CHARACTERIZATION, POLYMORPHISM ASSESSMENT, AND DATABASE CONSTRUCTION FOR MICROSATELLITES FROM BAC END SEQUENCES OF CATFISH: A RESOURCE FOR INTEGRATION OF LINKAGE AND PHYSICAL MAPS

Except where reference is made to the work of others, the work described in this thesis is my own or was done in collaboration with my advisory committee. This thesis does not include proprietary or classified information.

_____

Benjaporn Somridhivej

Certificate of Approval:

_____	_____
Yolanda J. Brady Associate Professor Fisheries and Allied Aquacultures	Zhanjiang (John) Liu, Chair Alumni Professor Fisheries and Allied Aquacultures
_____	_____
Nannan Liu Associate Professor Entomology and Plant Pathology	George T. Flowers Interim Dean Graduate School

CHARACTERIZATION, POLYMORPHISM ASSESSMENT, AND DATABASE CONSTRUCTION FOR MICROSATELLITES FROM BAC END SEQUENCES OF CATFISH: A RESOURCE FOR INTEGRATION OF LINKAGE AND PHYSICAL MAPS

Benjaporn Somridhivej

A Thesis

Submitted to

the Graduate Faculty of

Auburn University

in Partial Fulfillment of the

Requirements for the

Degree of

Master of Science

Auburn, Alabama
May 10, 2007

CHARACTERIZATION, POLYMORPHISM ASSESSMENT, AND DATABASE CONSTRUCTION FOR MICROSATELLITES FROM BAC END SEQUENCES OF CATFISH: A RESOURCE FOR INTEGRATION OF LINKAGE AND PHYSICAL MAPS

Benjaporn Somridhivej

_____
Signature of Author

_____
Date of Graduation

VITA

Benjaporn Somridhivej, daughter of Kosol Somridhivej and Smurchai Somridhivej, was born on August 26, 1974 in Bangkok, Thailand. She graduated from Bodindecha (Sing Singhaseni) high school, Bangkok, Thailand in 1993. She entered Kasetsart University, Bangkok, Thailand where she graduated with a Bachelor of Science in Fisheries Biology in 1997. After working as a fisheries biologist at the Department of Fisheries, Bangkok, Thailand for five years, she enrolled in Graduate School at Auburn University to pursue a Master of Science degree in Department of Fisheries and Allied Aquacultures in May 2005.

THESIS ABSTRACT

CHARACTERIZATION, POLYMORPHISM ASSESSMENT, AND DATABASE CONSTRUCTION FOR MICROSATELLITES FROM BAC END SEQUENCES OF CATFISH: A RESOURCE FOR INTEGRATION OF LINKAGE AND PHYSICAL MAPS

Benjaporn Somridhivej

Master of Science, May 10, 2007
(B.Sc., Kasetsart University, 1997)

71 Typed Pages

Directed by Zhanjiang (John) Liu

To apply genome-based technologies for genetic improvements using marker-assisted selection, genome research involving genetic linkage mapping and physical mapping is required, and integration of genetic and linkage maps would significantly enhance the capacities for genome research. In catfish, the major aquaculture species in the United States, linkage and physical maps have been constructed. However, integration of genetic linkage and physical maps demands large-scale, genome-wide hybridizations, or genetic mapping of polymorphic markers derived from bacterial artificial chromosome (BAC) clones whose location is known from the physical map.

In this work, we identified a large number of microsatellites from BAC end sequences of channel catfish, characterized the microsatellites, tested their utility for linkage mapping in a resource family used for genetic mapping, and constructed a web-searchable database for BAC end sequences, their linked microsatellites, microsatellite primers, PCR conditions, and polymorphic information. A total of 2,744 distinct BACs containing microsatellites were identified. Of these, 1,100 had sufficient and complex flanking sequences for PCR primer design. We have tested 500 primer pairs and found 211 (42.2%) were polymorphic and segregating in the resource family used for genetic mapping. These microsatellites represent a major fraction of co-dominant polymorphic markers identified to date in catfish, and should be a valuable resource for genetic mapping to increase linkage map resolution, and for integration of genetic linkage and physical maps.

ACKNOWLEDGEMENTS

The author would like to thank her major professor, Dr. Zhanjiang (John) Liu, for his instruction, support, enthusiasm and guidance throughout the learning processes. The author would also like to extend her thanks to lab members, Dr. Huseyin Kucuktas, Ping Li, Dr. Puttharat Baoprasertkul, Dr. Eric Peatman, Xu Peng, Wang Saolin and Jason Abernathy for their assistance and invaluable friendship. Moreover, the author is especially indebted to Dr. Yolanda Brady and Dr. Nannan Liu for taking the time to be on her thesis committee. Special thanks go to all her friends that gave her support and help. Finally, the author would like to thank her parents and all members of her family for their understanding, encouragement, and support.

Style manual used Aquaculture

Computer software used Microsoft Word 2002, Microsoft Excel 2002, Adobe Photoshop 6.0, FASTPCR and Msatfinder

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

Catfish is the most important cultured fish in the U.S. and accounts for over 50% of all U.S. aquaculture production. The catfish industry is valued at 2 to 3 billion dollars and production last year exceeded 700 million pounds. In the State of Mississippi and Alabama, catfish is one of the top agricultural commodities and it is extremely important for development of rural America in the Southeast and other areas in the U.S.

Despite the development of the aquaculture and catfish industry in the U.S., a large trade deficit, 8 billion dollars annually, exists for seafood products. It is disappointing that aquaculture products are the third largest trade deficit contributor following petroleum and automobiles. It is time to address our aquaculture problems, especially considering the collapsing natural fisheries. According to USDA estimates, the U.S. demand for seafood is increasing steadily and wild fisheries will be able to supply only 25% to 30% of the additional demand. Trade deficit for seafood products is expected to increase. One way of increasing seafood supply is to increase marine fishing, but the world fish stocks are in crisis. Almost two-thirds of marine stocks in the Pacific and Atlantic Oceans are being fully exploited or have already been overfished.

Future projections predict a steadily widening gap between the world's demand for fish and the ability of the oceans to meet it . The solution lies in development of aquaculture.

Several problems severely limit development of the catfish industry. Diseases cause the largest amount of loss in catfish industry. Superior brood stocks resistant to major diseases are desperately needed. Although a rich resource exists among Ictalurid catfish for resistance to major diseases such as *Enteric septicemia* of catfish (ESC) and columnaris, for fast growth and for high carcass yield, genomic research is required to introgress these genes for combined benefits. Resistance- and carcass yield-linked markers are especially needed for marker-assisted selection.

The major problems of the catfish industry are related to the low profit margins. As a matter of fact, catfish producers lost money in the last a couple of years because of very low catfish prices. Among many things, improving performance and production traits could potentially reduce production cost and, therefore, increase profit margins. Of many performance and production traits, the most important ones include growth rate, feed conversion efficiency, disease resistance, tolerance to low dissolved oxygen, tolerance to low water quality, processing yields, and seinability. Much progress has been made in improving these traits through various means. Disease resistance was improved through interspecific hybridization (Dunham et al., 1990; Dunham, 1996), intraspecific crossbreeding and strain selection (Wolters and Johnson, 1994). Efforts to improve growth rate were made through selection (Bondari, 1983; Dunham and Smitherman, 1983a), intraspecific crossbreeding (Dunham and Smitherman, 1983b; Bondari 1983), and interspecific hybridization (Dunham et al., 1990; Dunham, 1996); tolerance to low dissolved oxygen was improved through interspecific hybridization (Dunham et al., 1983b; Dunham, 1996); seinability was improved by interspecific hybridization (Dunham, 1996) and strain selection. It is believed that traditional selection

methods will continue to make major contributions to improving the genetic quality of broodstocks. However, several limitations of the traditional selection demands development of new selection approaches. For instance, a selected broodstock may not harbor the desired gene; accurate measurement of quantitative traits is difficult and progeny testing of the selected traits may require great efforts; selection for some important traits such as disease resistance and carcass yield may be impractical. Challenging fish with disease in a production environment is not desirable. Direct selection for carcass yield is lethal to the broodstocks. All these limitations demands novel approaches such as marker-assisted selection (MAS) in aquaculture species.

Marker-assisted selection is a selection procedure based on the presence or absence of specific DNA markers that have been previously identified to be linked to the performance traits under consideration. While traditional selective breeding is based on phenotypic observations, MAS is based on DNA markers. For instance, if a specific DNA marker is already known to be linked to disease resistance, then brood fish can be selected based on if the fish harbors this specific marker without any disease challenges. MAS offers a more accurate selection of the desired genotype by using the linkage information of a molecular marker and a certain phenotype.

Marker-assisted selection requires linkage information of molecular markers and the performance or production traits. Such information can be obtained through linkage mapping of quantitative trait loci (QTL) because most, if not all, of these traits are controlled by many genes and, therefore, are inherited in a quantitative fashion. In doing so, a large number of molecular markers are needed to construct a genetic linkage map. Certain genomic information is helpful for the development of molecular markers

including genomic sizes (both physical and recombination sizes), polymorphic rates, and availability and applicability of various molecular makers in the species of interest. Channel catfish has 29 pairs of chromosomes (LeGrande et al., 1984) and a genome size of 1.0×10^9 base pairs (Tiersh et al., 1990; Tiersh and Goudie, 1993).

Advances in molecular biology and instrumentation facilitated the rapid development of molecular markers. A molecular marker is a site of heterozygosity for some type of neutral DNA variation because it easily detected and numerous in a genome. When mapped by linkage analysis it fill voids between genes of known genome. DNA marker is a specific, unique sequence of DNA that can be detected, identified and tracked a location on the chromosome. At the beginning, genes were used as marker on genetic mapping but the problem is map based on genes is not very detailed. To dated, several types of DNA markers have been used in mapping including allozyme markers, restriction fragment length polymorphism (RFLP), randomly amplified polymorphic DNA (RAPD), amplified fragment length polymorphism, microsatellites, and single nucleotide polymorphism (SNP).

Allozyme markers are type I markers and should be highly useful as anchorage points for comparative mapping. Variation is detected at the protein level, and the markers are co-dominant. However, total number of polymorphic loci is small and polymorphic rates are low at each locus.

Restriction fragment length polymorphism markers (RFLPs) are co-dominant markers. They are easy to score. However, they have low levels of polymorphism; they are time-consuming and laborious; probes and/or sequence information are required. Because of these limitations, RFLP markers are rare in catfish.

Randomly amplified polymorphic (RAPD) is a method of creating genomic fingerprints using short arbitrary primers and PCR. It is suitable for, but not limited to, species of which little molecular genetic information is known (Welsh and McClelland, 1990; Williams et al., 1990). It is technically easy and highly economical; and polymorphism levels are very high. However RAPD has low reproducibility because of low annealing temperature used during PCR. RAPD markers are inherited as dominant markers. In catfish, Liu et al., (1999a) found that polymorphic rates of RAPD are low among strains of channel catfish, but high between channel catfish and blue catfish. More than 600 RAPD markers were identified in catfish (Liu et al., 1998a).

Amplified fragment length polymorphism (AFLP) is a PCR-based fingerprinting technique (Vos et al., 1995). In contrast to RAPD, AFLP uses long primers during PCR and, therefore, is much more reproducible. AFLP is robust, reliable, powerful, economical, and applicable to all species as previous genetic information is not required. However, AFLPs are inherited as dominant markers.

Microsatellites are simple sequence repeats (SSR) of 1-6 bp long. Microsatellites have even distributions on all chromosomes though abundance varies with species. Microsatellites are highly polymorphic and co-dominant markers making it highly reliable and useful for linkage mapping. Microsatellite loci are generally short so genotyping can be facilitated by PCR. Liu et al. (1999b) found that most of the microsatellite-flanking sequences have been conserved across the genus borders of the Ictalurid catfish. This will allow the development of comprehensive linkage map using interspecific hybrid system and various types of markers (Liu and Feng, 2001).

Microsatellites are abundant and distributed on all chromosomes. These microsatellites have high level of polymorphism (Litt and Luty1989; Weber and May 1989; Tautz 1989) and are co-dominant markers; therefore, they are highly reliable. The Microsatellite analysis requires a primer pair for each marker locus, but these primer sequences can easily be shared throughout the world and rapidly be constructed by using a DNA synthesizer. The microsatellite loci is generally short, which makes it easy to generated by PCR and the result obtained can be observed by electrophoretic method. The variation of microsatellite loci is considered by differences in the number of repeating units in DNA segments. Microsatellites have proven to be very useful for many purposes; namely, estimating genetic variation in natural populations (Bruford and Wayne 1993), studying paternity, identifying any individual, genetic and linkage mapping.

Genome research also requires understanding of the physical organization of the genome. Most often, such understanding is achieved through physical mapping using bacterial artificial chromosome (BAC) libraries. BAC contain large genomic DNA insert of 100-250 kb, and therefore, a single genome equivalent can be included in approximately 5000-6000 BAC clones. A typical BAC library contain 6 to 15X genome coverage of genomic DNA. In catfish, a BAC library, CHORI 212, was constructed and characterized (Wang et al., 2005). Fingerprints of BAC clones would allow them to be arrayed into contigs: a series of BAC clones overlapping one another spanning a large segments of the chromosome. The catfish physical map contains approximately 3000 contigs (Xu et al., in review).

BAC end sequences of microsatellites are more suitable than other resources because they not only provide an unbiased survey of genomic sequence, but also allow an overall glance at the types and relative abundance of microsatellites in an organism. BAC end sequences can be used for identifying conserved synteny for comparative genome analysis to observe evolution, construct physical map and integrate genetic linkage map.

The genetic linkage mapping is a map that show genetic distance of gene related to each other in each chromosome. This distance is called centimorgan (cM). The genetic linkage map can be analyzed by using recombinant frequency from linkage analysis. A recombinant frequency (RF) of 1% is equivalent to 1 cM. To construct genome maps, genetic linkage mapping techniques such as marker segregation followed by analysis of recombination frequency can be used.

Physical mapping is a map of the position of a cloned genomic fragment that identifies landmark on DNA by using molecular biology techniques. The purpose is to identify a set of overlapping cloned fragments that together encompass an entire chromosome or an entire genome (Griffiths et al., 1999).

Integration of genetic and linkage maps can be approached in two different ways. First, DNA markers that have already been mapped to genetic linkage maps can be used as probes to hybridize to high-density BAC filters. This approach can be made more effective by the adoption of two dimensional hybridizations (Han et al., 2000; Gardiner et al., 2004), but can be complicated by the presence of repetitive sequences, gene families, and pseudo-genes associated with the probes. While efforts have been devoted to hybridization studies in catfish (Bao et al., 2005; Peatman et al., 2006), several major

technical problems limit large-scale, genome-wide hybridization of microsatellite markers to BAC contigs. Second, polymorphic DNA markers can be developed from the known locations on the physical maps, but there is no polymorphic markers available before my work for this purpose.

The objective of this work is to generate polymorphic markers derived from BAC clones that have been physically mapped so that these polymorphic markers can be mapped genetically on the linkage map. Specifically, the objectives of this study are:

a) To identify microsatellites from BAC end sequences through data mining;
b) To characterize microsatellites identified from the BAC end sequences concerning repeat types, microsatellite repeat numbers, location within the BAC end sequences, flanking sequences, and a distinct set of BACs containing microsatellites;
c) To test polymorphism of BAC-derived microsatellites in our resource family used for the construction of the genetic linkage map by using PCR analysis and determination of their segregation among individuals of the resource family; and
d) To develop a database for the BAC-anchored microsatellites, making them a useful resource for the integration of the genetic linkage and physical map

II. CHARACTERIZATION, POLYMORPHISM ASSESSMENT, AND DATABASE CONSTRUCTION FOR MICROSATELLITES FROM BAC END SEQUENCES OF CATFISH: A RESOURCE FOR INTEGRATION OF LINKAGE AND PHYSICAL MAPS

1. Introduction

The major objectives of structural genomics are to elucidate genome structure, organization, and evolution (O'Brien et al., 1991). These issues are approached by linkage and physical mapping, genome sequencing, and comparative genome analysis. Linkage and physical mapping, in particular, provide a framework for the understanding of genome organization and set the foundation for whole genome sequencing. Thus, linkage maps have been constructed from various aquaculture species including rainbow trout (Young et al., 1998; Sakamoto et al., 2000; Nichols et al., 2003; Danzmann et al., 2005), Atlantic salmon (Moen et al., 2004; Gilbey et al., 2004), tilapia (Kocher et al., 1998; Agresti et al., 2000; Lee et al., 2005), channel catfish (Waldbieser et al., 2001; Liu et al., 2003), European sea bass (Chistiakov et al., 2005), sea bream (Franch et al., 2006), common carp (Sun and Liang, 2004), shrimps (Moore et al., 1999; Wilson et al., 2002; Li et al. 2003; Pérez et al., 2004; Li et al., 2006), oysters (Yu and Guo, 2003; Hubert and Hedgecock, 2004; Li and Guo, 2004), scallops (Li et al., 2005;

Wang et al., 2005), and abalone (Sekino et al., 2006; 2007; Liu et al., 2006). Similarly, efforts toward the construction of physical maps have been made in aquaculture species including the construction of large insert BAC libraries in Atlantic salmon (Thorsen et al., 2005), tilapia (Katagiri et al., 2001), and catfish (Quiniou et al., 2003; Wang et al., in review), and the construction of BAC-based physical maps in Atlantic salmon (Ng et al., 2005), tilapia (Katagiri et al., 2005), and channel catfish (Xu et al., in review).

In genetic linkage mapping, genome organization is characterized by the analysis of marker relationships. Markers on the same chromosome tend to segregate together as they are physically linked to one another. However, recombination frequency increases as the distances among markers increase. The recombination frequency, therefore, has been used to order markers on a chromosome. Use of a large number of markers, therefore, allows construction of detailed genetic linkage maps that can place genetic markers on the genome. Genetic maps, however, are purely based on genetic distances in relation to genetic recombination frequency. In spite of the generally parallel relationship between genetic distance and physical distance, recombination frequency can vary greatly among organisms, or among various genome regions within an organism. In addition, once a trait is mapped genetically, the only information known is the distance of this trait to certain markers. Without a physical map, further studies and analysis of the gene controlling the trait is hindered.

In contrast to the situation of genetic linkage mapping, physical maps are constructed using physical pieces of DNA. In most cases, whole genome physical maps are constructed using large-insert BAC contigs, where overlapping BAC clones are ordered by their overlapping patterns of restriction enzyme fingerprints. With a well-

developed physical map, accurate distances between any of the BAC clones can be obtained. However, physical maps lack genetic information concerning performance traits that can only be mapped genetically. Therefore, the integration of genetic and physical maps becomes essential for the identification and analysis of the genes underlining performance traits. Once the traits or traits-linked markers are mapped to physical maps, the exact DNA sequences between a set of markers mapped in the proximity of performance traits can be decoded by DNA sequencing. For aquaculture research, linkage mapping allows connection of performance or production traits with genomic regions, while physical mapping establishes the relationships of physical segments of DNA for further characterization. Integration of genetic linkage map with physical map would allow performance traits to be placed on physical intervals of DNA segments, whereby candidate genes can be identified and characterized.

Integration of genetic and linkage maps can be approached in two different ways. First, DNA markers that have already been mapped to genetic linkage maps can be used as probes to hybridize to high-density BAC filters. This approach can be made more effective by the adoption of two dimensional hybridizations (Han et al., 2000; Gardiner et al., 2004), but can be complicated by the presence of repetitive sequences, gene families, and pseudo-genes associated with the probes. While efforts have been devoted to hybridization studies in catfish (Bao et al., 2005; Peatman et al., 2006), several major technical problems limit large-scale, genome-wide hybridization of microsatellite markers to BAC contigs. Second, polymorphic DNA markers can be developed from the known locations on the physical maps. In this approach, polymorphic markers such as microsatellites can be identified from BAC clones that are already fingerprinted for the

construction of physical maps. The polymorphic markers can then be genetically mapped by analysis using the resource families that were constructed for linkage mapping.

In channel catfish, various markers have been developed including microsatellite markers (Liu et al., 1998; Tan et al., 1999, for review, see Liu, 2003). In particular, the identification of a large number of microsatellites from expressed sequence tags (ESTs) (Serapion et al., 2004) has allowed the construction of a gene-based genetic linkage map useful for comparative genome analysis (Liu et al., in review). Linkage maps have been constructed using microsatellite markers (Waldbieser et al., 2001; Liu et al., in review). Recently, a BAC contig-based physical map has been constructed using the CHORI 212 BAC library (Xu et al., in review). As discussed above, integration of the genetic linkage map and the physical map would significantly enhance the capacities in catfish genome research. However, BAC-anchored markers have been lacking for map integration. In order to develop BAC-anchored microsatellite markers, we have initiated a BAC end sequencing project (Xu et al., 2006). Over 20,000 BAC end sequences have been produced (Xu et al., 2006). In this project, our objectives were to characterize microsatellites identified from the BAC end sequences, to test their polymorphism in our resource family used for the construction of the genetic linkage map, and to develop a database for the BAC-anchored microsatellites, making them a useful resource for the integration of the genetic linkage and physical maps.

2. Materials and Methods

2.1. Mining microsatellites from BAC end sequences

The FASTA file of the BAC end sequences was downloaded from NCBI and stored on the local computer for microsatellite mining. A Perl-based script *Msatfinder* (freeware, downloaded from http://www.genomics.ceh.ac.uk/msatfinder/) was used for microsatellite mining from the FASTA file (Thurston and Field, 2005). *Msatfinder* was run on the Linux operation system (Fedora Core 5). *Msatfinder* examines sequence files in GenBank, FASTA, EMBL and Swissprot formats, and determines the number, type and position of microsatellite repeats. The parameters were set at default for microsatellite searching: minimum-repeat number was set at eight for mono-and di-nucleotide repeats, and at five for tri- to hexa-nucleotide repeats. As mononucleotide repeats are not useful for mapping, they were manually excluded from the search output file.

Searches were conducted following Msatfinder Manual (http://www.genomics.ceh.ac.uk/msatfinder/msatfinder_manual.html) as the following: The search results were stored in the directories, Counts, Fasta, Flank_tabs, MINES, Msat_tabs Primers and Repeats. The first five directories included the motif, type, sequence, primer information and database files for each individual microsatellite. All the summary files were stored in the Repeats directory containing seven files. The *index* file is a handy summary of the results with the total motif types and numbers found in the total sequences. The *sequence* file contains the information on each sequence, including number of microsatellites found, GC content and so on. The *repeats* file contains the

details of every individual microsatellite, including the type and number of motif, the location of the microsatellites, plus similar genomic information to the sequence file. Both this file and the sequence file may easily be imported into Excel, or imported into a database. The *type.count* file showed the number of microsatellites found, categorized by motif type (mono-, di-, tri-, tetra-, penta-, and hexa-nucleotide repeats), and the number of repeat units. The *motif.count* file is similar to the *type.count* file, which shows the number of microsatellites found categorized by the bases/amino acids in the motif. These are ordered by the total base content only, thus AAT would be counted the same as TTA (exact summaries are available in the Counts/directory). The *primers.csv* file is a tabular summary of the information in all the primer files in the Primers/ directory. The *errors* file contains details of anything that looked unusual, e.g. very short sequences, features that did not match the sequence, and the summary of motif type and numbers.

2.2. Characterization of microsatellites

The *index* file was used to summarize the total motif types and numbers from the BAC end sequences and the distribution of motif types within the di-, tri- and tetra-nucleotide repeats. The *repeat* file with the microsatellite information was imported into Excel to get all the microsatellite information. All mononucleotide microsatellites were removed before analysis. First, the microsatellite-containing BAC end sequences were searched to determine whether they had sufficient flanking sequences for primer design by harboring at least 50 bp flanking sequences on either side of microsatellites. The file was sorted by start points, microsatellites were excluded if the start position began from 1-50 bp. The lengths between the end of microsatellites and the end BAC end sequences

were calculated by using the total lengths of sequence minus the stop positions of the microsatellites. Microsatellites were also excluded if the length was less than 50 bp after the microsatellites. All the remaining microsatellites were regarded as microsatellites with enough flanking sequences for the design of primers.

The distinct BACs with microsatellites with sufficient flanking sequence were identified by using linux command *uniq*. The resulting unique set of BAC end sequences containing microsatellites were used to design primers using *Msatfinder*. Only a fraction of these so-called microsatellite-containing BAC end sequences with sufficient flanking sequences supported successful primer design as many flanking sequences contain sequences of low complexity that prohibit generation of PCR primers using *Msatfinder*.

2.3. Assessing the utility of BAC-anchored microsatellites for linkage mapping

The usefulness of the identified microsatellites depends on their polymorphism. For genetic linkage mapping, their usefulness can be tested in the resource families. We have tested a fraction of the identified microsatellites in one of resource families, F_1-2 x channel catfish-6. PCR primers stored in the *primer.csv* file generated by *Msatfinder* provide five pairs of primer sequences. The first pair of primers was selected and purchased from Sigma Genosys (The Woodlands, TX) if the G/C content was 40 % to 60%, and the PCR product length was 100-300 bp. In cases where first pair of primers had a very low G/C content, or the anticipated PCR products were long (>300 bp), the second to the fifth pair of primers was evaluated until the criteria were met. The PCR

condition was optimized by FASTPCR according to the primers generated from *Msatfinder*.

PCR amplification was conducted using a thermocycler (Eppendorf AG, Brinkmann Instruments, Inc., Westbury, New York) using the following amplification profiles: 1X PCR buffer, 2 mM $MgCl_2$, 0.2 mM each of dNTPs, 4 ng upper PCR primer, 6 ng lower PCR primer, 1 pmol labeled primer, 0.25 units of JumpStart *Taq* polymerase (Sigma, St. Louis, MO), 20 ng genomic DNA, in a total reaction volume of 5 µl. After an initial incubation at 94°C for 90 seconds, PCR was carried out at 94°C for 30 seconds, 30 seconds at the appropriate annealing temperature specific with primer sets (50°C to 60°C, see database for details), 72°C for 45 seconds, for 35 cycles. Upon the completion of PCR, the reaction was incubated at 72°C for an additional 10 min. The PCR products were analyzed on 7% sequencing gels using a LI-COR automated DNA sequencer.

After gel electrophoresis, the positions of both alleles from the male and the female were determined, and their segregation was confirmed by genotyping eight fish of the mapping population. Upon genotype calling and determination of allele segregation, polymorphism in the resource family was determined.

2.4. Development of a database for BAC end sequences and their associated microsatellites

A database for BAC end sequences and microsatellites were developed based on the Apache/Mysql/PHP/CGI platform. The microsatellite information was categorized to six Excel sheets for the database. Sheet one contains the ID information including the GenBank BAC end sequence accession ID and Microsatellite ID. Sheet two contains the BAC end sequence information including GenBank BAC end sequence accession ID, AU ID (BAC clone name) and BAC end sequence. Sheet three contains the microsatellite information including Microsatellite ID, motifs, numbers of repeats, total length, position, and primer ID (if available for design primers). Sheet four contains the PCR primer information including primer ID, upper primer sequences and lower primer sequences. Sheet five contains the PCR condition information including primer ID, annealing temperature, cycles, and other reaction information. Sheet six contains polymorphism information including microsatellite ID, polymorphism information, and linkage group ID (will be amended when available). The primary key microsatellite ID, the foreign keys accession ID and primer ID were used to establish the relationship among the sheets in the database. This database will be available for query of the microsatellite information, including the BAC end sequences, motif type and numbers, primers, PCR conditions, and polymorphism information related to the linkage group.

3. Results

3.1. Identification of microsatellites in BAC end sequences

A total of 20,366 BAC end sequences we previously generated (Xu et al., 2006) was used as the source for the identification of BAC-anchored microsatellites. A total of 5,553 microsatellites (including multiple microsatellites per BAC end sequence) was identified from the 20,366 BAC end sequences. Of these, some BAC end sequences harbor more than one microsatellite; also as both BAC ends were sequenced, some BACs harbor microsatellites in both BAC end sequences. For the purpose of linkage mapping, we are interested in mapping only one microsatellite per BAC. A total of 3,652 distinct BAC clones were found to harbor at least one microsatellite. In order to be useful for mapping, the microsatellite-containing BAC end sequences must have sufficient flanking sequences for the design of PCR primers. Analysis using *Msatfinder* revealed that 605 BAC end sequences harbored microsatellites at the very beginning of the BAC end sequences, and 1,296 BAC end sequences harbored microsatellites at the end of the BAC end sequences (Table 1). These will not be useful for testing as markers unless additional sequencing is conducted. After eliminating these, a total of 2,744 distinct BAC end anchored microsatellites had sufficient flanking sequences for primer design (Table 1, for details, Table 4).

3.2. Characterization of the BAC-anchored microsatellites

The majority of the microsatellites identified from the BAC end sequences were dinucleotide repeats (63.5%), while the tri- and tetra-nucleotide repeats accounted for

22.0% and 14.5%, respectively (Figure 1). Of the dinucleotide repeats, the most abundant types were AC (27.1%), AT (27.0%), and GT (23.6%), while AG (14.0%) and CT (8.1%) were much lower; and the CG type was very rare (0.06%) (Figure 2). As the BAC end sequences were obtained from both strands of the catfish DNA and the true orientation of the BAC end sequences were unknown, the four distinct dinucleotide repeat types are AC/GT: 50.7%, AG/CT: 22.1%, AT/TA: 27.0%, and CG/GC: 0.06%. Clearly, the AC/GT type of dinucleotide repeats were the most abundant type in the catfish genome, accounting for over 50% of all dinucleotide repeats.

The tri-nucleotide repeats are uneven in distribution, with ATT (35.8%) and AAT (27.8%) being most abundant (Figure 3). These two types of tri-nucleotide repeats accounted for over 63.6% of all tri-nucleotide repeats. It is apparent that all A/T-rich repeat types were more abundant than G/C-rich repeat types. For instance, after the most abundant ATT and AAT (both are 100% A/T repeats), all tri-nucleotide repeats with two of their three bases of the repeats being A or T had a representation of at least 2.6%, whereas all G/C-rich tri-nucleotide repeats with two of their three bases being G or C were all below 0.6% of the tri-nucleotide repeats with the exception of AGG (1.5%) (Figure 3).

Figure 1. Percentage of di-, tri-, and tetra-nucleotide repeats identified from BAC end sequences of catfish.

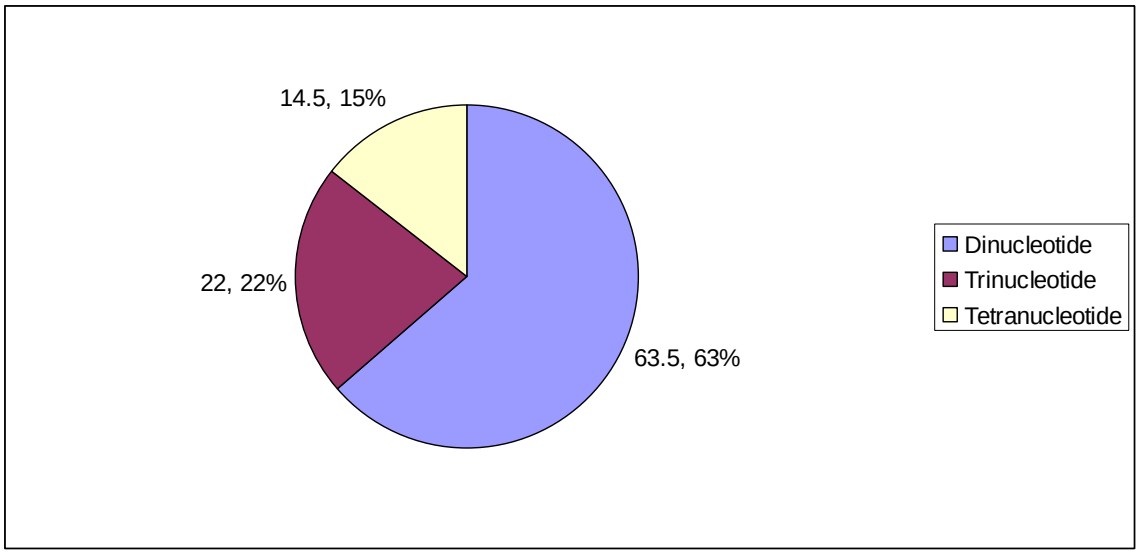


Figure 2. Distribution of various types of dinucleotide repeats identified from BAC end sequences of catfish. Note the low representation of G/C-rich types.

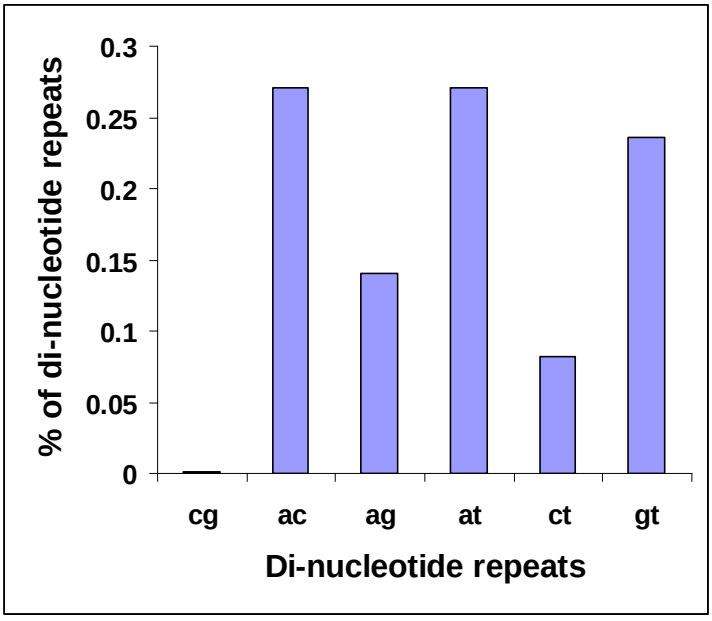
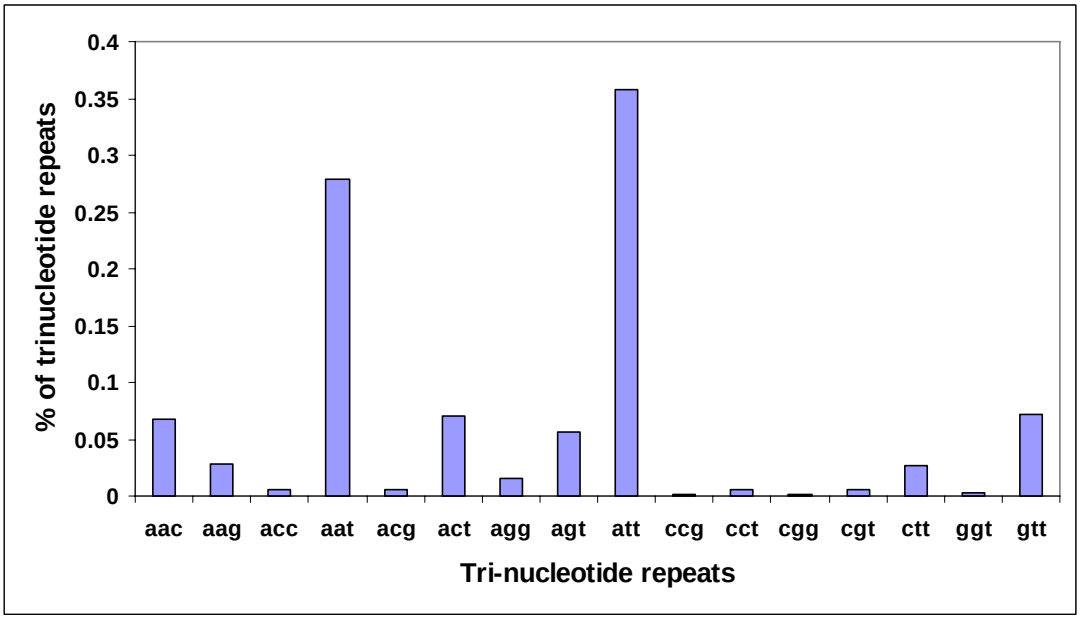


Figure 3. Distribution of various types of trileotide repeats identified from BAC end sequences of catfish. Note that A/T-rich types are highly abundant.

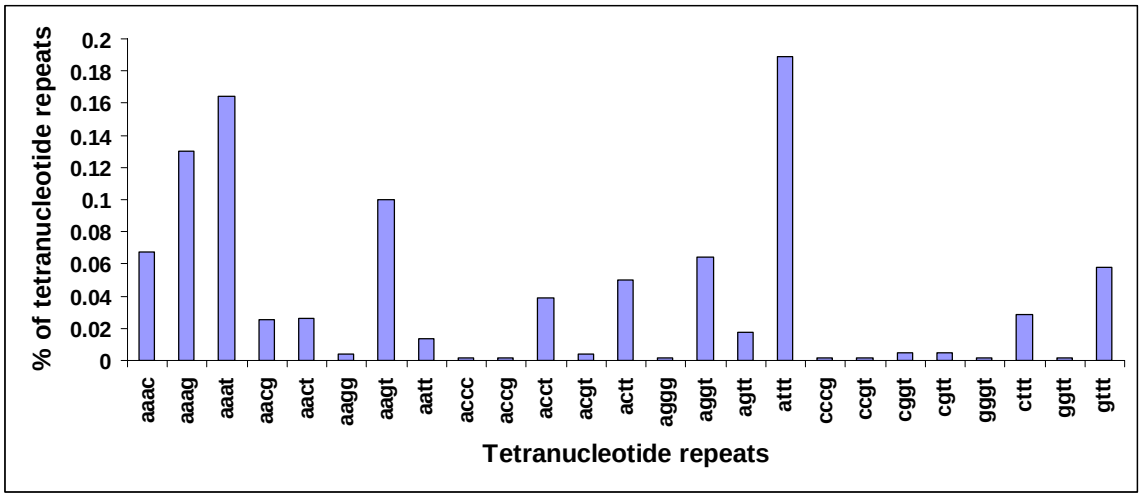


Very similar to the situation of the trinucleotide repeats, the distribution of tetranucleotide repeats was not even (Figure 4). They were most abundant with AAAT (18.9%) and TTTA (16.4%). In general, it was also true that tetra-nucleotide repeats with greater A/T had a greater representation. For instance, tetranucleotide repeats with at least three bases being A or T accounted for almost 80% of all tetranucleotide repeats with AAAG (13%), AAGT (10%), AAAC (6.7%), GTTT (5.7%), CTTT (2.8%), AACT (2.6%), AGTT (1.7%) among the most abundant types. The only exception appeared to be AATT which accounted for only 1.3% of all tetranucleotide repeats. G/C-rich tetranucleotide repeats were rare with many types not detected at all (Figure 4). Microsatellites with repeats longer than five bases were found rare in the catfish genome and therefore they were not characterized.

3.3. Assessment of the utility of the BAC-anchored microsatellites for linkage mapping

In order to be mapped on the genetic linkage map, microsatellites must be polymorphic in the resource families used for genetic linkage mapping. To assess the proportion of the BAC-anchored microsatellites useful for linkage mapping, PCR analysis was conducted using the parents of the mapping population. A total of 500 pairs of primers were ordered for testing. As shown in Table 2, of the 500 microsatellites tested, 211 (42.2%) were polymorphic within the ($F_1$2 x Channel 6) resource family.

Figure 4. Distribution of various types of tetranucleotide repeats identified from BAC end sequences of catfish. Note that A/T-rich types are highly abundant.



It seems that tri-nucleotide microsatellites produced the highest percentage of polymorphism within the resource family. Of the 349 tested dinucleotide repeats, 133 (38.1%) were polymorphic; of the 80 tested tri-nucleotide microsatellites, 46 (57.5%) were polymorphic; and of the 71 tested tetra-nucleotide microsatellites, 33 (46.5%) were polymorphic in the resource family (Figure 5). While polymorphic levels were similar (and also in some cases the numbers were too small to make a meaningful assessment, Table 3) among various types of trinucleotide and tetranucleotide microsatellites, it appeared that CT (48.5%) and AG (43.9%) types of dinucleotide repeats were most polymorphic, whereas the AT type (20.9%) of dinucleotide repeats were least polymorphic in the resource family (Figure 6).

3.4. Database construction for the BAC-anchored microsatellites

A web-based searchable database was constructed for the BAC end sequences, and their associated microsatellites. Information included in the database included BAC clone name, BAC end sequences, GenBank accession number, microsatellite motifs and location, microsatellite primer name, primer sequences, and PCR conditions.

Figure 5. Comparison of polymorphic rates of di-, tri-, and tetra-nucleotide repeats within the resource family.

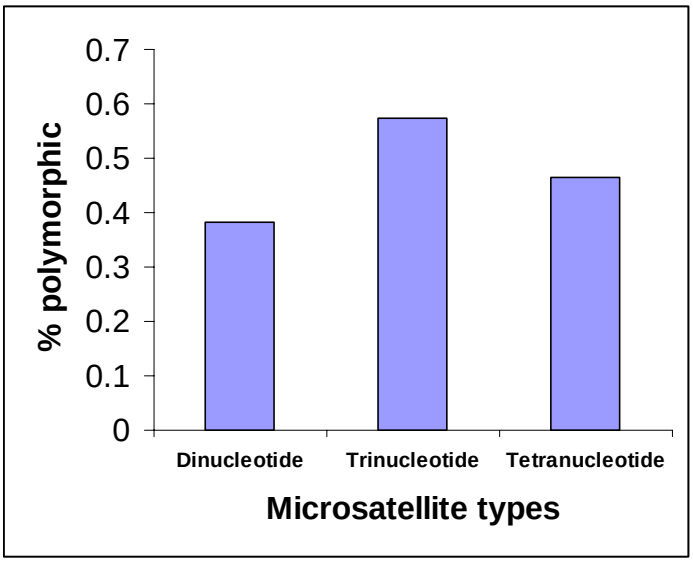
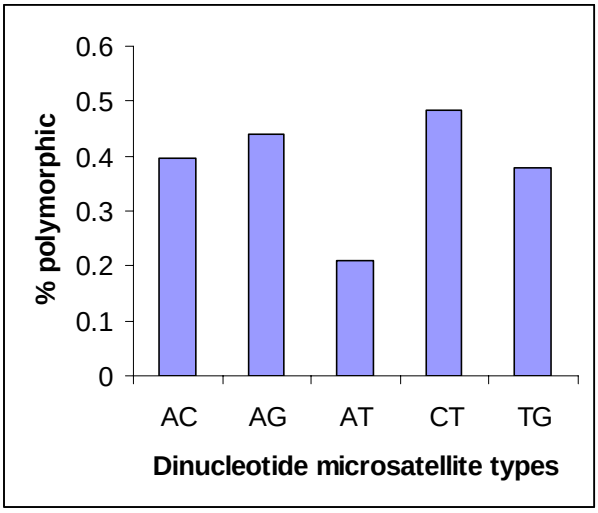


Figure 6. Comparison of polymorphic rates of various dinucleotide repeats within the resource family.



4. Discussion

In this work, a large number of microsatellites were identified from BAC end sequences of channel catfish. These microsatellites represent a major fraction of microsatellites in catfish identified to date. These microsatellites will be significant not only as potential polymorphic markers for genetic mapping, but also as the marker resource for integration of genetic linkage and physical maps as they were developed from BAC clones that are already fingerprinted for the construction of a physical map (Xu et al., in review). Mapping of these BAC end-derived microsatellites will not only add additional markers on the linkage map thereby increasing map resolution, but may also improve the coverage of the linkage map because BAC end sequences are more randomly distributed along the genome than are gene-containing regions.

Earlier efforts in microsatellite marker development in aquaculture species were accomplished by the construction of microsatellite-enriched libraries (e.g., Liu et al., 1999, Carleton et al., 2002; Coulibaly et al., 2005). However, recently, it has been shown that the identification of microsatellites through data mining is a very effective way for marker development (Serapion et al., 2004; Ju et al., 2005; Schwenkenbecher and Kaplan, 2007, Garnica et al., 2006; Blenda et al., 2006; Perez et al., 2005). In many instances, microsatellite markers were identified using EST resources. Here we demonstrate that data mining is also very effective using BAC end sequences for the purpose of identifying microsatellite markers. The limiting factor is the availability of BAC end sequences. In catfish, we previously generated 20,366 BAC end sequences. In order to integrate the linkage and physical maps to the fullest extent, microsatellite markers need

to be developed from as many BAC clones as possible among those BACs that have been fingerprinted for the construction of the physical map (Xu et al., in review). Efforts for the sequencing of additional BAC end sequences are ongoing in our laboratory. To fully integrate the physical map with the genetic linkage map, multiple polymorphic markers are needed from a single contig to both integrate and orient the linkage map with physical map.

Analysis of the utility of the BAC anchored microsatellites for linkage mapping was determined by testing the polymorphic status in the resource family used in catfish linkage mapping. In spite of the large numbers of microsatellites identified from BAC end sequences, their utility for linkage mapping depends on the nature of flanking sequences to support PCR primer design, the amplifiability of the designed microsatellite primers for the generation of PCR products with high fidelity, and the polymorphism of the microsatellites within the resource family. Clearly, the largest loss of the number of microsatellites useful for linkage mapping resulted from the flanking sequences with low sequence complexities. Of the 2,744 distinct BAC harboring sufficient flanking sequences as defined by the presence of at least 50 bp flanking sequences on either side of the microsatellites, only 1,100 (40%) supported primer design using *Msatfinder*. The next major reduction of useful microsatellite for linkage mapping resulted from the lack of PCR products or PCR products without fidelity, or the lack of polymorphism within the resource family. Of the 500 pairs of PCR primers tested, 211 microsatellites (42.2%) were polymorphic in the resource family. It appeared that the trinucleotide (57.5%) and tetranucleotide repeats (46.5%) had a higher level of polymorphism in the resource family than the dinucleotide repeats (38.1%). Among dinucleotide repeats, it appeared

that the AT repeats had the lowest polymorphic rate in the resource family tested. Such information concerning repeat types and polymorphic rates will allow us in the future to pick the microsatellites most likely to be polymorphic as our BAC end sequence resource expands. Obviously, the tri-, and tetra-nucleotide repeats are favored because of their greater polymorphic rates and much reduced problems in stutter bands, a common problem for dinucleotide repeats. The tested polymorphic microsatellites are ready for mapping.

Based on this polymorphic rate, an estimated 460 polymorphic microsatellites will be available for linkage mapping from the present set of BAC end sequences. In spite of their significance for linkage mapping and for the integration of the linkage and physical maps, many more BAC-anchored markers are required for full integration of linkage and physical maps. Catfish has 29 pairs of chromosomes, the estimated 460 markers will provide approximately 16 markers per chromosome, or approximately one marker per 8 cM. It is obvious that many more markers are needed to bring a greater level of map resolution for detailed analysis of aquaculture traits. From the perspective of physical mapping, the current assembly of the BAC contig-based physical map has over 3000 contigs. Therefore, just one marker per contig requires 3000 polymorphic markers, and multiple markers per contig are needed to orient the contigs on linkage maps. Clearly, more BAC ends should be sequenced. Additional efforts are ongoing in our laboratory in BAC end sequencing, and in refinement of the physical map to bring the number of contigs to a smaller scale.

The distribution of microsatellites in the catfish genome is highly biased toward A/T-richness. This is particularly true for tri- and tetra-nucleotide repeats as almost all

microsatellites with a higher A/T have a larger representation than the G/C-rich microsatellites. This is probably due to the fact that the catfish genome is AT-rich, estimated to be 60.7% (Xu et al., 2006).

Long term genome research requires establishment of various databases such that linkage information, BAC clones, their associated sequences and markers can be easily accessed and tracked. In this work, we have constructed a database presenting BAC end sequences, microsatellite location, microsatellite types, microsatellite primer location and sequences, PCR conditions, and polymorphic information in the resource families. This database can be amended upon generation of additional information related to linkage and physical maps. The microsatellites developed from BAC end sequences, along with this database, will provide a valuable resource for the integration of genetic linkage and physical maps in catfish.

III. Conclusions

My thesis project has four objectives:

a) To identify microsatellites from BAC end sequences through data mining;

b) To characterize microsatellites identified from the BAC end sequences concerning repeat types, microsatellite repeat numbers, location within the BAC end sequences, flanking sequences, and a distinct set of BACs containing microsatellites;

c) To test polymorphism of BAC-derived microsatellites in our resource family used for the construction of the genetic linkage map by using PCR analysis and determination of their segregation among individuals of the resource family; and

d) To develop a database for the BAC-anchored microsatellites, making them a useful resource for the integration of the genetic linkage and physical maps.

I have used the BAC end sequences generated in our laboratory (Xu et al., 2006), and mined for microsatellites using 20,366 BAC end sequences. I have identified a total of 2,744 distinct BACs harboring microsatellites, and thus the first objective was successfully reached.

I have characterized the identified microsatellites. Over 60% of all microsatellites identified were dinucleotide repeats, of which the major types were AC and GT types. Of the tri-, and tetranucleotide repeats, the A/T-rich types were more abundant.

I have assessed the utility of the microsatellites by testing their polymorphism in the resource family. Of the 500 pairs of tested primer pairs, 211 were polymorphic (42.2%). These polymorphic microsatellites will be useful for genetic linkage mapping. Mapping of these microsatellites will greatly enhance the map resolution. More importantly, once these microsatellites are mapped, many contigs will be anchored to the linkage map, allowing partial integration of the genetic linkage and physical maps. However, as the estimated number of polymorphic microsatellites is approximately 460 out of this batch of BAC end sequences, they are not enough to fully integrate the linkage map with physical map. This is because we have 3000 some contigs, and we need at least one microsatellite per contig to anchor the contigs to linkage maps. Moreover, multiple microsatellites per contig may be needed to orient the contigs. The limiting factor is the availability of BAC end sequences. I suggest that BAC end sequencing should be expanded and actually that is ongoing in our laboratory. Once more BAC

end sequences become available, similar work should be conducted to significantly increase the number of BAC-anchored microsatellites.

A database has been constructed contained all useful information such as BAC end sequences, GenBank accession numbers, microsatellites, their location, motif types, microsatellite primers, and PCR conditions. Obviously, due to lack of funding, microsatellites were not mapped yet, but that should be of highest priority. Once the microsatellites are mapped, the database should be amended accordingly.

Acknowledgements

This project was supported by a grant from USDA NRI Animal Genome Tools and Resources Program (award #2006-35616-16685). We are grateful for an equipment grant from the National Research Initiative Competitive Grant no. 2005-35206-15274 from the USDA Cooperative State Research, Education, and Extension Service.

References

Agresti, J.J., Seki, S., Cnaani, A., Poompuang, S., Hallerman, E.M., Umiel, N., Hulata, G., Gall, G.A.E., May, B., 2000. Breeding new strains of tilapia: development of an artificial center of origin and linkage map based on AFLP and microsatellite loci. Aquaculture 185, 43-56.

Bao, B., Peatman, E., Li, P., He, C., Liu, Z., 2005. Catfish hepcidin gene is expressed in a wide range of tissues and exhibits tissue-specific upregulation after bacterial infection. Developmental and Comparative Immunology 29, 939-950.

Blenda, A., Scheffler, J., Scheffler, B., Palmer, M., Lacape, J.M., Yu, J.Z., Jesudurai, C., Jung, S., Muthukumar, S., Yellambalase, P., Ficklin, S., Staton, M., Eshelman, R., Ulloa, M., Saha, S., Burr, B., Liu, S., Zhang, T., Fang, D., Pepper, A., Kumpatla, S., Jacobs, J., Tomkins, J., Cantrell, R., Main, D., 2006. CMD: a Cotton Microsatellite Database resource for Gossypium genomics.
BMC Genomics 7, 132.

Bondari, K., 1983. Response to bidirectional selection for body weight in channel catfish. Aquaculture 33, 73-81.

Bruford, M.W., Wayne, R.K., 1993. Microsatellites and their application to population genetics studies. Current Opinion in Genetics and Development, 3, 939-43.

Carleton, K.L., Streelman, J.T., Lee, B.Y., Garnhart, N., Kidd, M., Kocher, T.D., 2002. Rapid isolation of CA microsatellites from the tilapia genome. Animal Genetics 33, 140-144.

Chistiakov, D.A., Hellemans, B., Haley, C.S., Law, A.S., Tsigenopoulos, C.S., Kotoulas, G., Bertotto, D., Libertini, A., Volckaert, F.A., 2005. A microsatellite linkage map of the European sea bass *Dicentrarchus labrax L*. Genetics 170, 1821-1826.

Coulibaly, I., Gharbi, K., Danzmann, R.G., Yao, J., Rexroad, C.E., 3rd, 2005. Characterization and comparison of microsatellites derived from repeat-enriched libraries and expressed sequence tags. Animal Genetics 36, 309-315.

Danzmann, R.G., Cairney, M., Davidson, W.S., Ferguson, M.M., Gharbi, K., Guyomard, R., Holm, L.E., Leder, E., Okamoto, N., Ozaki, A., Rexroad, C.E., 3rd, Sakamoto, T., Taggart, J.B., Woram, R.A., 2005. A comparative analysis of the rainbow trout genome with 2 other species of fish (Arctic charr and Atlantic salmon) within the tetraploid derivative Salmonidae family (subfamily: Salmoninae). Genome 48, 1037-1051.

Dunham, R.A., 1990. Genetic Engineering in Aquaculture. AgBiotech News and Information. 2, 401-406.

Dunham, R.A., R.O. Smitherman, 1983a. Response to selection and realized heritability for body weight in three strains of channel catfish, *Ictalurus punctatus*, grown in earthen ponds. Aquaculture. 33, 88-96.

Dunham, R.A., R.O., Smitherman, 1983b. Crossbreeding channel catfish for improvement of body weight in earthen ponds. Growth. 47, 97-103.

Dunham, R.A., R.E., Brummett, M.O., Ella, R.O., Smitherman, 1990. Genotype-environment interactions for growth of blue, channel and hybrid catfish in ponds and cages at varying densities. Aquaculture. 85, 143-151.

Dunham, R.A., R.O., Smitherman, C., Webber, 1983. Relative tolerance of channel x blue hybrid and channel catfish to low oxygen concentrations. Progressive Fish Culturist. 45, 55-56.

Dunham, R.A., 1996. Contribution of genetically improved aquatic organisms to global food security. International Conference on Sustainable Contribution of Fisheries to Food Security. Government of Japan and FAO, Rome, Italy, 50 pp.

Franch, R., Louro, B., Tsalavouta, M., Chatziplis, D., Tsigenopoulos, C.S., Sarropoulou, E., Antonello, J., Magoulas, A., Mylonas, C.C., Babbucci, M., Patarnello, T., Power, D.M., Kotoulas, G., Bargelloni, L., 2006. A genetic linkage map of the hermaphrodite teleost fish *Sparus aurata L*. Genetics 174, 851-861.

Frelichowski, J.E., Jr., Palmer, M.B., Main, D., Tomkins, J.P., Cantrell, R.G., Stelly, D.M., Yu, J., Kohel, R.J., Ulloa, M., 2006. Cotton genome mapping with new microsatellites from Acala 'Maxxa' BAC-ends. Molecular genetics and genomics 275, 479-491.

Gardiner, J., Schroeder, S., Polacco, M.L., Sanchez-Villeda, H., Fang, Z., Morgante, M., Landewe, T., Fengler, K., Useche, F., Hanafey, M., Tingey, S., Chou, H., Wing, R., Soderlund, C., Coe, E.H., Jr., 2004. Anchoring 9,371 maize expressed sequence tagged unigenes to the bacterial artificial chromosome contig map by two-dimensional overgo hybridization. Plant Physiology 134, 1317-1326.

Garnica, D.P., Pinzon, A.M., Quesada-Ocampo, L.M., Bernal, A.J., Barreto, E., Grunwald, N.J., Restrepo, S., 2006. Survey and analysis of microsatellites from transcript sequences in Phytophthora species: frequency, distribution, and potential as markers for the genus. BMC Genomics 7, 245.

Gibley, J., Verspoor, E., McLay, A., Houlihan, D., 2004. A microsatellite linkage map for Atlantic salmon (*Salmo salar*). Animal Genetics 35, 98-105.

Griffths et al., 1999., An introduction to genetic analysis. pp. 445.

Han, C.S., Sutherland, R.D., Jewett, P.B., Campbell, M.L., Meincke, L.J., Tesmer, J.G., Mundt, M.O., Fawcett, J.J., Kim, U.J., Deaven, L.L., Doggett, N.A., 2000. Construction of a BAC contig map of chromosome 16q by two-dimensional overgo hybridization. Genome Research 10, 714-721.

Hong, J.M., Chae, S.H., Oriero, N., Larkin, D.M., Choi, C.B., Lee, J.Y., Lewin, H.A., Bae, J.H., Choi, I., Yeo, J.S., 2005. Identification and chromosomal localization of repeat sequences through BAC end sequence analysis in Korean cattle. Journal of Genetics 84, 329-335.

Hubert, S., Hedgecock, D., 2004. Linkage maps of microsatellite DNA markers for the Pacific oyster *Crassostrea gigas*. Genetics 168, 351-362.

Ju, Z., Wells, M.C., Martinez, A., Hazlewood, L., Walter, R.B., 2005. An in silico mining for simple sequence repeats from expressed sequence tags of zebrafish, medaka, Fundulus, and Xiphophorus. In Silico Biology 5, 439-463.

Kantety, R.V., La Rota, M., Matthews, D.E., Sorrells, M.E., 2002. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Molecular Biology 48, 501-510.

Katagiri, T., Asakawa, S., Minagawa, S., Shimizu, N., Hirono, I., Aoki, T., 2001. Construction and characterization of BAC libraries for three fish species; rainbow trout, carp and tilapia. Animal Genetics 32, 200-204.

Katagiri, T., Kidd, C., Tomasino, E., Davis, J.T., Wishon, C., Stern, J.E., Carleton, K.L., Howe, A.E., Kocher, T.D., 2005. A BAC-based physical map of the Nile tilapia genome. BMC Genomics 6, 89.

Kocher, T.D., Lee, W.J., Sobolewska, H., Penman, D., McAndrew, B., 1998. A genetic linkage map of a cichlid fish, the tilapia (*Oreochromis niloticus*). Genetics 148, 1225-1232.

Lai, C.W., Yu, Q., Hou, S., Skelton, R.L., Jones, M.R., Lewis, K.L., Murray, J., Eustice, M., Guan, P., Agbayani, R., Moore, P.H., Ming, R., Presting, G.G., 2006. Analysis of papaya BAC end sequences reveals first insights into the organization of a fruit tree genome. Molecular Genetics Genomics 276, 1-12.

Lee, B.Y., Lee, W.J., Streelman, J.T., Carleton, K.L., Howe, A.E., Hulata, G., Slettan, A., Stern, J.E., Terai, Y., Kocher, T.D., 2005. A second-generation genetic linkage map of tilapia (*Oreochromis spp.).* Genetics 170, 237-244.

LeGrande, W., R.A., Dunham, R.O., Smitherman, 1984. Comparative karyology of three species of north American catfishes (siluriformes: Ictaluridae: Ictalurus) and four of their hybrid combinations. Copeia. 1984, 873-878.

Li, L., Guo, X.M., 2004. AFLP-based genetic linkage maps of the Pacific oyster *Crassostrea gigas* Thunberg. Marine Biotechnology 6, 26-36

Li, L., Xiang, J.H., Liu, X., Zhang, Y., Dong, B., Zhang, X.J., 2005. Construction of AFLP-based genetic linkage map for Zhikong scallop, *Chlamys farreri* Jones et Preston and mapping of sex-linked markers. Aquaculture 245, 63-73

Li, Y.T., Byrne, K., Miggiano, E., Whan, V., Moore, S., Keys, S., Crocos, P., Preston, N., Lehnert, S., 2003. Genetic mapping of the kuruma prawn Penaeus japonicus using AFLP markers. Aquaculture 219, 143-156

Li, Z., Li, J., Wang, Q, He, Y., Liu, P., 2006. AFLP-based genetic linkage map of marine shrimp Penaeus (Fenneropenaeus) chinensis. Aquaculture 261, 463-472

Litt, M., Luty, J.A., 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. American Journal of Human Genetics, 44, 397-401.

Liu, X., Liu, X., Guo, X., Gao, Q., Zhao, H., Zhang, G., 2006. A preliminary genetic linkage map of the Pacific Abalone *Haliotis discus* hannai Ino. Marine Biotechnology (NY) 8, 386-397.

Liu, Z. J., Li, P., Argue, B., Dunham, R.A., 1998a. Inheritance of RAPD markers in channel catfish (*Ictalurus punctatus*), blue catfish (*I. furcatus*) and their F1, F2 and backcross hybrids. Animal Genetics, 29, 58-62.

Liu, Z., 2003. A review of catfish genomics: progress and perspectives. Comparative and Functional Genomics 4, 259-265

Liu, Z., Karsi, A., Li, P., Cao, D., Dunham, R.A., 2003. An AFLP-based genetic linkage map of channel catfish (*Ictalurus punctatus*) constructed by using an interspecific hybrid resource family. Genetics 165, 687-694.

Liu, Z., Tan, G., Li, P., Dunham, R.A., 1999. Transcribed dinucleotide microsatellites and their associated genes from channel catfish *Ictalurus punctatus*. Biochemical and research communications 259, 190-194.

Liu, Z.J., Tan, G., Kucuktas, H., Li, P., Karsi, A., Yant, D.R., Dunham, R.A., 1999b. High levels of conservation at microsatellite loci among Ictalurid catfishes. Jounal of Heredity 90, 307-312.

Moen, T., Hoyheim, B., Munck, H., Gomez-Raya, L., 2004. A linkage map of Atlantic salmon (*Salmo salar*) reveals an uncommonly large difference in recombination rate between the sexes. Animal Genetics 35, 81-92.

Moore, S.S., Whan, V., Davis, G.P., Byrne, K., Hetzel, D.J.S., Preston, N., 1999. The development and application of genetic markers for the Kuruma prawn *Penaeus japonicus*. Aquaculture 173,19-32

Mun, J.H., Kim, D.J., Choi, H.K., Gish, J., Debelle, F., Mudge, J., Denny, R., Endre, G., Saurat, O., Dudez, A.M., Kiss, G.B., Roe, B., Young, N.D., Cook, D.R., 2006. Distribution of microsatellites in the genome of *Medicago truncatula*: a resource of genetic markers that integrate genetic and physical maps. Genetics 172, 2541-2555.

Ng, S.H., Artieri, C.G., Bosdet, I.E., Chiu, R., Danzmann, R.G., Davidson, W.S., Ferguson, M.M., Fjell, C.D., Hoyheim, B., Jones, S.J., de Jong, P.J., Koop, B.F., Krzywinski, M.I., Lubieniecki, K., Marra, M.A., Mitchell, L.A., Mathewson, C., Osoegawa, K., Parisotto, S.E., Phillips, R.B., Rise, M.L., von Schalburg, K.R., Schein, J.E., Shin, H., Siddiqui, A., Thorsen, J., Wye, N., Yang, G., Zhu, B., 2005. A physical map of the genome of Atlantic salmon, *Salmo salar*. Genomics 86, 396-404.

Nichols, K.M., Young, W.P., Danzmann, R.G., Robison, B.D., Rexroad, C., Noakes, M., Phillips, R.B., Bentzen, P., Spies, I., Knudsen, K., Allendorf, F.W., Cunningham,

B.M., Brunelli, J., Zhang, H., Ristow, S., Drew, R., Brown, K.H., Wheeler, P.A., Thorgaard, G.H., 2003. A consolidated linkage map for rainbow trout (*Oncorhynchus mykiss*). Animal Genetics 34, 102-115.

Nonneman, D., Waldbieser, G.C., 2005. Isolation and enrichment of abundant microsatellites from a channel catfish (*Ictalurus punctatus*) brain cDNA library. Animal Biotechnology 16, 103-116.

O'Brien, S.J., 1991. Molecular genome mapping lessons and prospects. Current Opinion in Genetics & Development 1, 105-111.

Peatman, E., Bao, B., Peng, X., Baoprasertkul, P., Brady, Y., Liu, Z., 2006. Catfish CC chemokines: genomic clustering, duplications, and expression after bacterial infection with *Edwardsiella ictaluri*. Molecular Genetics Genomics 275, 297-309.

Perez, F., Erazo, C., Zhinaula, M., Volckaert, F., Calderon, J., 2004. A sex-specific linkage map of the white shrimp *Penaeus (Litopenaeus) vannamei* based on AFLP markers. Aquaculture 242, 105-118

Perez, F., Ortiz, J., Zhinaula, M., Gonzabay, C., Calderon, J., Volckaert, F.A., 2005. Development of EST-SSR markers by data mining in three species of shrimp: *Litopenaeus vannamei*, *Litopenaeus stylirostris*, and *Trachypenaeus birdy*. Marine Biotechnology (NY) 7, 554-569.

Quiniou, S.M., Katagiri, T., Miller, N.W., Wilson, M., Wolters, W.R., Waldbieser, G.C., 2003. Construction and characterization of a BAC library from a gynogenetic channel catfish *Ictalurus punctatus*. Genetics Seletion Evolution 35, 673-683

Sakamoto, T., Danzmann, R.G., Gharbi, K., Howard, P., Ozaki, A., Khoo, S.K., Woram, R.A., Okamoto, N., Ferguson, M.M., Holm, L.E., Guyomard, R., Hoyheim, B.,

2000. A microsatellite linkage map of rainbow trout (*Oncorhynchus mykiss*) characterized by large sex-specific differences in recombination rates. Genetics 155, 1331-1345.

Schwenkenbecher, J.M., Kaplan, R.M., 2007. Development and characterization of microsatellite markers for the canine hookworm, *Ancylostoma caninum*. Parasitology Research 100, 1015-1021.

Sekino, M., Hara, M., 2007. Linkage maps for the pacific abalone (genus Haliotis) based on microsatellite DNA markers. Genetics 175, 945-958.

Sekino, M., Kobayashi, T., Hara, M., 2006. Segregation and linkage analysis of 75 novel microsatellite DNA markers in pair crosses of Japanese abalone (*Haliotis discus hannai*) using the 5'-tailed primer method. Marine Biotechnology (NY) 8, 453-466.

Serapion, J., Kucuktas, H., Feng, J., Liu, Z., 2004. Bioinformatic mining of type I microsatellites from expressed sequence tags of channel catfish (*Ictalurus punctatus*). Marine Biotechnology (NY) 6, 364-377.

Sun, X.W., Liang, L.Q., 2004. A genetic linkage map of common carp (*Cyprinus carpio L.*) And mapping of a locus associated with cold tolerance. Aquaculture 238, 165-172

Tan, G., Karsi, A., Li, P., Kim, S., Zheng, X., Kucuktas, H., Argue, B.J., Dunham, R.A., Liu, Z.J., 1999. Polymorphic microsatellite markers in *Ictalurus punctatus* and related catfish species. Molecular Ecology 8, 1758-1760.

Tautz, D., 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Research 17, 6463-71.

Thorsen, J., Zhu, B., Frengen, E., Osoegawa, K., de Jong, P.J., Koop, B.F., Davidson, W.S., Hoyheim, B., 2005. A highly redundant BAC library of Atlantic salmon (*Salmo salar*): an important tool for salmon projects. BMC Genomics 6, 50.

Tiersch, T. R., C.A., Goudie, 1993. Inheritance and variation of genome size in half-sib families of hybrid ictalurid catfishes. Journal of Heredity 84, 122-125.

Tiersch, T. R., B. A. Simco, K. B. Davis, R. W. Chandler, S. S. Wachtel, G. J. Carmicheal 1990. Stability of genome size among stocks of the channel catfish. Aquaculture 87, 15-22.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van der Lee, M. Hanes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau, 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research 23, 4407-4414.

Waldbieser, G.C., Bosworth, B.G., Nonneman, D.J., Wolters, W.R., 2001. A microsatellite-based genetic linkage map for channel catfish, *Ictalurus punctatus*. Genetics 158, 727-734.

Wang, L.L., Song, L.S., Chang, Y.Q., Xu, W., Ni, D.J., Guo, X.M., 2005. A preliminary genetic map of Zhikong scallop (*Chlamys farreri* Jones et Preston 1904). Aquaculture Research 36, 643-653

Weber, J.L., May, P.E., 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. American Journal of Human Genetics 44, 388-96.

Welsh, J., M., McClelland, 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Research 18, 7213-7218.

Williams, J.G.K., A.R., Kubelik, K. J. Livak, J. A. Rafalski, and S. V. Tingey 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research. 18, 6531-6535.

Wilson, K., Li, Y.T., Whan, V., Lehnert, S., Byrne, K., Moore, S., Pongsomboon, S., Tassanakajon, A., Rosenberg, G., Ballment, E., Fayazi, Z., Swan, J., Kenway, M., Benzie, J., 2002. Genetic mapping of the black tiger shrimp *Penaeus monodon* with amplified fragment length polymorphism. Aquaculture 204, 297-309

Wolters, W. R., and M. R., Johnson, 1994. Enteric septicemia resistance in blue catfish and three channel catfish strains. Journal of Aquatic Animal Health 6, 329-334.

Young, W.P., Wheeler, P.A., Coryell, V.H., Keim, P., Thorgaard, G.H., 1998. A detailed linkage map of rainbow trout produced using doubled haploids. Genetics 148, 839-850.

Yu, Z.N., Guo, X.M., 2003. Genetic linkage map of the eastern *oyster Crassostrea* virginica Gmelin. Biological Bulletin 204, 327-338

Table 1

A summary of the microsatellites identified from BAC end sequences

BAC end sequences	20,366
Microsatellites found	5,553
Microsatellites at the beginning of BES	605
Microsatellites at the end of BES	1,296
BES with microsatellites and enough flanking sequences for primer design	3,652
Distinct BAC clones harboring microsatellites with enough flanking sequences for primer design	2,744

Table 2

Assessing the utility of the BAC-anchored microsatellites for linkage mapping. *Number of distinct BAC end sequences with at least 50 bp flanking sequences both upstream and downstream of microsatellites. Quality flanking sequences were defined as sequences that support primer design using Msatfinder.

Number of distinct BAC harboring microsatellites with sufficient flanking sequences for primer design*	2,744	
Number of distinct BAC harboring microsatellites with quality flanking sequences allowing for primer design**	1,100	40.1%
Primer pairs designed and purchased	500	45.5%
Number of polymorphic microsatellites in resource family (F_1-2 x Channel-6)	211	42.2%

Table 3

The number and polymorphism tested from various types of microsatellites

Microsatellite types	Number of microsatellite primer pairs tested	Number of polymorphic microsatellites	% polymorphic
AC	129	51	39.5
AG	41	18	43.9
AT	43	9	20.9
TC	33	16	48.5
TG	103	39	37.9
Sub-total	349	133	38.1
AGG	7	3	42.9
AAT	27	14	51.9
AAC	3	1	33.3
ATC	2	2	100
CTG	1	0	0
GGA	1	0	0
ATG	7	4	57.1
ATT	27	18	66.7
GTT	1	0	0
TCC	1	1	100
TTC	1	1	100
TGG	1	1	100
GTT	1	1	100
Sub-total	80	46	57.5
AACA	8	3	37.5
AATA	13	7	53.8
TAAC	2	1	50
AATC	2	1	50
TAAT	1	0	0
GACA	1	0	0
ATCT	4	2	50
ATGG	5	0	0
TCCA	4	3	75
TATT	21	9	42.9
TGTT	6	4	66.7
TTTC	2	1	50
TGAA	2	2	100
Subtotal	71	33	46.5
Total	500	211	42.2

Table 4

Microsatellites identified from BAC end sequences. P indicates polymorphism, NP indicates no polymorphism, empty cells indicate no PCR products were generated with fidelity such that genotyping was not possible.

BAC name	Locus name	Microsatellite position	Motif	Primer sequences (5' to 3')	Annealing temperature (°C)	Polymorphism
AU01010A1A06.f1	AUBES1073	350	$(TA)_{25}$	TGCTACTCTGTTGGTGCCAG GACACCAAAATGTGAAGGGTGTTCTC	50	
AU01010A1E12.f1	AUBES1074	350	$(AACA)_5$	GTCCAGTGTTTGGTAGCCAC GACCAACCAACTTTGAACCACATTGC	50	
AU01010A2A02.r1	AUBES1075	230	$(AATA)_9$	GTGCTCTGTTAGCTGGAGTG GACAAGCAAGCCTGGACCATGAC	55	P
AU01010A2E09.r1	AUBES1076	610	AATA(5)	AGCTACATAGCTGGGGAGTC GACCAGAACCACTGTGTCCACAG	50	
AU01010A2G11.r1	AUBES1077	440	$(AC)_{31}$	TCTACTGCTGCCTGTGAACG GACGCTGAAACAGACTGTGGACAC	50	P
AU01010B2H08.r1	AUBES1078	360	AATA(5)	TGGGTTGTGTGATGTGGCTC GACGGAAAAGCTGTTTATACCTGCTGG	50	P
AU01010B2A03.r1	AUBES1079	550	$(ATG)_7$	CGTTTCATTCCTCTTATGCCAGC GACGTTTCATACATGATCCAGGCCATC	55	P
AU01010B2H10.r1	AUBES1080	310	$(ATT)_9$	CACCTCCATGCCACCAGAGG GACCTGAAGCACTTCGGTCAACTC	50	P
AU01010B2A02.f1	AUBES1081	260	$(ATT)_{13}$	CATCAACTACAATATCAGCCGCAG GACTGGAGGCGACAGGCAGGTGG	50	P
AU01010B2C03.f1	AUBES1082	280	$(ATT)_{15}$	TTGGGGCCTGTGGGGCTTGG GACAGAAGTGTTCAGCCTGTTGG	55	P
AU01012A2A05.r1	AUBES1083	264	$(GA)_{28}$	ACAGGACGATGCTGGCAGTG GACTCGACACCAACATGACCGAC	55	P
AU01012B1E03.r1	AUBES1084	590	$(TTTG)_9$	ACCATCGTGTATCGCGGACG GACACGGAGTTGCAGTCACCACG	55	P
AU01012B2C01.f1	AUBES1085	347	$(TTTA)_6$	CGATCCTGTCCGGCGTTCTG GACAAACCCGGTGACACGACTGC	55	
AU01018B1F11.r1	AUBES1086	280	$(TG)_{13}$	TTGCCAAGAACCCGTTGAGC GACTGGGAAGCATTTGGTTTGGTC	55	P
AU01018B1F03.r1	AUBES1087	490	$(ATA)_{13}$	TCAAGTGCAGAAATTACTGCCAC GACACCTTCAAGGGTGCGAAGAG	50	
AU01018B1D04.r1	AUBES1088	200	$(AC)_{19}$	ACTGAACCAGAGCAGAGTCC GACGGTACGTTCAATACTTCTGGCAC	55	P
AU01018B2F12.r1	AUBES1089	490	$(AT)_{16}$	CTTATTTCCCCTACAGTGTGTGTG GACGTAGAACCCATCACCCTTTGG	50	
AU01018B2C07.r1	AUBES1090	350	$(CT)_{13}$	GTGATGAGTCAATGCAACTCAGG GACACAGACGCATGACAGCTTCC	55	P
AU01018B2B09.r1	AUBES1091	430	$(TG)_{18}$	ACGGTCCTACACACTCCAGG GACGTGTGACGAGTGGCTGAAGC	55	P
AU01018B1E06.f1	AUBES1092	470	$(TG)_{11}$	TCATGGTTACAGGCTTGCAG GACCACAGGCTCACCGAAACTGG	55	P
AU01010A1A12.f1	AUBES1093	570	$(TG)_{15}$	AGCCACATAAAAGCCTGTCC GACGGAGCACTAACACAGACACC	55	P
AU01010A1E01.f1	AUBES1094	400	$(TG)_{18}$	AAGCCAACCCCAAAGCCTCG GACATCCCAGAGGAGAATGCTGC	55	
AU01010A1G11.f1	AUBES1095	580	$(TTTA)_5$	ATCACGCACACCCCAAACAC GACTCCTCCCTGCCTGGCATGAG	55	P
AU01010A2C11.r1	AUBES1096	280	$(AC)_{19}$	AATCTGCCACTGCTGTTGAG GACGTCAAGCACATGGCATGACC	55	P
AU01010B1C06.r1	AUBES1097	220	$(TG)_{15}$	CTTCGGTCTTCTCGAAAGTGG GACGACAGTGCAGCGTAGTGGAG	55	P
AU01010B1H11.r1	AUBES1098	660	$(AG)_8$	GGGGTGTGTGTGCGTTTAGG GACCCACCAAAAGTACACGATGCTC	55	
AU01010B2B06.r1	AUBES1099	230	$(GA)_{29}$	AGTTGTTGGTCAGCGCCAGG GACACATCAGCCAGCCCTGTGTG	55	P
AU01010B2H02.r1	AUBES1100	300	$(ATAA)_4$	GCTACATGCTGTGAAGCTCC GACGTTTTACTTTGAGTGTCGTGACTACC	50	
AU01010B2A11.f1	AUBES1101	640	$(AACA)_5$	CCCACTGAATGTGTTGCTCG GACTAATCAGGCCCGGTTGCGTC	50	
AU01012B2E08.f1	AUBES1102	394	$(TG)_9$	ACCTTGATAAGCACGTCAACG GACAATCTCACCGTGGCCTTGAG	55	
AU01010A2A08.r1	AUBES1103	360	$(TC)_{17}$	GGAGTGAGCTGTGTGCCCTG GACTGGCAAAGTAGCACTGTGTC	50	
AU01010B2C09.r1	AUBES1104	363	$(ATAA)_9$	CATTATGGCGGGTCATGTGC GACTTGCTTGGTAATAGTCAGCGTGTG	55	NP
AU01012A2C01.r1	AUBES1105	375	$(AACA)_6$	TGTCAAAGGCACACACAACG GACCCTGCTGGACTGAGGGGCTC	50	P
AU01012A2G03.f1	AUBES1106	690	$(AG)_{17}$	CCGTCAACGACAGCAACAGC	55	NP

AU01012B1D03.r1	AUBES1107	510	$(ATT)_{19}$	GACCCGATACATGCTGGAGCCAC AGGTCATTCTGGGGTCACTC GACTGGCTGAGTATCGGCTATGC	50	P
AU01012B1C08.f1	AUBES1108	80	$(AC)_{29}$	GTCGCTGAGCAGCAGCTCTC GACCACGGTCAAAACCCTCTGTTC	50	
AU01012B2A01.r1	AUBES1109	109	$(TG)_{12}$	AGAGGCATCCCAAGTGGCTG GACGTGGATTTCCTGTAACCGTGG	50	P
AU01012B2D01.r1	AUBES1110	465	$(AGA)_{8}$	AAACTAGCACGCCAAGTAGC GACAGCTTTGTGTACGGTCGCTG	50	
AU01012B2F02.r1	AUBES1111	385	$(AC)_{9}$	AGATAACGGAGAGCAGCCAC GACTCCTCCCGCAACTGCCGCTG	57	P
AU01012B2F08.r1	AUBES1112	161	$(GT)_{16}$	CTTGGCACTTCACCCGCCAG GACACCTGGAGCTGCTCAGCGTG	57	P
AU01012B2G07.r1	AUBES1113	386	$(AT)_{19}$	GCCATGTTCAGGTAACGTGG GACGCAAAGTGTCATTTCTTCAGTGTCG	50	
AU01018A1C03.r1	AUBES1114	358	$(TG)_{20}$	AAACTTACCCTCGGCGTGTC GACCTCAGAGTGTTCCAAAGCCTG	55	
AU01018A1F06.f1	AUBES1115	196	$(ATC)_{10}$	ATATGTGTATGTGGCGTGCAG GACAGGAGCAAATGCTCAAGGTG	50	P
AU01018A2C01.r1	AUBES1116	380	$(AC)_{8}$	TGGTGTGTCCAGGGTTGTGC GACACAAACCGCACTTCACCGAG	57	P
AU01018A2C02.r1	AUBES1117	109	$(ATT)_{15}$	GCCCAAACATACTGGCTACC GACAGGGCAATGAGCGTTTCCTG	50	P
AU01018B1A07.f1	AUBES1118	350	$(GAT)_{6}$	AGTCTCACAGATAGTCCTGGTG GACAGTAAGTCAGTATGTAAGCTCCCAG	50	P
AU01018B1E06.f1	AUBES1119	470	$(TG)_{11}$	TCATGGTTACAGGCTTGCAG GACCACAGGCTCACCGAAACTGG	50	P
AU01018B2E07.f1	AUBES1120	510	$(AC)_{11}$	TAAGTGAGGGAGCCGGAATC GACCCATAACGCTTCCAGAGTGAC	55	P
AU01007A1C11.f1	AUBES1142	350	$(AC)_{17}$	TCCCTAGTGCCTCGTGTGTG GACTGTAGACAGCAGCGAGCCTG		
AU01007A1H05.f1	AUBES1143	200	$(GA)_{15}$	CTTACACACACTAGCTTGCACC GACGTCTCCAGTGTATGTGAGCAC	55	
AU01007A2D08.r1	AUBES1144	280	$(TC)_{9}$	ACATCAGCAAAGGCTTGACAG GACTCACAGCAACTTGCCAAAGAG	55	
AU01007A2D11.f1	AUBES1145	150	$(GA)_{16}$	AGGAGGTCTGATGGTTGTGG GACGCTGATGTCCGATTGCCCAC	55	P
AU01007B1F03.r1	AUBES1146	450	$(GA)_{12}$	AGTGGCTGCTCTGAGGCGTG GACTCAAAACCAAAAGCAGGTCAGAC	55	P
AU01007B1C07.f1	AUBES1147	430	$(TAAA)_{9}$	TCAAGACAGGAGCCAACCTG GACTCCTTTAGCCTGGGCCAGAC	55	P
AU01007B1H06.f1	AUBES1148	200	$(TC)_{13}$	CCCTCTTAAACGTGCGCGTG GACCATACACACACGGACACATCC		
AU01007B2B12.r1	AUBES1149	360	$(TG)_{11}$	GAGGCCCAGTACAACGTACC GACACAGCTACACACCCACAATG	55	
AU01007B2E01.r1	AUBES1150	250	$(TG)_{12}$	ACATGCCCCTCTAGCACCAC GACAGCCATCTGTGTGTGGGGAC	55	
AU01007B2E05.r1	AUBES1151	120	$(TC)_{22}$	GATTGTGAGGTAGGCACTGC GACACAGAGGTGACTCAGGGCTG	55	P
AU01007B2H08.r1	AUBES1152	250	$(TG)_{8}$	GCCTCCATGTTGACGCACAC GACAGAGTCGTTACTGACCGCAC	55	NP
AU01007B2A11.f1	AUBES1153	200	$(TAC)_{9}$	AAGGAGCTGTCCTGTTCACC GACGTTTATGGGTAACCCTGTCAAGG		
AU01007B2H03.f1	AUBES1154	120	$(TTTA)_{5}$	TCTAAGTCTTATACCTGGGGTTGTC GACTTCTGCTCCAGGGGTGCTTC		
AU01008A1B11.r1	AUBES1155	280	$(ATG)_{10}$	CATTGTCTGCCAACCGAAGC GACTCCACAAATGATCCGAAGTGC		
AU01008A1C06.r1	AUBES1156	200	$(CA)_{10}$	CGCGCTTTATTGTGAGCATGAC GACGAAGCCCTTTCTCAGGAACG		
AU01008A1F09.r1	AUBES1157	540	$(TG)_{9}$	TGGAGCCACTGTACTCGTGC GACACACACACTCAGCCTGTTGG	55	
AU01008A1H05.r1	AUBES1158	580	$(TG)_{15}$	TAAGGAGTGACAGGCGGCAC GACGATTACGGCGTGATTGGCAC	55	
AU01008A1H07.r1	AUBES1159	170	$(GA)_{17}$	ACATCATGCACATGCAGAGC GACCAGGTAAACGACGTGGTCTG	55	
AU01008A1A06.f1	AUBES1160	390	$(TTTA)_{4}$	ACTGTGCCATACGTCTCTCTG GACAGAACAGTCATTGCAGGTGTC	55	P
AU01008A1G06.f1	AUBES1161	340	$(AC)_{8}$	CAGGTCAGCAAGGGGGTTCG GACTCTGATAAGAGCAGGGGTGAC	55	
AU01008A1H03.f1	AUBES1162	95	$(CT)_{12}$	TGTTCATGGCTTGCGGTCAG GACGGGTGTATCCCACTGCTCAG		
AU01008A2A10.r1	AUBES1163	450	$(ATGG)_{5}$	TGGGACGATGCACGTTCCTG GACGTACACCCTGAAGGGCTTTG	55	
AU01008A2B06.r1	AUBES1164	170	$(ATT)_{8}$	CTGGGTTTAGGGGTGGAAGC GACTGATGGCCCGTTGTGGTGTG	55	P
AU01008A2B11.r1	AUBES1165	420	$(AAT)_{12}$	AATGGCAGAAGGTTTCCACC GACTCCGAACCTACGAGACAAGC		
AU01008A2E03.r1	AUBES1166	310	$(AC)_{8}$	CTGGGATTACTTGCTCACTGG GACATACAGGGCTCGCTTCACAG	55	
AU01008A2E08.r1	AUBES1167	400	$(AAAT)_{5}$	GGGCATTAACATTTGACCGAGG GACGCCTCAACAATTTGGTGTGG	55	
AU01008A2F09.r1	AUBES1168	180	$(TG)_{19}$	CAGAGCCAAAGTTCCCTCTG GACTTTCTTTTGAGAGTCCAGTGTGC	55	P
AU01008A2G03.r1	AUBES1169	240	$(TCCA)_{5}$	TGCCTTGTACCCGATGCTCC GACACCTCACACCTCGGAGACAG	55	
AU01008A2G06.r1	AUBES1170	150	$(TATT)_{5}$	CCTATAAGAAGTTCACTACAGCCTGTC GACCTGGAAGCAATGTGTTGTGC	50	
AU01008A2H03.r1	AUBES1171	250	$(AC)_{8}$	GTCAAAGATACAGTGGAACTGAGC GACACATAATCTAGGTTCGCCTCTGG	50	
AU01008A2D03.f1	AUBES1172	380	$(TC)_{10}$	AGTGTTTCCTTGGCGGTGTC GACTCCCTGTCCACCCTCCATCC	55	

AU01008A2H08.f1	AUBES1173	210	$(CA)_{12}$	GATTGTGAAGCAGTGGCAGC GACAGCGGGAGTAACGTGTTGTC	55	
AU01008B1B03.r1	AUBES1174	55	$(ACAA)_{13}$	CGTGTTTGTTTTCGTGCAATGC GACATGTGGGCTAACTGCACTGG	55	
AU01008B1G01.r1	AUBES1175	550	$(GA)_{12}$	TAAGGGGTGTGTGTGCGTTC GACAACTCCTGGAGGGCATCCTC	55	
AU01008B1A06.f1	AUBES1176	370	$(TG)_{18}$	AACACACTCACAGAAGCCAC GACAATAGAGGGGGCATGACCTC	55	
AU01011A1A02.f1	AUBES1177	383	$(AC)_{13}$	GGACATCCTTGAAGGACGTG GACTGGTTACGCTGCTCTGATGC	55	P
AU01011A1B12.f1	AUBES1178	71	$(CA)_{14}$	TCCTGGTCAGGGTTGTGGTG GACTCCGGCCAGCGCCATAGACG	55	P
AU01011A2C01.r1	AUBES1179	218	$(AC)_{17}$	ACGTGCATTAGTGGCCCCTG GACACTCCATTGGTTTGAGGCAC	55	P
AU01011A2F02.r1	AUBES1180	362	$(TTTG)_{7}$	GGACACATATCATGGCTTGTGC GACACCATGTTGAACTGCTCTTCAC	55	
AU01011A2F09.f1	AUBES1181	265	$(AC)_{20}$	ACTGAAAGCATCTGTTGCACG GACGACCTTGATAAGGCGACACTC	55	
AU01011B1A10.r1	AUBES1182	175	$(TG)_{10}$	AGGCACGCAAAGGCCCAACC GACTTCGGTACGGACGCAGTGTC	55	P
AU01011B1C04.r1	AUBES1183	222	$(TG)_{9}$	TGACGTATGGTTGGCAACAC GACTTATTTGGGGAACGGCAGTG	55	
AU01011B1E02.r1	AUBES1184	262	$(TG)_{10}$	CACTACACACTTGACCATCGAAC GACCTCCAGAACTGAAGCAGCAC	55	NP
AU01011B1H04.r1	AUBES1185	313	$(CA)_{27}$	TCATGCCATTTAGCGGCCTG GACTGTTTACCGGCTGTGCGAGC	55	P
AU01011B2C12.f1	AUBES1186	377	$(AC)_{9}$	AGACCCCAGCCAAGCTGTCC GACGATTAGGGCACTGAAGTCACG	55	
AU01021A1C09.r1	AUBES1187	62	$(TG)_{10}$	TCCCACACAGCAGCCTCCAC GACTGCAAGCTATCGCACACCAC	55	
AU01021A2F02.r1	AUBES1188	479	$(AC)_{33}$	CGAGTCCTTTAGAGGCCCAG GACCTCTCACTGTCCTCCACTGC	55	NP
AU01021A2C05.f1	AUBES1189	70	$(AAT)_{6}$	GTGTATTTAATCCTGGGGTGAGTTG GACCGAAGTCCATAGGGGTTTCC	55	
AU01021A2F03.f1	AUBES1190	145	$(CA)_{9}$	AGCAAAGTCATGCCTGCATG GACGACAAACGTGTGCAAGTTGG	55	
AU01021A2F07.f1	AUBES1191	438	$(TC)_{20}$	ACTCCAGCAAAACATGCAGC GACAATTCACTCACACAGCTCACAC	55	
AU01007A2E08.f1	AUBES1225	100	$(AC)_{15}$	ATTGTGATAGTGTGCTGTGGTG GACTCCTGAGCAAACTGAGCCTG	50	
AU01008B1D04.f1	AUBES1226	400	$(GA)_{29}$	AGGGAGAGCGAGTCGCTGTG GACACCTGCTGCTGAGGCTGACC	55	P
AU01028B1F11.f1	AUBES1227	170	$(CA)_{8}$	GCTCCCAATAGTAAGCCTAACTGG GACAGGGCACAGGCAAGCCACAC	55	P
AU01023A1G09.f1	AUBES1228	170	$(AAAT)_{6}$	ACAGCCTGAAGACCCGTGAG GACACTCAGAACCTACAGCCAACTC	50	
AU01023B1G08.r1	AUBES1229	120	$(AG)_{19}$	AACCCATCAGACACGCTCAC GACCTGGAGACAGCGCGAGGGAG	50	
AU01022B2E01.f1	AUBES1230	215	$(TG)_{21}$	CCAACAGTCATCTGTCTGAGC GACTGTGCTCCCGTGGACCTCAG	55	P
AU01025B1A10.f1	AUBES1231	135	$(AC)_{21}$	GTCGCATCATTTTGATTGCAGC GACGGACGGGCTCCTGTTGGACG	57	P
AU01018B1F11.r1	AUBES1232	185	$(TG)_{13}$	TTGCCAAGAACCCGTTGAGC GACTGGGAAGCATTTGGTTTGGTC	50	
AU01022B2B07.f1	AUBES1233	135	$(AC)_{16}$	ATGGCTATGGGACTAGGTGC GACACAAGCACATACACACGAGC	55	
AU01032B1E03.r1	AUBES1234	225	$(TG)_{18}$	GGTCAGACATATTCCTCCAAAGC GACTCGATTATCGGTATCGGCTGAG	50	
AU01036A1G04.f1	AUBES1243	70	$(AT)_{10}$	CGGCAAGTCGGGCGAGTTTC GACAGCATAAACAAGTAGCAGACAGC	55	P
AU01036B1G11.r1	AUBES1244	220	$(AC)_{11}$	TCCTCCTCAGCAGGGGTGAG GACCAGAGGCAATGATGTGGTCC	55	P
AU01014B2F05.r1	AUBES1258	330	$(AC)_{28}$	AGAAAGCAGCTTGCAGATGC GACGCACAAAAGTTCAGGCCATG	50	
AU01001B1F04.r1	AUBES1259	125	$(GA)_{15}$	TTTGTGTGGTGCTATGCTGC GACGGGGTTATCAGAGTGGTCCTG	55	
AU01010A2A04.f1	AUBES1260	70	$(TAA)_{18}$	CATGGGAGTGTGTGCATGTG GACGTCTGCCTCTGATGGAGTCG	50	
AU01001B2A07.r1	AUBES1261	258	$(AC)_{21}$	AGAAAGCAGCTTGCAGATGC GACGCACAAAAGTTCAGGCCATG	50	
AU01008B2F01.f1	AUBES1262	460	$(TA)_{8}$	AGTCCTAACTGATATTCAGTCCAGG GACCGGCTCGAAGGCCAACATCC	50	
AU01033A2D02.f1	AUBES1263	535	$(GT)_{14}$	CTGACCTGAGCACTGGTGTG GACAGACGTATGGGCGAGATCAG	57	P
AU01024B2C11.f1	AUBES1264	450	$(ATG)_{10}$	AAGGCAGCAAGGTGAGAAGC GACCGGCTCAGGTCGCTCATACG	57	P
AU01033A2D11.f1	AUBES1265	75	$(GAC)_{6}$	ACACACGGAGGGGTCAGAGG GACTGTACGTTGGCTCGTTGACC	57	P
AU01030B2E04.f1	AUBES1266	330	$(TGA)_{15}$	TAACAGCCGCTCCTGCATCG GACGCTATTATGTGTCTACTGCCAATCC	50	
AU01001B2H05.r1	AUBES1267	479	$(TG)_{9}$	CTCCGCCGTTCAGTGCGAGG GACTGCCAACATAAAGAGAAGCTGG	55	P
AU01011B2H06.r1	AUBES1268	480	$(AGC)_{6}$	ACTCCGCCGTTCAGTGCGAG GACTGCCAACATAAAGAGAAGCTGG	50	P
AU01010B2G07.f1	AUBES1269	500	$(AT)_{10}$	ATGGTGATGGAGCCTGAGAC GACAGTCAAAGGCTTTTCGTCGTC	50	
AU01030B1D09.r1	AUBES1270	225	$(GATG)_{5}$	CAGCCACTCCTCGGTGTGTG GACCAGGTCGCTTGTGGGATGAG	50	P
AU01034B1D08.f1	AUBES1271	210	$(AC)_{28}$	GTTTCGCCTTTGACTGCGTC GACTGCCCTGTGGAGGTTGTTGG	55	
AU01025A2E09.f1	AUBES1272	295	$(AAT)_{12}$	TCCTCATCTGGGTGACAACG	55	

AU01028B1C09.r1	AUBES1273	135	$(CT)_8$	GACTAGGGTCCACAGCCAGCCTG TGTGAGAACAGTGACCGTGC	57	
AU01007B1B11.f1	AUBES1274	183	$(CA)_{30}$	GACTCGTAGATGGAGGCTCCGAG GAAAGAAGCAGAACGAGAGAGG GACGTGGTGAGTTTCACTGACTCC	55	P
AU01001B1D04.r1	AUBES1275	90	$(AAT)_7$	TCGTAGTGATCCACCACCTG GACCTCTAACTAACCAACCCACAGC	50	
AU01008B1A06.f1	AUBES1276	412	$(AC)_{14}$	AACACACTCACAGAAGCCAC GACTGAAAGAGGCATGGGCTGTG	55	
AU01014A1E03.r1	AUBES1277	133	$(AATA)_5$	GAATTTACGGGAGGCCGCTG GACACGGAGACTCCATGCTGAGC	57	P
AU01028B1B05.f1	AUBES1278	265	$(GT)_8$	TCAACCCTGCAACGTCACAC GACTGTCGTTCAGAAAGCCGCTG	55	
AU01022B1G11.f1	AUBES1279	305	$(TCTA)_6$	TCGTGAGCAGAACCCTGGAC GACCATGCAGCACATGGCCGTGG	55	
AU01033B1A02.f1	AUBES1280	485	$(AAT)_{11}$	CTTGCCACTCCACCCGTTAG GACGCGTTTGAGTTCTCCAAACACC	50	NP
AU01019A2C04.r1	AUBES1281	515	$(AAT)_{12}$	ACCATAATCTTGCCACAAGCAG GACTGGAGAGAGTCAACACGCAG	50	P
AU01002A1H01.f1	AUBES1282	115	$(GGA)_7$	CTGTGACTCCCTCAGACTGC GACTCATTCTGCCACCTGCTGAC	50	P
AU01006B1B04.f1	AUBES1283	314	$(AC)_8$	TTGGAGGACTGTTGGGTCAG GACACAAAACGCAGTGGGGTCTG	50	P
AU01011A1D09.f1	AUBES1284	116	$(TTTA)_6$	CTGAGGTCTTCAACTGCCAG GACTCAGAACCAAACAGATGCAGC	55	
AU01023A2H05.r1	AUBES1285	175	$(AT)_8$	GTTCAGTGAGATGTCGCCTG GACCCTATCCCGTACAGGTACACG	50	P
AU01033B1D05.r1	AUBES1286	210	$(AG)_{11}$	GAATGGTATCCTCGCCAAGC GACGGACATTCTTCAGCAGCCAG	50	P
AU01028A1C10.f1	AUBES1287	270	$(TC)_{19}$	TTGGGTCAGTTGACGCCTGG GACGGGCGGATCAGAACCGTCTC	50	P
AU01020B1A04.f1	AUBES1288	72	$(TAAC)_5$	ACTAGAACTCTGGCCTAGCAC GACCTTTCAGTGAGGGCGTTTCC	50	
AU01019A2A03.r1	AUBES1289	495	$(TG)_{18}$	AGCTGCAAAACCTACAGCAG GACTAACGGCACATCCCCAAGTG	50	
AU01022B2A03.f1	AUBES1290	404	$(CCA)_7$	GTGTAAATGTCCAAACATGCACG GACACAAGTGGGAAACACTGTGG	50	
AU01026B1E04.r1	AUBES1291	319	$(GA)_9$	CCCACAGCAGCCAGATGGAC GACTGGCAGGGAAACCCAAGCAC	55	
AU01029A2E09.f1	AUBES1292	298	$(GA)_{13}$	CACCCAGAACTTGCACACAC GACTGCTGGTAGGACTGGTTTGG	55	
AU01002B1A01.f1	AUBES1293	422	$(TA)_{11}$	TGGTCATGTGACTGTGATGC GACCTGAAGCAGAGTAGCCAACAC	50	NP
AU01011B2D09.r1	AUBES1294	175	$(AC)_9$	CTGTTGCACCACTAGGGAGC GACTCGCAACGACTGCACTGGAC	55	
AU01003A1B02.r1	AUBES1295	165	$(GAA)_9$	GTCCGTCTACGAACAGGGTG GACTGTGGGTTCTCTCGCATCAG	55	NP
AU01025A1D09.f1	AUBES1296	335	$(ATAG)_4$	CAGACAGTGAACAGGAACTTGC GACACAACAGCACAGTCTGCCAG	55	
AU01005B2H04.r1	AUBES1297	412	$(TC)_{11}$	TCTTCCGTCTCTCCGCACAG GACCTGCGAGGACTACGGCTCTG	55	
AU01021A2F07.f1	AUBES1298	285	$(TG)_{19}$	ACTCCAGCAAAACATGCAGC GACAATTCACTCACACAGCTCACAC	50	P
AU01021B2E09.f1	AUBES1299	214	$(TC)_{11}$	TCCAAAACCATCCGCGACAG GACAGCAGTTTAGACATGGCTGC	55	P
AU01021B2H07.f1	AUBES1300	280	$(AAAC)_5$	CTATGTCAAGACTATGGCGATGTTG GACCTATGTTCAGATGACCCATGAAGG	55	P
AU01006A2B03.f1	AUBES1301	435	$(TA)_{35}$	GCAGTCATCCAGAGTCCCAG GACTCCTTCACTGCTGTCTGAGC	55	P
AU01010A2G11.r1	AUBES1302	330	$(AC)_6$	TCTACTGCTGCCTGTGAACG GACGCTGAAACAGACTGTGGACAC	55	
AU01018A1E03.r1	AUBES1303	490	$(TG)_{31}$	GAACAAGTGCTGCACGGAGC GACAGTTCCCTGGCTCGGTTTGG	55	
AU01007A2H08.f1	AUBES1304	290	$(TC)_9$	ACACAACATGGACAACGAGTC GACTCAGCTCCGTCACCTCCGAG	55	
AU01003A1D11.f1	AUBES1305	430	$(CA)_{16}$	AACACAGCCTGCCTCTCATC GACGTAGGCGGTCTGCCTCTCAG	50	P
AU01019B2H05.r1	AUBES1306	130	$(AC)_{16}$	TCAGCTCCAGCGCAACGAGG GACCACTCCCGGAATACACGCAG	55	P
AU01021A2F02.r1	AUBES1307	440	$(GA)_8$	CACAACATGGACAACGAGTCC GACGTCCTCCACTGCTGTCTGTG	50	NP
AU01028A1E08.r1	AUBES1319	185	$(CA)_{12}$	AGAGAAACCCTAAACTCACAGTCC GACGTGTTTTCTGAGCCAGGAGG	50	P
AU01028A1C09.f1	AUBES1320	210	$(TG)_{16}$	CACACTGACGTACATACGTGC GACGTTTGGTCAGGGCTGATGTC	55	P
AU01009B1D04.r1	AUBES1321	220	$(TG)_{13}$	AGGTAACAGCATCCACCCAG GACTCGGAACCTTCGAGTTCACG	50	P
AU01023A1C05.r1	AUBES1322	365	$(TC)_{20}$	CTCAGAGTACACACAACCATCG GACGGGAGATACGTTCAGTCCGAC	50	P
AU01023B2F01.f1	AUBES1323	413	$(TG)_{22}$	CCAGTGTTTTCAGCCATGTGAG GACATGACCCTCTGCGGAACCAG	55	P
AU01031A2H01.r1	AUBES1324	306	$(CA)_{29}$	AAGTCTCTGGTGTCGTGTGC GACGGTGAGTTCCTAATGCTCGTG	55	
AU01027A2B01.f1	AUBES1325	110	$(AC)_{30}$	AAGAATCACAGCCCAGATGC GACGTTTAGTCGCATACTGTTGCTCAC	50	P
AU01004A2E04.f1	AUBES1326	430	$(AC)_{30}$	TTAGGATCGGTTAGCCGCAG GACTCAAGGCCATTTCTGCACTG	50	P
AU01031A1C08.f1	AUBES1327	373	$(TG)_{19}$	CTCGCATCAGCACCATCTCG GACAGGGTCCTCCAACAGGCTCG	55	
AU01022A1H01.f1	AUBES1328	229	$(AC)_{23}$	AGGTACTCCGACCTCCACAG GACACACAGAACGAGACAGGTGAG	50	P

AU01014A2F08.r1	AUBES1329	75	$(TG)_{19}$	TTGGCAGGACAGGTCAGTCC GACAACTGTAGCGCACGCTGTTG	50	
AU01033A2G04.f1	AUBES1330	300	$(AC)_{10}$	CTTACCCATGTATCATTGGGACC GACAGCCTATCCCAGGGGACTTG	50	P
AU01004A2H08.r1	AUBES1331	480	$(AC)_{30}$	TGCACAGTTAAATCATCACCAGC GACAGCATCAGCATGGACCTCAG	50	P
AU01022A1D08.f1	AUBES1332	236	$(GA)_{19}$	CAGATACAAATAGGAGGCACACG GACGTAAAGTGTCCCTGAGCTGTG	55	P
AU01014B1H11.r1	AUBES1333	111	$(CA)_{12}$	GGATTGTACGTTCTGCTTGACG GACTTGTGAGGGTCCTCATGCTC	55	P
AU01002B2G03.f1	AUBES1334	103	$(CA)_{13}$	GTTCACGAGGACGTGGGTCG GACCCCAAGTGCCTGAGGCGAGC	55	P
AU01014B2G09.r1	AUBES1335	203	$(TG)_{11}$	GTGTTGTACGTTTGGGCTAGAG GACAGGTCCTGACTATGTGCTGAG	55	
AU01030B1A07.r1	AUBES1336	397	$(TAA)_{11}$	AGTCTCCCAGCACCGGCACG GACCACGCGCTCGAACCCAGAGC	55	
AU01031A1H06.f1	AUBES1337	140	$(AC)_{8}$	TGAAGGGTGAATGAATGGAGC GACTGTATGTGGGACCACTGTCC	55	
AU01025A1E11.r1	AUBES1338	295	$(AC)_{10}$	GACCACATGACGGCTTCC GACTTTCCAAGCCCGCTGAG	55	P
AU01029A1C02.f1	AUBES1339	278	$(CT)_{23}$	TCTCCACATCTGACACCTGAC GACAGTGGTAGATCACCTTGGGTC	50	
AU01028B1B06.f1	AUBES1340	205	$(GA)_{21}$	GTAGGCAGCGTTCCAGC GACTGCTGGTAGGCCCAGTG	55	
AU01019A2B04.r1	AUBES1341	295	$(TG)_{9}$	GATTTCACATTTTCTGGCAGTGC GACCAGTAAAACACAGATGGTGTCTGAC	50	
AU01006A2H12.f1	AUBES1342	140	$(TAA)_{16}$	AGGGCATCCACAGCTTCAGG GACTGAGAGCCCAGGTGTCTGTC	55	
AU01006B2C09.r1	AUBES1343	372	$(TCTT)_{6}$	AACATGGCCCTGTGGTCAGG GACCAGCATAGGCGTCTGGCAGG	55	
AU01010B2D02.f1	AUBES1344	206	$(GA)_{8}$	CATGACTTTTGCAGGTCCTCC GACGTGGTCATCTGTGGTGTCTG	50	NP
AU01002B1E01.f1	AUBES1345	365	$(TTA)_{11}$	TGGATCAATCTCAATCAGGTCAGG GACGCTCATTCACAGGGACTTCAC	50	P
AU01011B1H04.r1	AUBES1346	291	$(GA)_{14}$	TGCCATTTAGCGGCCTG GACTTTACCGGCTGTGCGAG	50	
AU01011A2F09.f1	AUBES1347	283	$(GAA)_{6}$	AGCATCTGTTGCACGCTG GACAGCAAGACTGGGGTGC	50	
AU01006A1C09.r1	AUBES1348	119	$(TG)_{9}$	GCGCAAGTTTAAGGATGTGTGC GACCCCATTCACCTCGACATGG	50	
AU01013A1B03.f1	AUBES1349	308	$(AC)_{13}$	TGTATTGGCACCCCTTTCC GACGTCCTTCCCCGCTCTG	50	
AU01014B1E02.r1	AUBES1350	205	$(TTTA)_{9}$	AGCGTGGTTCACACTGC GACCACAGGAGTTCAGCATCAGC	55	
AU01028B2A03.f1	AUBES1351	200	$(ATT)_{8}$	GCATTCCACATGCTCACCTC GACACTGATCTCGTCCACGGTG	57	P
AU01031B2B07.f1	AUBES1352	335	$(TG)_{15}$	AAGCTGAACGTCGTTCCAC GACGCAGTCCAGATTGTGTGACG	50	
AU01021B2H09.f1	AUBES1353	235	$(TG)_{9}$	TCACCAGCTTGCTCTGAGAC GACGTCCTCCTGAAGTCCAGACG	50	P
AU01032A2H07.r1	AUBES1354	385	$(CA)_{30}$	ACAACGCTCAGTTGCTGGAC GACACACTGAAGCGGAACGATGG	55	P
AU01025B1H02.f1	AUBES1355	175	$(TG)_{22}$	CACCGCAGTCGGAATCCTGG GACCACAGACACGGAGACGCCTG	55	
AU01032B1F09.f1	AUBES1356	142	$(CA)_{21}$	GCGCAAGTTTAAGGATGTGTGC GACCCCCATTCACCTCGACATGG	50	
AU01020B1H06.r1	AUBES1357	330	$(GA)_{19}$	GTATTCTTCACTTAGGGCAAGGTC GACGTATGCTGCTGATGCTCAGG	50	
AU01010B2B03.f1	AUBES1358	478	$(TG)_{11}$	TGACTGTGTTTGCCCAGGTG GACCAGTGGAATGTCCTCACAAGG	55	
AU01011B2C12.f1	AUBES1359	420	$(AC)_{27}$	ACACTTCTCAGGCTCTCCAG GACGATTAGGGCACTGAAGTCACG	55	
AU01011A2C01.r1	AUBES1360	340	$(AC)_{19}$	CCCCTGAACGGAGCAG GACACTCCATTGGTTTGAGGCAC	55	
AU01024A2D06.f1	AUBES1361	284	$(TTA)_{8}$	GTTTGAGTCGGCAGCACTAC GACTATTCTGGCTGGAGGCTACG	55	
AU01034A1A01.f1	AUBES1362	323	$(CA)_{8}$	TCTGCACCTTCACGCAG GACCACTGGGCACAGAGCAC	55	
AU01034B1B02.r1	AUBES1363	304	$(TG)_{24}$	TGGACGGAATGGTCTGGAG GACTTTTCTCCTGCCGGTGG	55	
AU01034B1F10.r1	AUBES1364	190	$(GA)_{11}$	GTCAACGCCGAGGTCAC GACGGAAGTCTAAGGCTGTGTCAC	50	
AU01032A2F10.r1	AUBES1365	322	$(TG)_{10}$	GCTTGTTCCATCAATAGCCAGC GACTCTCAAAATGGTGCTGGAAGTG	50	
AU01018A2C02.r1	AUBES1366	380	$(AC)_{8}$	TCAAAGCAGCAGCCTTCTC GACACTCTCCAGGGGTGACG	55	
AU01021B2F01.r1	AUBES1367	172	$(TA)_{25}$	AAACTGCTACTGCACTGCTC GACCCTCGTCAATGCTGTAGTCC	55	
AU01030B2G12.r1	AUBES1368	114	$(GA)_{9}$	TGTTGAATGAGGCTGTCGTG GACGAAATCCAGTCAGGGTCGTC	55	
AU01026B2E02.r1	AUBES1369	188	$(GA)_{8}$	CAGAGAAATCTGTCTTTGTGCTCC GACTGAACGCACACGCTGAC	55	
AU01021A1C09.r1	AUBES1370	83	$(TAAT)_{5}$	ACAGCAGCCTCCACAAC GACGGCCAGGTCCGTGTAC	55	
AU01029A1H05.f1	AUBES1371	367	$(TG)_{15}$	CACAGAGGTTCTGCCATTACG GACCATCTCTGCCTCCAAAGCTG	50	
AU01022B1H04.f1	AUBES1372	283	$(CA)_{9}$	TGCACACATACATGCTGCTG GACCCTGTGGTTTGGTGGGTG	55	
AU01021A1H07.f1	AUBES1373	303	$(CA)_{16}$	TGAGCGTGTGCCAGTG GACCATCGGGCAGGTCCTG	55	
AU01022A1F10.r1	AUBES1374	51	$(CA)_{26}$	CTCCCGCACAACAGACG	55	

AU01019B1B07.r1	AUBES1375	116	$(CA)_{16}$	GACGAGCTGAGCTGAGCTGC CATTCCCTGCGAGTCTGC	55	
AU01033B1A08.r1	AUBES1376	121	$(TG)_{16}$	GACGTCCACCCGCACAGAC TTCTTCAACTGTGTGGATGAGC	55	
AU01025B2F07.r1	AUBES1377	142	$(CA)_{19}$	GACCAGTCCTGAACTCACACTGG CACATACACACCTCCATGAGC	55	
AU01034A1G06.r1	AUBES1378	205	$(GGA)_{7}$	GACCAGTTACAAGGGGTTTCCCAG CAGTCCTGACTTGCCAGG	55	
AU01020B2G10.f1	AUBES1379	254	$(TAA)_{12}$	GACCCAGTGGACAAAGCCTGC GGAAACAGACTCCACACTGAG	55	
AU01023A2D07.f1	AUBES1380	80	$(CA)_{30}$	GACGTCTGGTGAACGGGTTTGAG CTCAGAGACCAGCAACACTG	50	
AU01023B1F03.r1	AUBES1381	205	$(CA)_{13}$	GACCCTTAATCCACACTGCATGGAG ACCTGAATTAGGAGGCTACCTG	55	
AU01018A1B08.r1	AUBES1389	423	$(AACA)_{10}$	GACAGTTCAGTTCCGCAAGCTG TCATGGCTCCAAGGTTGC	55	
AU01018A1E03.r1	AUBES1390	492	$(TG)_{19}$	GACTGCCATCTTGCCATTCCTG CAAGTGCTGCACGGAGC	55	
AU01018A1E05.r1	AUBES1391	495	$(TG)_{17}$	GACGGCTCGGTTTGGCTCTG TCATGCCAGTGCATCACAG	55	
AU01018A1E08.r1	AUBES1392	187	$(GA)_{19}$	GACCAGGTTCTCCGGTTTCCTC AGAGAAACCCTAAACTCACAGTCC	55	
AU01018A1G08.r1	AUBES1393	400	$(TGTT)_{6}$	GACACCGTGTTTTCTGAGCCAG GAATGCCTGACTCTGGGAG	55	
AU01018A1C04.f1	AUBES1394	310	$(TTTC)_{8}$	GACTGGTGCTCGGGAGTCTC TGGGTTGTAGAGGTATCCTGC	55	
AU01018A1D06.f1	AUBES1395	251	$(AC)_{20}$	GACGGGCATAATGCTTTTGCAGC CTTCCTAATCTCTTGTGGACAGC	55	
AU01018A1D08.f1	AUBES1396	139	$(AACA)_{5}$	GACGTTCTGGGGTCGCCATTAC TGTAACCCTAGCCAGCTACAG	55	P
AU01018A2B05.r1	AUBES1397	155	$(TG)_{14}$	GACGCTGCTACTCTCGCATGAC TCCCTCTGACTACACCAGC	55	
AU01018A2B06.r1	AUBES1398	452	$(AC)_{14}$	GACGACATATTGGGCACCCCTG TTGTGAAGCTGGTGGACG	55	
AU01018A2C01.r1	AUBES1399	160	$(CA)_{8}$	GACACAGAAAGCGTTCAGCAGC TGTGTCCAGGGTTGTGC	55	
AU01018A2C01.f1	AUBES1401	470	$(TTAT)_{25}$	GACAACCGCACTTCACCGAG GCTAGTTCTCCAGGATGCAAC	55	
AU01020B1A04.f1	AUBES1402	73	$(AT)_{35}$	GACGTTAGCGACGACAGTGTTGG ACTCTGGCCTAGCACCATGT	55	
AU01021A1B12.f1	AUBES1403	241	$(GA)_{11}$	GACACCTTTCAGTGAGGGCGTTT CCCGAGGGTAAAAATATGGA	60	
AU01021A1C04.f1	AUBES1404	209	$(TC)_{20}$	GACCAGCCTGTATATTCCATGCAGA TTTGTTAAATGGCCGCTAGG	55	
AU01021A2F07.f1	AUBES1405	290	$(TC)_{20}$	GACTGCTCAACTCCTGGTACTGC CAAAACATGCAGCCTGAGAG	55	
AU01021B1D01.f1	AUBES1406	178	$(CTAA)_{14}$	GACTCACACAGCTGCTGATCAAA TTCGTTAGTTAGTTCGTTCGTTC	60	P
AU01022A2B12.f1	AUBES1407	160	$(AC)_{23}$	GACCCCCCAAGAACTTGAGGTAA TCAGGTTTGCACATCACTCTG	55	P
AU01022B2D01.r1	AUBES1408	30	$(GT)_{12}$	GACCAACCCCTCATTCACGAGAT CCCATGTGGGTTATTTCCAC	55	
AU01023B1B07.r1	AUBES1409	348	$(AG)_{17}$	GACGATGTGTTGTGAACGATGTCA TTTTTACACGCCTTCCCAAG	60	NP
AU01023B2G08.f1	AUBES1410	468	$(AC)_{19}$	GACGAAGTGCTTTGGATGGAACC GGGAATCGTTACGTGCTGTT	55	
AU01024B1H05.r1	AUBES1411	314	$(TG)_{14}$	GACAGACCAGATGCATAGGTGAGC ACGCTGTTAGGGGGTTGAT	60	P
AU01024B2A09.r1	AUBES1412	108	$(ATT)_{11}$	GACGTGATGAGGAAAGGGACAGC TGCCACCAATAACAGACAACA	55	
AU01026A1C02.f1	AUBES1413	563	$(ATGG)_{8}$	GACGTCCTAAGGCCGGGAAATAG CGTGGATACTGCTCTGCGTA	60	
AU01026A2C11.f1	AUBES1414	370	$(ATGG)_{7}$	GACCCCACGACCCTGTAGGATAA GCGTCTCTTTGCTTTTCTCG	55	
AU01026B1A11.f1	AUBES1415	319	$(AT)_{29}$	GACCTGGGATAGGTTCCATGCTC GAATGCCCTAAGTGGTCATGT	55	
AU01027A1B06.r1	AUBES1416	368	$(GT)_{15}$	GACTGGCCATACTTATACTCTTACTCAAG GGTGCCATCACATGTCTCAC	60	NP
AU01030B1A06.r1	AUBES1417	319	$(TAT)_{10}$	GACTCCTTCATGGTGGAAACAAA TTGTGTTGGCACACAGATCA	60	P
AU01032A2G10.r1	AUBES1418	239	$(TG)_{12}$	GACTCTCCGCTAAGAGCTCACTTG TGTCAACATGTTAAGCACACTAGC	55	
AU01018B1A07.f1	AUBES1419	197	$(ATG)_{6}$	GACTGATGGGGAAGCTGAGAGTT GCCACATTTCAATTTGGGCTC	53	
AU01018B2B09.r1	AUBES1420	315	$(TG)_{17}$	GACGTCAGTATGTAAGCTCCCAGC ACGGTCCTACACACTCCAG	55	
AU01018B2C07.r1	AUBES1421	220	$(TC)_{14}$	GACGTGACGAGTGGCTGAAGC GTGATGAGTCAATGCAACTCAGG	55	
AU01018B2E07.f1	AUBES1422	414	$(CA)_{11}$	GACACAGACGCATGACAGCTTC GACTGCTGAATTTAAGTGAGGGAG	55	
AU01019A1E10.f1	AUBES1423	409	$(TG)_{12}$	GACCCATAACGCTTCCAGAGTGAC GCAGCTCTGCACGCAC	55	
AU01019A1G12.f1	AUBES1424	173	$(AC)_{25}$	GACACTGTGCAAAACAGCCTCG TGTCCGTAGATACTGGTGGAC	55	
AU01019A2A03.r1	AUBES1425	496	$(AC)_{9}$	GACGGAAGCTGTCACAAGGTTGC TCGCTGTGCAGCAAGTC	55	P
AU01027B2C09.f1	AUBES1426	393	$(AATA)_{5}$	GACAAACATGATACCACACTGTCTGC TTGGCTTCATAACAATTCCAAA	55	
AU01029B1F07.f1	AUBES1427	316	$(TAA)_{8}$	GACCCTAACCAGCTTCCCACAAA CGCTCCTACTGTGGTGATTG	55	
				GACTTGGTTTATCAGCGGGACAT	55	NP

AU01030A1F10.f1	AUBES1428	369	$(GT)_{13}$	TGATGGAGGAGTGAATGCAA GACGTCTTTCGCACCCTGTCTGT		
AU01032A1D01.f1	AUBES1429	459	$(GT)_{10}$	CATCACAGGCCACGACTG GACTCCAAAGACATGCGCTGTAG		
AU01032A2C10.r1	AUBES1430	332	$(CA)_{22}$	TCAAGCACTGGTAAAGAACTGG GACGGTTCTGTGTGCCTGTCAGA	57	P
AU01032B1F12.r1	AUBES1431	394	$(CT)_{10}$	AAGGCCTCTTACCTAAACAGCA GACTTGAAACTGGACAGCACTGG	55	P
AU01032B2A11.r1	AUBES1432	476	$(GA)_{28}$	CATTCATTCATTCAGTCGTTCG GACTCACCTAAAAGAATCGGCTCA	55	
AU01019A2A11.f1	AUBES1433	445	$(GTT)_{6}$	TGAAGGTGCATTTGCATTGT GACCAGCAGGGCACATTTTCAG	55	P
AU01020A1F12.r1	AUBES1434	96	$(AC)_{9}$	GAACATCACATCAAGTGGAGGA GACCTGGGTCTCCTTCAGCATCT	55	P
AU01020A2C05.f1	AUBES1435	450	$(CA)_{17}$	GCTGATGCCATGCTAGTGTT GACGTGACATGGCTCTGCTAGGC	55	P
AU01020B1C12.f1	AUBES1436	145	$(TTA)_{14}$	CCCCTAGGGAACCTGAACAT GACGAACCACTGTCTTGCAATGTG	55	P
AU01020B2B03.f1	AUBES1437	455	$(TATT)_{6}$	TTACGTTCCATGACAGTGACG GACATATGAAGAGGCCCGTGAGA	55	NP
AU01020B2G10.f1	AUBES1438	255	$(TAA)_{12}$	TGGAAACAGACTCCACACTGA GACGGCTCAGACTCTCCTGTCAGA	55	NP
AU01021A1E06.f1	AUBES1439	180	$(AC)_{20}$	GGCTCCCGAGGTTTAGAGAT GACGAAAGGGCCTCGTCTTGAG	55	NP
AU01021A2G02.r1	AUBES1440	450	$(AC)_{14}$	TGCTGGACTCAACTCACAAA GACCCTTGCCAAGTGTGTGTAGAA	55	
AU01021B2C07.f1	AUBES1441	340	$(TA)_{13}$	TGGTACAGAGAGAAGGGGACA GACCATCCAGACCTCTGAGGACATA	55	
AU01022B1G08.f1	AUBES1442	187	$(TA)_{29}$	TGGCTATAGTCAGGGGTAAGAGA GACTCGAGTAAATGATGTTAGCTGAGG	55	
AU01022B2E01.f1	AUBES1443	217	$(GT)_{22}$	CTGTCTGAGCTGGAATTGGA GACGACCAGAATGCCTGCAGATTA	53	
AU01023A1E08.r1	AUBES1444	415	$(TG)_{14}$	CCTACGCACTTGACACCTTG GACACATGGGCACGTGTGATGTA	53	P
AU01025A1C11.r1	AUBES1445	241	$(AC)_{23}$	AAGAGGCAACACGGAGTGAT GACCAGGACCAAGCTTCACTGAG	55	P
AU01025A1E02.r1	AUBES1446	96	$(TG)_{8}$	GAGAAACCAGGCTTCAGCTC GACCGTCCTCAGACGGTTTCAGA	57	
AU01025A2D08.r1	AUBES1447	359	$(AAC)_{8}$	ATACCAGCGTTTCCCAAATG GACTGTGGGATCTTGATTGTTGG	55	
AU01027B2C11.f1	AUBES1448	181	$(GT)_{9}$	GGTCAGTGTCCCCTCAGAGT GACATGCAAGCAAATCGAAATGG	55	P
AU01028B2E03.r1	AUBES1449	318	$(GT)_{12}$	TGATCTGTGATCTTCTGCACTG GACCAGCTATAGTGCGCGTGTGT	55	
AU01032B2C11.r1	AUBES1450	139	$(TA)_{12}$	CGTTCATTTTGCTGAACGAG GACGGAAGCAGGAGCATCAGAAA	55	
AU01032B2H09.f1	AUBES1451	446	$(TA)_{33}$	GGAACCCCAGGCTACATCTT GACCACACCCTTCTGTCACTTCG	55	
AU01021B2A12.r1	AUBES1452	111	$(CA)_{8}$	GCATACTGTATTTGGGCATGG GACTGCATGTAATGGATGCGTCT	55	
AU01019B2G05.r1	AUBES1531	108	$(AT)_{14}$	ATTACGCACTCTCGGACTCG GACGGAGGAATGCAACAGGTACAA	57	P
AU01020B1G11.f1	AUBES1532	544	$(TA)_{28}$	CTTCCCCTTTTTCCAATCCT GACCCGGACTGATAACATCAAGACA	57	
AU01023B2F11.f1	AUBES1533	478	$(AAT)_{23}$	AAGCCTTCCTGTCTGTCAAAG GACCAGCAAACAATCTGATGTGGA	57	
AU01025B1F11.r1	AUBES1534	404	$(AT)_{8}$	CCCGCAGGAATTCTATAAAGG GACGCAGGAATGTCCAGAACACA	55	
AU01028A2D06.f1	AUBES1535	240	$(GT)_{13}$	ATGCACACGCATACACACG GACCCTCAGTAACTGGCAATCACA	57	NP
AU01029A2A02.f1	AUBES1536	314	$(AC)_{11}$	TGCTGGCTGTAATTTGAACA GACTCCATGGGGTGTACTTGTCC	55	
AU01029B2A11.f1	AUBES1537	186	$(TA)_{8}$	CCTAGGAACCAACATCGTTGTAA GACGCGGTACTGTACTTCCATCCA	57	
AU01030B1G03.r1	AUBES1538	193	$(AT)_{30}$	TGAGGGGTGTACTCACTTTTG GACTTGCATCGGTGCATCTCTAA	57	
AU01030B2B01.f1	AUBES1539	162	$(TG)_{9}$	GAGGCTCACTCCTCCATCTG GACCTCAAGACACGGTGACCAAA	57	P
AU01031A1D05.r1	AUBES1540	174	$(CA)_{18}$	CAGTAGGTGGAATGGCCAAA GACGGTACACCAGTCCATTGCAG	55	
AU01031A2A11.r1	AUBES1541	86	$(ATT)_{18}$	GCTGTAAGCAGCCATGTTGA GACTGAGAAGCTGTTTTTAAGGTGCT	55	NP
AU01032A1C10.f1	AUBES1542	457	$(TA)_{31}$	GCAACCAGGATCTTGTGTGAA GACATGGGTTTGGTTGCCAAGTA	55	
AU01029B2H01.r1	AUBES1543	111	$(TC)_{8}$	ACACTGGTGGGTTGTGACCT GACAACGATGGAAGCAAGTCCAG	55	NP
AU01019A2C08.r1	AUBES1544	95	$(ATTT)_{6}$	AATGCAAGTGGTACAGCCCTA GACTTGCCTATAGTTTACGATCACAAT	55	P
AU01021B1E11.f1	AUBES1545	493	$(AC)_{15}$	CGTAGCTGCCTATCTGCCTTT GACTTGATTAGGGCTGAGCAATG	55	
AU01024B2C11.f1	AUBES1546	79	$(GT)_{18}$	ACAGGCCATCAAATTCCTCA GACCAAGGGGAGTGAAATGGTGT	55	
AU01025B1A12.r1	AUBES1547	168	$(TATT)_{5}$	TGGCCCAGTGAAATCTGTTT GACCTGGAAGCAATGTGTTGTGC	55	P
AU01026A1A11.f1	AUBES1548	172	$(AG)_{10}$	GAACGGCATGCTCTATGACA GACTGAAGATGGACTGCTTTGCTT	55	
AU01028B1B09.r1	AUBES1549	283	$(CA)_{23}$	CCTCCCACCAGATCAGTGAA GACAGGTGCAGCACTCCTAAACG	55	
AU01028B1F04.r1	AUBES1550	460	$(TC)_{8}$	AAGGAGCTGAGATCTGCTTGG GACGCGTGGCAATAATATAGATGTCG	55	
AU01029A1A06.f1	AUBES1551	231	$(AATA)_{7}$	AGGGTTGGAAGCAGAGTTGA	55	

				GACGCTCAAAGGACAGCAGAACC		
AU01027A1F04.r1	AUBES1552	112	$(ATTT)_6$	GAAAGCGCAACACATTGAGA	55	
				GACGGAGTGGAACGCTGTGTTT		
AU01025B2A06.r1	AUBES1553	471	$(CTG)_6$	ACCAGCACGTCTTATCTCTGC	55	
				GACATCCTGCCGATGCAATTTTA		
AU01020A1B01.r1	AUBES1554	342	$(ATT)_{13}$	TGTGGGCACAAATGTGTACTT	55	
				GACCCATGTGCTGAGTATGGCATT		
AU01023A1C07.f1	AUBES1555	237	$(AT)_{29}$	GGGGACAGCAAATTTACACTG	55	
				GACGGTGGATTCATTGGGTTTGA		
AU01026A1H11.r1	AUBES1556	439	$(AT)_{34}$	TGCACATATTCACTGTCCATAGC	55	
				GACTGCAGTTTAGCGTGTGCCTA		
AU01026B1F04.r1	AUBES1557	117	$(TG)_{10}$	CGGTTATCGGTGTCGACTG	55	
				GACTGTTGACACGTCCCATTCTT		
AU01028B2A05.r1	AUBES1558	227	(AATC)5	CCACATGAGTGGGAGTGATTT	55	
				GACTCTGCAACCTCCAGCCTACT		
AU01032B2A03.f1	AUBES1559	497	$(GT)_{16}$	GGTGGCACTTGCAAAACAT	55	
				GACCACAGACTCGTGGCTTTTTCT		
AU01032B2D09.f1	AUBES1560	320	$(GTT)_7$	AGCAGGTGAGAGTGCTTTGA	55	
				GACTTGTACAGTATTGTCATGGGTCTG		
AU01019A1B12.r1	AUBES1561	294	$(ATTT)_9$	CATTCATTCTGGTACAATGCAG	55	P
				GACCAAAGGACATTCATGTGCAG		
AU01019A1C10.f1	AUBES1562	189	$(GT)_{10}$	ATCACACGCCATCCATCAT	60	
				GACTTCTTTGCACTCATTCGTGTG		
AU01019B1G09.r1	AUBES1563	418	$(CA)_{16}$	GGTCAGATTAACCGCACTGA	57	P
				GACCGTGTTTGGAAGGCTGTTGT		
AU01021A1A06.r1	AUBES1564	286	$(ATTT)_9$	TGCACGTTGTTCCTTTCATT	55	P
				GACGCCTCGTAGAAACCAAGGTG		
AU01021B1B10.f1	AUBES1565	227	$(TG)_{10}$	CAGTTTCACACAAACCCTATCG	57	P
				GACGCCCTGACCTCTGATTCGTA		
AU01021B2D07.f1	AUBES1566	230	$(AC)_9$	AGTCTGACCCCAGCAACACT	55	
				GACTGCCTAATCAGTGTCGCAAT		
AU01022A1F04.r1	AUBES1567	443	$(TG)_{15}$	GGCATTTCATTTCAGGACTCA	55	
				GACGAGGAATGTCGTGATCTTTGC		
AU01022B1G07.f1	AUBES1568	119	$(TA)_{15}$	ACCCACGCCTTGACCTTT	55	
				GACCGTCGACATGGATCAGTCTT		
AU01023A1C10.r1	AUBES1569	441	$(CA)_{19}$	CCTCTTGAAAATGAGCAAACG	55	
				GACTGCATGATTGCCTCATACG		
AU01023A1E06.r1	AUBES1570	84	$(TCAA)_7$	GCAGACTCAAAACACAGCAAA	55	P
				GACATGGCACACAAAAGCATGAC		
AU01024B1C02.f1	AUBES1571	147	$(GT)_{11}$	GAGGGGTTCAGAGCATGTTT	60	P
				GACCACACACACATTCCTTTCCAA		
AU01025B2F04.r1	AUBES1572	327	$(GT)_{15}$	CTCAGCACCATATGCCACAC	57	P
				GACGGCTGTTTGCCTCTTAGCAG		
AU01026A1F08.f1	AUBES1573	390	$(CA)_{25}$	AAACAGTTCAGACTTCAGTGCTC	55	
				GACCAGAATGCACGCTGAAAGAG		
AU01027B1G10.f1	AUBES1574	178	$(CA)_{10}$	AAATGCCGAGTCAGGAGTGT	55	
				GACGGTTTCACTGTGCGTTGATG		
AU01028A1C07.f1	AUBES1575	484	$(AT)_9$	TGCATGGGGAATCTTTTCAT	55	
				GACGCCTTCACACGATGTCAAAA		
AU01028A2D07.r1	AUBES1576	186	$(AC)_{11}$	CTGAAGCTCCATCCCAACAG	57	
				GACAACCCGAGCTGGAGTGATTA		
AU01028B1C08.f1	AUBES1577	487	$(TAAA)_5$	GGAGTCGCAGTCCTAAGTAGC	55	P
				GACGCACAGAAGGGCATTCTACA		
AU01029A1A11.f1	AUBES1578	331	$(GA)_{28}$	CACACTCCTACAGCCCTGCT	55	
				GACGAAACATCAGCACCAGCACT		
AU01029B1C07.r1	AUBES1579	111	$(TTA)_{18}$	CAACTGGTGACCTGCAGAAA	55	
				GACGGCAGTAACACCATCAGAGGA		
AU01030A1H03.r1	AUBES1580	522	$(ATCT)_6$	AGTACCACGGCTGTTTGAGC	55	
				GACTCAGTACACACAGGCTCTATCCA		
AU01031A1B05.f1	AUBES1581	232	$(TATT)_9$	TGCCTTAGTGTTGCTTCACAG	55	
				GACGATGGTGTCATTAGACAGTGCAA		
AU01032A2E08.r1	AUBES1582	227	$(AT)_{29}$	TTTTCGGCCCAATTACAGAG	55	
				GACGCGAGATGCCTTGCTATTTT		
AU01032B1H12.r1	AUBES1583	371	$(TG)_{20}$	TTTCCCACTCACCCATTCAC	55	
				GACCTCCCAATGGCAGGCTAAC		
AU01032B2D06.f1	AUBES1584	201	$(GT)_{13}$	TCCTTGCAATAACAGGGATGT	55	
				GACAAACGCATCGCTTCCATTTA		
AU01018B1C05.f1	AUBES1585	272	$(AC)_{18}$	TCCTGAAAACATGTCAGTTGG	55	
				GACCTGGAATACACCCTGCATGA		
AU01019B1G08.r1	AUBES1586	399	$(ATA)_{12}$	CAAAACGCCAATGACATGAT	55	
				GACTCCTCTCATCATCACCACAGT		
AU01020A2D03.f1	AUBES1587	272	$(AT)_{31}$	TGTTACCACCCCGGAAATAA	55	
				GACCAGGAACTTGTATTGCGGAAG		
AU01021A1E04.r1	AUBES1588	411	$(CA)_{23}$	TTTTGGTCCAGAACATGGAG	55	
				GACCCTGTCATTCCCTGCTTAGTG		
AU01022A1G01.f1	AUBES1589	73	$(GT)_{15}$	TCGGCAAAATCCACAAAAGT	60	P
				GACTTTGAAATCGCATTGTAGGC		
AU01023B2G12.f1	AUBES1590	232	$(ATA)_{12}$	CGCTTTCATTCGAAGCAACT	55	P
				GACTGTCAACTTGGACCTAATGTGC		
AU01029A1C01.r1	AUBES1591	550	$(ATT)_{10}$	GCCAAAATGCCGGACTATCT	55	P
				GACTGCAGTTGAGCTCTCGGTAA		
AU01029B1A08.r1	AUBES1592	314	$(CA)_{12}$	GGTGCGTTCAAGAGAAAGGA	55	P
				GACCCCTTTAAAGAGGGCTTTTCC		
AU01029B2E02.f1	AUBES1593	422	$(GT)_{10}$	CATTGCCTCGTCCAGAGAATA	55	P
				GACAATGGCATTTGGCTGAAGAG		
AU01020B1E04.f1	AUBES1594	134	$(GT)_{24}$	CCCTCAGTGTTGTGAACCTGT	55	
				GACAAATTAGGCCACGTGTAGGG		
AU01021A1F09.f1	AUBES1595	283	$(AAAT)_5$	CATGGGCATGATCACAGACT	55	P
				GACGGGTGCAATTTTTCACATGG		
AU01021B1B01.f1	AUBES1596	327	$(GT)_{12}$	GTCCTGTGACCATGTGACCA	55	
				GACGAAAGACCACCAGTGTGCTG		

AU01023A1A10.r1	AUBES1597	352	$(GTA)_{12}$	AAAGCGATTCCCCATCATC GACTGGAGCAAACAACGTTTGAG	55	P
AU01023A1E07.f1	AUBES1598	107	$(TAT)_{8}$	TTGCATCAAACCTGAAATGC GACTGAAGTGACCTTGGGTGTCA	55	
AU01023B2B03.f1	AUBES1599	213	$(GT)_{18}$	CGCTCATGCTCATCACCTC GACTGAGGAATAATGCCACCACA	55	
AU01024A1A03.r1	AUBES1600	367	$(GA)_{27}$	CCGGGGCTTCTATAGCACAT GACCTCCCTTCACAAGCTGTTCA	55	P
AU01028B2D02.f1	AUBES1601	398	$(TG)_{20}$	CAGCACATCCTTCTGAGTGC GACTCCCTGCATTCCTCTCAGTT	55	
AU01029A2B08.f1	AUBES1602	482	$(CA)_{11}$	CTGCCATCTCCAAATTGTCC GACAGGAGCGTGGAGCCTATACC	55	NP
AU01029A2F05.f1	AUBES1603	236	$(AC)_{13}$	CTGCAGCAGAACAGCACATT GACAGCAGCCCGCATTCTATGTA	55	
AU01029B2G10.f1	AUBES1604	149	$(TC)_{20}$	TTTTTGGCGGACGAACAC GACTCCTCAGCCCACACTTCCTA	55	
AU01031A2E08.f1	AUBES1605	85	$(AG)_{11}$	AGCCAGATCCGATCACTCAG GACGGGTTAGGCGTTAGGGGTTA	60	
AU01031A2G03.f1	AUBES1606	335	$(AC)_{16}$	ACAGCCAGATGATTTCCAGTT GACGGTTAAACAGCTAGGTGCACTG	55	P
AU01032B1F02.f1	AUBES1607	352	$(TA)_{32}$	CCAGTCCGACATAGTGAGGA GACCCACCATGTGCCCAGTCTAT	60	P
AU01021A2E04.r1	AUBES1608	433	$(TG)_{9}$	CCACTTCACACTGCCGTCTA GACCCCTACTTGTGCCTGAGAGTG	60	P
AU01021A2E10.f1	AUBES1609	363	$(ATG)_{9}$	CCGGCTCTAATGATGCAGTT GACAATTGGGATGAATGGATGGA	60	P
AU01021B1A10.f1	AUBES1610	466	$(GT)_{12}$	CGCTCACTACATAGGGCATGA GACACTCGCTGAAGAAGGCATTT	60	P
AU01023B1C07.f1	AUBES1611	611	$(AC)_{12}$	TAGCCCGTACGTGTTTATGC GACGTGATCGAGGCTATGCCATT	60	
AU01024A2B01.r1	AUBES1612	148	$(CA)_{17}$	GTCTCTTTTCGGTCCAGACG GACTACCAGCCTTCCAAGCATTC	60	P
AU01024B1A09.f1	AUBES1613	324	$(TATT)_{8}$	TCACGTGACCACACGTTACA GACTGTCCTGAATTGCGTAGTCG	55	P
AU01024B1G07.f1	AUBES1614	91	$(CA)_{9}$	AAGGCTGGACAAGCAATGTT GACCCCTAACTGCTAAGCCATCA	60	NP
AU01024B1H04.f1	AUBES1615	77	$(GT)_{20}$	TCCCTTAAAGCCCTCAATCA GACTGATGCCTGGCTGAGAGATA	60	P
AU01026A1G06.r1	AUBES1616	257	$(CT)_{8}$	TGGATCAAAGTCCCCAATTC GACATGGATCTGGCACAATGGAT	60	P
AU01026A1G03.f1	AUBES1617	122	$(AG)_{16}$	AAACTGCGTCGAGTTCCACT GACGCGCTCCTCAGTCTCTCATT	60	P
AU01018B2A04.f1	AUBES1618	138	$(AC)_{27}$	GAGTTCGGAGAAAGCACACC GACGCTTCATCCACCTACACATGC	60	P
AU01020A1A11.f1	AUBES1619	525	$(ATAA)_{5}$	GCGAGATACTGCCGTTTGAT GACCACCGGAGACAATGTACTGG	55	P
AU01020B1H12.f1	AUBES1620	225	$(GAA)_{7}$	GTGGAATAATCACGGCTTCC GACCACGTGTTTTAGCCTGTCCA	55	P
AU01021A1G03.f1	AUBES1621	299	$(TG)_{8}$	GGGAATACTTTGTGGGTAGTGC GACCCCTGACCAGGATACAGTGG	55	P
AU01021A2F06.f1	AUBES1622	209	$(ATTT)_{6}$	GCAGGCACTCCACAACATTA GACCCATGCAGTAAGGGGTTCAT	55	NP
AU01022B2E04.f1	AUBES1623	199	$(GT)_{9}$	CCTGACCTGCACACTCATTC GACAGGGAGTGCAAGTTGTGGAA	55	NP
AU01023A2H12.r1	AUBES1624	432	$(CA)_{11}$	CCTCTTTTAGGTCGGCTGAA GACAGCACTAAGCACAGGTGCAA	55	P
AU01026A2C01.r1	AUBES1625	263	$(TGG)_{7}$	TCTGACTGCTCGGGTTTACA GACCCACCATGTCCGTGACAATA	55	P
AU01027A2E03.r1	AUBES1626	175	$(AT)_{9}$	TGGTACACGATCATCTTCCTGA GACAGTGATTGCACATTCACAAGG	55	P
AU01029B1F01.r1	AUBES1627	563	$(ATGG)_{7}$	GCCAAACAGCGACAACTCTT GACTGGGATAGGCTCCAGGTTC	55	NP
AU01030A1C12.r1	AUBES1628	94	$(TATT)_{9}$	GAAGTATGCATGGGGATTGG GACACCACTCACCTGTGCCTGAA	55	
AU01031A2A05.r1	AUBES1629	425	$(AG)_{9}$	CAGAGCACTTGCATCAGGAG GACCACGCCTACAAAACTCCGTA	55	P
AU01032A1H08.f1	AUBES1630	405	$(CT)_{11}$	TTCCCTGTCTGAGCGAGTCT GACGAGCGCGAAGGTAAGAGTTG	55	P
AU01019B2H06.r1	AUBES1631	252	$(AG)_{22}$	CTGTAAGCTCACTGCCACCA GACCAGTGTGAGGTGAAAGCACTG	55	
AU01020A1A03.f1	AUBES1632	273	$(TG)_{15}$	AAAGCCGGTACCTCATTCCT GACTCTGCACAGCATCACTCCAT	55	
AU01020A2E12.r1	AUBES1633	221	$(CA)_{11}$	CACAGTGCTTTGTTGTGACG GACGTACCCCAGGTGTGTTTGCT	55	
AU01021B1C06.r1	AUBES1634	230	$(CA)_{27}$	CCCTGGCGTTTTCAGTAGAA GACCGAGGACCTGGATCAGACTC	55	
AU01023A1B02.f1	AUBES1635	74	$(AC)_{16}$	CAGTGGAATGTCCTCACAAGG GACACTGTGTTTGCCCAGGTGTC	55	
AU01024A1A04.r1	AUBES1636	148	$(AT)_{17}$	GGTGCAAGGAAAATGACAGG GACGTCCTTCAAGCTGCCAGTGT	55	
AU01019A2F09.f1	AUBES1721	329	$(CA)_{22}$	AGTTGGAGCCAGGTAAGTGC GACACCATCGCACCTAGCAAAAC	55	
AU01020A1A08.f1	AUBES1722	203	$(ATT)_{12}$	CGACATTGAGGTTTGGAGGT GACGAGGTGAGGAGTGGCCATAA	55	P
AU01020A1C03.f1	AUBES1723	487	$(TTGT)_{6}$	CATGAGCCTGACACTGGAGA GACCGGACGCTCCACATAATCTA	60	P
AU01021A1A05.f1	AUBES1724	225	$(TCA)_{11}$	GCAGTGTGAAGCTATGTCATGTT GACGTGACTGGAAGCATGGGAAT	55	P
AU01022A2H05.r1	AUBES1725	233	$(TTAT)_{8}$	TGTCATGTCAGTTGGAAGCA GACCCATGTATCAGGTTTGCACATT	50	P
AU01023A1F08.f1	AUBES1726	216	$(CA)_{19}$	CAGTAGGTGGAATGGCCAAA	60	P

AU01023A2C12.r1	AUBES1727	110	$(GT)_{12}$	GACGGTACACCAGTCCATTGCAG TCGATCCGTGGCCTAAATAC	60	P
AU01023B1C08.f1	AUBES1728	401	$(CT)_{20}$	GACCAAAACAGTCCTGGCTGACA GCATGAATGAGCTCACAAGC GACTTGGCCTCTAAATTTGTGCTC	50	P
AU01024A1B03.f1	AUBES1729	194	$(AT)_{30}$	CAATGTAGCCTTCGGACAGC GACCTTGCTCCACAGCAAGAGC	55	
AU01026A1G04.r1	AUBES1730	104	$(TAT)_{18}$	CAGCCAGGAGTCCAAACTGT GACTCAAACCAGTGTGCTTTCTCC	60	P
AU01026B2G04.r1	AUBES1731	200	$(TG)_{16}$	CTACAGGTTCCCCATGGTTG GACGCGCTTAGCCACAATCATCT	55	P
AU01026B2F08.f1	AUBES1732	199	$(TG)_{12}$	AGGGGACTCGAGGCAATC GACTGTTTAGAGAGGTGCCTTGC	55	
AU01027A2E05.r1	AUBES1733	421	$(AT)_{11}$	AACTGACCGGAACTACCTGTG GACCGCATCCCAGCGTACAATTA	50	P
AU01028A1C01.r1	AUBES1734	98	$(CT)_{16}$	ATCATGCCTTTCGACGTCTC GACGTTCACCGATCTCACTGCTG	55	
AU01030B2B08.r1	AUBES1735	386	$(TC)_{13}$	AACTGAGCCAAGCAAACTGC GACCCTTTAAGGCCATCTGTTCC	60	
AU01030B2C05.r1	AUBES1736	259	$(AG)_{9}$	GGTGAACAATGGCTGGAGTT GACCCCACCCTATGTCCTTAGCA	50	P
AU01031A1F10.f1	AUBES1737	251	$(TTA)_{6}$	AAAGCAAGCAGTCATCACGA GACTGCTCAGCACTGAAATCACA	60	P
AU01031B2C07.r1	AUBES1738	345	$(TG)_{24}$	GCAAGATGGGATTCCAGTGTA GACACTTCCCAAAAACCTGCTGA	60	P
AU01032B2A06.f1	AUBES1739	203	$(GT)_{13}$	TCCTTGCAATAACAGGGATGT GACAAACGCATCGCTTCCATTTA	55	
AU01032B2G01.f1	AUBES1740	126	$(TTA)_{16}$	CAGTGTCAGCATCCAAGAGG GACAACCCTGAACTTAAAACCCTGA	55	
AU01019B1D12.r1	AUBES1741	483	$(TG)_{7}$	GGCTACACACACTTCCCATATC GACGAGGGGCAGGTTAATACGAA	60	P
AU01020A1A02.f1	AUBES1742	173	$(ATA)_{8}$	TGCGTAGGTCAGATCCCTCT GACGCCCAGGGGTCCAATTAC	60	P
AU01020A1B07.f1	AUBES1743	427	$(GT)_{9}$	CGCTGAAGAGCAGAGAGGTT GACTGGCTAACGACATGTGACCT	60	NP
AU01020A1F05.f1	AUBES1744	276	$(TAGA)_{7}$	CATCCTTCCTTTCCTTCACG GACTTTGATGCTGCTTTCACCTG	60	P
AU01021B1F04.f1	AUBES1745	245	$(AC)_{5}$	AATTGGGCTGCTCTCAGTGT GACCACAGGACAAAGCTCGTTCA	60	NP
AU01021B2A03.r1	AUBES1746	135	$(TG)_{19}$	CAGGCATGCGAGAGTGTATC GACAGCATGACGCACAGTTCTTG	60	
AU01022A1E09.r1	AUBES1747	97	$(TCC)_{8}$	GCAACATGACAACTCGGCTA GACTTCTGCTCAGAACCCTTTGC	60	P
AU01028A2F04.r1	AUBES1748	326	$(GTTT)_{7}$	CACCTGCAACTGCACTATGA GACTGACTGTGAGTAGTTTCCCTGCT	60	P
AU01030A1E12.f1	AUBES1749	437	$(TC)_{15}$	CTGTCTCGAAGCTGTGCTTG GACACCTTGGAGAGGGAACGTCT	60	P
AU01032A1F10.r1	AUBES1750	326	$(CA)_{13}$	CTGCAGCCAGCTCTTTCTTT GACTCACCATATGACGGAATGGA	60	
AU01032A2B02.f1	AUBES1751	485	$(CA)_{29}$	AAATGGAGCGTTATGGGATG GACCTGCATGTACTGCCCATTTC	57	
AU01018A1C09.r1	AUBES1752	188	$(TA)_{37}$	CTCTTCCCCACCTCCAAGTA GACGGGAGTACAGTAAACCACTTGC	57	
AU01023B1D06.r1	AUBES1753	380	$(AG)_{27}$	CATTGGGAAAAGCCCTCTAA GACCGTTGAATCATTTGGATTGC	57	
AU01024A1G05.f1	AUBES1754	154	$(TAA)_{12}$	TGTAAAGTGCCTTGAGAAGCTG GACAGAGAGGCCTTTTCACAGCA	60	P
AU01024B2B05.r1	AUBES1755	211	$(AC)_{20}$	TCTCTCAGAGGGCATGTCTG GACCGACAGCTTTGTGTGGAAAA	60	NP
AU01024B2D02.r1	AUBES1756	134	$(AT)_{36}$	CCGGCATTTCACGGATATAG GACAGCAACGTAACAGCCAGTCA	55	
AU01026A2A05.f1	AUBES1757	49	$(CA)_{10}$	CGAGACACTGAACCCCAAGT GACCGCTTTACGCTGGTTCTCAT	60	
AU01027B2A02.f1	AUBES1758	130	$(AC)_{10}$	AGCTGGGACAAAAGTCTTGG GACGCCAACCTTCCTTTGCTCTA	60	
AU01027B2B05.f1	AUBES1759	150	$(CT)_{13}$	TGGAAGATGGTACGGGAGAC GACGCCCTCTGGCAATAACAAGT	60	P
AU01028B1A03.r1	AUBES1760	244	$(GT)_{8}$	GCGGTTTGCCTTACACAATC GACGACTTCGCGTTCCGTAGTTC	60	NP
AU01028B1D05.r1	AUBES1761	201	$(TAT)_{10}$	GCTGCATGTTCCTGAGGAGT GACGAAACACTGGGGATGAGGTG	60	P
AU01030B2C11.f1	AUBES1762	359	$(TTAT)_{7}$	CCCAAACTGACGGAGTGAAT GACCGCTTTCCGTCACCTAAATG	60	P
AU01032A2G02.f1	AUBES1763	487	$(AC)_{40}$	TCATGGCATGCATACTAACACA GACCTGCATGTATGTGGCTCCAG	60	P
AU01032B1H08.f1	AUBES1764	431	$(AAT)_{16}$	AGCACGGAATGACTGCTTTT GACCAATGTCGAGGCCAATCTG	57	P
AU01032B2H02.f1	AUBES1765	427	$(ATTC)_{9}$	GGGACCAGGGTAAAGCAGTT GACGCCCTTGTTTGTTCATGTCC	57	P
AU01019B1B01.f1	AUBES1859	319	$(AG)_{11}$	TCTGATGCCAGAACTTGTGC GACTCACACCATGTGCTTCAACA	55	P
AU01019B1D10.f1	AUBES1860	119	$(TCCA)_{5}$	CTGGGATATGCTCCAGGTTC GACGTTGCCAGTGGAACATCTCA	55	P
AU01020B1E09.f1	AUBES1861	357	$(AT)_{11}$	GCATAGGAGTGGAGCTTCAAA GACGCCATTTATATTGTGCCTGTTG	55	
AU01021B1G12.f1	AUBES1862	294	$(TA)_{32}$	TCAGCCTAAAGCTTTCAATTCC GACGGCTCCACGTTTTATGTCGT	55	
AU01021B2H11.f1	AUBES1863	164	$(CA)_{8}$	GGGTGCAGTCTATGCAGGTC GACAGTTGTTGTCAGGGGCTTCA	55	P
AU01022B1A02.r1	AUBES1864	173	$(GAA)_{13}$	TTCCCCGTATGAGTGTAGGC GACCGGTCCACAAGTGGTAAAGAA	55	

AU01022B2C01.f1	AUBES1865	451	$(TAT)_{14}$	CGTGTGGAGCAATTTGAGTG GACTGAAACTGGATCAGAGGCTTT	55	P
AU01023A1G06.r1	AUBES1866	156	$(TG)_{11}$	TGTCTCAATCCTTTGCATCA GACGCTGCATTCTAGTGCAGAACC	55	NP
AU01023B1E10.r1	AUBES1867	221	$(AAT)_{17}$	TCTGGCAACTTTTACAGTGCAT GACTGTTCACACACATTGCTGACA	55	P
AU01024B1A09.r1	AUBES1868	413	$(TTA)_{13}$	TCCTCTCCTGAGATCCTTAACA GACGTTCTCCAGGCCCAGTCAT	55	P
AU01025A1E08.f1	AUBES1869	394	$(TAT)_{11}$	GAGCCGGGTTACACCTAACA GACTTGCTACAGCTGAGGCATGT	55	P
AU01025A2G01.r1	AUBES1870	188	$(AAC)_{6}$	AAGGGTCTTGCCCTCTGAAC GACGGTGAAGCCCAATGATGCTA	55	NP
AU01025A2E02.f1	AUBES1871	409	$(CA)_{13}$	CATGGACGCATTGTAGTGTTG GACTTGGATAGCTCACGGTGTATG	55	P
AU01025A2G04.f1	AUBES1872	245	$(AC)_{18}$	TCCTGAAAACATGTCAGTTGG GACCTGGAATACACCCTGCATGA	55	P
AU01025B1H07.f1	AUBES1873	109	$(CT)_{19}$	TTGCCTGCTGAACATGATTG GACCTGCCCTCTGATACCTGCAT	55	P
AU01026A1E04.r1	AUBES1874	265	$(AC)_{11}$	CCCACTGTTCAACAACATGC GACGGATGGGACGTCAATCCAT	55	P
AU01026A1D01.f1	AUBES1875	249	$(TA)_{35}$	CACGTTTCTGTTTAACGAGCAC GACCGTGGTACAGGGACAAAGGT	55	NP
AU01026A2A04.r1	AUBES1876	322	$(AG)_{10}$	GGTGCTGAAGTGCCAAGACT GACCCAGCTGAAGTGAGATGGTG	55	P
AU01026A2E12.r1	AUBES1877	246	$(TCCA)_{7}$	TGAAGGCAGGATAAGCGGTA GACCGGCCGTTTATAGCTTCTGT	55	P
AU01026A2G10.r1	AUBES1878	427	$(AG)_{16}$	GAGTCCCTGCTTGCACTCTT GACTGAGGATCAGGCAACATCAG	55	P
AU01026B1F12.r1	AUBES1879	409	$(TG)_{33}$	GGTGCATTAACCGTTTCTCTG GACTGAAGGTTGACAGCATCAGG	55	P
AU01027A1F03.r1	AUBES1880	201	$(AC)_{40}$	ATGGGAACCTTTGAAGCTGA GACTTCAGGGTGGTTGTAGAATGC	55	
AU01027B2B03.f1	AUBES1881	328	$(ATA)_{11}$	ATCGGGCCTTGAGGTAGATT GACTGTCACTTAAATCACGGAGGTT	55	P
AU01028A1F11.r1	AUBES1882	491	$(TGAA)_{8}$	TTGATGATGGTGCTGGAGAA GACAACAAAGCTTGGCCTTATGC	55	P
AU01028A2F12.r1	AUBES1883	319	$(TG)_{8}$	CCTGCGAAGTTTTCCTGAGT GACACTCCGGAACCTTGATTTCC	55	NP
AU01028B1H01.f1	AUBES1884	118	$(ATCC)_{7}$	CCCGTGACCCTGTACGATAA GACCCCAGAAAAGGGATCTTGGT	55	P
AU01028B1H10.f1	AUBES1885	354	$(TC)_{20}$	CATCAGGCTTTGAGCAACTG GACATCCACCCCCTTGTCTGACT	55	P
AU01029A1A12.r1	AUBES1886	255	$(GT)_{30}$	GTTGCACTTGATGCAAAGGA GACGGCTCTTGACCTGAATTGTG	55	P
AU01029A2C07.r1	AUBES1887	242	$(ATA)_{10}$	CAGCACAATGCAGTTTTGAA GACCATGGAAGGAGTCTCCAGTG	55	P
AU01029B1H08.r1	AUBES1888	294	$(AG)_{31}$	GTGACGGAGCCTGTCTCTCT GACCCTGTTCCCAGATCAGAAGC	55	P
AU01030B1D06.f1	AUBES1889	201	$(TA)_{33}$	GCTTTTGCAGATACCCAGAAA GACTCCGTTAATAATCGGCTGAGA	55	
AU01030B1E03.f1	AUBES1890	328	$(AT)_{27}$	GCCCCATCCTGATTTCTTCT GACTCACACTTGCCCAGTTGGTA	55	NP
AU01030B2F12.f1	AUBES1891	330	$(CA)_{17}$	GCAGAGAGTCATGTAGGGTGTG GACCGCACGACTCGGACAGTAAT	55	P
AU01031A1D03.r1	AUBES1892	275	$(AT)_{10}$	TGCCATCAAGCGTTAGCATA GACGCATCACCATTCGTGGTCTA	55	P
AU01031B2D10.f1	AUBES1893	354	$(TAA)_{14}$	CACTGAAGACATTTGGGTTTGA GACGCACACCAGTGGTTTCTTTCT	55	NP
AU01032A1B12.r1	AUBES1894	96	$(AT)_{8}$	TGACATGCAGTCTTGCTGAAG GACAGGTGACGTGGCAATTAAGC	55	NP
AU01032B1C11.f1	AUBES1895	428	$(GT)_{9}$	CCTCACCTGGAAATCCCATA GACGCATGGCAGCTCTGCTACTA	55	NP
AU01032B1H04.f1	AUBES1896	229	$(AAT)_{6}$	TGGTTTACTTGGGACCATCTT GACTTCATTCAGCTTTGCGTCAT	55	P
AU01019A2D06.f1	AUBES1947	343	$(CA)_{14}$	TGGACTCTGCCTTTTGATCC GACCAGATCCTGATCCCTGATGG	60	P
AU01019B1F09.f1	AUBES1948	372	$(AC)_{11}$	GGAACTAAGAGGCCCAAACC GACCCACTCCAACCATAACACACC	55	P
AU01020B1E12.r1	AUBES1949	343	$(CA)_{14}$	CATGCTGCTGATACGACTCC GACAGACCTCCATAGGCCACGTC	55	P
AU01020B1F04.f1	AUBES1950	306	$(TTC)_{13}$	CTGGCCACAGAGCAGAGAG GACCTACACCATGAAGGGCCAGT	55	P
AU01020B2D10.f1	AUBES1951	597	$(GA)_{10}$	CCACCCCTCCCTTTGTTTAT GACCTGCTGAGATATGGAGGAGGA	55	P
AU01021A1B04.r1	AUBES1952	192	$(ATT)_{14}$	GCGAACGTGCTAACCACTAA GACCCCTGTTGCTACCCGTGTT	55	P
AU01021A1G02.r1	AUBES1953	324	$(CA)_{30}$	TGGCATCTTTGACTTGTGGA GACAGGATGCCTACCCATCACAG	55	P
AU01021A1D05.f1	AUBES1954	92	$(CA)_{18}$	TGCCCTAGATGTGTCAGTGTG GACCAGTGCATCACAGGGCACTA	55	
AU01021B1A04.f1	AUBES1955	541	$(TTA)_{6}$	GGAACTCGCTGAGCCTTTTT GACCCACACATGCTTCCAATGAG	55	P
AU01021B2D08.f1	AUBES1956	200	$(CA)_{20}$	CCCAGTCCAAAGGCATACAC GACCAACCGATTGCAGGGTACA	55	P
AU01022A2H04.r1	AUBES1957	193	$(TA)_{22}$	CCCCTCGATCTATGCTCACT GACTTTGCATGGTACCTTCATGG	55	P
AU01022B1F05.f1	AUBES1958	404	$(CA)_{11}$	TCCGACCATATTGTGTGTGC GACGTGTTGGCAAGGTAAGAAGACC	55	
AU01022B2B02.f1	AUBES1959	279	$(GT)_{11}$	GCGATGTTCTTCTGGGTTTC GACGCAGCTGCCTATCTGCATTT	55	P
AU01023A1D10.r1	AUBES1960	72	$(AC)_{35}$	AGCCACCAGAAGGCTAAAGG	55	P

AU01023B2D05.f1	AUBES1961	386	$(AG)_{15}$	GACTGAATTAGCCACGACAGACAA CCATCATGTCTTGCATGCTC GACACAACCACTGGCCAATCTTC	55	NP
AU01026A1F07.f1	AUBES1962	457	$(AAG)_{8}$	GTAAGGAGCACGGGAATGAA GACCTCCGTTACGCACAGAACAA	60	
AU01026B1C03.f1	AUBES1963	200	$(AT)_{10}$	GAATACTTTCCGGATGCACTG GACTCCTGCTCTTGCTGTCTTGA	55	P
AU01027B1G12.f1	AUBES1964	537	$(CA)_{11}$	GCTCCCCATCTCTCTGTTTG GACTCCTGTTTTCCCTTCACAGC	55	P
AU01028B1B01.r1	AUBES1965	51	$(ATA)_{20}$	CCCAACAAATGGCAGGTACT GACTATGGCCAAGGCCGATGTAT	55	P
AU01029A1B01.f1	AUBES1966	404	$(GATG)_{5}$	GCCAGCTCATGAAGTGTTTG GACAATGCTCCCTGGGGTAGG	60	
AU01029B2G12.f1	AUBES1967	363	$(TC)_{21}$	CCCTACTATTTTAACAGGGGTGTG GACTGGTCACAAGTCCTCACAGG	60	P
AU01030A1F10.r1	AUBES1968	444	$(TAA)_{22}$	CGCATTGAGCAACTGTGTCT GACGCCTGGTTTTACGCTGACAT	55	P
AU01030B2A05.r1	AUBES1969	370	$(TG)_{12}$	AAGGCGCGAGTCTAGACAGA GACCAGCTCGGTGTTGTTGTAGC	55	P
AU01032A1H06.f1	AUBES1970	125	$(TGTT)_{5}$	GGTGTGAGTCCATACCAGCA GACCAGTCGAGCCTGGTGAAGTA	55	P
AU01032B1D02.f1	AUBES1971	321	$(TA)_{32}$	CCAGTCCGACATAGTGAGGA GACCCACCATGTGCCCAGTCTAT	55	P