Amplicon Sequencing Template Preparation

The DNA sample preparation procedure for Amplicon Sequencing consists of a simple PCR amplification reaction, but uses special Fusion Primers (Figure 1-1). The 3'-part of each primer is designed to anneal with a specific sequence on either side of the target of interest, on the initial (e.g. complex) DNA sample, delineating the margins of the amplicon that will be produced (which should be no longer than 500 bp). This requires detailed knowledge of the target sequence, in particular the sites targeted by the primers. The 5'-part of the primers exists in two types, "A" and "B", to match other components of the Genome Sequencer 20 System. The design of the Fusion Primers is described in detail in Appendix.

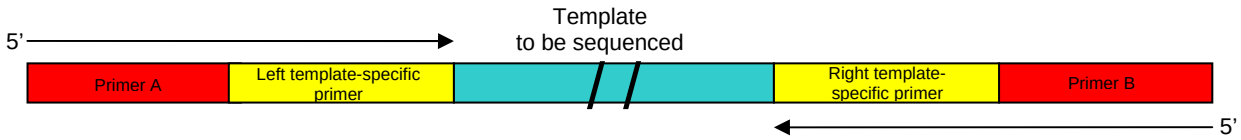


Figure 1-1: Schematic representation of an amplification product generated by the Amplicon library preparation procedure described in this Guide. The composite primers each comprise a 20-25 bp target-specific sequence region at their 3'-end; and a 19 bp region (Primer A or Primer B) that will be used in subsequent clonal amplification and sequencing reactions, at their 5'-end.

The Amplicon library preparation procedure takes full advantage of the emulsion-based clonal amplification (emPCR) feature of the Genome Sequencer 20 System, allowing for single molecule sequencing without cloning the target sequences into bacteria. Two emPCR kits are available for Amplicon sequencing, one for sequencing a library from Primer A (GS emPCR Kit II) and the other for sequencing from Primer B (GS emPCR Kit III).

Amplicons for sequencing with the Genome Sequencer 20 System should be no longer than 500 bp, as longer amplicons do not amplify well under the experimental conditions of emPCR. Make sure to design your Fusion Primers accordingly.

Remember, however, that in a highly variable target, such as rapidly mutating viruses, it may be difficult to find appropriate sequences for the template-specific parts of the Fusion Primers, especially within the distance restriction imposed by amplification and read length considerations. The choice of appropriate PCR primers for the generation of the Amplicon library is critical for a successful experimental design, as studies aimed at identifying and quantitating sequence variants will be at best as accurate and unbiased as the original amplification.

Your experimental design may require the monitoring of multiple amplicons (e.g. various exons of a gene of interest, various markers associated with a disease, etc.). For consistency and

maximum control of the amplification reaction, it is recommended to amplify each target separately (as its own "Amplicon library") rather than to combine multiple targets and primer pairs in a single library (multiplex PCR). However, multiple Amplicon libraries (each comprising a single peak) can be mixed and processed together through emPCR, and sequenced together on a single PicoTiterPlate device, or even in a single PicoTiterPlate region.

Sample

The quality and quantity of the DNA sample are critical to the success of this procedure. Any contamination inherent in the starting material will be directly reflected in the quality of the output library. Since it includes an amplification step, this procedure requires less input DNA than the standard sstDNA Library Preparation procedure, though the amount required will depend on the nature of the experiment. For example, if you are searching for low abundance sequence variants out of a complex sample (such as genomic DNA), you should start with 10 – 50 ng of DNA (equivalent to ~3,000-15,000 haploid human genomes). If your starting material is cloned into a plasmid or is a PCR-generated DNA fragment, 1 – 5 ng is usually sufficient.

Ideally, the DNA will have been verified to ensure the material is derived from the target organism and contains no other contaminating DNA. At a minimum, it is recommended that the DNA sample meet the following criteria:

- DNA should be non degraded, and containing no particulate matter
- Input DNA size should be sufficient to support amplification of the target(s)
- $OD_{260/280}$ ratio of 1.8 or above
- Concentration of 5 ng/µl or above, in TE (0.5 ng/µl for cloned or PCR-generated targets)

Because DNA quantitation using OD_{260} is variable and depends on DNA purity, it is recommended to verify the input DNA concentration and integrity by densitometry, e.g. on a 1 – 2% agarose gel using a DNA Mass Ladder, or by fluorometry, e.g. using the Quant-iT™ PicoGreen® dsDNA Assay Kit (Invitrogen).

454 LIFE SCIENCES

Required Materials

Equipment	Quantity Required	Source	Ref. Number
Thermal Cycler	1	Many possible	N/A
96 well 0.2 ml block for Thermal Cycler	1	Many possible	N/A
BioAnalyzer	1	Agilent	2100
Microcentrifuge (1000-16000 RCF)	1	Eppendorf	5415D
Magnetic Particle Concentrator	1	Invitrogen	120-20D

Lab Supplies/Consumables	Quantity Required	Source	Ref. Number
DNA 1000 LabChip	1	Agilent	5067-1504
Quant-iT™ PicoGreen® dsDNA Assay Kit	1 assay	Invitrogen	P-7589
1.7 ml microcentrifuge tubes	1 bag	Dot Scientific	RA-1700-GMT
0.2 ml Tubes with caps	1 bag	Dot Scientific	620-PCR
Full set of micropipettes, 2-1000 µl	1 ea.	Rainin	RL series
Pipette Tips	1 box ea.	Rainin	RT-LxF series

Reagents	Quantity Required	Source	Ref. Number
FastStart High Fidelity PCR System	NA	Roche	03 553 426 001 03 553 400 001 03 553 361 001
dNTPs (10 mM each)	1 µl	Pierce	NU606001
Fusion Primer A* (10 µM)	1 µl	IDT	NA
Fusion Primer B* (10 µM)	1 µl	IDT	NA
Ampure beads (SPRI)	72 µl	Agencourt	000130
TE 1X, pH 8.0	5 ml	Fisher Scientific	BP2473-1
Ethanol (70%, v/v)	400 µl	Many possible	N/A
Molecular Biology Grade Water	41 µl	Many possible	N/A

Table 1-1: Materials required to prepare an Amplicon DNA library, but not supplied. Items identified under "Source" (and corresponding Ref. Numbers) are examples only.

454 LIFE SCIENCES

1. AMPLICON LIBRARY PREPARATION PROCEDURE

1. In a 0.2 ml microcentrifuge tube, add the following reagents, in the order indicated:

41	µl	Molecular Biology Grade Water
5	µl	10X FastStart High Fidelity Reaction Buffer with 18 mM $MgCl_2$
1	µl	dNTPs (10 mM each)
1	µl	Fusion Primer A (10 µM)
1	µl	Fusion Primer B (10 µM)
1	µl	FastStart High Fidelity Enzyme Blend (5 U/µl)
50	µl	final volume

2. Mix by vortexing, spin down briefly, and add the sample DNA. The amount of DNA to use as input depends on the nature of the sample. In all cases, however, the volume should not exceed 2 µl.
    a. For a complex DNA sample (e.g. genomic DNA), use 10 – 50 ng of DNA, in no more than 2 µl.
    b. For a template cloned in a plasmid or for PCR-generated template DNA, use 1 – 5 ng of DNA, in no more than 2 µl.
3. Place the tube in a thermocycler and set-up and launch an amplification program appropriate for your sample such that the total amount of product at the end of the reaction does not exceed 10^{11} molecules. For guidelines on the thermocycling program, see the Appendix.
4. For amplified fragments larger than 100 bp, purify the amplified DNA using SPRI size exclusion beads, as follows (for smaller amplicons, see "Note" below):
    a. After the amplification program completes, vortex and briefly spin down the amplification reaction.
    b. Transfer 45 µl of the amplified DNA to a fresh 1.7 ml microcentrifuge tube.
    c. Add **exactly** 72 µl of Ampure SPRI beads. Vortex to mix.
    d. Incubate 3 – 5 minutes at room temperature (22ºC).
    e. Using a Magnetic Particle Collector (MPC), pellet the beads against the wall of the tube (this may take several minutes due to the high viscosity of the solution).

Note: Leave the tube of beads in the MPC during all wash steps.

    f. Remove the supernatant and wash the beads **twice** with 200 µl of 70% Ethanol.
    g. Remove all the supernatant and allow the SPRI beads to air dry completely. The drying time can vary due to environmental conditions and the amount of residual fluid left in the tube. The tube may be placed in a heating block set to 37ºC to help speed drying; the beads are dry when visible cracks form in the pellet.
    h. Remove the tube from the MPC, add 15 µl of 1X TE Buffer, and vortex to resuspend the beads. This elutes the amplified DNA from the SPRI beads.
    i. Using the MPC, pellet the beads against the wall of the tube once more, and transfer the **supernatant** containing your Amplicon library to a fresh microcentrifuge tube.

Note: The SPRI beads are not appropriate for amplicons smaller than 100 bp. If your target(s) is (are) smaller, use an alternative method such as gel electrophoresis to purify them.

5. Run a 1 µl aliquot of the amplicon library on a BioAnalyzer DNA 1000 LabChip, to assess the quality of the amplification product. (See section 1.3 for example traces of a high quality Amplicon library, and one with primer dimer contamination.)
    a. Ensure that the amplification product is of the expected size; the size displayed on the BioAnalyzer typically does not diverge from the calculated size by more than a few bp, for amplicons in the 100-200 bp range.
    b. Examine the trace closely for extraneous products, such as primer dimers, as contaminants could seriously reduce the number of useful reads you will get from sequencing your Amplicon library. Repeat the SPRI bead purification if primer dimers are present.
6. Quantitate your library by fluorometry, using the Quant-iT™ PicoGreen® dsDNA Assay Kit (Invitrogen), following the manufacturer's instructions.

Whereas various quantitation methods are possible, fluorometry has been found to be highly reproducible and to provide consistent emPCR and sequencing results. It is especially important to accurately determine the concentration of individual Amplicon libraries when they are to be pooled prior to emPCR, e.g. to monitor multiple targets together; this will ensure even representation of all the targets in the sequencing reaction.

7. Given the library concentration (in ng/µl) measured by fluorometry, calculate the equivalence in molecules/µl, using the following equation:

$$\text{Molecules/µl} = \frac{(\text{Sample conc.; ng/µl}) \times (6.022 \times 10^{23})}{(656.6 \times 10^{9}) \times (\text{amplicon length; bp})}$$

… where 6.022×10^{23} is Avogadro's number (molecules/mole), and 656.6 is the average molecular weight of nucleotide pairs, in g/mole.

APPENDIX

1.1 Fusion Primer Design

The primers used to generate Amplicon libraries are composed of two parts fused together (see Figure 1-1). The 5'-part is a 19-mer, dictated by the requirements of the Genome Sequencer 20 System for binding to the DNA Capture Beads, for annealing the emPCR Amplification Primers and the Sequencing Primer, and ending with the sequencing key "TCAG". There exist two kinds of such primers, termed "Primer A" and "Primer B", for use with the GS emPCR Kits II and III respectively, allowing the sequencing of the insert from either end. The exact sequences are as follows:

Primer A Sequence

5’ GCCTCCCTCGCGCCATCAG 3’

Primer B Sequence

5’ GCCTTGCCAGCCCGCTCAG 3’

The 3’-part of each primer is typically 20-25 nt in length (may vary), and is designed to anneal to either side of the target to be sequenced. The normal constraints of primer specificity and annealing conditions apply. It should be noted that amplicons should be no longer than 500 bp because templates longer than this do not amplify well in emPCR.

Example of a Fusion Primer Pair Design

The following is a typical Fusion Primer pair design. The template of interest is HLA SNP DD14 with the SNP highlighted in red.

DD14

AGATGTAGCCCTTGAAATGTCATAAATATAGATTTTTGCTTCTGATTCAATCTGACGATCTCTG
TCTTCTAACCTATGTTCAATTCATATGGTAGTCAAAGTGAGCAAACT**T**GTTTCTGCAAGAGACA
AACACTGAAGCCTCAGTGGTTTAACAAAACACAGGTTTATTTTTTAGCCACGTGTAGTTCAAGG
CAGGTTGG

```
OLIGO         start  len      tm     gc%   any    3' seq
LEFT PRIMER      51   24   59.06   45.83  4.00  0.00 TCTGACGATCTCTGTCTTCTAACC
RIGHT PRIMER    193   20   60.32   55.00  6.00  2.00 GCCTTGAACTACACGTGGCT
SEQUENCE SIZE:  200
INCLUDED REGION SIZE: 200

PRODUCT SIZE: 143, PAIR ANY COMPL: 3.00, PAIR 3' COMPL: 2.00

    1 AGATGTAGCCCTTGAAATGTCATAAATATAGATTTTTGCTTCTGATTCAATCTGACGATC
                                                >>>>>>>>>>

   61 TCTGTCTTCTAACCTATGTTCAATTCATATGGTAGTCAAAGTGAGCAAACTTGTTTCTGC
      >>>>>>>>>>>>>>

  121 AAGAGACAAACACTGAAGCCTCAGTGGTTTAACAAAACACAGGTTTATTTTTTAGCCACG
                                                      <<<<<<<

  181 TGTAGTTCAAGGCAGGTTGG
      <<<<<<<<<<<<<
```

Fusion primers:

Fusion Primer A-DD14 GCCTCCCTCGCGCCATCAG TCTGACGATCTCTGTCTTCTAACC
Fusion Primer B-DD14 GCCTTGCCAGCCCGCTCAG GCCTTGAACTACACGTGGCT

To obtain sequence reads starting from Primer A choose the GS emPCR Kit II (Amplicon A, Paired End) and for reads starting from Primer B choose the GS emPCR Kit III (Amplicon B).

1.2 Amplification Program

Create an appropriate thermocycling program for usage in section 1, step 3. For example:

- 1X (3 minutes at 94°C) – Hotstart Initiation
- 20-30X (30 seconds at 94°C, 45 seconds at 57°C, 60 seconds at 72°C) – Amplification
- 1X (120 seconds at 72°C) – Final Primer Extension

Note: The number of cycles (20-30X) and the annealing temperature (57ºC) are examples only, and may need optimization with each primer set; the total amount of product at the end of the reaction should not exceed approximately 10^{11} molecules. In particular, in certain samples, such as tumor-derived DNA, locus-specific chromosomal amplifications may be present; this could significantly increase the effective number of amplification templates (copies per genome), thereby reducing the number of amplification cycles necessary to reach the proper amount of amplified DNA.

1.3 Example of an Agilent 2100 Trace of an Amplicon Library (Single Amplicon)

Figure 0-1 shows a typical Agilent 2100 DNA 1000 LabChip profile for 1 µl of an amplicon library comprised of a single 177 bp amplicon, showing the expected single sharp peak. If unintended amplification products are observed, such as primer dimers (Figure 0-2), optimization of the PCR conditions should be considered, or the SPRI bead purification repeated: these contaminating products must be removed or they will interfere with sequencing. The peaks at 15 and 1500 bp are internal markers.

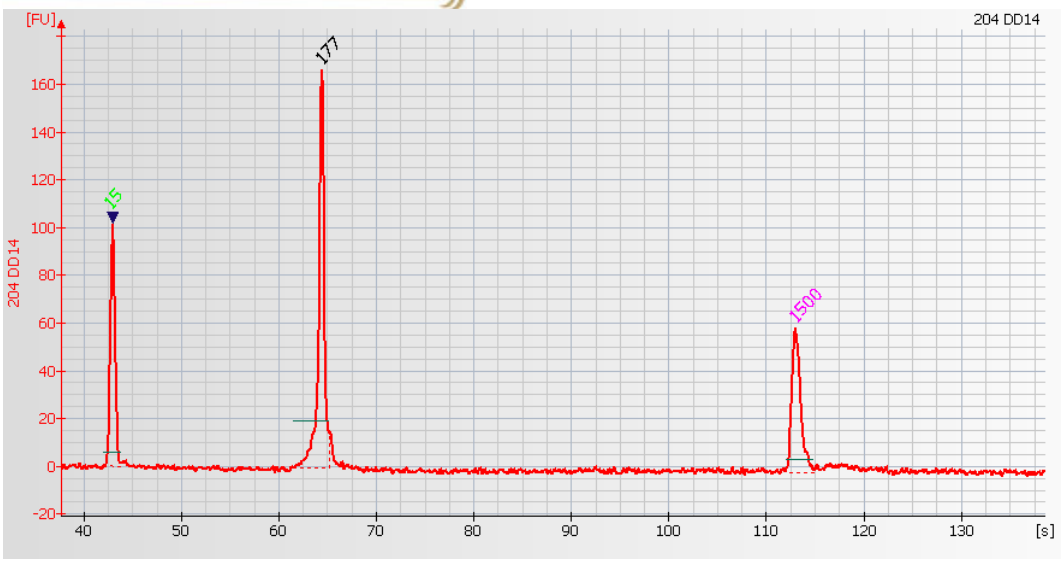
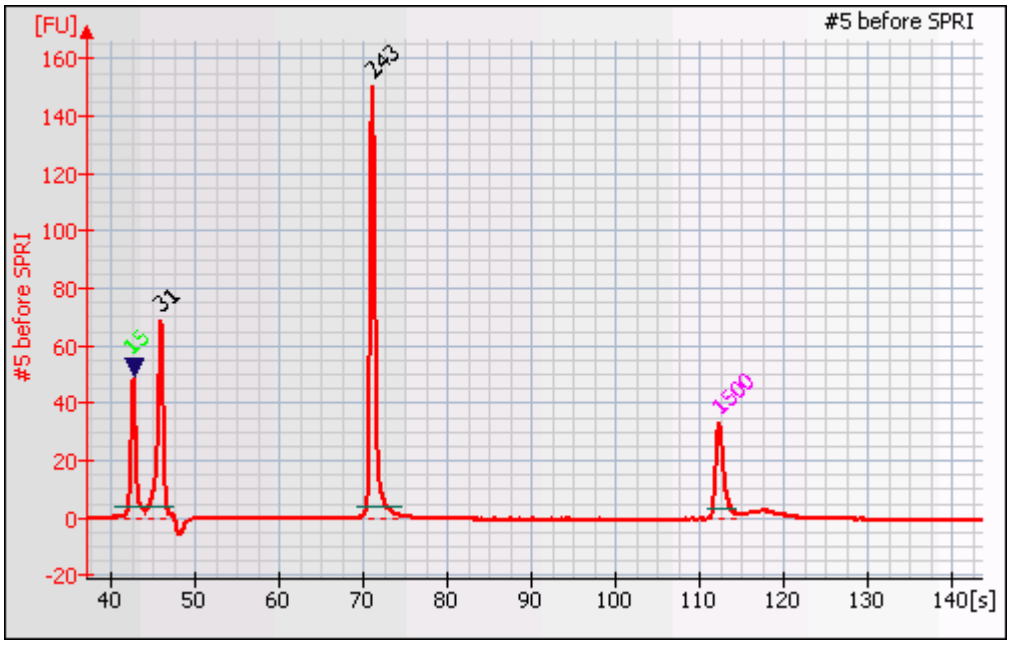


Figure 0-1: DNA 1000 LabChip profile of an Amplicon library comprised of a single amplicon of 177 bp.

A



B

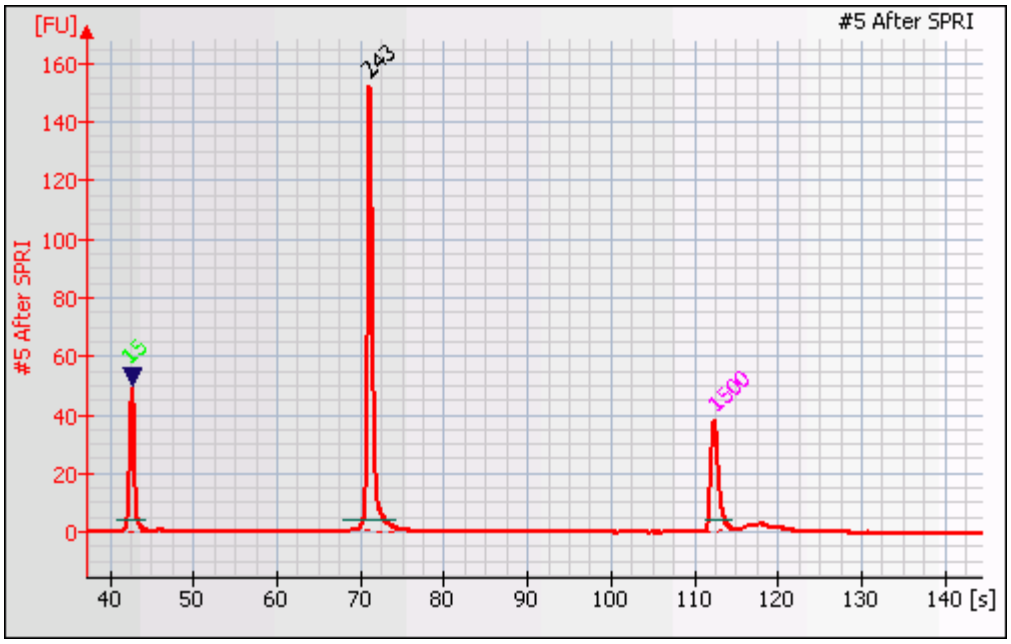


Figure 0-2: (A) DNA 1000 LabChip profile of an Amplicon library comprised of a single amplicon of 243 bp, showing a sizable adaptor dimer peak at 31 bp. (B) Same library after an additional purification on Ampure SPRI beads