Aiding DNA Amplification of GC-rich Regions in the Human Genome for Illumina Sequencing

BROAD INSTITUTE

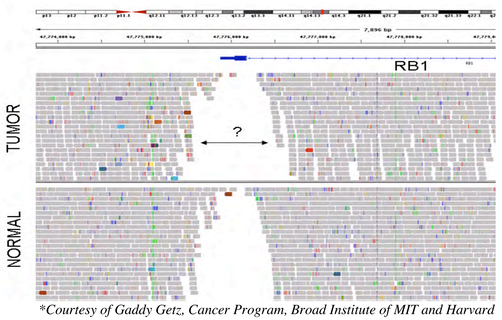
Wei-Sheng, Daniel Aird, and Andreas Gnirke

Genome Sequencing and Analysis Platform, Broad Institute of MIT and Harvard, Cambridge, MA

Introduction

The Human Genome Project was launched in 1990 and completed in 2003. However, a number of regions of the genome (approximately 0.7% of the whole genome) were not successfully sequenced because of issues caused by GC-rich regions of DNA and highly repetitive sequences. A new sequencing technology, Illumina, has greatly reduced the cost and effort of genome sequencing, yet still encounters the same problems with GC-rich areas of the genome.

Polymerase chain reaction (PCR), a critical step in the DNA library construction process, amplifies GC-rich regions very poorly in comparison to regions that are ~50% GC. This uneven amplification results in missing reads of GC-rich parts of the genome in sequencing stage. For example, the tumor suppressor gene RB1's promoter is of high GC content, which is not amplified well and therefore results in sequencing gaps, as shown below.



**Courtesy of Gaddy Getz, Cancer Program, Broad Institute of MIT and Harvard*

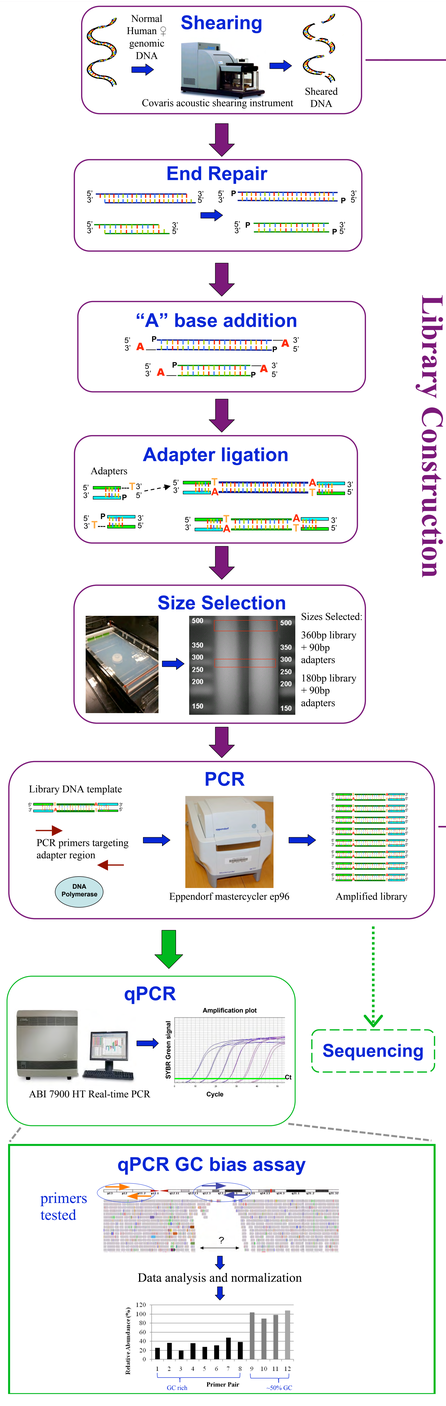
GC-rich regions of DNA are hard to amplify because Gs and Cs have three hydrogen bonds while As and Ts have two. In addition, after the two strands of DNA denature (separate because of heat), GC-rich regions tend to form secondary structure upon cooling. This secondary structures inhibits PCR amplification of these GC-rich regions.

Project Aim:
Find new PCR conditions that amplify every part of the genome evenly regardless of the GC content it possesses

Combinations of PCR conditions Tested

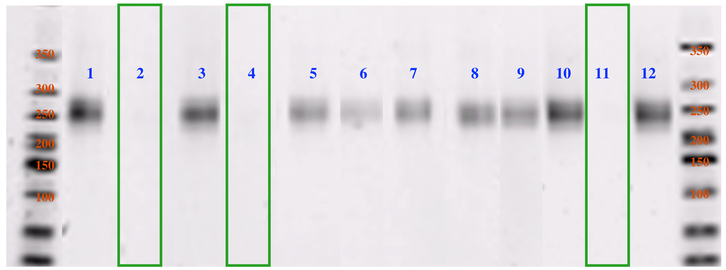
Enzyme	Additive	Denaturation Time
Phusion (Standard)	Betaine	10" (Standard for Phusion)
Platinum Taq	DMSO	30" (Standard for others)
HF Platinum Taq	SSB protein	
AccuPrime P-taq	None	80" (extended)
AccuPrime Pfx		

Methods



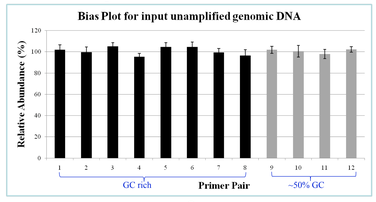
Results

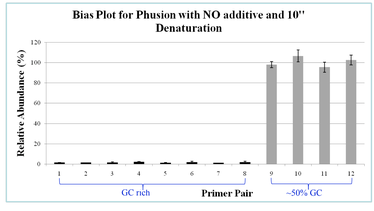
Gel Results after PCR: Effectiveness of Amplification of Whole Libraries

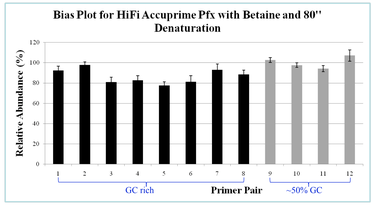


1: HF AccuTaq no additive 30" den
2: HF AccuTaq 2M Betaine 30" den
3: HF AccuTaq no additive 80" den
4: HF AccuTaq 2M Betaine 80" den
5: HF AccuPfx no additive 30" den
6: HF AccuPfx 2M Betaine 30" den
7: HF AccuPfx no additive 80" den
8: HF AccuPfx 2M Betaine 80" den
9: HF P-Taq no additive 30" den
10: Phusion no additive 10" den
11: Phusion 5% DMSO 10" den
12: Phusion 2M Betaine 10" den

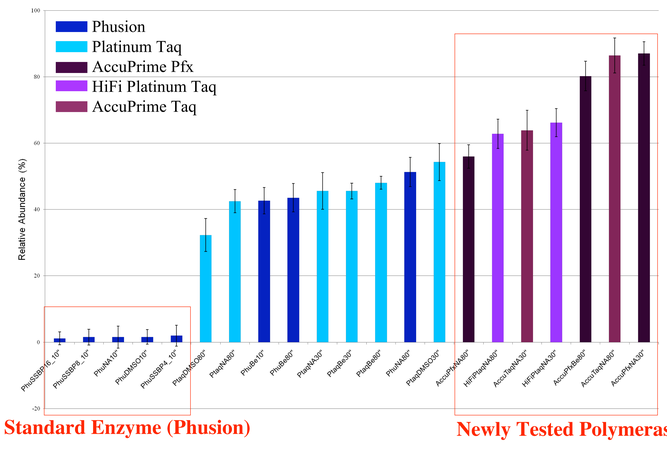
GC Bias Plots after qPCR: Effectiveness of Amplification of 8 GC-Rich Regions







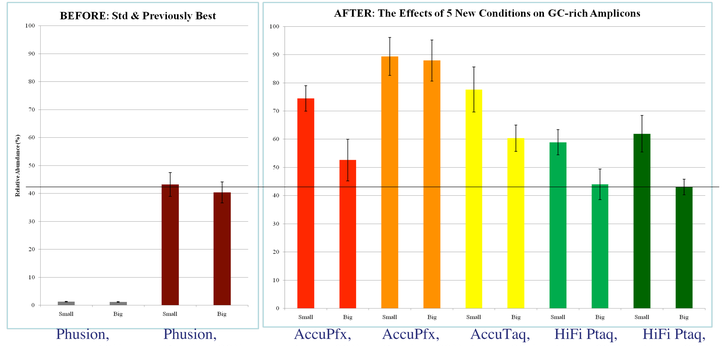
qPCR Results Averaged from the 8 Regions: Overall Effectiveness of Amplification of GC-Rich DNA



Standard Enzyme (Phusion)

Newly Tested Polymerases

Five Best Conditions Tested against Two Insert Sizes



Phusion, None, 10''
Phusion, Betaine, 80''
AccuPfx, None, 30''
AccuPfx, Betaine, 80''
AccuTaq, None 80''
HiFi Ptaq, None, 30''
HiFi Ptaq, None, 80''

Conclusions

- The best conditions generated a relative abundance of 80% to 90%, while the standard PCR condition yields only 1.2% to 1.6%. Therefore, we see a 50-75 fold improvement of GC-rich regions in PCR.
- Larger libraries (360bp) lost ~15% abundance compared to smaller libraries (180bp) under some conditions.
- The combination of betaine and longer denaturation helped the PCR process most, amplifying all regions of DNA within a genome more equally.
- AccuPrime Pfx with betaine and long denaturation was the best condition among all, and worked well for both insert sizes.
- We determined three DNA polymerases that did a much better job in comparison to Phusion, the standard enzyme currently used in the PCR process: Platinum Taq, AccuPrime Taq, and AccuPrime Pfx.

Future Directions

DNA library products of the best conditions should be sequenced on Illumina to see if the reduction of GC-bias applies to the whole genome. We then can select the most promising candidate for further testing on larger scale.

Acknowledgements

Special thanks to:
Andreas Gnirke and Daniel Aird for guidance and mentorship;
Megan Rokop, Kate MacSwain, and Allison Martino for making a very supportive team of program staff;
Dawn Ciulla for providing different DNA polymerases; and
Xian Adiconis for providing reagents for library construction.