Isolating Microsatellite Sequences for Development of Microsatellite Primers for Cave and Surface Amphipod *Gammarus minus*

by Andrew Frank

Honors in Biology

Advisor – David Carlini, PhD

American University

Washington, DC

Spring 2011

Abstract

Recent genetic studies of the cave and surface amphipod *Gammarus* minus have revealed hydrological differentiation of populations in the Karst geography of Virginia and West Virginia, as opposed to previously indicated. This indicates that cave populations are more closely related to their neighboring above ground population rather than fellow cave populations. However, these genetic studies were not able to elucidate the fine scale geographical structure of these populations at the underground-aboveground interface. In this study, microsatellite containing DNA sequences are isolated *de novo* from *G. minus* genomic DNA using a selective hybridization technique. A number of genomic DNA fragments containing microsatellite repeats were isolated from a large pool of candidate fragments via sequencing of candidate fragments. These microsatellite containing genomic fragments serve as a starting point of the future primer design for a number of different *Gammarus minus* microsatellite loci.

Introduction

Subterranean ecology offers complex insight to evolutionary processes, especially processes of convergent evolution and stabilizing selection. These habitats offer the dual benefit of having both a limited number species and readily identifiable selection pressures (Culver 1976). While a number of vertebral organisms have been studied in context of cave evolutionary processes, much work is still being done on various genera of invertebrates that inhabit subterranean ecosystems. Work started primarily in the late 1960s and early 1970s by T.C. Barr, David Culver, and John Holsinger, includes studies of a variety of invertebrate cave inhabitants, including both terrestrial and aqueous forms (Culver, 1976).

Notable work has also been done on aquatic subterranean invertebrates, with studies being done on a variety of forms native to the habitat (Culver, 1995). Specific focus was also put on the order amphipoda, which included a variety of members that were either eutroglophilic or troglobilic. Combined with the isopods, these two orders consist of more than 90% of the biomass present in a large number of caves (Culver, 1976), providing ample subjects for evolutionary and ecological studies.

Amphipods in general consist of a variety of species at different stages of adaptation to the cave ecosystem (Culver, 1976), but are not immediately identified as an apt model for cave evolutionary history studies. Species tend to exhibit traits specific to a subterranean environment, and are largely invariant, making it impossible to ascertain what the original non-troglobilic form resembled (Culver, 1995). However, the species *Gammarus minus* represents a morphological mid-point of cave adaptation. This species is often seen both in aboveground cave springs and streams and in subterranean waters, occupying both habitats with seeming equal effectiveness. *Gammarus minus* also lacks distinct characteristics typical of their underground relatives. This eutroglophilic amphipod therefore presents an opportunity for the study of the evolution of troglomorphy, and specifically the study of how a population transitions from the surface to the underground.

The Evolution and Phylogeography of Gammarus minus

Not all populations of *Gammarus* minus prove to have characters amenable to evolutionary studies. Only select populations of Gammarus *minus* have been a primary model used in cave ecology and evolution. This is due to the unique populations within the Greenbrier Valley of West Virginia and within Tazewell County, Virginia (Carlini, 2009). Although its *G. minus*'s geographic distribution consists of carbonate rock formations in states like Pennsylvania, Tennessee, and Indiana, the populations within Greenbrier Valley and Tazewell County have an additional morphological variety (Carlini, 2009). This variety exists only within subterranean ecosystems, and possesses reduced eyes with few ommitidia, a larger body, longer antenna, and a loss of the characteristic brown pigmentation (with select subpopulations possessing a bluish hue) (Fong 1989).

Only a limited number of studies have explored the genetics and phylogeographical distribution of the endemic Virginia and West Virginia *Gammarus minus* populations. The first of these studies, conducted by Daniel Fong in 1989, demonstrated that morphological evolution of these *G. minus* populations was consistent with a pattern of negative pleiotropy, where reduced eyes were the result of selection for longer antennae. Further genetic studies were done in 1992 by Kane et. al., confirming a hypothesis put forth by Culver et. al. in 1990, which stated the morphology of cave dwelling *Gammarus minus* populations was the result of independent invasions of each cave system, followed by morphological convergence driven by the selective force outlined by Fong (Kane, 1992).

In addition to supporting the independent invasion hypothesis, Kane and later publications by Culver et. al. put forth that the genetic connection between these cave and surface *Gammarus minus* populations has largely been degraded due to multiple local extinctions followed by repopulation via colonization. The most recent genetic study on these *Gammarus minus* populations was done by David Carlini et. al. in 2009, seeking to improve the resolution of the Kane study, as the allozyme data put forth by Kane et. al. was insufficient to conclusively elucidate the historical relationship between surface populations and cave populations. In this study, significant differentiation was found between variability of the populations' cytochrome c oxidase subunit 1 (CO1) DNA sequence, with cave populations possessing significantly less variation. By ruling out the possibility of CO1 being linked to troglomorphic characteristics, and thus under selection pressure, Carlini et. al. showed that cave populations were far more prone to genetic bottlenecks, consistent with local extinctions, and that the populations still held their hydrological associations with their immediately neighboring spring population (Carlini et. al., 2009). In addition, codon biases within the cave populations were limited, again pointing towards a founder effect within the cave populations (Carlini et. al., 2009).

While the Carlini et. al. study was able to spatially structure populations based upon the CO1 locus, and was able to elucidate the lack of variation in the locus within the cave populations, the resolution of the CO1 locus was unable to ascertain the fine-level phylogeographical structuring within the cave systems, which is critical to understanding the pattern of bottlenecks and reinvasions as described by Culver and Kane. As of yet, no data show how genetic variation

and differentiation differ in *Gammarus minus* at various depths within the cave systems. Modern genetic analysis techniques offer the possibility of characterizing this phylogeographical structure, particularly through the use of microsatellite genetic markers, which has been shown to provide highly resolved local phylogeographical data (Goldstein et. al., 1997).

Microsatellites as Genetic Markers

The use of microsatellites as genetic markers in evolutionary biology studies began in the early 1980s, and rocketed in popularity with the advent of PCR and rapid assessment of microsatellite genotypes (Ellegren, 2004). Early studies found novel sequence motifs in DNA, within non-coding regions, assembled of seemingly repeated units. Further research showed that every base combination possible in groups of 2, 3, and 4 are vastly overrepresented in the human genomes, with later studies confirming the presence of these repeated motifs in the genomes of other organisms (Ellegren, 2004).

Given that almost every microsatellite region occurred within non-coding regions (with the rare exception of repeated motifs occurring in genes like the Huntington's gene), these DNA regions were recognized for their value in estimating genetic relationships between organisms (Ellegren, 2004). In addition, microsatellite loci are differentiated only by number of repeated di-, tri-, or tetranucleotide repeats, making alleles easy differentiable, and mutation rates easily determined (Avise, 2004). Finally, most microsatellites are intraspecifically polymorphic, making them excellent potential genetic markers for studies requiring high resolution in a localized intraspecific region (Tatutz, 1989).

Although microsatellites have the same generalized mutation mechanism, which adds or subtracts whole repeat motifs, the rate of mutation has been shown to differ among loci, among alleles, and among species (Ellegren, 2004). The polymerase slippage model commonly applied to microsatellites readily explains this variable mutation rate, but it also indicates that the mutation rate seen in microsatellite regions may not correlate directly with the mutation rate of the entire genome (Ellegren, 2004).

The intrinsic polymorphism of microsatellites was shown in 1989 by Diethard Tautz using PCR technique on *Drosophila melanogaster*, human, and whale's DNA, and was one of the first papers to show the usefulness of PCR to analyze microsatellite alleles. Tautz used known primers for intergenic regions in these genomes, and subsequently sequenced the regions to scan for repeated motifs. From the sequence results, Tautz showed the hyper-variability of alleles within a population.

However, until 2002 it was unclear whether or not population genetics studied done with microsatellites were significantly affected by homoplasy, where two identical microsatellite alleles are analogous, not homologous. Under the stepwise mutation model, which dictates that repeated microsatellite motifs are added or subtracted in whole numbers, homoplasy seems a distinct possibility in the model. This issue was addressed by Estoup et. al. in 2002, which showed that size homoplasy (microsatellite alleles with identical number of repeats) is largely compensated for the significant hyper-variability of microsatellites, as observed by Tautz. Additionally, analysis of multiple loci and multiple repeat motifs practically eliminate the possibility of size homoplasy

affecting population genetics studies in all but populations with severe restraints on microsatellite allele size (Estoup et. al., 2002).

Microsatellite Population Genetics Studies in Invertebrates and Crustaceans

Microsatellite regions have been shown to be powerful Mendelian genetic markers, and provide revealing data for intraspecific analysis and interspecific analysis for closely related species (Jarne et. al., 1996). While a large portion of microsatellite studies have been performed on vertebrate populations, the viability of microsatellite markers for defining invertebrate species relationships is strong. A number of studies in recent years have shown the efficacy of microsatellites in expanding upon and resolving previous allozyme data and mitochondrial DNA sequences in invertebrates and crustaceans.

Studies of marine fishes with large geographical ranges face problems much like the current state of *G. minus* genetic information, with allozyme data and mitochondrial sequence data providing a limited resolution of population structure. In their 1999 study, PW Shaw et. al. demonstrate the utility of microsatellite DNA markers for examination of subtle population genetic structuring in the marine squid *Loligo forbesi*. Unlike mitochondrial DNA, which has limited polymorphism, Shaw showed that microsatellite sequences demonstrated high levels of polymorphisms that far better accounted for population structure than counterpart allozyme data. Indeed, the Shaw et. al. study showed significantly more variability in the population structure of *L. foresi* than previously implied by the allozyme data. Shaw et. al. suggests that allozyme data cannot detect fine scale geographical differentiation, which correlates with the data seen for *Gammarus minus*.

A recent study by S Pascoal et. al. indicates similar patterns of microsatellites providing fine scale geographic resolution among *Carcinus maenas*, a European shore crab. Again, similar to *G. minus* and *L. forbesi*, *C. maenas* exhibited distinct differentiation over large geographical areas when analyzed under the mitochondrial CO1 DNA sequence, but characterization of fine-scale population structure was otherwise not possible due to lack of variation (Pascoal, 2009). Microsatellites again provided finer scale geographic information, but Pascoal et. al. describe significant difficulty using previously developed microsatellite loci primers for *C. maenas*. Because of this difficulty, Pascoal et. al. highlights the need for developing novel microsatellite primers when analyzing a new distinct population of a species. Difficulty was also attributed to the lesser degree of microsatellite sequence occurrence within invertebrate genomes, but other studies suggest that arthropod microsatellite occurrence is on par with vertebrates (Zane et. al., 2002).

Using Microsatellite Alleles to Elucidate the Evolutionary History of Gammarus minus

Microsatellites remain one of the most powerful tools in a population geneticist's arsenal for exploring subtle relationships either within species, or between closely related species, and represent one of the best molecular marker candidates for resolving the population structure of *G. minus*. Isolation of microsatellite DNA from cave and surface populations of *Gammarus minus* in the karst regions of Virginia and West Virginia is the next logical step in resolving

the evolutionary and geographical history of this model cave species. This study represents the first steps towards developing microsatellite genetic markers for use in the *G. minus* species complex within Greenbrier Valley and Tazewell County. By developing a genetic library of a number of microsatellite sequences from *G. minus* populations throughout the Karst topography of Virginia and West Virginia, this study makes it possible for future research to utilize these known microsatellite markers for development of specific primers and studies to elucidate the population genetic structure of cave adapted *G. minus* populations.

Materials and Methods

Sampling Locations and Specimen Storage

Approximately 200 cave specimens were collected live from Ward Spring and The Hole Cave in Greenbrier County, West Virginia. Sampling locations were selected to overlap with those from previous studies (Kane et. al. 1992, Culver et. al. 1995, and Carlini et. al. 2009). The Ward Spring locality was identified as a surface site, while The Hole Cave locality was identified as a cave site. Specimens were maintained live at 4˚ C in the water from the sampling location until returned to the lab.

Between sampling and DNA extractions, specimens were stored in a refrigerator for 4˚ C, typically in tubs with approximately 2 liters of Deer Park brand spring water, with approximately 50 specimens in each tub. *G. minus* specimens were fed with handful clumps of dead leaf litter collected from the Ward Spring locality.

Isolation of Genomic DNA

Genomic DNA was harvested from mixed populations of *G. minus* maintained in the laboratory for approximately two weeks. After two weeks, the animals were starved in four containers of thirty individuals were kept in a refrigerator held at 4˚ C for seven days, each approximately 1 cm deep with Deer Park spring water. No food was given during this time, but the water was changed after three days. This was done to reduce contaminants from the specimen's diet, and thereby prevent contamination of the DNA extraction. In each container there were approximately fifteen specimens from Ward spring, and fifteen specimens form The Hole Cave, in order to assure genomic DNA that was a blend between the two populations.

High molecular weight genomic DNA extractions were done on the seventh day of specimen starvation. Two extractions were performed, utilizing the two groups of thirty specimens from two of the four containers, again with an equal split of specimens from The Hole Cave and Ward Spring. Wet weights of the samples were each approximately 550 mg, after removal of trace amounts of water. Each group was then flash frozen with liquid nitrogen, and ground to a powder. Phenol extractions were then performed on each sample group, and resulting isolating genomic DNA was placed in the refrigerator for storage at 4˚ C.

Digestion of Genomic DNA, Ligation of Adaptor Oligonucleotides, and Size Selection

Genomic DNA digestions were performed utilizing the *MboI* endonuclease and corresponding appropriate buffer solution. The endonuclease was inactivated for 10 minutes at 65˚ C after digestions. The resulting digested genomic DNA was checked on a gel, revealing a characteristic smear.

Two oligonucleotides were utilized to create a sticky end adaptor, (5'-GGCCAGAGACCCCAAGCTTCG-3') and (PO_4- 5'GATCCGAAGCTTGGGGTC-TCTGGCC-3') (Bloor et. al., 2001), which were then ligated to the genomic DNA digests previously prepared. A Qiagen minElute kit was utilized to concentrate the resulting ligated genomic DNA to a volume of 20 μL. This volume was then loaded into a 1.5% agarose gel, and the region between 400 and 1000 bp was excised. A Qiagen minElute gel purification kit was then utilized to purify the gel extrations.

Captured of Microsatellite Containing DNA Fragments

To capture particular microsatellite containing genomic DNA fragments, a selective hybridization protocol was utilized via biotin capture with streptavidin-coated magnetic beads (Zane et. al., 2002). A protocol first used within the laboratory of Steve Kemp, and compiled by Paul Bloor, was followed (Bloor et. al., 2002). Dynal M2-80 streptavadin-coated dynabeads were utilized (size 280) in the procedure, along with a small Dynal magnet provided with the dynabead kit. 20 μL of the digested, adaptor-ligated, and size-selected DNA was required for microsatellite isolation. Initially, a $(CATA)_8$ probe was utilized for capture of microsatellite sequences containing CATA repeats, but later microsatellite isolation procedures also utilized the $(CAT)_8$ probe, the $(GT)_{12}$ probe, and the $(ACT)_8$ probe. Final rinsed samples were stored in the refrigerator at 4˚ C.

Transformation of Microsatellite Containing DNA into Vectors, Screening for Vector Insert, and Sequencing of Purified Microsatellite Containing DNA

From the rinsed samples of microsatellite containing DNA fragments, 2 μL of the solution was used in transformation into pGEM-T and pGEM-T easy bacterial vectors, using the pGEM-T cloning kit for each. JM109 competent cells from the pGEM-T cloning kit were used for the transformations. Competent cells were incubated overnight at 37˚ C on LB/Ampicillin plates. Two control plates were also incubated overnight at 37˚ C on LB/Ampicillin plates, one containing 10 μL of pUC19 vector with no vector insert, and one containing no plasmid. White colonies were harvested from the plates, and grown in overnight 3 mL liquid cultures. Qiagen MiniPrep kits were utilized on each liquid culture to purify the resulting DNA, with an elution volume of 20 μL of buffer AE. Cultures that utilized the pGEM-T kit were digested with SacI using 5 μL of the purified DNA to screen for insert, while cultures that utilized the pGEM-T easy kit were digested with EcoRI using 5 μL of the purified DNA to screen for insert. The digestions were then run out on a 1.5% agarose get to screen for insert.

Purified DNA product was sequenced based on presence of insert detected on the 1.5% agarose gel. Microsatellite containing DNA fragments isolated with the $(CATA)_8$ probe were sequenced directly using M13F/IRD-800

and T7/IRD-700 labeled sequencing primers (LI-COR Biosciences). The DNA was sequenced with the SequiTherm EXCEL II DNA sequencing kit (Epicentre Biotechnologies), run on a LI-COR 4300 DNA Analysis System automated DNA sequencer, with results read with the bundled eSeq software. Microsatellite containing DNA fragments isolated with the the $(CAT)_8$ probe, the $(GT)_{12}$ probe, and the $(ACT)_8$ probe were mailed-out to an outside company to be sequenced.

Analysis of DNA Sequences

A total of 39 vector clones were sequenced from a total of 238 potential vector clones. The clones sequenced via the LI-COR 4300 DNA Analysis System were analyzed utilizing the 700-channel. All sequences were manually edited to remove detectable vector DNA. The clones sequenced via the LI-COR 4300 DNA Analysis System were scanned for presence of adaptor sequences and pGEM-T vector sequences bordering the vector insert (5′-GCGGCCGCGGGAATTCGATT-3′) and (5′- GAATTCACTAGTGATT-3′). The clones that were mailed out were scanned for presence the EcoRI restriction sites of the pGEM-T easy vector sequence bordering the vector insert (5′-GAATTC-3′). For the full resulting sequences, see Appendix A.

Resulting DNA sequences were then analyzed via the Microsatellite repeats finder tool (BioPHP), specifying a minimum length of repeated sequences as 2, a maximum length of repeated sequences as 6, with a minimum of 3 repeats, a minimum of 6 tandem repeats, with a 0% allowance for mismatches. The sequences were also analyzed via the BLASTn tool to search GenBank for potential sequence matches. Results from The Microsatellite repeats finder tool are summarized in table 1, and BLASTn are summarized in table 2.

Results

Each of the 39 sequences was analyzed using the BioPHP tool Microsatellite repeat finder, and summarized in table 1. Of the 39 sequences, 8 were found to possess a repeat motif within the sequence. The clone (CATA)8-6 possessed the most uninterrupted repeat motifs, the motif TGTA, with 20 total repeats. This was followed by (GT)12-2, with 11 repeats of the motif CA. Of the 8 found to contain a repeat motif, two possessed motifs that matched the same original probe (the $(CATA)_8$ probe). Clones (CATA)8-2 and (CATA)8-11 each possessed repeats that matched the original probe. The clone (CATA)8-11 was found to have 5 different variations of the original probe's motif: CATATA, CATA, TACA, and CACATA. When analysis was done with the Microsatellite repeat finder tool to allow for 0.3% mismatches, there were 28 repeats, starting at the 281 bp position, with the sequence (5′-ATATACATACATACATATACATATAC-ATACATATACATACATATACATACATATACATACATATACATACATATA-CATACATATACATACATATACAT-ACATATACATACAT-3′). This allowance indicates that this sequence possesses the longest microsatellite repeat of all the sequenced fragments. Changing the mismatch allowance to 0.3% for every other sequence did not produce additional or changed repeat motifs for the remainder of the sequences analyzed (data not shown). Three sequences were also significantly shorter, specifically (CAT)8-12, (CAT)8-13, and (ACT)8-16.

Table 1 - Summary of clones selected for sequencing. Each sequence was searched individually for microsatellite repeat motifs. Below the repeat motif found is reported, along with the location of the repeat motif, and number of times the motif was repeated.

Clone	Sequence Length	Repeat Motif Found	Start of Repeat Motif	Number of Repeats
(CATA)8-2	571 bp	CATA	48 bp	6
(CATA)8-6	438 bp	TGTA	225 bp	20
(CATA)8-10	875 bp	-	-	-
(CATA)8-11	580 bp	CATATA	85 bp	3
(CATA)8-12	575 bp	-	-	-
(CATA)8-15	740 bp	-	-	-
(CATA)8-18	787 bp	-	-	-
(CATA)8-23	620 bp	-	-	-
(CATA)8-24	863 bp	GGAT	65 bp	8
(CATA)8-28	862 bp	GGAT	65 bp	8
(CATA)8-29	649 bp	-	-	-
(CATA)8-35	646 bp	-	-	-
(CATA)8-36	726 bp	ATAC	413 bp	4
(CATA)8-37	622 bp	-	-	-
(CATA)8-41	633 bp	-	-	-
(CATA)8-42	622 bp	-	-	-
(CAT)8-2	650 bp	-	-	-
(CAT)8-3	731 bp	-	-	-
(CAT)8-4	814 bp	-	-	-
(CAT)8-5	663 bp	-	-	-
(CAT)8-7	567 bp	-	-	-
(CAT)8-8	688 bp	-	-	-
(CAT)8-11	817 bp	-	-	-
(CAT)8-12	102 bp	-	-	-
(CAT)8-13	102 bp	-	-	-
(GT)12-1	721 bp	-	-	-
(GT)12-2	798 bp	CA	720 bp	11
(GT)12-3	674 bp	-	-	-
(GT)12-5	631 bp	-	-	-
(GT)12-6	548 bp	-	-	-
(GT)12-7	566 bp	-	-	-
(GT)12-9	705 bp	-	-	-
(GT)12-10	563 bp	-	-	-
(GT)12-11	1006 bp	-	-	-
(ACT)8-3	681 bp	-	-	-
(ACT)8-8	566 bp	GT	43 bp	3
(ACT)8-11	680 bp	-	-	-
(ACT)8-12	792 bp	-	-	-
(ACT)8-16	114 bp	-	-	-

Each of the 39 sequences was also analyzed using the BLASTn tool to scan GenBank for potential nucleotide matches, and summarized in table 2. 11 of the 39 sequences had greater than 50 significant matches according to their E-value. Additionally, Clone (CATA)8-6 had 31 significant results according to the results' E-values. The identify of the most significant result varied from sequence to sequence. The clones (CATA)8-23, (CATA)8-24, (CATA)8-28, (CATA)8-29, (CATA)8-35, (CATA)8-36, (CATA)8-37, (CATA)8-41, and (CATA)8-42 all had the majority of their sequence match existing pGEM cloning vectors (data not shown). The clones **(**CATA)8-2 and (CATA)8-10 had the most significant BLASTn match identified as a microsatellite sequence, but BLASTn only identified sequence concordance with the first 10 5' nucleotides and the last 10 3' nucleotides (data not shown). The clones (CAT)8-12 and (GT)12-2 both had at least 1 significant match, but neither identified as a microsatellite sequence or cloning vector, but clone (CATA)8-6 did identify as a sodium-potassium pump in its 31 significant results (data not shown).

Table 2 - Summary of clones selected for sequencing. Analysis of each sequence was also done via the BLASTn protocol to identify concordance with existing microsatellite sequences in GenBank. Reported is the number of BLASTn significant results, and the corresponding E-value and accession number of the most significant result.

Clone	Number of Significant BLASTn Results	E-Value of Most Significant BLASTn Result	Accession Number of Most Significant BLASTn Result
(CATA)8-2	>50	7.00E-39	EF089568.2
(CATA)8-6	31	2.00E-19	AC133579.10
(CATA)8-10	>50	6.00E-80	AM412177.1
(CATA)8-11	0	-	-
(CATA)8-12	0	-	-
(CATA)8-15	0	-	-
(CATA)8-18	0	-	-
(CATA)8-23	>50	0.0	HQ335169.1
(CATA)8-24	>50	0.0	GU370778.1
(CATA)8-28	>50	0.0	GU370778.1
(CATA)8-29	>50	0.0	X65309.2
(CATA)8-35	>50	0.0	X65309.2
(CATA)8-36	>50	2.00E-16	FN645959.1
(CATA)8-37	>50	0.0	X65309.2
(CATA)8-41	>50	0.0	X65309.2
(CATA)8-42	>50	0.0	X65309.2
(CAT)8-2	0	-	-
(CAT)8-3	0	-	-
(CAT)8-4	0	-	-
(CAT)8-5	0	-	-
(CAT)8-7	0	-	-
(CAT)8-8	0	-	-
(CAT)8-11	0	-	-

(CAT)8-12	1	5.00E-05	FJ465017.1
(CAT)8-13	0	-	-
(GT)12-1	0	-	-
(GT)12-2	2	2.00E-07	AL031291.3
(GT)12-3	0	-	-
(GT)12-5	0	-	-
(GT)12-6	0	-	-
(GT)12-7	0	-	-
(GT)12-9	0	-	-
(GT)12-10	0	-	-
(GT)12-11	0	-	-
(ACT)8-3	0	-	-
(ACT)8-8	0	-	-
(ACT)8-11	0	-	-
(ACT)8-12	0	-	-
(ACT)8-16	0	-	-

Discussion

The results gathered from the 39 sequences point towards the validity of considering microsatellite genetic markers to elucidate the population genetic relationships of the cave and surface populations of *Gammarus minus*. The presence of microsatellite sequences within these sequences provides evidence that while isolation of microsatellite sequences from *G.* minus is a challenge, it is similar in difficult to comparable species (Shaw et. al., 1999). Significant repeat motifs were found in clones (CATA)8-2, (CATA)8-6, (CATA)8-11, (GT)12-2, and (ACT)8-8. The greatest indication of this was the results seen from the $(CATA)_8$ probe clones, which contained the only microsatellite repeat motifs out of all the probes utilized during the study that matched the original sequence probed for. However, this was certainly not the case for the entire gamut of sequences that utilized the $(CATA)_8$ probe. For clones (CATA)8-6, and to an extent, (CATA)8-11, the repeat motif did not match the $(CATA)_8$ probe.

The clone (CATA)8-11 represents an interesting occurrence amongst the sequences, and hence prompted addition analysis allowing for a 0.3% mismatch when considering repeat motifs. This clone presented one of the most obvious cases of microsatellite repeats found during sequencing, since the sequences were visually apparent both on the sequencing gel (data not shown) and on the written sequence (see (CATA)8-11 in Appendix A). When analyzed allowing a 0.3% mismatch with its repeat motifs, it quickly grew to be the longest sequence of microsatellite repeat motifs, with 28 consecutive repeats, and it generally matched its corresponding probe. This radical change did not occur in any other sequence analyzed allowing for a 0.3% mismatch. A number of possibilities for this occurrence are possible. Considering the relatively older sequencing technology used to sequence the (CATA)8-11 DNA fragment, the sequence could simply be erroneous. Since the method of sequencing precludes quality scores on the sequence, it is difficult to ascertain how likely that certain base calls are correct, especially those within the large repeat motif.

Considering the data resulted in the capture of many microsatellite repeat motifs that were not probed for, it is a clear indication that the streptavidin-

coated bead step was one of the limiting factors in gathering microsatellite repeats that actually matched the probe used. The procedure used did not have any effective way to discern fragment composition immediately after isolation of microsatellite containing sequences via the magnetic beads, and instead results were only gathered once the fragments were cloned into the competent cells, making it difficult to judge how effective the magnetic-bead step was. Despite having a lower occurrence of microsatellite sequences matching the probe, the microsatellite sequences that were found provide important roots for further identification of *G. minus*'s microsatellite loci.

From the data gathered it is obvious that some of the sequences provided false positives for presence of true microsatellite repeats. For example, clones (CATA)8-24, (CATA)8-28, and (CATA)8-36 supposedly contain microsatellite repeat motifs within their sequence. However, when the sequences were compared to the GenBank database using the BLASTn protocol, these sequences were found to be highly concordant with existing pGEM vector sequence. This would be expected if no insert sequence was present in these sequences, and only the vector had been sequenced as a result. Therefore, these sequences can be considered false positives.

This is compared to the other sequences that contain microsatellite repeat motifs, whose BLASTn matches were primarily identified to have concordance with the first 10 5' nucleotides and the last 10 3' nucleotides, indicating that this sequence concordance was likely due to matching adaptor sequences rather than a greater match for the remainder of the sequence. However, the significance of (CATA)8-6's matches with sodium-potassium pumps remains difficult to ascertain. As with all of the $(CATA)_8$ probe sequences, it lacks a sequence quality score measure, so determining accuracy of the sequence is only possible if the fragment is sequenced. Finally, a great number of the clones sequenced contained no microsatellite repeat motifs, and did not have any significant BLASTn matches. These sequences are most likely the result of vectors containing fragments that did not contain microsatellite repeat sequences.

Conclusion

Beginning in the 1970s, the freshwater amphipod species *Gammarus minus* has served as a model for troglomorphic evolution in aquatic invertebrates. The unique populations of the Karst geography in Virginia and West Virginia contain two distinct morphological forms of the species, making possible morphological and genetic comparisons between populations within caves and the populations residing on the surface. Recent genetic studies done on these populations have revealed hydrological differentiation of populations, as opposed to habitation differentiation, and indicate that cave populations are more closely related to their neighboring above ground population rather than fellow cave populations. However, these genetic studies have not been able to elucidate the fine scale geographical structure of these populations at the underground-aboveground interface.

This study shows that use of microsatellite markers have the potential to serve as high-resolution genetic markers for characterizing the cave-surface interface. Here, 39 clones produced microsatellite repeat sequences for 5 of the clones, one of which containing 28 sequence repeats. Although screening for

microsatellite fragments, and thereby microsatellite repeat loci, did not yield high percentages of clones containing microsatellite motifs, this has been reported in previous studies as a typical result when screening for loci (Shaw et. al., 1999). With more of the 238 clones remaining to be sequenced, and more loci yet to be discovered, the results so far and the positive isolation of microsatellite sequences shows the promise of continued use of microsatellites in the *Gammarus minus* population in order to characterize population structure.

References

Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.

Avise, J.C. (2004). *Molecular Markers, Natural History, and Evolution*. Sunderland, Massachusetts: Sinauer Associates, Inc.

Bloor, P.A., Barker, F.S., Watts, P.C., Noyes, H.A., Kemp, S.J. (2001). *Microsatellite Libraries by Enrichment*. United Kingdom: Animal Genomics Laboratory, University of Liverpool.

Carlini, D.B., Manning, J., Sullivan, P.G., Fong, D.W. (2009). Molecular genetic variation and population structure in morphologically differentiated cave and surface populations of the freshwater amphipod *Gammarus minus*. *Molecular Ecology, (18)*, 1932-1945.

Culver DC (1976) The Evolution of Aquatic Cave Communities. *American Midland Naturalist*, **110**, 945-957.

Culver DC, Kane TC, Fong DW (1995) *Adaptation and Natural Selection in Caves: The Evolution of Gammarus minus*. Harvard University Press, Cambridge, Massachusetts.

Dieffenback, C.W., Lowe, T.M., & Dveksler, G.S. (1993). *Manual Supplement: General Concepts for PCR Primer Design*. Cold Spring Harbor: Cold Spring Harbor Laboratory.

Ellegren, H. (2004). Microsatellites: Simple Sequences with Complex Evolution. *Nature Reviews*, *(5)*, 435-445.
Goldstein, D.B., Pollock, D.D. (1997). Launching Microsatellites: A Review of Mutation Processes and Methods of Phylogenetic Inference. *Journal of Heredity, (88)*, 335-342.

Estoup A, Jarne P, Cornuet J (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology*, **11**, 1591-1604

Fong DW (1989) Morphological evolution of the amphipod *Gammarus minus* in caves: quantitative genetic analysis. *American Midland Naturalist*, **121**, 361-378.

Goldstein DB, Pollock, DD (1997) Launching Microsatellites: A Review of Mutation Processes and Methods of Phylogenetic Inference. *Journal of Heredity*, **88**, 335-342.

Jarne P, Lagoda PJL (1996) Microsatellites, from molecules to populations and back. *TREE*, **11**, 424-429.

Kane TC, Culver DC, Jones RT (1992) Genetic structure of morphologically differentiated populations of the amphipod *Gammarus minus*. *Evolution*, **46**, 353-365.

Mason, S.L et. al. (2005). Development and Use of Microsatellite Markers for Germplasm Characterization in Quinoa. *Crop Science, (45),* 1618-1630.

Pascoal S, Creer, S, Taylor MI, Queiroga H, Carvalho G, Mendo S (2009) Development and Application of Microsatellites in Carcinus maenas: Genetic Differentiation between Northern and Central Portuguese Populations. *PLoS ONE*, **4**, 1-6.

Shaw PW, Pierce GJ, Boyle PR (1999) Subtle population structuring within a highly vagile marine invertebrate, the veined squid *Loligo forbesi*, demonstrated with microsatellite DNA markers. *Molecular Ecology*, **8**, 407-417.

Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research*, **17**, 6463-6471.

Zane, L., Bargelloni, L., Patarnello, T. (2002). Strategies for microsatellite isolation: a review. *Molecular Ecology*, *(11),* 1-16.

Appendix A – DNA sequence data from sequenced clones (see table 1).
Sequences have been trimmed to remove flanking vector and adaptor sequences.

$(CATA)_8$-2

CGACCAGTAATTCCGCCAATATTCTGAAAAAGCGTATGTCCCAGTATTCACATACATACATACATACATAC
ATACTTACATACATACATACATAATCGCCTTGTCAAATGCCCTGTATCCTCTCTTCCCCAAGGCCTCAAGA
CCATCGTGGGCGCTGTGATAGAGTCCGTGAAGAATCTGAGGGACGTAATCATCCTCACCGTCTTCTCGCTG
TCCGTCTTTGCTCTTCTCGGCCTTCAGATTTACATGGGAGTACTCACTCAAAAATGCGCCAAGAAATTTGA
TTTTACCAAAGAGAACTGGACTCACGAAGAGTACTTCGTACATAACTCTAATGCCTGTGAGTATGGTACAC
CCTAGGGACACTTTTACTCCTCTACCTTCCAGGRGTTGTCACTTGAAGCTTCATATTCACGAAGTTCTTCC
TAATCATGCTATATATGGATTCATCTACACCTATCAATGCCGAGATTCTCATTTATTCTGATTCTTCCATA
TCAMTTTCATTGTCARGGGCGTGATATGGGAGRATMTTCCAGTCAWCCTTCTCAATATTTGTCTATAGCCT
MTT

$(CATA)_8$-6

GCCGCCATGSCCGGGGATTAACCGACGTTAAGTGCTTACGTTTTTGCACACGCTCCATTAGGCGCAATTTC
AAATGTGTTCTGCTGTTGTATTGAACTTCATCCCAGTGCTTAGATAATAGAGTACATTATTCAGGAGACAA
AAGTATTCCTGATATTTTTACCTCTAATGCATTGATGCCAAAAAAGCTCTAGGCTCACTTAGGTTTGTATT
GAGCAATACTAATATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTAT
GTATGTATGTATGTATGTATGTATATGTGTGTGTATATGTATGTTTATGTGTATATTATTTGCAGGTACTG
GTTGAGTATTGCGCGCTCCTTGTGTTGTTTTCATTGTGCTATATTCCAGGKCCTGCTATAATTTTTCAATT
TATTTATACACC

$(CATA)_8$-10

GGCCGCCATGGCCGGGGATTCAGGACCATATAATGTGAACTAAAGGTATGTTACTCGGTCAAGTGTTCGGA
AAACAGCGTTTTTAGAGCAATTCCCACGTTAGAATGAACATTAGTGCTGAGAACCTTAGAAGTGAACTTAA
AGTATGTTACCTGCTTAAGCATTCGGCAAACAGCGTATTTAGGGCAATAACCGCGTTTGAAGAGACACTAG
TGCTCAGGACCCTAAAATGAGAACTAATAATGTTTTACTCGAATAAGCGTTCGGCAAACGGCGTATTTTGT
GCAAAAACCGCGTTTTAAGAGTCACATGTAATCATCTCTTCGCTATTTGTTAACTGCTCTGAGTCGATGCC
AATAGAACACAGTGCGTTCGAAGCTTGGGGTCTCTGGCCAATCACTAGTGCGGCCGCCTGCAGGTCGACCA
TATGGGAGAGCTCCCAACGCGTTGGATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATAGCTTGGCGT
AATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCGTGTCWCACTCACACACAACAYACGCGGCGCG
CGCATATAGTGTAGAGNGTNGTGTSTCTAGTGAGWGAGMGAGAKCACAYACATAGCGTGCGNTNWNCNCCA
GCTCTACASWCGMGACACCTGTCGTGTCAGAGACACACACGCACNGNCCACGCGCGTAGAGAGNTNKNTNC
TATATGTGTSTGCGNTNCGTGWGATGATGTGGTGTGTSGTGGGGTRTWAGGAGTTAATGWAGAGATGAGVA
GCGCTAGGTTGTTTGTSCTAGTAATGTRWGTGAMTGTAKWGTCTGTTTGTCTGAGTGTTGTTTGAKGTTTG
TGTTGTGTTTDGTTRAATAATGT

$(CATA)_8$-11

TTCCTCACCTGTTCGTATTTATATTGCCTACACATATATACGTATACAACATATACATATATACATATACA
CATATTTACATATTCATATACATATACATATACATACAGACATACACATACAGACATATACATACAGACAT
ATACATACAGACACATACATACAGACATATACATACAGACATATACATACAGACATATACATACAGACATA
TACATACAGACATATACTGTACATACAGGCATATACTTACATACGTATACATACATACATATGCAAACATA
TACATACATACATATACATATACATACATATACATACATATACATACATATACATACATATACATACATAT
ACATACATATACATACATATACATACATATACATACATATWCATACATATWCATACATATWCATACATACA
TATACATACATACATTTWCATAAATATWCACATTTTTTTTTTTTTTTTTCTTAGGKSGCCCYGGGGSTCCT
CCTCCCCGTCTCTTTTCCCGCCTCCCMTTCTYCTTCCAGCCGCCTTCCCTCCGCGTTCACCTCGGCTGYAT
GWGTTTCTWGTC

$(CATA)_8$-12

TGTGAGAAATATCAGATAAATTCTTTCACCTGTTACTTTTCTCTCGCTCTATGATGGCAGTGTAAGAGTGG
AACAGACGCTTTATCAAATACATTAAAATACACCAAGACCAAGGGTTTGGATTCATCTGATTTGAAGCGCC
CTCTGTGCGATAAACCTAGGGATGTTAACCAATGAATTAATCTCTATTGTATTTCTATTGCAGAACAACAG
CTGAAATTAGCTGTTAAATAAAATAAAACGTTGTATCGTGGTCACTTATAAATATAAAGTAAGATATAGGC
TTTAAATTAATAAGGTATTAAGGTATATGGTGCTGCCACCTGGCAAACGGACCTAAATCGACGGACAAACT
CAAAATGAAGTCTTTTGCAAGAGCCTGCCCGTGCATCTTATTGTAGTTTATCGTCTCTGACATTGTGCTTC
AGCGAGACGCTTCGTTGCATCAGGCGACACMTTACTCTGCAGCGCGTATGATACTGTGACGGACACCATCC
GATAACAGGATCCGAAGCTTGGGGKCTTTTGGGCCAAATAGTGGGGSCGSCTGCAGGTGGACCATATGGGG
GGGSTCC

$(CATA)_8$-15

AGAACCATATAATGTGAACTGAAGGTATGTTACTCGGTCAAGTGTTCGGAAAACAGCGTTTTTAGAGCAAT
TCCCACGTTTGAATGACCATTAGTGCTGAGAACCCTAAAAGTGAACTTAAAGTATGTTACATGCTTAAGCA
TTCGGCAAACAGCGTATTTAGGGCAATAACCGCGTTTGAAGAGACACTAGTGCTCAGGACCCTAAAATGAG
AACTAATAATGTTTTACTCGAATAAGCGTTCGGCAAACGGCGCATTTTGTGCAAAAACCGCGTTTGAAGAG
TCACTTGTGATAAGGACCATATAATGTGCGCTAAAGGTATGTTACTCGGTCAACTGTTCGGAAAACGGCGT
TTTTAGAGCAATTCCCACGTTAGGATGAACATTAGTGCTGAGAACCTTAAAGTGAACTTAAAGTATGTTAC
ATGCTTAAGCTTTCGGCAAACAGCGTATTTAGGGCAATAACCSCGTTTGAAGAGACACTAGTGCTCAGGAC
CCTAAAATGAGAACTAATAATGTTTTACTTGAATAAGCGTTCTGCAAACGGCGTATTTCGTGCAAAAACCG
CGTTTTAAGAGTCACTTGTGATCAGGACCATATAATGTGCACTAAAGGTATGTTACTCGGTCAAGTGTTCG
GAAAACAGCTTTTTAGAGCAATTCCAAAGTTAGRATGAACATTAGCGCTAAGRRCCTTAAAAGTGGACTGK
AAGTATGTTACCTGCTTTAGCATTCGGCAA

$(CATA)_8$-18

GCCGCCATGGCCCGGGATTCCTGCCTGGTCATTGAGCTATTCGGGTAGATGGTCCTTGGTTGGTTGTCGTC
CCCTAGAAGTACGACCGTTGACGGGTCAATCGTTGGTAACGATGTAAAGGCAAGTTGACGCTGCGGTCGGA
GCGCGGGACAGTCCGTGAGAATATGCTTCGTTGTTGGAGGTACGTCATAGCACCSAGGGCAGGGGGTGAGG
GCTGTCTCACGGGCCCCATCAAATGTACATGTGTTAGTCGTGTGTAGCCAAGTCTCAGTCGGTTGATGATG
AACCCCTCGCGTCGTGTTAATTTCAGTTTTCTCCAGTTACCAGGGGTATTTTTKATAGAGAACAGGTGGCT
GTTTCCGCTGGCTCACTGTTCACCCTAACGGTTGTACAAAGCGTTCCTTATAACAGGAAACCAATCAGTGT
ATGGACAGCAGATACATTTTGGATTTCCAGACGCTGCTCACTTTGCGGCCTTAACGACCACTTCATTGTTT
GGTATTTTCCTTTGTCCCGGCACCCACAAAATTACAATATCGCGGGGTAGCGAGGTAAGTTGTAAATCGCG
TCTAGAGTCGTGCCATCAATGGATTTTKAGGGCGCACTGTTAAGAGGCCCTGTGTTGCACTCAGAGAGTCT
GTGCATGTAAGATAACTTGTACTGGAGTCTGTGTGGKGGCAAAAGATGAGACTGCATGCTATTATCAAAGC
ATGGAGTTCATCTGTATATACCSMTGTAATGCGAGGGGAGTGTTTCCGATTTKTTACTAGCTCCCMCGRAA
ACTGCT

$(CATA)_8$-23

GCCGCCATGGCCGGGAGAGCTCCCAACGCGTTGGATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATA
GCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCKTBKSACARTTCCACACAACATA
CGAGCCGGNAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGC
GCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGG
AGAGGCGGTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTG
CGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAA
GAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGGACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATA
GGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGSTCAAGTCAGAGGTGGCGAAACCCSACAGGACTAT
AAAGATACCAGGGGTTTCCCCCTGGRAGCTCCCTCGTGCGCTCTCCTGTTCC

$(CATA)_8$-24

AAGACGCTCAGACGGAGAAGGGGTTTCTAAAGTGGCCGCGAAATGGCTCAATAAACGCTGCACTCGGATGG
ATGGATGGATGGATGGATGGATGGATGAATCACTAGTGCGGCCCCTGCAGGTCGACCATATGGGAGAGCTC
CCAACGCGTTGGATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATA
GCTGTTTCCTGTGTGAAATTGTTATCGGGTCACAATTCACACAACATACGAGCCGGAAGCATAAAGTGTAA
AGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGG
GAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGC
TCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTC
AAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAGGCCAGCA
AAAGGCCAGGAMCSTAAAAAGGCCGCGTTGCTGGCCGTTTTTTCATATGCTCCGCCCCCTGACGAGCATCA
CAAAAATCGACGCTCAAGTCAGAGKGGCGAACCCGACAGATCTATAAAAGATAACMWGGCSTTTCCCCTGG
MAGCTCCTCGTGCGTGTGCYKTTCGMCCTGCSTTACCGATACTGTCGCTTTTCTTGGAGGTGSTTTTATGT
AGTKAGHTTAGTWGAGWAGTGTGTCAGTGGTTTGAGACCGTAGCGCCTGTWTGTATTGTTGYACGTAGAGT
TGATGGGMTGW

$(CATA)_8$-28

AAGACGCTCAGACGGAGAAGGGGTTTCTAAAGTGGCCGCGAAATGGCTCAATAAACGCTGCACTCGGATGG
ATGGATGGATGGATGGATGGATGGATGAATCACTAGTGCGGCCGCCTGCAGGTCGACCATATGGGAGAGCT
CCCAACGCGTTGGATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCAT
AGCTGTTTCCTGTGTGAAATTGTTATCGRGKGACAATBCACACAACATACGAGCCGGAAGCATAAAGTGTA
AAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCG
GGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCG
CTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACT
CAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAGGCCAGC
AAAAGGCCAGGAMCSTAAAAGGCCGCGTTGCTGGCGTTTTTTCATAGGCTCCSCCCCCTGACGAGCATAAM
AAAAATCGACGCTCAAGTSAGAGGTGGCGGAACCCACAGACTATATAGATAKCAGGSGTTCCTGRGTCTST
GGTTCTGTTCACTCCTACGTAYTTCYTTTCTKGAGGTGGTTTAWGTAGTGTAGTTAGTGGTGGGTGTAGTG
TKTGAGACCGTACGASTGGTWTGTATTGTTGTACGTAGAGTTGATGGGMTGTAGTGGGTGGTAGTGGTAGA
WGARTGTGTK

$(CATA)_8$-29

GGCCGCCATGGCCGGGGATAATYACTAGTGCGGCCGCCTGCAGGTCGACCATATGGGAGAGCTCCCAACGC
GTTGGATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTT
CCTGTGTGAAATTTTTATTKKKTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG
GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACC
TGTCGTGCCAGCTGCATTAATGAATCGGCCAACGSGCGGGGAGAGGSGGTTTGCGTATTGGGCGCTCTTCC
GCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGC
GGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGG
CCAGGGACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAA
ATCGACGCTCAAGTCAGAGGKGGCGAAACCCGACAGGACTATAAAGATACCAGGGGTTTCCCCCTGGGAGS
TCCCTCGTGC

$(CATA)_8$-35

GGCCGCCATGGCCGGGGATATCACTAGTGCGGCCGCCTGCAGGTCGACCATATGGGAGAGCTCCCAACGCG
TTGGATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTC
CTGTGTGAAATTGTTATSGCGKSACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGG
GGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCT
GTCGTGCCAGCTGCATTAATGAATCGGCCAAACGSGSGGGGAGAGGSGGTTTGCGTATTGGGCGCTCTTCC
GCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGC
GGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGG
CCAGGGACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAA
AATCGACGCTCAAGTCAGAGGTGGCGAAACCCSACAGRACTATAAAGATACCAGGSGTTTCCCCCTGGGAG
CTCCCTC

$(CATA)_8$-36

AGCCACCTTCTGGGAATATTGCAACTACGCACCTGATATTTAAGTGACTTGGTTAGTTCGCTCACCCCAGA
AAAGATACGTCACATATTTACATGAAGTACATCTTGKACAAGTATCCAAGAGCAAGTGAGTTCAAATTCTG
CGTGGCTCCACGTTTGCTTTTCTGAAAAGAACACCCAGATATTTACACAAAGTACATTCTGACAAGGATAA
GACAACAAAGAAAGCAACTATGCACGGACATACCTTCAATTACTGCATAAACACAAACTCTCGAGTGACCG
GTGTATAAATAAATTAAAAAAGTTATAGCAGGACCTGCAATATGGCACCGTGACAACACACAAGGAGCGCG
CAATACTAAAACAGCACCTGCAAGTAATATACACATATACATACATATATACATACATATACATACATACA
TACATATATATACGTACATACATACGCACATACATACATACATACACATACATWTWCACGTACATACGSRC
CACATACATWCATTTTTTCWTTGGCACACAGGGGTACCCCGSGGGGGCTTTTGCTGTTYCGCACTCACAGG
GRTCCGGGGSTTGGGGKTTTTKGGGSCCACTTGTGGGGGCGGCTGCAGGTGGGCCSTYTGGGGGGGGTSCC
TCCGMGTTGGGTGGATWGCTTGKGTGTTTTTTTGTGTGACCTAAATAGCTTGGGGTGRTMATGGKSATAGS
TGKKTTCTGTGTGGGA

$(CATA)_8$-37

CGGCCGCCATGGCCGGGGATAATCACTAGTGCGGCCGCCTGCAGGTCGACCATATGGGAGAGCTCCCAACG
CGTTGGATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTT
TCTKKTGAAATTGTTATKSSGGGACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGG
GGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCSCTTTCCAGTCGGGAAACCT
GTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCC
CTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCG
GTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGC
CAGGGACCCTAAAAAGGCCSCGTTGCTGGCGTTTTTCCATAGGCTCCCCCCCCCTGACGAGCATCACAAAA
ATYGACGCTYAAGTCAGAGGTGGCGAAACCCSACAGGACTATAAAGATACCAGG

$(CATA)_8$-41

GGCCGCCATGGCCGGGGATTATCACTAGTGCGGCCGCCTGCAGGTCGACCATATGGGAGAGCTCCCAACGC
GTTGGATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTT
CTGTGTGAAATTGTTATSGGGKRACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGG
GGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCT
GTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCG
CTTCCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGC
GGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGG
CCAGGGACCSTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGSTCCGCCCCCCTGACGRGCATCACAAA
AATCGACGCTCAAGTCAGAGGTGGCGAAACCCSACAGGACTATAAAGATACCAGGCGTTTCCCCC

$(CATA)_8$-42

CGGCCGCCATGGCCGGGGATAATCACTAGTGCGGCCGCCTGCAGGTCGACCATATGGGAGAGCTCCCAACG
CGTTGGATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTT
TCTKKTGAAATTGTTATKSSGGGACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGG
GGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCSCTTTCCAGTCGGGAAACCT
GTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCC
CTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCG
GTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGC
CAGGGACCCTAAAAAGGCCSCGTTGCTGGCGTTTTTCCATAGGCTCCCCCCCCCTGACGAGCATCACAAAA
ATYGACGCTYAAGTCAGAGGTGGCGAAACCCSACAGGACTATAAAGATACCAGG

$(CAT)_8$-2

GCAGAGACCCCAAGCTTCGGATCCATGTTTGCGTCTCTTACCTCTCTACGGGTCTCTCACTATGTTATATT
TATCTGACAGTTGTATGTCTTTCTTGCACTTGCATTTCACTAGAATCCACACATAATCCTCATACCCGTTC
ATACATATGTCTCTTAGGCTAAAATGTAATTATGGATACATTTATATTTTATTCTAATAATTTCATCAAAA
AAACGGAGTCTCATCGCTTCGGTCACTGAATAAAATTTACAGATGTGATGTCATGATATGAACAAACCTAT
TTATTTTTTCATAATAATCAGTTGATATTCTCTCATGCGATTGCATTAATATAACTGACTTAAAACACAAA
TTTTTCAATACCCAGTTATATTCTTTTTGACGTTACGCTAGATTCAGAGTACGATGAAACATATTCTGGCA
GTAACCATTTCAATAAAAGTGAAACCCAAAGAAATAATGTTCAACGACCCTGCTAAGATGAGCTCTCCATT
CACTGCCACACAGGTGGCTTACGGCATTTTGTCCACGAAACAATATACCTAACTTGAGATGTGAATTCATT
TCATAGGATTAAACATATATTTTTTGCGAAAGTAACGACAATACAATTAAGAATTCGTGATCCGAAGCTTG
GGGTCTCTGGA

$(CAT)_8$-3

CAAGCTTCGGATCACTTCATGATGGTGACAATACCATCAAAATGCTGCTGATGATGGCATGATGGTGGCAG
TGTCAACAATATGCTGCTGATGGTGGCAACTGTACCAACGACTTGCTGCTGATCATCTCTCGTAAACTATC
GTGTAGTTTACTGGTTCGATTTAGATTGACAACCGGTAATAGGTGGCATCCAGTATAACTGATTTTTTTGA
ATCAAACGTTTCTATAAATAGCGTTTGTTTTGCTGGCCACACGACGTCCTTGGTTTCAATGTCGCAAGAGA
GGGGAAATTTGTGTAAGCCTATCAGGTTATCTTTCACTCGACTTTCAGGACAGTCTAACGTCACTTTAGTC
TGTTCGGAAAAATAGACAAGTCGATGTTGGGTGGTTTCTGTTACAAAGGTCCCGGATATTTGTCCTGCCTT
CTTGTAATTACATTACTCTATCATTGCCGTACTGTTCTTCTGGAAACTTTCTCACTCACAAGTAGATTTTG
AGGTGTCATATATAGGCTCTACTTCCAAAAGTACATGATTACTAGGTGTAATCGAATACACATTACTTTTG
TCCGTAATTATATATGAAATGTAATCTTTCAAGATAGTGTTTCTTAATTCTGGGACTACTAGAGATACAGA
ATCCAATCGCAAAGGATGTGGAGTAATGCTGTATACTTCATAGTTTTGTGGATAGGTTAGAGGAATGATCC
GAAGCTTGGGGTCTCTGGCCA

$(CAT)_8$-4

GGCCAGAGACCCCAAGCTTCGGATCCACAGGCGACTGCAAATGCCTTCAATAATTATTTCAATAGTATTCC
TTGTACCCTAAGTGACTCTATTGTAAACAACGACATACCTTTTGGAAATTACTTAGAGAATGCAATCCCCG
AACAGCCAATTTTTAACCCTACGGACCAGAACGAAGTAAAATCAATCATAAATGGTTTAAAAAATTGTAGT
AGCACTGGCTGGGATAACATCCCGACCAATGTAATTTAGGCGGTTTCGAGTATTATTGTTGCTCCTTTGGC
TCGACTCGTCAATAAATCCCTCCAGCAAGGTATTTTTCCAGATGCACTCAAAATAGCTAAAATTAGTCCAA
TTTATACAAGTAGAAGCAAAATGGATGTACAAAATTATCGCCCAATATCAGTGCTGCCCGTCCTAAGCAAA
GTGTATGAAAAAGCATTTTATGCCAGATTAAATACCCATTTTGAATCTAACAACATTCTCAGTAAAAATCA
ATATGGCTTCCAGACTGGCAGAAATACCGAATATGCTTTGCTAAAATTTACTGGAGATGTTATTGATGGGT
TTGATGCTGGGGACGTCACTGTTGCAACGTTTATGGACCTCAGCAAAGCTTTCGATTGTGTGAACCATGAA
ATTTTGCTAGCGAAATTACAGCATTATGGAGTTAGCCCTGTTGGCCAAAACTGGATAAAGAGTTATCTTCT
TGATAGAAGACACTTGGTGAGTTGGAACAAAACCTCATCAACAACAAGTACGCTTAATATAGGAGTACCGC
AAGGGTCGATCCGAAGCTTGGGGTCTCTGGCCA

$(CAT)_8$-5

GGCCAGAGACCCCAAGCTTCGGATCGGAATCGGCTAAACAGAACGGTGTGAGAGTATTAGATAGAAACAAA
AATGAAATTTCGACTCTGTAAAGACGAAAACTGTATCCGATTTAGAACAAAATATTTATGTTCAATCAAAA
AAATGCAGGAAATTGTTGAATATACTCCATCCAACTCGTTTCCAAAAAGTTATCCTATAAATCCCAATCAC
ATTACCTACATATTAGGGCAGGATTTGCCAAGCGCACACTTGAAGGACAACTTTGACGTTCCCGTCCACTA
TCGCATACAATCCCGCGTTGTTCCCTGCTTTGTTGCCTTACCACACAACGATGACATGCGATGGAAAATAT
GTTTGTTCATAGTCAATTCATCGGCAGGTTTGGTTTCCATATGAATTTGCGTCTGACACATATTTTGCCAC
CGGAATAAAATATAATACAGTTATCTAAAGAAACACCGGAATCCACGATTTTTTTCTACTTCTCCCATCTA
CAAAACTTCCTTGCTCGCTCCAATGGAGCATCGGTGGCAGGCAAGCATGGGACGTCACTGCCAAGAAATGT
TGACCCGTAAGCCTCACAACCAGGGAGCATTTATCCTTCTGCTCCTGAAGCGGAGCTAATAAACCACACCA
ACTGATCCGAAGCTTGGGGTCTCT

$(CAT)_8$-7

GGCCAGAGACCCCAAGCTTCGGATCCCTCCATGTTCAGTGCAGAACTTCTGTCACGACGTCAGAAAAACGT
ACAGAAGGTGTTGCATCGGTTCCAAAGCAGGTTCCGAAGGGAGTATCTACTTATTCTCTGGCATGGTGCCG
TGGAAGCTAGGGGAATAGTTGCTCCCAAGGTGAGTGACACCGTTCAGATACATGGAATCTTGGCCGAATCA
TGGAACTTCACGAAGGTAAAGGCGGTCATGTCAGGAGTGTGAACCTCAAAACTCATGGAGGAGCAGTAAGT
CGACCAGTAACTCTGCTCTATCCGTTGGAGGTGCGAGTTGGGGAACCCGAAGAGAGGGTCCTGGAGCTGTC
TGGTCCAGACCTGGCACCATCGCGCCAGGAAGGTTGCGCGGACGATACTAAATCGTACCCCGTTGGTACCG
GCAAAGAAGCCGACCCCGTTCCAACAGCTGTAGGTGGGCTGGATGCCTCCAGACCCAAGAGGTTGGCGGCG
TTAGAAGCTCGTAGGCGGTTCATCACCCAATTACAATAGGTGATGATCCGAAGCTTGGGGTCTCTGGCCA

$(CAT)_8$-8

CTCGCTGCGTCCACCGTAATACAACTTCTTCGAAGCTAAGAGCCACGTCTCGTAACTGAAACTCAGCGTCC
TCCACTCAGTGAATACCTTCCCGACCTTAACAAAACGCCCCTACATCATGCATAAGCGTAACGTCTTGCCT
TGCATCAGGCTAACGTCTCAACCTCAGGGAAACGTTTCAACCTGACGGAAACGTTTCCCTATCAGTGAAAA
GTCTTACTGTCAGGGTAACGTCTCATGCTCAGCGTCTAACGTCTCATTCTCTATGTAACGTCTCGCCTTTA
ACATACCGTCTACCATATGGGTAACGTCTCACCGTGTGTGTGTTCTTCCCTTGCAGGTGGGCACCGGAGGA
GTCCTCACCTTCCCTTCTCCCCAAACTCGGCCCCTTTGGAGCAAAGAAAAATGGCGCGGCTCAGATTAATC
ACGGCGAGAGTAAGGCAAGTGCAGGAATTAATCTTGAGGGTAGTATAAAAGGAGGAGAGGAAGCGAGGCTT
CAAGAGTCATGAAGAGATGACAGTTTTTACTTCCATTGCTCATCATATCGTCTGTTTTGTGATTCATGTTC
AAGTGTTGCGAGTCTTTTTCATGTCCTCCTCCGGGTTAGTTTAAAGTTGTTCATATATTCTAAGGAACATT
GTTCAGGTACAACAGCGAGTAGCGATCCGAAGCTTGGGGTCTCTGGCCA

$(CAT)_8$-11

GGCCAGAGACCCCAAGCTTCGGATCCCCCGTGGCGAGGCAAAATGTGCCGTCCTTTAGCCCCAGGAGCGCA
CACCAGCCGCACGAAGTGGTGTCCTAAAATGGATTACAATTGTTGCAGGGATTTCTGCGGTGTTTGGCCTA
AAAATCAGGGCAGAAAAGAATGCATCCGTTGCTCTCTCGGATTTTAGTGGAATAAAGTGGTCGACAGGTTG
CGGGTATCTACATGATGTATATTCTACTGTAAATGCCTTATATGCTGCGTGTAGTTGATGTATGTTGTCTC
ATCAGTATGTTATTTGTATACTCAATATACGGTTCTTTATGTTGGTAGGTATGAGCATTTTCATATGTTGC
ATGTAATTTACATGTGAATTATTTTGAATACTTGACGCGTAATGTGCACTCCTCCTCACGGTATGAAAACT
TTGTAAGTTACCTGTAAACATTTAACTGTAAACTGAATAACTATGATAATGGTCGTTTATACGTGCTGCAC
GTCAGAAAATGGGATAAGCGCAGAGATTTGAAAACCACGTCTTCTTTAAATTCAGCATTTAAAAGCCTCGT
AAATCTTTGTTAAAGGGCCACTGTCGCGTAATAATTTGTTCTTGTATTTACCGAAGTGTATATGTATAAAA
ATATGTAAGTAATACTAAACAAATCTATTGAGCCGTTACTGGGATACAGCGATTTGTCTGTTAGAATTTCC
ACTTGACTTTCTTAGGAAACAACGCTTGTGACGTCACGAAAATAGAACTTTGAGAGGTGATATTACGAAAT
CAGTTTGTCCGATCCGAAGCTTGGGGTCTCTGGCCA

$(CAT)_8$-12

GGCCAGAGACCCCAAGCTTCGGATCTGTTTTCGTTATAGAGCTTATACCCCAGCAGAGCTGTTCTCGTCAC
AGAGCTCCGCATCCTGTTGAACTATTCTCGT

$(CAT)_8$-13

GGCCACAGACCCCAAGCTTCGGATCTGTTTTCGTTATAGAGCTTATACCCCAGCAGAGCTGTTCTCGTCAC
AGAGCTCCGCATCCTGTTGAACTATTCTCGT

$(GT)_{12}$-1

GGCCAGAGACCCCANGCTTCGGATCCGTATCATTTGTGCATTGTATTGTGTCTTTAACGTTGTCTACTTGT
CATGGTACCTAAATAAATTTTAAATCCTTAAATCCTCTATCCTAAAACGGGTTTGAGATGCAGAACAGACT
CAGGACCTAGAACGGGCCAGATATGCAGAACAGTCAGTCTCAGGGACCTAGAACGGGGCAGAGATGCAGAA
CAACCTCAGAGACCTAGAACGGGCCAGAGATGCAGAACAGCCTCAGAGACCTAGAACGAGCCAGAGATGCA
GAACAGCTTCATAGACCTCGAACGGGACAGAGATGCAGTACAGCCTCAGAGACATTGAATGGTCCATAGAT
GCAGAACAGCCTCAGAGACCTAGAACGGGCCAGAGATGCAGAACAGAGTAAGAGACCTAGAACGAGCCGGA
GATGCAGAACAGCTTCATAGACCTAGAACGGGTCAGAGATGCAGAACAGCCTCATAGACCTAGAACGGGCC
AGAGATGCAGAACAGTCTCAGAGACATGGAACGGGTCAGAGATGCAAAACAGCCTCAGAGTCCAAAAACGG
TCCAGAGATGCAGAACAGCCTCAGATACCTAGAACGGTCCACATATGAAGAACAGCCTCAGCTAGAGACCT
AGAAAGGTCCAGAGATGCATATCAGCATCAGTGACCTAGAATGGGACAGAGCTGCAGATCCGAAGCTTGGG
GTCTCTGGCCA

$(GT)_{12}$-2

GGCCAGAGACCCCAAGCTTCGGATCCTCCAAGAGAAATCTCTCAAACTGTATTCATGTATATGACCCACAT
GGTCTTTTAGACAATTTAACTTCATTTAGAATTACATTGAGCCTTCATACAGAATAGTTAGCTATCACAAG
GAGGCAATTTGTTTGTCCAATGCTCCACAAATAAAGAACTAAAATATAATTTGAATGATGACAGTAAAGCA
CCTTTAAATTAATCTCATATGACCTCAATTAGCATTGGTCTTATCCATCTCTCACCTTTGTCTCTCACCTT
TGCAGCCCCCCCAGCAGCACCGCAGACAGTGAGGAGCCACAGAGCTGGGGGAGGCCAGCACAGCCATGCGC
GCCAGTATAATATTGAGAACAGCAACGTAAACACAGGTTTAGGAACTGGAATTAGTGCGCACTGTTGGCTA
TCATTTAACATACACAACCTATGACACGAAAAAAAGAAAAAAAAAAAAAGAAAAAGGAGCATAACCCCCTG
TACACACAGATATGCACATTGTGCCACTTACAACACCCTCAGACGCACCCACACTGCAACACACACTTCAT
GTCTACACCACAGAATCACCCAGACACACAGTACAGGCTCAACACGAACAACAGGACAATCCAAACATGGC
ACACCACACAGCGAATCACAATCCACTGTACAACAACAGAGTTGCACACTGTGAAACAAGCGCAACACATC
CAGAAAACACCACACACACACACACACACACAAGACACTGCAACCATGACAGCCCCACAAGACACCCACGN
NNGGCGGAGCATCCGGA

$(GT)_{12}$-3

GGCCAGAGACCCCAAGCTTCGGATCCCCTCGGTAGGTCAGGAGTCCGTCGGAGTCCGTCGGGATCATCGGT
TACTTTGATTTATCGATGGTTCATGCATTTTTGTTTTGTGAAAAAAATCAGTTAAAACCAACAACCTTAAC
ATCGCGGAGCTAAAAGTATAGTTTGGTATAGTAAATAGTCATGGGACGGGCTGAAAAGTCCATTCAAGTAA
TATTTCAAGTGTACAGTAGTACTAGCCCAGCCTAGATGTGCTGCCTGGTTTGAGGGGGCTGCCAGCCGCTC
TGTTTCTACCTGTCTAACTCAGTAGAGCGATGGATTGGTCTCTCAGTGCTTGTCGAGCTGGTTCATGAAGG
TGTCGATGTTGTTACTATTCACCACATCATGTGGAAAACGGTTCCAGATTGTTGGTATCTTGACTGTAAAT
AAGTTTTGAGTAACTGATATTCTGAAGGGTTTCAGCTGAAGTTTTTGTCCATTATATTTATTCATTTGATT
CTAATCTCTATCAAAGAGACCTTCGGGGCTGGCTTTAGTAAACCCATTCAGATATTTAATGGTGTCTATGA
GCTGACTCCTAAGTCCTCTTCTTTCTAGTGATACAAGATTCAGTTTTTCAAGTTATTGGGGTCTTGGTTCT
TAGTTCTGTGATCCGAAGCTTGGGGTCTCTGGCCA

$(GT)_{12}$-5

GGCCAGAGACCCCAAGCTTCGGATCAACATTCGACTTACGAACTCAAATAAAGGATTCAGCAAGAGCCTGC
GCCTCTTGCTGTAATTTTTTGTCTTTGGTCATACCAATAAGCAAAACTTGCTCATTGACTGTTTTAACCTA
TTCCATCATTTTTTTGTCGTCTTGTTCTTTATGTTCAATATATATTAGCTCATTCTGCCGATTCGAAGTTT
TTTTACAAACTTTTTTTCTTCCTTCTATTTAACGAAAAAAATATCAGGTCGCCATTGACAAAGCTAGACTC
ATTATGCGCGATGTAGCTTGTGCACGCGGAGCGACTTTAATTTGAGGATGTGCCAAAGATTTCTGTTAACG
GAAAATATATTTTGGTCAGGGATATTGAGACTGGAGAAATAGAAGAGTTGTTCCCTTCAATTTTTCCATCA
ATTTGAGTTTTATCTTATTTTATTGCCATCACGTGAGTTTTGAATCCTAGTGGGTGGTACAATCCCCCCCC
CCCCTCACCCGGCCAGTGCAAGGCGACCTCACTTATGGTGGACATAAATGCATGTATTTTCATTTGGCACC
CGCTTACCAAATTCACTTATTTGGCATCCCCTTAGGCGATCCGAAGCTTGGGGTCTCTGGCCA

$(GT)_{12}$-6

GGCCAGAGACCCCAAGCTTCGGATCATAAGGCGAAATTACCCCTTAATTTGTATCTTTAGCTGGGTAGTGA
AGAATTTCCGTACGCAAATCGTGGGATTATCTCTGCATGCCGTTAAACAATCGTGGGATTAACTCTGCCGT
ACGACAATCGTGGGATTAACTCTGCCGTACGACAATCGTGGGATTATCTCTGCATGCCATTAAACAATCGT
GGAATTAACTCTGCCGTAAAACAATCGTGGGATTATCTCTGCATGCCGTTAAACAATCGTGGAATTAACTC
TGCCGTAAGACAATCGTGGGATTATCTCTGCATGCCGTTAAACAATCCTGAGATTAACTCTGCCGTACGAC
AATCGTGGGATTATCTCTGCATGCCGTCAAACAATCGTGGAATTAACTCTGCCGTACGACAATCATGGGAT
TTCCCTGAACCTAGATATGTGGTGAGTCCATACATCGTCCGTGTTCGATAAAACAAAATATAATGTTGTAT
CAATAAGTTATAACCATCCCGGTGGCGAGGGAGTGCAAACCGGTGTTAACC

$(GT)_{12}$-7

GGCCAGAGACCCCAAGCTTCGGATCGTAGTCTCTTTTATTGAAGTACCGTTCCATATATGTCGCGAATACG
TCCTTACAACCGTCTATTTCGATGAAACCCAGATAGAGGTTCATGTTCACGTTGACCTCTTTCCAGGCGTC
AATTTGTAACGCCTCAATGTCTTGCTTAGGACTGTACCCAATTTGGTGCTTCCATGATTGACTGACAGCAT
GTATCAAGCAGTGGCCATTTTTGGGAGTTGTCATTAGCTGTAATTTCAGATTGTTTAGAACACTATTGATT
TTATTTTCCCCTTGGTTTGAGTGATTTAAATTACCTGTCCCCGTACTTGATTGTAAACATCCTTCCTTAGG
TAAAATCAGATGTCTGCTTGCGGCTTGTGCAGTGGCAACGCTACCCTGACTAAGAACGTGGCTGTAAATAC
GTCCGCTTATGACCTGATTTCTGACGGAGGGGTTACTGGGCGTATTTTGTAGGTTGATGTTGCAACTGAGT
AGTTTTTCAGCAGGAGGTGTCGCCAAATGAGTCTTTGAATTGGCTTGCGAAGCTTGGGGTCTCTGGCCA

$(GT)_{12}$-9

GGCCAGAGACCCCAAGCTTCGGATCAAGCGCTAGCACTTTGTGACTGGTAGACGAATCTGCTCTAATTCTT
TGTCGTGCTCCGTTGCTAAGTCACTGAGTGTTTCATCGCTGTATAATTCATCTGTTCATGCATCTTTTGAA
TCTGGTTTTGCACAACCACCACACTTCGGATGTGGTTTTCGTCAAGTTCCAGATTCGTATTGAGAGCATCT
AAATGTTCTGGAGTGTTGAAGGGGTCCGCACTGCAACTCTGCTCCAACTTTGGAATGAGTGAAGTGATAAA
TCCCTTCCTCGTTCCCGAACTACGTGGAAGGTGTCTTGCCTCTTCGACCACATCTGGCGACAACTTGAAGG
TCTCTTACTATTTACCTTATCAGCAGCGTAGATGACATAGTTAACTTATTATAGTATATTAAGTAGGCATG
AAACGCACTAAAAAGTCCATTCAACTAATACTTCAAGTTTGCAACAGTACTAGCCCAGCCTAGATGTGCTG
CCAGGTTTGAGGGCTACCAGCCGCTGCAATGCTTCAACGGAGGTTGCCTCCTTGTGCGGCTGCTGTGTTTC
CACCGTTTTAACTCAGTAGAGCGGTGGATTGGACTTCCAGTGCTGGTCAAGCCTGTTTTTAGAGGTATTGA
TGTTGTTATTATTTACCACATCATGTTGTAAGCAGTTCCAGATCCGAAGCTTGGGGTCTCTGGCCA

$(GT)_{12}$-10

GGCCAGAGACCCCAAGCTTCGGATCCTAGTCTCTTTTATCGAAGTATCGTTCCATATATGTCAGGAATACG
TCCTTACAACCGTCTATTTCGATGAAACCCAGATAGAGGTTCATGTTGATGTTGACCTCTTTCCTGGCGTC
AATTAGTAACGCCTTAATGTCTTGCTTAAACCGTGCCTAAATTGGTGCTTCCATGTTTGACTGACAGCATG
TGTCAAGCACTGGCCATCTTTGGGAGTTGTCATTAGCTGTAATTTCAAATTGTTTAGAACCCTATTGATTT
TATTTTCCCCTTGGTTCGAGTGATATAAATGACCTGTCCCCGTACTTGATTGTAAACGCCCTTCCTTAGGT
GAAATCAGATGTCTGCTCGCGGCTTGTGCAGTGGGAACGCTAGCCTGACTGAGAACGTCACTGTGAAGACG
TCCGCTTATGACCTGATTTCTGACGGAAGGGTTACTGGGCGTATTTTGTAGGTTGATGTTGCAACTGAGTA
GTTTATCAGCAGGAGGTGTCGCCAAATGAGTCTTTGAATTGGCTTGCGAAGCTTGGGGTCTCTGGA

$(GT)_{12}$-11

GGCCAGAGACCCCAAGCTTCGGATCTAATGACAATACTTAGTACAGAAAATAGCATTTAGATGACAAATTT
TGTATCCACTAAAATGTTATGAACACCAAATTGCTATATAGTTAGTTGCATGAAAACACATTTCCAGATTG
TTCAGAACACAATTTTGACTGTTCTCGCATTGTTTTGGCATGCTGTGGCATTTGCGCCTCCAGTTTGGCTC
AGCGGTTTGGCCTTGCTGGTGTGTTTGTTGGTCATGAGGCTACCAAGGTGAGAAGCCAATTGATAGGATGA
ATGAACATGAACACTTTCCAATCTGGTAATTGTGAGTAAAGTATCTCCCACAGAATTGGCATAGGGCTCTC
CTAAACATTTGACCTAATTCTTAAAATACTCTCAATAGGTAAATGTAGGTGTCTGTGGTAAAGTATTTTGA
ATATAGGTACTGCTTTTCAAGTTTAAAATTGAGTTATTTCAAACTAAATTTGCAAAGGTATAGTCAATTGT
ATATAATATACAAAAAATGTAATTATTTGGTAATACTCTCAGGAAATAAAATTATCATGCCTTGGAATCAC
TAAGACAATTTTATGCACAGCAAAAACATATGAATTCTAAAAAGTGTTTATTTCAATCTGATGCTAATAAA
AGTGTTAAAAGTATGAAGAAAATGGATGAAAGTATTGCTGGTATCAGAAGATTTTCCACTGATACCGATAC
TCCAAAAAAGAATCGATACCACCAACACTAATACCAATACATTGGTAGCTGCATATCTAGTTCATATAGCT
TCAGCTTCCATACTCACACGTCAAACTACAAAGGGTGCATCAGTCNNCTTTTCTTGTCAAGATCTGAAGCA
TTTTGAAAACAGTCTTTGGTACAAGGGATGACAAACAAGAACTGAAAAATGGCCAAGTTCAGAGCTGAATT
TTGGATTTATGAAAACAAATTTCAAAAAATCTAAAGTGGANCTCTGCGTGTGATGCTGATCCGAAGCTTGG
GGTCTCTGGCCA

$(ACT)_8$-3

GGCCAGAGACCCCAAGCTTCGGATCACGAAGCATGCACAACCTAGAAATAATCTCGCACGGACCTAATTAT
ATGCCGAGTATTTGCCTTTTATCTCCTACTCTCATACTCTCATACTCTCCTATTATCCTTCCACTTTCAAT
AATTGTTCCTATATTTCTGGTCACTCTGACCCCCATGATTATCTTCCTGCACTCTTACCCTTATGGCATCG
TCCTCCTCCTTCATCGCTTTTCACCCTTCAATAATTGTCGACAGAAAATACGGTGCGTTGTGATGGTTGAG
AAGCCGATAACTCGAGCATACGAAAACAAGTTGCGGCGTTTAAGAAAATATTCGAGTTTCCGATACGCGAA
GATTGCCACAATAAACAATGTTGGTCGCAAAAAAACAGTATCTAGACACTTTTTATTGATTAGTTTGTTCG
ATGTTTGGAGGAGAGTTTTGGACTAAATACTCTGCAAATGTATCCAGATACTGTATATTACTCAACACTGA
CAACCGCACAGTCGTGACTGGAAGGGCTTCATTGGAGGCAAACAACTGGCGCTGGGTACGACCTCTTATTA
TTTGTTTGTGCATGCATTGATGCCTTAGACCAAACACTTCTTGTTAATACCAGGTCATGGAACATAAACAA
TATGTTATAAATTTAAGATCCGAAGCTTGGGGTCTCTGGCCA

$(ACT)_8$-8

GGCCAGAGACCCCAAGCTTCGGATCCTGGGGCTGTTGGTAGTGGTGTGTACTGTGTGACCCTGGGACCGTT
GGTAGTGATGTGTGCTGTGTGACACCGGGACTGTTGGTAGTGATGTGTACTGTGTGACCCTGTGACTGTTG
GTAGTGATGTGTACTGTATAACCCTGAGACTGTTGGTAGTGATGTGTACTATGACCCTGGGACCGTTGGTA
GGGACGTGTACTGTGTGACCCTGTGACTGTTGGTAGTGATGTGTGCTGTGTGTCCCTGTGACTGTTGGTAG
TGATGTGTGCTGTGTGACCCTGGGACTGTTGGTAGCGATGTGTACTGTGACCCTGGGACCGTTGGTAGTGT
TGTGTACTGTATGACCTTAGGACTGCTGGTAGTGATGTGTGCTGCGTGACCCTGTGACTGTTGGTAGTGAT
GTGTACTGTATAACCCTGAGACTGTTGGTAGTGATGTGTACTGTGACCCTGGGACTGTTGGTAGTGATATG
TGCTGTGTGACCCTGGGACTGTTGGTAGTGATGTGTGTTGTGAGATCCGAAGCTTGGGGTCTCTGGCCA

$(ACT)_8$-11

GGCCAGAGACCCCAAGCTTCGGATCAAAATGTCTATTTCACAGTCGGGTAGCCTCGAACCTACTACCTTCA
GCTTCCAGTGGAAGTCTATTGTCAGGAGTGACCGACGCGCAACCCACTCTGCCACTGAGCTGGTGGAGGGG
GTAAGATTTAACTCAATAAATGTGTAAATCAAATTGTCTGTCTCATTTAAAGCTTTTTACTGATTATCGTT
CCAGTTCAACTTCAAAACATTGTGTGTAGTAACCACTTCACTGCATAGACTCAATGTGTGAGTCACAACCC
GTTCACTGCGTAGACTAACTGCATGAGTCACAACCAGTTCACTGCGTAGAGTCACTTTGTAAGTCACAACA
TCTCACAACATCTCACTCTCCATCCTCTTGACGAGAGCTGTCATGCTCTGGGCTTGTTAGAAACGTGCCAC
TAATTCCGATAATATAATAATACAAGAAGGAAGTACCCGTGAATATTTCCGTGTTAGGAGTAGCGTACACA
AATTGGGTTATTACATGGTTGCATTTGTAAATGCATACAGATATCAAATTAATTCGCTTCATTACAAAGTG
AAGCTGATAAAAATGTCAACTATTCTTCAATTCCCTGAAATATATACCATAGCACATAAAGTTACTGAACC
CCCAATTTGAATCATGATCCGAAGCTTGGGGTCTCTGGCCA

$(ACT)_8$-12

GGCCAGAGACCCCAAGCTTCGGATCATTTCTAGTGACCTTGAATTGTACGACTCTGAAAAACCGCCATCGT
TGTTTACCTTTCGACTCTAACGCTGGAGCTCCTGTATGGGGACTCTACCTGTGTAGCTATTGAGCCAGCTG
TCCAACCTATTAGTGCGAGTTTATGTGGACTCACACAGGGTACATTGATACCGTGGTAACTAGTTTACCTG
ATTGGGTAAACGACCGTGACGTTTAATTTGTCCTATAGGAATATTGGCATAACCTAGTATCAATTTTGCAG
TGGTGTTGGCGTCCCCCTGACCCTGATAGTTACTTGGTACTTATTATTATCGGCCTTGGAGGTGGGAGACA
TACCACGCACTAGTTGTATCTTTACTTTGTTACCTAATTATAGGCAACATTAGGTAGTGTACGGATACTCC
ACGCCAGTCGGTTATCGGCATCGACTACCGTGTTTTCAATCTATGGTCGGACTTAGGTAGGGTAACTATAC
CTCACTCTAGTCACGTATCACTACCGCGGTGATTAGTCACATCCAACATACCTATTGGTGCCGCCTGCATT
GCCTTACCGTGGGTAGTTTACCTATTTATATTGACATTGAAAAGGCTACGGTGGTGAGCAAGGAGGCACTC
ACGGTGCTACTGCTTTCATGATTAGGTGGTTGGTCAGTGGGCATTTCTTANGTGGCGGCCTATAAGTAACT
ACTACCTGTAGCCAATACCTTTGGTCGGGACCACAGGTAGGCATCTACAAAATTCGGATCCGAAGCTTGGG
GTCTCTGGCCA

$(ACT)_8$-16

GGCCAGAGCCCCCCCGCTTNGAAAACATCCTACATCCAAAAATATCCCAAACTCAATACCAATACATAAAA
CTAACTTACTACTAATAGGTCGAAAAAACATAGACTCACACAC