University of Georgia
Department of Environmental Health Sciences &
Savannah River Ecology Laboratory

454 Sequencing and Microsatellite Development FAQs

1. What is 454 sequencing?

454 sequencing refers to a method of high throughput massively parallel next-generation sequencing (NGS). There are several instruments capable of NGS, one of which is the Genome Sequencer FLX developed by 454 Life Sciences, which is now owned by Roche. 454 developed the first commercial NGS platform and its main advantage over other NGS platforms is read length (hundreds of bp vs. tens of bp). See http://www.454.com/about-454/index.asp for more specific information and multimedia presentations.

2. What can be sequenced using 454 technology?

Pretty much any DNA or cDNA. For the purposes of microsatellite development the starting material would either be genomic DNA or PCR products.

3. What is the basic procedure for 454 sequencing?

To use 454 the starting material needs to be fairly small. Genomic DNA would be fragmented into 300-800bp pieces. PCR products in that range do not need to be further fragmented. Very briefly, adaptors are ligated to single-stranded fragments, the fragments are immobilized onto DNA capture beads, the captured fragments are amplified via emulsion PCR on the bead, the beads are transferred to a sequencing plate with millions of wells that each hold a single bead (with it's amplified fragments), and sequencing reactions take place in the wells. Each bead on the plate yields one sequencing read, which derives from a single molecule (just like a bacterial clone).

4. What is considered a 454 run?

One run of a full plate on a 454 instrument running standard chemistry typically generates > 400,000 sequence reads with an average read length of ~250bp resulting in a total of ~100 Mbp of sequence. With the new Genome Sequencer FLX Titanium Series reagents one full run generates about 1 million reads with an average length of about 400 bp.

5. What is meant by a partial run?

The plate used for sequencing can be partitioned into 2, 4, 8, or 16 sections. Some labs providing 454 sequencing service will offer these sections independently to researchers that do not need a full run, but there have to be enough other users who want identical-sized runs.

6. Why use 454 sequencing for microsatellite development?

The current standard protocol for developing microsatellites involves creating a DNA library enriched for repeats and then cloning and sequencing the library. The cloning and

sequencing is time consuming and expensive. Depending upon the species, 60-90% of clones will be positive for a microsatellite. Sequencing a 96 well plate will therefore yield 58 – 87 microsatellites. Within these only ~ 50% will have enough flanking sequence for primer design and only a portion of those will be successfully optimized for PCR and be polymorphic. Thus, only ~10-25% of clones result in a working microsatellite locus. Using 454 eliminates the need to clone and results in vastly more loci being sequenced at a much lower cost per sequence (about $0.01 per sequence). The number of loci sequenced will depend upon methodology (see below), run size, and efficiency of enrichment (both in terms of the proportion of DNA fragments with repeats and the proportion of the total from the genome still remaining in the enrichment).

7. Will microsatellite development be faster with 454 sequencing?

Not necessarily. The actual process of preparing samples for sequencing takes several days (about the same as cloning & capillary sequencing). Some researchers (Abdelkrim et al 2009) have suggested that microsatellites can be sequenced and optimized in two weeks using 454 technology. However, the queue for 454's tends to be many weeks, particularly when multiple projects are required in order to fill a complete 454 run. This long wait time has led many institutions (including UGA) to purchase these instruments, even though the instrument costs more than half a million dollars, and accessories add several hundred thousand more to that. Most institutions also favor their own researchers when scheduling machine time. Our experience and perception from discussions with many people is that most research cores with 454's returning excellent results have queues of at least 1-2 months.

8. With 454 sequencing is it still necessary to enrich genomic DNA for microsatellites before sequencing?

It is not mandatory to create an enriched library. Microsatellites are frequent enough in the genome that random sequencing with enough depth will result in a relatively small numbers of markers being obtained. However, these will tend to be dominated by dinucleotides, which are the most common but also the hardest to genotype, and loci with small numbers of repeats, which yield the lowest levels of polymorphism. The small number of markers obtained by brute force will certainly work for some types of projects on some species, but this approach will often fail to achieve a good number high quality markers. By enriching the genomic DNA first we will not only sequence more loci but can preferentially select for tri- and tetranucleotide repeats.

9. What is the best enrichment strategy?

We are currently working on this question. There are many strategies out there. Very simple approaches will increase the percent of sequences with microsatellites. However, using simple strategies and low numbers of repeat types will yield limited results. We are working toward methods that yield more information than has generally been possible previously to truly take advantage of the potential of NextGeneration DNA sequencing.

10. What is the difference between standard FLX chemistry and titanium chemistry for microsatellite development?

As mentioned above the titanium chemistry will produce more reads and longer reads. For microsatellite development it is critical to sequence enough DNA flanking the repeat to design primers. Although standard FLX chemistry will yield microsatellite DNA sequences, the yield of useful markers is much higher with longer read lengths achieved using Titanium chemistry. It is not simply a linear increase in useful markers with increasing read lengths.

11. How much will it cost to develop microsatellites using 454 technology?

The honest answer is not much less than with standard protocols. However, the end product will be substantially improved—with many more loci to choose from. A full 454 run with titanium chemistry is more than $10,000 and it can only be divided so many ways (i.e., current approaches don't allow researchers to get $10 or $100 worth of sequences). A major reason that the costs won't go down tremendously is that 1/16th [titration] runs aren't done very often when using Titanium chemistry. Thus, 1/16th runs are a specialty item one has to wait for 15 other people wanting done. This increases wait time. There are also specific costs with each library preparation, no matter how many sequences one gets. For most researchers a full or even a half run would be huge overkill, especially if the DNA is enriched. We will add unique ID_tags to each DNA sample & pool them thereby allowing for cost-sharing across species projects (similar to Roche's MID tagging approach). We expect that this will allow us to virtually divide a quarter run on a regular basis.

12. How soon will we offer microsatellite development with 454 sequencing?
We expect to add this option in July of 2009. We are currently experimenting with DNA from multiple species to fine tune the process and determine how to maximize output while minimizing time and cost. We will post data covering the number of sequence reads, % positive, and repeat breakdown at that time. If you would like to join the queue for this option, please let us know.

13. How will using 454 technology affect other options available?
All of our current options will remain available. The biggest difference is that for those researchers choosing the 454 approach we will guarantee a substantially larger number of loci for which primers can be designed. In addition, our genotyping options will increase. Currently we test 24 loci per project. With projects run on a 454 we will still start with 24 loci to test, but will add additional options for testing more loci. Details will be made available shortly.

14. Why would you choose traditional sequencing over 454?
We will continue to offer our standard cloning and sequencing approach. The primary reason for this is to avoid queues while waiting for enough projects to make use of a ¼ 454 run. We understand that under certain circumstances waiting even one month to initiate marker development will adversely affect research plans. In addition, in some situations researchers desire a stock of the clone for other uses and we can continue to provide that option. However, it is clear that using 454 will improve the end product without increasing cost and will become the suggested method.

References:

Abdelkrim J, BC Roberston, JL Stanton, and NJ Gemmell (2009) Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. BioTechniques 46: 185-192.