Microsatellite discovery by deep sequencing of enriched genomic libraries

Quentin C. Santana[1], Martin P. A. Coetzee[1], Emma T. Steenkamp[2], Osmond X. Mlonyeni[1], Gifty N. A. Hammond[3], Michael J. Wingfield[1], and Brenda D. Wingfield[1]
[1] Department of Genetics, Forestry and Agricultural Biotechnology Institute, University of Pretoria, South Africa, [2]Department of Microbiology and Plant Pathology, Forestry and Agricultural Biotechnology Institute, University of Pretoria, South Africa, [3]Department of Biochemistry, Forestry and Agricultural Biotechnology Institute, University of Pretoria, South Africa

BioTechniques 46:217-223 (March 2009) doi 10.2144/000113085
Keywords: microsatellites; 454 sequencing; simple sequence repeats; FIASCO; ISSR-PCR

Robust molecular markers such as microsatellites are important tools used to understand the dynamics of natural populations, but their identification and development are typically time consuming and labor intensive. The recent emergence of so-called next-generation sequencing raised the question as to whether this new technology might be applied to microsatellite development. Following this view, we considered whether deep sequencing using the 454 Life Sciences/Roche GS-FLX genome sequencing system could lead to a rapid protocol to develop microsatellite primers as markers for genetic studies. For this purpose, genomic DNA was sourced from three unrelated organisms: a fungus (the pine pathogen *Fusarium circinatum),* an insect (the pine-damaging wasp *Sirex noctilio*), and the wasp's associated nematode parasite (*Deladenus siricidicola)*. Two methods, FIASCO (fast isolation by AFLP of sequences containing repeats) and ISSR-PCR (inter-simple sequence repeat PCR), were used to generate microsatellite-enriched DNA for the 454 libraries. From the resulting 1.2–1.7 megabases of DNA sequence data, we were able to identify 873 microsatellites that have sufficient flanking sequence available for primer design and potential amplification. This approach to microsatellite discovery was substantially more rapid, effective, and economical than other methods, and this study has shown that pyrosequencing provides an outstanding new technology that can be applied to this purpose.

Introduction

Microsatellites, or simple sequence repeats (SSRs), are DNA sequences that consist of tandem repeats of 1–6 nucleotides, found at varying frequencies in the genomes of just about every known organism and organelle (1). They belong to a class of highly mutable genomic sequences known as variable number of tandem repeat (VNTR) elements (2,3) that show extensive levels of intraspecific polymorphisms in both eukaryotic (4–6) and prokaryotic (7,8) genomes. Because of their ease of use, co-dominance, and high levels of polymorphism (9), microsatellites have been particularly valuable in genome mapping, forensics, paternity testing, population genetics, conservation or management of biological resources, and molecular typing of microbial strains (9–12).

Both the identification and development of microsatellite markers represent significant challenges. This is especially true in the case of organisms for which there are little or no sequence data and where the development of microsatellite markers requires the protracted steps of generating clone libraries and sequencing them (9,13). For species with known genome sequences, in silico scanning of genome databases using bioinformatics tools can be used to identify microsatellites and to design primers targeting these regions (5,11). However, genome sequences are available for relatively few eukaryotes and providing these is generally beyond the limited budgets of most research programs. Also, microsatellite loci in some instances cannot be employed across distantly related species (14) and they usually need to be identified and characterized de novo for each species, which can be a time intensive and expensive exercise. In general, the success with which microsatellite markers are obtained and the size of clone libraries to be constructed are related to the frequency of occurrence of microsatellite sequences in the genome of interest (15,16). However, the frequency of microsatellites observed in the genomes of plants, animals, fungi and prokaryotes has been reported to be significantly different (5), and in some cases researchers have reported extreme difficulty in obtaining any microsatellite sequences (17).

A number of methods are available to identify microsatellites (17). Of these, the most commonly used methods employ targeted enrichment of DNA for microsatellites (16,18). One is known as inter simple sequence repeat PCR (ISSR-PCR) (19). In this procedure, ISSR primers, which contain microsatellite motifs and three anchoring nucleotides at the 5′ terminal end, are used to amplify regions of the genome that are thought to be abundant in microsatellites. The PCR products are cloned and subsequently sequenced to determine the presence of microsatellite sequences (20). More recently, DNA enrichment strategies involving hybridization with probes containing microsatellite sequences to genomic DNA fragments have been introduced (16). After exclusion of the non-hybridized DNA that presumably lacks repeat regions, the

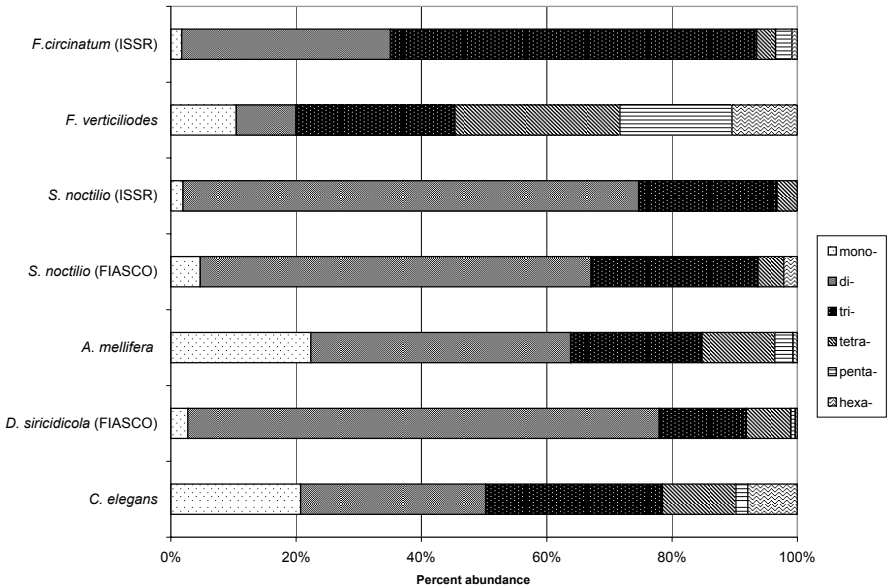

Figure 1. Distributions of microsatellite repeat classes for the organisms of interest. Percentages were calculated according to the total number of microsatellites for a respective class divided by the total microsatellites identified for that organism.

remaining microsatellite-rich fragments are cloned and sequenced to identify the microsatellite sequences. One of these approaches, known as fast isolation by AFLPs of sequences containing repeats (FIASCO), also uses amplified fragment length polymorphism (AFLP) (21) to aid in the enrichment process. Both the ISSR-PCR and FIASCO methods have been widely used in contemporary studies to isolate microsatellites from a wide variety of different eukaryotic species (22–25).

Development of new microsatellite markers has been streamlined to some extent by optimizing the numerous steps in microsatellite identification and subsequent sequencing throughput, to make the process cheaper, more efficient and more successful (15,16). However, using the currently available methods, certain factors—such as cloning efficiency, the necessity to sequence large numbers of cloned fragments, and the need for a multitude of hybridization probes for enrichment—limit the success rate of microsatellite isolation. The recent appearance of next-generation sequencing such as Roche 454 genome sequencing (26), which uses pyrosequencing, raised the question of whether this could facilitate more effective production of microsatellites.

Materials and methods

In this study, the 454 Life Sciences/Roche GS-FLX genome sequence system (Roche Applied Science, Penzburg, Germany) (26) was used for the identification of microsatellite sequences, directly from microsatellite-enriched genomic DNA. To provide a broadly applicable test, we evaluated this method on three unrelated eukaryotes with little genome information available: *Fusarium circinatum* (a fungal ascomycete), *Sirex noctilio* (a hymenopteran insect), and *Deladenus siricidicola* (a tylenchid nematode).

DNA extraction and microsatellite enrichment

DNA was extracted from *F. circinatum*, *S. noctilio* and *D. siricidicola* as previously described (27,28). The methods used to enrich genomic DNA for microsatellites were adopted from existing ISSR-PCR (19) and FIASCO protocols (16,27), but without cloning the microsatellite-containing DNA. ISSR-PCR enrichment was used for *F. circinatum*, the FIASCO enrichment method was used for *D. siricidicola*, and both enrichment methods were used for *S. noctilio*. For the ISSR-PCR, the following primers were used: ISSR1 (5′-DDB$(CCA)_5$-3′), ISSR2 (5′-DHB$(CGA)_5$-3′), ISSR3 (5′ YHY$(GT)_5$G-3′), ISSR4 (5′-HVH $(GTG)_5$-3′), ISSR5 (5′-NDB$(CA)_7$C-3′), ISSR6 (5′-NDV$(CT)_8$-3′), and ISSR7 (5′-HBDB$(GACA)_4$-3′). For FIASCO, the following probes were used: $(CAC)_7$, $(AAG)_7$, $(TCC)_7$, $(CA)_{10}$, $(CT)_{10}$, $(AG)_{10}$, and $(GACA)_5$.

Pyrosequencing

For each microsatellite-enriched genomic DNA pool, 5 µg were analyzed on the Roche 454 GS-FLX platform at Inqaba Biotech (Pretoria, Gauteng, South Africa). For *F. circinatum*, a single-lane sequencing run using portioned sections of the PicoTiterPlate was performed. Single runs were carried out for the *S. noctilio* and *D. siricidicola* DNA libraries, each using one section of the PicoTiterPlate. Sample preparation and analytical processing such as base calling, were performed at Inqaba Biotech using the manufacturer's protocol.

Microsatellite discovery

Sequence reads for *F. circinatum*, *S. noctilio*, and *D. siricidicola* were assembled using the ContigExpress component of the Vector NTI software package

Table 1. Sequence Data Generated and Microsatellites Characterized from Deep Sequencing of Libraries

Species	Sequence data generated (Mb)	Sequence reads	Contigs and singletons	Sequences containing microsatellites[a]	Sequences with amplifiable microsatellite sequences[b]
F. circinatum (ISSR)	1.67	8644	1692	1644 (97%)	231 (14%)
S. noctilio (FIASCO)	1.47	7016	1840	463 (25%)	336 (18%)
S. noctilio (ISSR)			1013	512 (50%)	159 (15%)
D. siricidicola (FIASCO)	1.22	6388	1040	421 (40%)	296 (28%)

[a]Sequences that contain microsatellites that are larger than 100 bp in size. (Percentages in parentheses show amount from original contig and single reads.)
[b]Microsatellites that are not within 20 bases of the terminal regions of the sequence. (Percentages in parentheses show the proportion of the amplifiable microsatellites relative to the total number of microsatellite-containing contigs/singletons.)

(Invitrogen, Carlsbad, CA, USA). After sorting the assembled contigs and remaining single reads according to size, all contigs or single reads shorter than 100 bp were discarded from further study as they can complicate subsequent primer design.

All contigs or sequences longer than 100 bp were searched for microsatellites using the MSatFinder interface (www.genomics.ceh.ac.uk/msatfinder). For this purpose, perfect microsatellite repeats were identified using a regex-directed search engine. The minimum number of repeat motifs used to flag a sequence as containing a microsatellite was 12 repeats for mononucleotide motifs and 5 repeats for the remaining repeat classes (di-, tri-, tetra-, penta-, and hexanucleotides). The search was performed for sequences only containing microsatellites such as mono- to hexanucleotide repeats, per the definition of a microsatellite region (10). Tab-delimited files were generated from the searches using the MSatFinder web interface, and converted to spreadsheet files for use in Microsoft Office Excel 2007 (Microsoft Corp., Redmond, WA, USA) for subsequent data analysis including sorting according to size, determination of repeat class numbers, and average repeat length, using the data filter command in Microsoft Excel.

Sequences were filtered according to the position of the microsatellite contained within the sequence data. This was done in Microsoft Excel, as the results file from the MSatFinder contained the start and stop positions of the microsatellites. Since primers should be located in regions flanking the microsatellite sequences, those sequences starting within 20 bp from the 5′ terminal region or ending within 20 bp of the 3′ terminal region of the sequence were discarded.

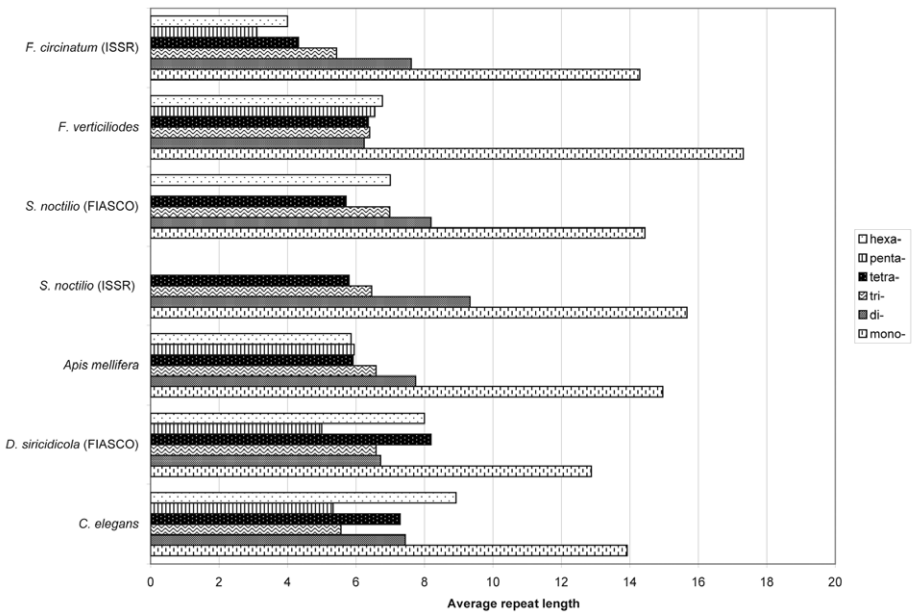

Figure 2. Average repeat length of all microsatellites discovered for each repeat unit class. Repeats were measured as the total number of times a motif was repeated. The average was taken for all microsatellites identified in a specific repeat class.

Table 2. Polymorphic Microsatellite PCR Primers Designed for *F. circinatum*

Locus Name	GenBank accession no.	Repeat sequence	Primer name	Primer sequence	Annealing Temperature (°C)
FCM-2	FJ436307	$(TTTC)_6$	FCM-2A	CGGAAGCAATCAGGACATTT	50
			FCM-2B	GAGCATGATGTCTCTCGAAGC	
FCM-3	FJ436308	$(CATGAG)_6$	FCM-3A	CAGTATGATAAGGCACCCATGT	53
			FCM-3B	GACTGACCCCTTGCCCTTAT	
FCM-4	FJ436309	$(TTCTT)_5$	FCM-4A	TGGTCCCGGCTCATTTACTA	50
			FCM-4B	AAAAGAAGACCCGCCTGATG	
FCM-6	FJ436310	$(GTGC)_7$	FCM-6A	GATGGAGATGAATGGGGAAA	50
			FCM-6B	GCCTCAGGTTGGTCTGGTTA	
FCM-7	FJ436311	$(AGGAGA)_5$	FCM-7A	ACGGCAGTGAAAAGAAGCAT	50
			FCM-7B	CAAGACCCTCTTGGCATCTC	
FCM-9	FJ436312	$(CAACGA)_8$	FCM-9A	CGACGACGACGACAACGAC	60
			FCM-9B	CTCCTCTTTGGCCCTCTTG	
FCM-16	FJ436313	$(CATCCA)_{18}$	FCM-16A	CGGATGAGAAAGCGAGAGAG	50
			FCM-16B	GGTGGATCACAGACCACAAA	
FCM-19	FJ436314	$(A)_{13}$	FCM-19A	GCGTCTTCCTCTGCCATTT	50
			FCM-19B	TAAGATTGAGGTTGTGCGGTTG	
FCM-20	FJ436315	$(ACTGT)_9$	FCM-20A	GCTGATCGAAGCCAATCG	50
			FCM-20B	TGACTACGCCAGAAGAGACG	
FCM-23	FJ436316	$(CACTT)_8$	FCM-23A	GGAGGTTATTGTCCGTCTCAA	50
			FCM-23B	ACTGAGGTGTGCCAAGCTGT	
FCM-24	FJ436317	$(TGATTG)_7$	FCM-24A	GACAGTTAGTCAGTCT-TAGTCTCG	45
			FCM-24B	GTCTCTTGCAGTCACAATCAC	
FCM-25	FJ436318	$(TGTCT)_6$	FCM-25A	TGATTCCTCCTGCCTCATTC	50
			FCM-25B	TAGGGCGATGTCTCTGGTTT	
FCM-26	FJ436319	$(A)_{12}$	FCM-26A	CAAACCGGCCAGAGACAG	45
			FCM-26B	TTCTCCAACCACCCTTGAAC	

This step was performed to streamline future design of primers for amplification of microsatellite markers.

To compare the sequence similarity between the microsatellites discovered in *S. noctilio* using ISSR-PCR versus FIASCO methods, the sequences that were amplifiable from each method (159 in ISSR-PCR; 296 in FIASCO) were used in ContigExpress to construct contigs. Sequences with >95% sequence similarity including the microsatellite regions were assembled into a contig. Manual searching through these contigs was performed to determine if they contained shared sequences generated from the two methods, which allowed for an estimation of sequence similarity between the two methods.

To determine whether the technique displayed any bias toward a certain microsatellite repeat class, the distribution of microsatellites from genome sequences of taxa related to the studied organisms were downloaded from GenBank (www.ncbi.nlm.nih.gov). These organisms were the insect *Apis mellifera*, the nematode *Caenorhabditis elegans*, and the fungus *F. verticillioides*. Their sequences were searched for microsatellites using the MSatFinder web interface with the same selection criteria as that for *F. circinatum*, *S. noctilio*, and *D. siricidicola*.

Comparison between next-generation sequencing and traditional cloning and sequencing

The number of microsatellites generated following traditional cloning and sequencing was determined for *F. circinatum* to assess the efficiency of next-generation sequencing. The pooled ISSR amplicons for this fungus were cloned using the pGEM-T easy cloning kit (Promega, Madison, WI, USA). Cloned inserts (≥100 bp) were amplified from a total of 100 recombinant colonies using vector-specific primers and standard PCR conditions. PCR products were purified with the Invitek MSB Spin PCRapace cleanup kit (Berlin, Germany) and sequenced using Applied Biosystems' BigDye version 3.1 sequencing kit and ABI 3130xl Sequencer (Foster City, CA, USA). The resulting sequences were analyzed and filtered as described in the "Microsatellite discovery" section.

Microsatellite primer design and testing for *F. circinatum*

Following successful identification of microsatellites in *F. circinatum*, sequences were analyzed for primer design. In order to increase the possibility of targeting polymorphic loci, sequences were chosen that had a higher-than-average repeat number in comparison to the average repeat number for each microsatellite repeat class for *F. circinatum*. Primers were developed using the Primer3 web-based interface (primer3.sourceforge.net) (29), and then tested for reproducible amplification using standard PCR conditions, with annealing temperatures altered according to primer sequence. Primer pairs were tested for their ability to amplify polymorphic bands by using them in PCR reactions that included eight *F. circinatum* isolates from different geographic regions. For this purpose, amplicons were separated and visualized using 12% PAGE (30) and Syber Gold (Molecular Probes, Inc., Eugene, OR, USA). To confirm the identity of the microsatellites, amplicons were purified and sequenced as described in the previous paragraph.

Results

Microsatellite enrichment and DNA pyrosequencing

To enrich genomic DNA for microsatellites, the ISSR-PCR (19) and FIASCO (16,27,31) methods were used. FIASCO enrichment using di-, tri-, and tetranucleotide repeat probes of the extracted *S. noctilio* and *D. siricidicola* genomic DNA, produced amplicons with a wide range of sizes, which appeared as smears following agarose gel electrophoresis. ISSR-PCR enrichment of the *F. circinatum* and *S. noctilio* DNA using primers with di-, tri-, and tetranucleotide repeats produced a defined number of amplicons that could be visualized as distinct bands on agarose gels. After pooling the enriched DNA for the respective individuals, it was subjected to Roche 454 GS-FLX pyrosequencing, which generated 1.67 Mb of sequence for *F. circinatum*, 1.22 Mb for *D. siricidicola*, and 1.47 Mb for *S. noctilio* (Table 1). These sequences represented large numbers of individual sequence reads, the majority of which could be assembled into contigs (Table 1).

Microsatellite discovery and analysis

Analysis of the assembled contigs and single reads (≥100 bp) revealed that the fungal data set contained the most (97%) microsatellite regions (Table 1). Only 41% of the sequences generated for the insect and nematode harbored microsatellites. However, a much higher proportion of the animal microsatellites was suitable for subsequent conversion to amplifiable markers. At least 70% of the nematode and 45% of the insect microsatellites were situated toward the middle of contigs or sequence reads, while most of the fungal microsatellites (85%) were located within 20 bp of the ends of a contig or sequence read. Overall, 495 of the *S. noctilio* sequences harbored microsatellite motifs that were flanked by regions suitable for designing primers for PCR amplification. For *D. siricidicola*, 296 such motifs were identified, while 231 were indentified for *F. circinatum*. For *S. noctilio*, more potentially amplifiable microsatellites were identified using FIASCO than ISSR-PCR enrichment (Table 1), and of these, only 10% shared high sequence similarity, suggesting that the two enrichment protocols targeted different microsatellite regions.

Analysis of the relative abundance of the various potentially amplifiable microsatellite classes revealed similar trends in the three species examined (Table 1 and Figure 1). In all cases the di- and trinucleotide repeat classes were much more abundant than the mono-, tetra-, penta-, and hexanucleotide classes. In general, the average length of the microsatellite repeat motif also decreased with motif complexity (Figure 2). The only exception was for the nematode *D. siricidicola* that had, on average, more repeats in the tetra- and hexanucleotide motifs than in the di- and trinucleotide motifs. These results were comparable to those obtained from the analysis of complete genome sequences for related organisms (*A. mellifera*, *C. elegans*, and *F. verticillioides*) using similar filtering and microsatellite search criteria (Figures 1 and 2).

Comparison between next-generation sequencing and traditional cloning and sequencing

To evaluate the efficiency of our new method for microsatellite discovery, we compared it to the traditional approach that involves cloning and Sanger sequencing. Of the 100 cloned inserts considered, all harbored microsatellites at their 3′ and 5′ prime ends. Sequence analysis using the same criteria as described in "Comparison between next-generation sequencing and traditional cloning and sequencing" (in "Materials and methods") revealed that eight of the cloned inserts harbored microsatellite motifs that are potentially amplifiable. Of these, four represented dinucleotide motifs and four represented trinucleotide repeats.

Microsatellite markers for *F. circinatum*

To determine whether the microsatellite discovery procedure described here allows

for the identification of regions that can be converted to amplifiable microsatellite markers, we used the sequence information for *F. circinatum*. We designed 28 primer pairs to amplify microsatellite regions, 19 of which allowed amplification of single fragments in the expected size range. Their evaluation on a collection of *F. circinatum* isolates using PAGE rendered 13 primer pairs that yielded polymorphic amplicons for the isolates (Table 2). Sequencing of the amplicons (GenBank accession nos. FJ436307–FJ436319) showed that all of the primer pairs amplified the targeted microsatellite loci.

Discussion

This study marks one of the first that employs Roche 454 technology for sequencing DNA libraries enriched for microsatellites. Our approach of deep sequencing enriched nematode, insect and fungal genomic DNA facilitated the identification of large numbers of microsatellite-containing sequences. In fact, the number of sequences produced using this method resulted in the identification of more microsatellite sequences (by at least 1 order of magnitude) than what is usually generated in studies based on regular cloning technologies (16). This method substantially increased the throughput of microsatellite discovery by excluding the time-consuming steps of cloning and subsequent clone sequencing.

The frequency and length of the various microsatellite classes identified for the organisms in this study closely resembled those of related taxa for which whole-genome sequence information is available (Figure 1). The microsatellite classes in *A. mellifera*, *C. elegans*, and *F. verticillioides* have frequencies and corresponding repeat numbers that are not substantially different to those identified for *S. noctilio*, *D. siricidicola* and *F. circinatum*, respectively (Figures 1 and 2). These relative abundance values are also comparable to those observed in a previous survey of eukaryotic genomes (5). For example, trinucleotide motifs were found to be more abundant in fungi than animals, while dinucleotide motifs were more abundant in animals than in fungi (5). *Sirex noctilio* was the only exception as we found no pentanucleotide microsatellite motifs (Figure 1) (5), which could be a function of the insect's evolutionary history (15). The new method presented here, therefore, is not biased for specific microsatellite repeat classes.

The difficulties associated with developing microsatellite markers for eukaryotes (16) and fungi in particular (11,17,32) are well documented. These difficulties mostly arise from the small genome size and/or the low genomic abundance of microsatellite loci (11,17,32). The fact that we did not encounter pentanucleotide microsatellites for *S. noctilio* (Figure 1) is probably a manifestation of the latter problem, where microsatellites are not only relatively rare, but certain classes of microsatellites also have limited distribution. Despite these potential limitations, the new technique allowed us to identify a large number of potentially amplifiable microsatellite motifs for the two animal and the fungal representatives (Table 1). For *S. noctilio* and *D. siricidicola,* 16–29% of the sequenced contigs/singletons harbored potentially amplifiable microsatellite motifs, which is comparable to those reported in many animal microsatellite discovery studies (33–36). For the fungus, at least 12% of our contigs/sequence reads harbored microsatellites from which primers could be designed, which is within the range that has been reported previously for fungi (11,17,32). Our findings for *F. circinatum* are also consistent with those of previous studies, where no amplifiable microsatellites were found using Sanger sequencing of cloned ISSR amplicons (37) and FIASCO-enriched (unpublished data) libraries. Therefore, the relatively low number of microsatellites identified for this fungus most likely reflects the limited representation of microsatellites in the genome of *F. circinatum*.

The enrichment procedure employed in any microsatellite study is an important consideration. Our results clearly showed that ISSR-PCR targets significantly more microsatellites than FIASCO, although the latter method yielded more sequences that are potentially transformable into PCR markers (Table 1). This may be due to the fact that ISSR-PCR employs primers that contain microsatellite sequences, the application of which results in sequences containing microsatellites at their terminal ends. As such, these terminal microsatellites lack suitable flanking regions for primer design and thus, are not directly convertible to amplifiable markers. For this reason, we excluded all contigs and sequence reads with microsatellites located only in the terminal regions. Although considerable sequence information was lost in this way, large numbers of sequences—comparable to those identified using FIASCO—remained, from which primers could potentially be designed. We also performed direct comparisons of the two methods in terms

of frequency distribution of microsatellite class and microsatellite length for the two *S. noctilio* genomic DNA libraries, enriched using the two methods. In this case, no significant differences ($P > 0.05$) were found (Figure 1). This is in contrast to anecdotal evidence (17) suggesting that ISSR-PCR enrichment preferentially leads to the isolation of short microsatellites (17). We therefore believe that either enrichment protocol can be used to characterize microsatellites for any organism of interest. However, we suggest that the more labor-intensive FIASCO method only be used in cases where specific microsatellite classes are needed, as enrichment with specific microsatellite probes will facilitate their quick identification. When there is no need for the identification of a specific microsatellite class, ISSR-PCR is an excellent enrichment option, as numerous microsatellites and microsatellite classes can be identified rapidly. Furthermore, in cases where microsatellite loci are exceptionally rare, the ISSR-PCR–based deep sequencing approach will be valuable. Even if microsatellites are not situated toward the middle of contigs, genome walking procedures could still be used to convert the terminal microsatellites to amplifiable markers.

Eliminating the construction of microsatellite-enriched clone libraries and substituting Sanger sequencing with pyrosequencing significantly enhances the cost-effectiveness of microsatellite discovery. For *F. circinatum*, a total of 231 amplifiable microsatellites were identified from the pyrosequencing data after one sequencing run using pooled ISSR-PCR products. In contrast, only eight amplifiable microsatellites were discovered following Sanger sequencing of 100 cloned ISSR-PCR fragments. To obtain the same number of microsatellites as generated from our pyrosequencing data, approximately 2800 additional clones would have to be sequenced. However, in reality, the number of clones will be substantially greater since our estimate assumes a 100% sequencing success and the presence of usable microsatellite sequences in all clones. At the time of this study, the cost of generating the 1692 unique sequences that were obtained for *F. circinatum* through pyrosequencing would have been 62% greater if Sanger sequencing was employed. When only amplifiable microsatellites are considered, pyrosequencing resulted in a 276% cost reduction per microsatellite. As technology develops, the price of identifying new microsatellites with pyrosequencing will be further reduced, allowing for even cheaper development of microsatellites.

In order to show that the microsatellite regions that were sequenced were also useful for the development of polymorphic markers, we synthesized primers flanking 29 microsatellite regions in *F. circinatum*. Of these, 19 produced single amplicons, and their application on a set of only eight *F. circinatum* isolates showed that 13 primer pairs targeted polymorphic microsatellite loci (Table 2). Thus, in a space of 2 months, we were able to identify more PCR-based molecular markers than have previously been available for this species (37,38). What makes the technique described here particularly powerful is that while 29 microsatellite regions were chosen for analysis, a further 212 microsatellite regions were available for testing, without requiring any further cloning or sequencing. In addition, we were able to be selective regarding the regions used to develop primers. This is in contrast to previous studies with severe limitations in the amount of available sequence (37). Therefore, despite the fact that no sequence data are available for a specific species, our approach of microsatellite enrichment and deep sequencing genomic DNA libraries makes it possible to rapidly develop many of these powerful genetic markers.

The results of this study clearly demonstrate that sequencing essentially all enriched PCR amplification products for microsatellite regions using 454 genome sequencing technology represents a superior alternative to conventional screening of clone libraries using Sanger sequencing. This methodology for microsatellite discovery is much more robust than cloning and sequencing individual fragments, which is substantially more time consuming and delivers few microsatellite markers from a large library. It is possible to generate high-confidence sequence data that can be used to identify microsatellite regions in approximately one week. This technique also enables the sequencing of larger amounts of microsatellite-containing DNA that might have been lost due to inherent inefficiencies in the cloning of these fragments (17). This considerable saving of time and effort makes the development of microsatellites a much more certain and reasonable exercise for most research groups.

Acknowledgements

We thank the members of the Tree Protection Cooperative Programme (TPCP), the Department of Science and Technology (DST) & National Research Foundation (NRF) Centre of Excellence in Tree Health Biotechnology (CTHB), and the Technology and Human Resources for Industry Programme (THRIP) initiative of the Department of Trade and Industry (DTI) of South Africa for funding. We also thank Dr. Bernard Slippers for providing ideas and support that contributed substantially to this study.

The authors declare no competing interests.

References

1. **Chambers, G.K. and E.S. MacAvoy.** 2000. Microsatellites: consensus and controversy. Comp. Biochem. Physiol. B Biochem. Mol. Biol. *126*:455-476.
2. **Buschiazzo, E. and N.J. Gemmel.** 2006. The rise, fall and renaissance of microsatellites in eukaryotic genomes. Bioessays *28*:1040-1050.
3. **Tautz, D. and M. Renz.** 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Res. *12*:4127-4138.
4. **Katti, M.V., P.K. Ranjekar, and V.S. Gupta.** 2001. Differential distribution of simple sequence repeats in eukaryotic genome sequences. Mol. Biol. Evol. *18*:1161-1167.
5. **Tóth, G., Z. Gáspári, and J. Jurka.** 2000. Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res. *10*:967-981.
6. **Weber, J.L.** 1990. Informativeness of human (dC-dA)n.(dG-dT)n polymorphisms. Genomics *7*:524-530.
7. **Field, D. and C. Wills.** 1996. Long, polymorphic microsatellites in simple organisms. Proc. Biol. Sci. *263*:209-215.
8. **Gur-Arie, R., C.J. Cohen, Y. Eitan, L. Shelef, E.M. Hallerman, and Y. Kashi.** 2000. Simple sequence repeats in *Escherichia coli*: abundance, distribution, composition, and polymorphism. Genome Res. *10*:62-71.
9. **Jarne, P. and P.J.L. Lagoda.** 1996. Microsatellites, from molecules to populations and back. Trends Ecol. Evol. *11*:424-429.
10. **Hennequin, C., A. Thierry, G.F. Richard, G. Lecointre, H.V. Nguyen, C. Gaillardin, and B. Dujon.** 2001. Microsatellite typing as a new tool for identification of *Saccharomyces cerevisiae* strains. J. Clin. Microbiol. *39*:551-559.
11. **Lim, S., L. Notley-McRobb, M. Lim, and D.A. Carter.** 2004. A comparison of the nature and abundance of microsatellites in 14 fungal genomes. Fungal Genet. Biol. *41*:1025-1036.
12. **Luikart, G., P.R. England, D. Tallmon, S. Jordan, and P. Taberlet.** 2003. The power and promise of population genomics: from genotyping to genome typing. Nat. Rev. Genet. *4*:981-994.
13. **Queller, D.C., J.E. Strassmann, and C.R. Hughes.** 1993. Microsatellites and kinship. Trends Ecol. Evol. *8*:285-288.
14. **Barbará, T., C. Palma-Silva, G.M. Paggi, F. Bered, M.F. Fay, and C. Lexer.** 2007. Cross-species transfer of nuclear microsatellite markers: potential and limitations. Mol. Ecol. *16*:3759-3767.

15. **Selkoe, K.A. and R.J. Toonen.** 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecol. Lett. *9*:615-629.
16. **Zane, L., L. Bargelloni, and T. Patarnello.** 2002. Strategies for microsatellites isolation: a review. Mol. Ecol. *11*:1-16.
17. **Dutech, C., J. Enjalbert, E. Fournier, F. Delmotte, B. Barrès, J. Carleir, D. Tharreau, and T. Giraud.** 2007. Challenges of microsatellite isolation in fungi. Fungal Genet. Biol. *44*:933-949.
18. **Hamilton, M.B., E.L.E.L. Pincus, A. Di Fiore, and R.C. Fleischer.** 1999. Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. BioTechniques *27*:500-507.
19. **Zietkiewicz, E., A. Rafalski, and D. Labuda.** 1994. Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. Genomics *20*:176-183.
20. **Van Der Nest, M.A., E.T. Steenkamp, B.D. Wingfield, and M.J. Wingfield.** 2000. Development of simple sequence repeat (SSR) markers in Eucalyptus from amplified inter-simple sequence repeats (ISSR). Plant Breed. *119*:433-436.
21. **Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Friters, J. Pot, et al.** 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. *23*:4407-4414.
22. **Barnes, I., M.N. Cortinas, M.J. Wingfield, and B.D. Wingfield.** 2008. Microsatellite markers for the red band needle blight pathogen, *Dothistroma septosporum*. Mol. Ecol. Resour. *8*:1026-1029.
23. **Luque, C., L. Legal, H. Staudter, C. Gers, and M. Wink.** 2002. ISSR (inter simple sequence repeats) as genetic markers in Noctuids (Lepidoptera). Hereditas *136*:251-253.
24. **Pfunder, M. and J.E. Frey.** 2006. Isolation of microsatellite markers for *Contarinia nasturtii*, a European pest invading the New World. Mol. Ecol. Notes *6*:191-193.
25. **Squirrell, J., P.M. Hollingsworth, M. Woodhead, J. Russell, A.J. Lowe, M. Gibby, and W. Powell.** 2003. How much effort is required to isolate nuclear microsatellites from plants? Mol. Ecol. *12*:1339-1348.
26. **Margulies, M., M. Egholm, W.E. Altman, S. Attiya, J.S. Bader, L.A. Bemben, J. Berka, M.S. Braverman, et al.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature *437*:376-380.
27. **Cortinas, M.N., I. Barnes, B.D. Wingfield, and M.J. Wingfield.** 2006. Polymorphic microsatellite markers for the Eucalyptus fungal pathogen *Colletogloeopsis zuluensis.* Mol. Ecol. Notes *6*:780-783.
28. **Moller, E.M., G. Bahnweg, H. Sandermann, and H.H. Geiger.** 1992. A simple and efficient protocol for isolation of high molecular weight DNA from filamentous fungi, fruit bodies, and infected plant tissues. Nucleic Acids Res. *20*:6115-6116.
29. **Rozen, S. and H. Skaletsky.** 2000. Primer3 on the WWW for General Users and for Biologist Programmers, p. 365-386. *In* Krawetz, S. and S. Misener (Eds.), Bioinformatics Methods and Protocols: Methods in Molecular Biology, Humana Press, Totowa, NJ.
30. **Sambrook, J. and D.W. Russell.** 2001. Molecular Cloning: A Laboratory Manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
31. **Cortinas, M.N., N. Koch, J. Thain, B.D. Wingfield, and M.J. Wingfield.** 2004. First record of the Eucalyptus stem canker pathogen, *Coniothyrium zuluense* from Hawaii. Australas. Plant Pathol. *33*:309-312.
32. **Karaoglu, H., C.M.Y. Lee, and W. Meyer.** 2005. Survey of simple sequence repeats in completed fungal genomes. Mol. Biol. Evol. *22*:639-649.
33. **Kankare, M., M.K. Jensen, K.M. Kester, and I.J. Saccheri.** 2004. Characterization of microsatellite loci in two primary parasitoids of the butterfly *Melitaea cinxia, Cotesia melitaearum* and *Hyposoter horticola* (Hymenoptera). Mol. Ecol. Notes *4*:231-233.
34. **Tsuchida, K., T. Saigo, S. Tsujita, K. Takeuchi, N. Ito, and M. Sugiyama.** 2003. Polymorphic microsatellite loci for the Japanese paper wasp, *Polistes chinensis antennalis* (Hymenoptera: Vespidae). Mol. Ecol. Notes *3*:384-386.
35. **Johnson, P.C.D., L.M.I. Webster, A. Adam, R. Buckland, D.A. Dawson, and L.F. Keller.** 2006. Abundant variation in microsatellites of the parasitic nematode *Trichostrongylus tenuis* and linkage to a tandem repeat. Mol. Biochem. Parasitol. *148*:210-218.
36. **Plantard, O. and C. Porte.** 2003. Isolation and characterization of microsatellite loci in the sugar beet cyst nematode *Heterodera schachtii.* Mol. Ecol. Notes *3*:139-141.
37. **Britz, H., B.D. Wingfield, T.A. Coutinho, and M.J. Wingfield.** 2002. Sequence characterized amplified polymorphic markers for the pitch canker pathogen, *Fusarium circinatum.* Mol. Ecol. Notes *2*:577-580.
38. **Wikler, K. and T.R. Gordon.** 2000. An initial assessment of genetic relationships among populations *Fusarium circinatum* in different parts of the world. Can. J. Bot. *78*:709-717.

Received 18 September 2008; accepted 15 December 2008.

Address correspondence to Quentin C. Santana, Department of Genetics, University of Pretoria, Hatfield, Pretoria, 0002, South Africa. email: quentin.santana@fabi.up.ac.za