomes and genomes. *Nucleic Acids Res.*, **34**, e84.
- Ruan, Y., Ooi, H.S., Choo, S.W., Chiu, K.P., Zhao, X.D., Srinivasan, K.G., Yao, F., Choo, C.Y., Liu, J., Ariyaratne, P. *et al.* (2007) Fusion transcripts and transcribed retrotransposed loci discovered through comprehensive transcriptome analysis using Paired-End diTags (PETs). *Genome Res.*, **17**, 828–838.
- Ng, P., Wei, C.L. and Ruan, Y. (2006) Paired-end diTagging for transcriptome and genome analysis. In Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (eds), *Current Protocols in Molecular Biology*, 2006, Unit 21.12. John Wiley and Sons, Inc, New York.
- Chiu, K.P., Wong, C.H., Chen, Q., Ariyaratne, P., Ooi, H.S., Wei, C.L., Sung, W.K. and Ruan, Y. (2006) PET-Tool: a software suite for comprehensive processing and managing of Paired-End diTag (PET) sequence data. *BMC Bioinformatics*, **7**, 390.
- Pruitt, K.D., Tatusova, T. and Maglott, D.R. (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.*, **35**, D61–D65.
- Hsu, F., Kent, W.J., Clawson, H., Kuhn, R.M., Diekhans, M. and Haussler, D. (2006) The UCSC known genes. *Bioinformatics*, **22**, 1036–1046.
- Gerhard, D.S., Wagner, L., Feingold, E.A., Shenmen, C.M., Grouse, L.H., Schuler, G., Klein, S.L., Old, S., Rasooly, R., Good, P. *et al.* (2004) The status, quality, and expansion of the NIH full-length cDNA project: the mammalian gene collection (MGC). *Genome Res.*, **14**, 2121–2127.
- Hubbard, T.J., Aken, B.L., Beal, K., Ballester, B., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cunningham, F., Cutts, T. *et al.* (2007) Ensembl 2007. *Nucleic Acids Res.*, **35**, D610–D617.
- Boguski, M.S., Lowe, T.M. and Tolstoshev, C.M. (1993) dbEST—database for ‘expressed sequence tags’. *Nat. Genet.*, **4**, 332–333.
- Korf, I., Flicek, P., Duan, D. and Brent, M.R. (2001) Integrating genomic homology into gene structure prediction. *Bioinformatics*, **17** (Suppl 1), S140–S148.
- Guigo, R., Dermitzakis, E.T., Agarwal, P., Ponting, C.P., Parra, G., Reymond, A., Abril, J.F., Keibler, E., Lyle, R., Ucla, C. *et al.* (2003) Comparison of mouse and human genomes followed by experimental verification yields an estimated 1,019 additional genes. *Proc. Natl Acad. Sci. USA*, **100**, 1140–1145.
- Parra, G., Agarwal, P., Abril, J.F., Wiehe, T., Fickett, J.W. and Guigo, R. (2003) Comparative gene prediction in human and mouse. *Genome Res.*, **13**, 108–117.
- Burge, C. and Karlin, S. (1997) Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.*, **268**, 78–94.
- Elbing, K. and Brent, R. (2002) Growth in liquid media. In Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (eds), *Current Protocols in Molecular Biology*, 2002, Unit 1.2.1. John Wiley and Sons, Inc, New York.