BioSpotlight

Mining Many More Microsatellite Markers

Microsatellite sequences consist of 1–6 nucleotide tandem repeats and are found in nearly all known organisms. Because they mutate at a high rate and are inherited in a simple Mendelian pattern, microsatellites are valuable for genome mapping, forensics, population genetics, conservation and evolutionary studies. Bioinformatics tools are available to identify microsatellites in those few species with extensively sequenced genomes, but for most species, the markers must be developed de novo. Identifying and developing microsatellite markers is time-consuming and labor-intensive, requiring construction of a genomic library enriched for repeated motifs, isolation and sequencing of candidate clones, primer design, PCR amplification, and testing for polymorphisms in unrelated individuals. The time requirements and expense prohibit many labs from pursuing such studies. To simplify this process and make it accessible to more laboratories, three independent groups, working in the labs of N. Gemmell (Otago University, Dunedin, New Zealand), M. Bunce (Murdoch University, Perth, Australia), and M. Wingfield (University of Pretoria, Pretoria, South Africa), present the first examples of microsatellite discovery using

next-generation sequencing technology. Gemmell's group sought to eliminate the most intensive steps by using the Genome Sequencer FLX (GS-FLX) system to conduct shotgun sequencing of the endangered blue duck of New Zealand. One run produced enough random sequences that they were able to identify repeated motifs using the MSATCOMMANDER software. They carried out a meta-analysis of papers describing microsatellite marker isolation and showed similar levels of efficiency between this next-generation sequencing method and traditional methods, but with costs reduced 3–5 times and projects completed in a much shorter timeframe. The Bunce group used the same method to characterize a polymorphic microsatellite from an extinct New Zealand moa, using the first set of microsatellite primers developed directly from ancient DNA (aDNA) templates. Next-generation sequencing is particularly useful for approaches with aDNA, whose poor preservation hinders successful library enrichment and construction using traditional methods. Wingfield's group, seeking to develop microsatellite markers from a fungus, a wasp, and a parasitic nematode, took the method one step further by enriching for microsatellites by either of two methods prior to GS-FLX sequencing: fast isolation by AFLP of sequences containing repeats

(FIASCO) or inter simple sequence repeat PCR (ISSR-PCR). The results showed that ISSR-PCR targeted significantly more microsatellites than FIASCO, but that FIASCO enrichment returned more sequences potentially amplifiable by PCR. Compared to previous studies, the greater number of sequences returned by either method offered more flexibility in the choice of regions used for developing microsatellite amplification primers. When compared directly with the traditional cloning and sequencing approach, their method led to the identification of at least an order of magnitude more microsatellite sequences and showed a 276% cost reduction for the isolation of a comparable number of amplifiable microsatellites. Additionally, their method showed no bias for specific microsatellite classes since similar microsatellite frequencies and lengths were observed in related organisms for which whole genome sequences are available.

See “Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing,” “Identification of microsatellites from an extinct moa species using high-throughput sequence data,” and “Microsatellite discovery by deep sequencing of enriched genomic libraries,” on pages 185, 195, and 217, respectively.

Fishing for Novel SNPs

Single nucleotide polymorphisms (SNPs) are important genetic markers for identifying candidate genes responsible for complex phenotypic traits and diseases. The simplest method for identifying novel SNPs is to compare the sequence of an amplified target region of interest with genomic sequence information in a database, but this can be problematic for organisms whose genomes have not been extensively sequenced. While there are alternative methods for isolating random SNPs for these organisms, they require a fair amount of effort and can be inefficient. In this issue, J.-Y. Xu, G.-B. Xu, and S.-L. Chen at the Yellow Seas Fisheries Institute (Qingdao, China) demonstrate their method to facilitate the discovery of novel SNPs in the genome of an organism that has not been extensively sequenced—in their case, *Cynoglossus semilaevis*, the half-smooth tongue sole. DNA from ten individuals was pooled together, digested with *MseI*, ligated to adapters and PCR-amplified. This DNA was then denatured and reannealed, and heteroduplexes with SNP mismatches were nicked on one strand by limited CEL I nuclease digestion. DNA synthesis with the strand-displacing *Bst* DNA polymerase allowed the incorporation of dUTP-biotin in the nicked fragments, and this biotinylated DNA was isolated with magnetic streptavidin beads, re-amplified, cloned, and sequenced. Primers based on these sequences were then used to amplify and clone the putative SNP-containing genomic regions from four individuals in order to identify and validate putative novel SNP sites. Of 10 such fragments, 9 contained a total of 18 validated SNPs. This cost-effective and rapid method should enable the isolation of novel SNPs from virtually any organism.

See “A new method for SNP discovery” on page 201.

Selected and written by Patrick C.H. Lo, Ph.D. and Kristie Nybo, Ph.D.

BioTechniques 46:151 (March 2009) doi 10.2144/000113097