Identification of *Nostoc punctiforme* akinete-expressed genes using differential display

Claudia Argueta, Kamile Yuksek, Ronak Patel, and Michael L. Summers*

California State University Northridge, Department of Biology, 18111 Nordhoff St. Northridge, CA 91330-8303

*corresponding author. Tel.: +1-818-677-7146; fax.: +1-818-677-2034. E-mail address: michael.l.summers@csun.edu

Running title:
N. punctiforme akinete-expressed genes

Abstract

Akinetes are spore-like resting cells formed by certain filamentous cyanobacteria that have increased resistance to environmental stress. They can be found at low frequencies in dense cultures experiencing low light or phosphate limitation, but also form at high frequencies in a *zwf* mutant strain of *Nostoc punctiforme* following dark incubation in the presence of fructose. The wild-type strain is capable of facultative heterotrophic growth under these conditions and does not form akinetes. To identify genes associated with akinete development, differential display was used to amplify and compare cDNA from a wild-type and *zwf* mutant strain of *N. punctiforme* following a switch to dark heterotrophic conditions. Screening of candidate genes by reverse transcriptase real-time quantitative PCR and subsequent testing for akinete-specific expression using GFP transcriptional-reporter plasmids lead to the identification of three novel akinete-expressed genes. The genes identified from the screening encoded for proteins homologous to an aminopeptidase (*aapN*), a zinc protease (*hap*), and an ABC-type transporter (*aet*). Expression of *hap* was also increased in developing hormogonia, a transient type of differentiated filament capable of gliding motility. Transcriptional startsites for akinete-espressed genes were determined using RACE, and promoter regions were compared to orthologs in other filamentous cyanobacteria to identify putative regulatory sequences.

Keywords: filamentous cyanobacteria, akinetes, hormogonia, *aapN*, *aet, hap*, aminopeptidase, ABC transporter, zinc protease

Introduction

Nostoc punctiforme is a filamentous cyanobacterium capable of differentiating its normal vegetative cells into nitrogen fixing heterocysts, motile hormogonia, or spore-like akinetes (Meeks *et al*., 2002). Like others comprising the subset of strains capable of akinete-formation in the order Nostocales and Stigonematales, akinete induction by *N. punctiforme* usually occurs in cultures exposed to light limitation or phosphate starvation (Campbell *et al*., 1996; Wong and Meeks, 2002), although low intracellular ATP levels may not be the direct signal for their formation (Argueta and Summers, 2005). The size and position of akinetes relative to heterocysts vary between strains, with those of *N. punctiforme* appearing only slightly larger than vegetative cells, forming first midway between heterocysts within the filament. Subsequent akinetes then form adjacent to the first and spread out toward the heterocysts. In filaments grown with nitrate or ammonia that lack heterocysts, akinetes appear randomly in the filaments (Meeks *et al*., 2002). As akinetes mature, they become more rounded than vegetative cells, and separate from the filament as cell-cell contacts diminish.

Akinetes contain a thickened envelope composed of polysaccharides and glycolipids (Cardemil and Wolk, 1981; Soriente *et al*., 1993), and usually contain refractile granules of storage material for carbon in the form of glycogen, and nitrogen in the form of cyanophycin (co-polymers of arginine and aspartate). The importance of cyanophycin for akinete germination has been demonstrated in a *Nostoc ellipsosporum* arginine biosynthesis mutant incapable of accumulating this polymer (Leganes *et al*., 1998). These characteristics, and others yet to be discovered, are likely to be required for functional akinetes capable of surviving environmental extremes that kill normal vegetative cells, and their subsequent germination. Although not heat resistant like the endospores of Gram positive bacteria, cyanobacterial akinetes have been reported to survive 5-7 years of desiccation (Sili, 1994; Yamamoto, 1975), months of cold (4°C) dark conditions (Sutherland *et al*., 1979), and have been isolated from sediments as old as 64 years (Livingstone and Jaworski, 1980). Little is known about the molecular basis for such resistance to environmental extremes.

The little that is known about regulation of akinete differentiation has been gleaned through effects of genes originally identified for their role in heterocyst differentiation. Overproduction of DevR, a response regulator component of a two-component regulatory system involved in heterocyst polysaccharide synthesis, causes increased akinete induction in *N. punctiforme,* possibly due to crosstalk (Campbell *et al*., 1996). A second transcriptional regulatory protein, HetR, was shown to be essential for both heterocyst and akinete formation in *Nostoc ellipsosporum* (Leganes *et al*., 1994), however a *hetR* mutant strain of *N. punctiforme* was able to form cold-resistant akinete-like cells that lacked obvious granulation typical of akinetes (Wong and Meeks, 2002). In addition, the *hepA* gene, shown to be essential for formation of a polysaccharide layer in heterocysts (Wolk *et al*., 1994), was also required for normal akinete envelopes of *Anabaena variabilis* (Leganes, 1994). These results support a common envelope and possible storage materials between heterocysts and akinetes, but gives no clear evidence for a common regulatory pathway controlling their differentiation.

Prior to this report, the only study focusing specifically on identification of genes involved in akinete formation or function utilized a proteomic approach (Zhou and Wolk,

2002). A highly expressed protein from *Anabaena cylindrica* akinetes was identified, and the orthologous gene from *Anabaena* sp. strain PCC 7120 was used to design primers for amplification of *avaK* gene from the genome of *A. variabilis*. An *A. variabilis* strain bearing the *avaK* promoter region fused to a GFP transcriptional reporter exhibited high levels of expression in akinetes as compared to heterocysts or vegetative cells, confirming this gene as an akinete-marker gene (Zhou and Wolk, 2002). The ortholog of *avaK* in *N. punctiforme* is similarly regulated (Argueta *et al.*, 2004), but its function remains unknown.

Although the germination of akinetes has received much attention (Chauvat *et al.*, 1982; Sutherland, 1985; Sili, 1994; van Dok, 1996), the molecular genetic study of akinete induction has been hindered by their normally slow and asynchronous formation. Recently a model system has been developed for *N. punctiforme* that can be used to induce synchronous akinete formation (Argueta and Summers, 2005). The system utilizes a glucose-6-phosphate dehydrogenase mutant (*zwf*) lacking the first enzyme of oxidative pentose phosphate pathway, the major route of carbon catabolism in cyanobacteria required for dark heterotrophic growth (Summers *et al.*, 1995). When the *zwf* mutant is placed in the dark in the presence of fructose, the entire culture differentiates into akinete-like cells that have increased resistance to cold, desiccation and lysozyme treatment, have surface polysaccharides with staining properties like those of akinetes, and strongly induce the akinete-marker gene *avaK* at 3-4 days following induction (Argueta and Summers, 2005). The wild-type strain grows heterotrophically under these conditions and does not form akinetes.

This work describes a molecular genetic approach to identify genes expressed in akinetes to better understand akinete differentiation and the molecular basis of their resistant properties. Differential display was used to identify genes induced in *zwf*-induced akinetes of the *N. punctiforme* model system relative to non-differentiating cells. Three novel genes were confirmed to be similarly induced in wild-type akinetes, thus confirming the usefulness of the model system. Analysis of the three akinete-gene promoters reveal conserved regions that will help focus further studies on the details of their regulation.

Materials and Methods

Strains and culture conditions. *Nostoc punctiforme* ATCC 29133 and the *zwf* mutant UCD 466 (Hagen and Meeks 2001) were grown in 125 ml Erlenmeyer flasks containing 50 ml AA/4 supplemented with 5 mM MOPS buffer (pH 7.5), 2.5 mM NH_4Cl, 2.5 mM $NaNO_3$, and 2.5 mM KNO_3 for photoautotrophic growth, and 10 μg/ml ampicillin for plasmid selection. Mixotrophic and heterotrophic growth conditions were achieved by the addition of 50 mM fructose to liquid medium. Cultures were incubated at 25°C on a rotary shaker (120 rpm) with 17-19 μmol photons m^{-2} sec^{-1}. Dark conditions were maintained by covering flasks with foil and incubated at room temperature on a shaker with reduced shaking (40 rpm). Transformed *Escherichia coli* TOP10F' (Invitrogen, Carlsbad, CA) or DH5α MCR were grown using Luria Broth liquid or solid media containing 30 μg/ml Kanamycin or 100 μg/ ml Ampicillin at 37°C.

Akinete and hormogonium induction in WT reporter strains. Nitrogen fixing cultures in stationary phase (~80 μg Chl*a* ml^{-1}) were collected by centrifugation, washed with and transferred to an equal volume of AA/4 containing 5 mM MOPS but lacking phosphate. Reporter strains were microscopically observed every 2-3 days for 4-5 weeks after induction. Akinetes usually occurred after two weeks. The occasional culture that differentiated into hormogonia after transfer to AA/4 medium lacking phosphate were not used in akinete induction studies.

Differential display. RNA extraction was performed as described previously (Summers et al. 1995). Contaminating DNA was removed with RQ1 RNase free DNase (Promega, Madison, WI) as per manufacturers' instructions and column purified using a Zymo RNA clean-up kit (Zymo Research, Orange, CA). Control RNA was similarly extracted from a wild-type photoautotrophiclly grown culture following incubation with 150 μg/ml rifampicin for 2 hours. Akinete RNA was harvested from wild-type cultures 14 days following phosphate starvation and incubation in low light (7-9 μmol photons m^{-2} sec^{-1}) with gentle shaking at 60 rpm. Synthesis of cDNA to be used in differential display occurred in a 20 μl reaction with 200 ng RNA, denatured at 70°C for 10 minutes, and cooled to 50°C prior to the addition of 40 pmol of an arbitrary 10-mer primer (Table 1), 500 μM dNTPS, 1X first-strand buffer, and 50 U SuperScript II (Invitrogen, Carlsbad, CA). The 10 different arbitrary primers (50.1-50.10) were obtained as a set (Genosys, The Woodlands, TX) and contained a 50% G+C content that most closely matched that of the *N. punctiforme* genome. The reaction was carried out in a Perkin-Elmer DNA Thermal Cycler 480 machine ramped from 50 to 30°C over 15 min, 1 hr at 37°C, followed by a 3 min at 95°C. Samples were stored at -80°C prior to amplification.

The PCR was conducted by adding 3 μl of cDNA to a 20 μl reaction containing 1X PCR buffer, 2.5 mM $MgCl_2$, 20 μM dNTPS, 200 pmol of the same arbitrary 10 base primer used in RT reaction, 6% DMSO, 0.5 μM labeled (γ-^{35}S)dATP (2000 Ci/mmol; NEN life Science Products, Inc. Boston, MA) and 2 U Taq polymerase (Promega, Madison, WI). The PCR was performed using 40 cycles of 94°C (30 sec), 40°C (2min), 72°C (1min), followed by a single cycle at 72°C (10 min). Four microliters of PCR products were mixed with 4 μl denaturing loading dye (95% formamide, 20 mM EDTA, 0.05% bromophenol blue and xylene cyanol), heated for 2 minutes at 100°C then resolved by electrophoresis (4.5% acrylamide gel, 7 M urea, 1X TBE). Gels were soaked in 10% methanol before drying on Whatman 3MM paper at 80°C for 1hour. Autoradiographs of the dried gels were used as template to excise differential display products of interest from the gel. Products were eluted from gel slices by boiling in 50 μl 10 mM Tris buffer (pH 8.0) for 10 min and DNA precipitated with sodium acetate and ethanol in the presence of 1 μl 20mg/ml glycogen. Pellets were reconstituted in 20 μl 10 mM Tris (pH 8.0) and re-amplified using the PCR protocol described above with omission of labeled dATP .

Cloning, Sequencing, and Reporter construction. PCR products from differential display fragments were cloned into pCR2.1-TOPO (Invitrogen, Carlsbad, CA) and transformed into $CaCl_2$ competent *E. coli*. Replicate clones (3-6) from each re-amplification were sequenced at the Cal State University, Northridge DNA Sequencing facility on an ABI 377 automated sequencer using M13 forward and reverse primers. Similarity searches of

the sequences were performed against databases available at National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/BLAST/), and the Department of Energy Joint Genome Institute (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi).

Gene-specific primers were designed for selected genes manually using the on-line *N. punctiforme* genome sequence. Promoter regions of specific genes were amplified by PCR from the *N. punctiforme* genome using primers containing added restriction enzyme sites (Table 1), and cloned into the transcriptional-reporter vector pSUN202 (Argueta et al. 2004). Correct orientation was confirmed by using promoter specific forward primer and the GFP reverse primer anchored in the plasmid, followed by sequencing of the resulting PCR product. Reporter plasmids were transformed into *N. punctiforme* by electroporation as previously described (Summers et al. 1995).

cDNA Synthesis for RT QPCR and RACE. Twenty microliter reverse transcriptase reactions contained 1μg RNA, 40 pmol of a gene-specific reverse primer, 500 μM dNTPS, 1X first-strand buffer, and were subjected to 70°C for 10 min, then 42°C for 5 minutes, prior to the addition of SuperScript II reverse transcriptase (100 U in RT QPCR and 200 U in RACE reactions). Reactions were further incubated at 42°C for 60 min, stopped at 70°C for 15 minutes, and stored at -80°C prior to use.

Real Time Quantitative PCR. Two microliters of a gene specific RT reaction was added to 26 μl reactions containing 1x PCR buffer, 2 mM $MgCl_2$, 20 pmol of each forward and reverse gene specific primer (P1/P2; Table 1), 0.2 mM dNTP, 0.25 X SYBR Green and 1 U Hot Start Taq (TaKaRa, Shiga, Japan). These reactions were amplified on a Cepheid System Smart Cycler I using an initial denaturation at 95°C for 150 seconds, followed by 35 cycles of 95°C (15 sec), 56°C-60°C (primer-set-specific temperatures)(20 sec) and 72°C (15 sec). Ribosomal RNA was used to normalize for RNA and RT efficiency between samples, and the averages of results from triplicate samples were used in the data shown in Fig. 1.

Rapid amplification of cDNA ends (RACE). Ligation anchored PCR was used to map the transcription start sites for our genes of interest. Single stranded cDNA was ligated to a 5'-phosphorylated, 3' blocked (inverted dT) anchor oligonucleotide (DT88; Table 1) using T4 RNA ligase (Promega, Madison, WI). Ligation and PCR protocol were performed as previously reported (Kaebernick et al. 2002, Cunnac et al. 2004) with minor modifications. Briefly, removal of RNA was accomplished by 10 min of alkaline hydrolysis at 65°C, by addition of 1μl 0.5 M EDTA and 2.1 μl 1M NaOH to RT reactions and subsequent neutralization by 1 ml DNA Binding buffer supplied with Zymo DNA Clean & Concentrator Kit (Zymo Research, Orange). The entire volume of eluted cDNA was used in 40 μl ligations containing 1X ligase buffer, 20 pmol DT88 (anchor), 20% PEG, 100 U T4 RNA ligase incubated at 18°C for 18 hours. Anchor ligated cDNA (1μl) was used directly in a 50μl PCR reaction containing 10 pmol each forward (DT89) and reverse gene specific (RACE P1) primer, and Hot Start Taq (TaKaRa, Shiga, Japan) following manufacturers instructions. Touch down PCR conditions did not deviate from the Cunnac (2004) protocol and included a 500-fold dilution of the first PCR prior to the second PCR which used DT89/RACE P2 primer pairs. RACE PCR products were visualized by agarose gel electrophoresis, purified, and sequenced to determine the

transcription start sites for each gene. Sequences were manipulated using DNA Strider (Marck, 1988) and BioBike (Massar *et al*., 2005). Sequence features were identified using Wisconsin Genetics Computer Group (GCG) package of programs (Devereux *et al*., 1984), and BLAST tools available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/).

Microscopy. Epifluorescence microscopy was performed using a Zeiss Axiolab microscope containing a 100X oil immersion objective. Fluorescence images were obtained using a longpass blue excitation filter (395 nm) and a green bandpass (509 nm) filter set (Omega Optical) and captured with a DVC 1312 high resolution digital camera. Uniform exposure times were used for all photos to allow crude comparisons of relative GFP expression between photos.

Results and Discussion

Isolation of putative akinete-specific gene fragments using differential display.

RNA harvested from wild- and *zwf* mutant strains of *N. punctiforme* incubated under mixotrophic conditions, and at various time points following a shift to dark heterotrophic conditions were compared using differential display. In these experiments each primer was used individually to generate both the random cDNA fragments from total RNA, and as forward and reverse primers that generated the radiolabeled PCR products for gel separation. The population of radiolabeled fragments thus represented a "snapshot" of transcription under each condition.

Ten different arbitrary primers were initially screened for production of a usable number of radiolabeled PCR fragments. The 4 arbitrary primers listed in Table 1 each produced between 30 and 300 fragments ranging from 70-500 bp and were therefore used in subsequent experiments. A total of 21 gel fragments more highly expressed in mutant RNA samples from later time points after darkening were excised, re-amplified, cloned and sequenced. Several genes were represented by multiple differential display fragments, resulting in a total of 18 unique genes that were more highly expressed in the mutant following darkening.

Using total RNA for differential display is problematic using prokaryotic samples due to the large abundance of stable rRNA, and absence of poly-A tails on mRNA for convenient reverse transcription of full-length cDNA. Compared to preliminary experiments, the frequency of selecting rRNA sequences for cloning was reduced by inclusion of a control lane containing cDNA generated from a culture treated with the transcription inhibitor rifampicin. Due to the short half-life of most prokaryotic mRNA, PCR products arising from these samples were interpreted as arising from rRNA, and identical sized bands were avoided in adjacent experimental lanes in the excision step. Misalignment of autoradiograph templates used for excision of gel fragments, or minor variation in reaction condition between samples, may have contributed to the false positives produced using the differential display procedure.

Screening, confirmation, and identification of akinete-expressed genes

Each of the 18 different genes identified from differentially-expressed fragments

were tested by real-time quantitative PCR (RT Q-PCR) using gene-specific primers (Table 1). This screen was conducted to assure the gene represented by the excised and cloned fragment exhibited similar expression patterns as observed in differential display gels. Four genes were confirmed to exhibit between 3.7 and 112-fold increased relative expression in the developing akinetes of the mutant strain by day 6 after a shift to dark heterotrophic conditions (Fig. 1), and were selected for further study. No corresponding increases were observed in the heterotropically growing wild-type strain for these four genes (maximum 1.2-1.7-fold relative increase at day 6).

The first putative akinete-specific gene (NpF0062) encoded an N-terminal permease fused to a C-terminal ATP binding domain characteristic of ABC (ATP-binding cassette)-type transporters *(Dassa and Bouige, 2001)*. RT Q-PCR results showed an 11-fold increase in expression occurring after 3 days in the *zwf* strain (Fig 1). The promoter region (Fig. 2) was amplified from genomic DNA, cloned into the GFP reporter plasmid pSUN202 and transformed into *N. punctiforme*. Following induction of akinetes in transformed strains by phosphate limitation and visualization by eipfluorescence microscopy, GFP expression from akinetes was confirmed (Fig. 3A). Not all akinetes in a filament expressed the reporter at the same time, indicating transient production and subsequent degradation of GFP may occur. The NpF0062 *a*kinete *e*xpressed *t*ransporter gene was therefore designated *aet*.

The inferred Aet amino acid sequence is most closely homologous to the COG 1132 group of proteins *(Marchler-Bauer and Bryant, 2004)* containing multidrug transporters for export of hydrophobic compounds and components of lipopolysaccharides. The NpF0062 encoded protein was 26% similar and 47% identical over 583 amino acids (E value 1e-47) to MsbA from *E. coli*. MsbA is responsible for transport of lipopolysaccharides to the outer membrane, most likely transporting the lipid A core moiety. When expressed in a Gram positive organism lacking LPS, MsbA conferred resistance to multiple drugs *(Woebking* et al.*, 2005)*. LmrA, the MsbA homologue in *Lactococcus lactis*, was 23% similar and 44% identical to NpF0062 over 575 aa (E-value 2e-29). This Gram positive ABC transporter was able to functionally complement a *msbA* mutant in *E. coli (Reuter* et al.*, 2003)* and confer resistance to multiple antibiotics *(van Veen* et al.*, 1996)*. NpF0062 protein is also 26% similar and 47% identical over 589 amino acids (E-value 1e-43) to the human multidrug ABC transporter MDR1 *(Chen* et al.*, 1986)*. Together, these results suggest NpF0062 may either play a protective role, or be involved in assembly of the thicker protective envelope of akinetes, or both.

The second identified gene (NpR4070) was homologous to the beta subunit of a group of heterodimeric proteins comprising the M16 family of zinc-dependent proteases *(Marchler-Bauer and Bryant, 2004)*. RT Q-PCR analysis indicated this gene had a large112-fold increase in the *zwf*-induced akinetes, with increased expression at 1 day and large increases occurring after 3 days (Fig. 1). Transcriptional reporter analysis of the promoter region identified in Fig. 2 showed strong expression in phosphate-induced wild-type akinetes (Fig. 3C), that paralleled the high levels of induction in *zwf*-induced akinetes by RT Q-PCR analysis.

Occasionally hormogonia formed in flasks following transfer to phosphate limited media. These flasks were not used in akinete-induction experiments, but were observed under epifluorescence microscopy for gene expression in the small cells comprising the

motile hormogonial filament. Results indicated that NpR4070 was also expressed in hormogonia (Fig. 3D), unlike the other genes included in this study (Fig. 3B and F). This is the first evidence of a link between akinete and hormogonia development in any cyanobacterium. Based on reporter results and similarities to genes with known functions, we have designated the NpR4070 gene as *hap* (*h*ormogonium /*a*kinete-expressed *p*rotease).

The Hap protein belongs to a group including the mitochondrial processing peptidases that cleave N-terminal signal sequences from proteins transported to the mitochondria. The Hap protein is most similar to the beta subunit that consists of two structurally related domains that hold the substrate protein like a clamp, and contains the conserved HxxEH zinc binding motif characteristic of the active domain *(Taylor* et al.*, 2001)* in its N-terminal region. The alpha subunit of this heterodimeric protease is termed a processing enhancing protein that lacks a zinc binding motif, but is required for protease function *(Hawlitschek* et al.*, 1988)*. The downstream ORF in *N. punctiforme* (NpR4071) was found to be highly homologous to other alpha peptides. The similarity between the alpha and beta subunits in *N. punctiforme* is 46%, and similar to the 48% similarity between these two subunits in yeast mitochondrial peptidases. Rather than using primary amino acid sequence for substrate recognition, peptidases in this group depend upon secondary structures such as alpha helices containing amino acid residues at specific spatial positions for substrate recognition *(Taylor* et al.*, 2001)*. It has been hypothesized that this group of proteases are evolutionarily related to core proteins required for assembly of the cytochrome bc_1 complex used in electron transport chains *(Braun and Schmitz, 1995)*. A large number of homologues occur in many types of bacteria, but the function of these proteases remains to be determined. Due to its expression in both hormogonia and akinetes, this protease may be involved in general intracellular protein recycling rather than regulation of cell development.

The third identified gene (NpF6000) encodes for a probable nitroreductase protein involved in reduction of nitrogen containing compounds that exhibited only a small 3.7-fold increase in *zwf*-induced akinetes (Fig. 1). The reporter plasmid containing the promoter for this gene did not show GFP expression in akinetes, due either to the low level of induction, or the possibility that transcription of this gene was an artifact associated with *zwf*-induced akinetes. RT Q-PCR analysis of the upstream gene, NpF5999, was also conducted to check for similar levels of induction due to possible cotranscription. NpF5999, located 196 bp from nitroreductase ORF, was found to be expressed in a pattern similar to nitroreducase but with approximately 35-fold induction in *zwf*-induced akinetes at 6 days following darkening (Fig. 1). The NpF5999 gene encoded a predicted aminopeptidase belonging to the M28 peptidase family, and was included with the above-mentioned group of genes for further study. A transcriptional reporter plasmid bearing strain containing the NpF5999 promoter (Fig. 2) confirmed GFP expression in akinetes (Fig. 3E), and NpF5999 was designated *aapN* due to its role as an *a*kinete *a*mino*p*eptidase in *N. punctiforme*.

Aminopeptidases are a type of exopeptidases that remove amino acids from the N-terminal end of polypeptide substrates. The M28 family of metallo-peptidases include bacterial and eukaryotic enzymes that contain two zinc atoms in their active site. When aligned with the aminopeptidases of *Aeromonas proteolytica*, and the Ywad peptidase from *Bacillus subtilis*, the AapN protein was found to contain identical amino acids

involved for 7 of the 8 histidine, glutamate, and aspartate residues involved in zinc stabilization and active site formation (Chevrier *et al*., 1996; Fundoiano-Hershcovitz *et al*., 2005). The one non-identical active site amino acid residue contains a conserved substitution of a glutamate for an aspartate, supporting the assignment of AapN as a member of the M28 protease family. Exopeptidases are known to play a role in many biological functions including breakdown of exogenous peptides, turnover of normal endogenous proteins, and elimination of abnormal proteins (Gonzales and Robert-Baudouy, 1996). Work on the M28 family of proteases has focused on enzymology and crystal structure determination, but little is known about their actual physiological role in bacteria.

The GFP gene reporter strains for the genes described above were induced to form heterocysts following transfer to liquid medium lacking combined nitrogen. None of these exhibited fluorescence from heterocysts. Such a negative result supports a probable unique role for the encoded proteins in akinetes, and reduces the likelihood that they are involved in general protein recycling or transport of extracellular matrix material shared with heterocysts.

Transcriptional startsite mapping and promoter analysis.

Random amplification of cDNA ends (RACE) analysis was performed on RNA to map the transcriptional startsites of akinete-expressed genes identified above. To make this an inclusive promoter analysis, *avaK*, the only previously described akinete-marker gene originally identified in *A. variabilis* (Zhou and Wolk, 2002), was included since this gene has also been demonstrated to be expressed in akinetes of *N. punctiforme* (Argueta *et al*., 2004). The nitrate reductase gene was also included to look for similarities in promoter motifs with the other akinete-expressed genes in case GFP expression was not sufficient to report differential gene transcription.

Single translational starts were observed for *aet, aapN, avaK*, and the nitroreductase genes, whereas two different promoters were identified for the *hap* gene (Figures 2 and 4). RACE analysis was repeated using several different types of RNA samples with identical results, indicating there is a basal level of transcription from these promoters in photoautotrophic, mixotrophic, and dark hetertotrophic growth conditions. No additional startsites were identified in wild-type or *zwf*-induced akinetes, indicating that transcriptional increases were due to activation or de-repression of existing promoters.

The genomes of two closely related filamentous heterocyst forming bacteria have been sequenced (Kaneko *et al*., 2001)(Accession number NC_007413), and used here to identify putative sequences used in transcriptional regulation. Akinetes are only found in heterocyst-forming cyanobacteria, although not all heterocyst forming bacteria can produce akinetes (Castenholz and Waterbury, 1989), especially those in long-term culture (Herdman, 1988). *A. variabilis* is capable of forming nearly all cells into akinetes when aged on agar plates, and has been used to study the macromolecular components of akinetes (Simon, 1980; Cardemil and Wolk, 1981). *A.* PCC 7120 has been cultured continuously for many years, and although it cannot form akinetes, it has the *avaK* akinete-marker gene and all four genes referred to in this work. Similar gene organization in many of these genes provide further evidence that the original isolate

indeed did have the potential for akinete development. Due to the constraints on random mutation in DNA sequences required for regulation, we feel relative comparisons of intergenic regions for orthologous loci between these strains will give valuable insight into the location and sequence of DNA regions required for binding transcriptional regulatory factors. Because of the relatively recent cultivation of *A.* PCC 7120, it is unlikely that random mutations that caused the loss of akinete formation would have completely obliterated smaller binding sites required for their transcription. Thus *A.* PCC 7120 likely retains vestigial intergenic sequences that can be utilized in a bioinformatics approach to identify putative regulatory regions involved in transcription.

The *aapN* gene contains a single transcriptional startsite with a 210 nucleotide untranslated leader sequence extending to the putative TTG translational start codon. The upstream 106 bp region between the *aapN* startsite and translational start for the upstream divergently transcribed *dnaK* gene was 83-84% identical between *N. punctiforme* and upstream region of orthologous locus from *A. variabilis* (Ava0920) and *A.* PCC 7120 (All2989). This region contains two different inverted repeats. The *aapN* transcriptional startsite (bold) is contained within an imperfect inverted repeat with the sequence CCTCAATG–N_{13}–C**A**TTAAGG. The second imperfect inverted repeat contains the putative ribosome binding site of *dnaK* and has the sequence ACTCTT-N_6-AAGTGT.

Two different promoters were identified for the *hap* gene, with leader sequences of 234 (P1) and 31 (P2) nucleotides. The P2 promoter region contains 2 conserved regions between the –10 region and –52 with high identity to analogous regions in orthologous genes in *A. variabilis* (Ava3995) and *A.* PCC 7120 (Alr0710) (Fig. 4). The upstream promoter (P1) is absent in *A. variabilis* and *A.* PCC 7120, likely due to a genome rearrangement resulting in loss of the upstream serine/threonine protein kinase gene along with 287-289 bp of 5'-, and 229 bp of 3'-flanking DNA from this locus. *N. punctiforme* still retains this kinase and the 3'-flanking DNA containing the P1 promoter. Because RACE analysis is not quantitative, the contribution of the individual promoters to cell-type specific expression patterns cannot be determined from this data. Current work is underway to determine if the P1 promoter is a homogonia-induced promoter. The P1 promoter contains two inverted repeat sequences centered at –28 and –45 bp from the transcriptional start (Fig. 4) that could be a potential target for DNA binding proteins.

The nitrate reductase gene startsite was also identified to see if conserved sequence elements were present, and if these were common to those found in other genes confirmed to be akinete specific by GFP reporter assays. A large region of identity between of orthologous locus from *A. variabilis* (Ava2154) and *A.* PCC 7120 (All4884) was found extending on either side of the –10 region, and another from –36 to –28 (Fig. 4). Neither of these exact sequences was found in other confirmed akinete-specific promoter regions, although the sequence AATG adjacent to the –10 region was also present in the *aapN* promoter region. Identical inverted repeats with the sequence TTGTAGAGACG-N_{11}-CGTCTCTGTAA were found centered at –149 of the nitrate reductase promoter and –150 of *hap* P2 promoter.

The *avaK* promoter region contains two motifs common in the promoter regions of orthologous genes from the from *A. variabilis* (Ava1657) and *A.* PCC 7120 (All4050). The first is located from –50 to –34 and the second from –82 to –66. The adenine-rich sequence AAGAAAAATA centered at –89 is also found centered at +7 in *N.*

punctiforme, however this direct repeat is not found in these areas in orthologous genes from *A. variabilis* and *A.* PCC 7120.

The *aet* promoter region, like that of the *hap* P1 promoter, was located near the stop codon of the upstream gene and had no similarities to intergenic regions from orthologous genes of *A. variabilis* (Ava2224) and *A.* PCC 7120 (All4955). The sequence CAAAAAAAGA centered at +43 near the Shine Dalgarno sequence of *aet* is also present at –2 of the nitrate reductase gene overlapping the transcriptional start region. RACE analysis gave no indication of co-transcription of any of the genes in this study with upstream genes, or from unidentified promoters lying further upstream for any of the mapped startsites, with the exception of *hap* P2 caused by transcription initiation from the P1 promoter.

In summary, differential display was successfully used to increase the number of known akinete-expressed genes from the one previously identified (*avaK*) to a total of four, illustrating the usefulness of this approach in the study of gene regulation when no DNA microarray is available. However, due to the inherent randomness of gene identification based on arbitrary priming, and failure to find *avaK* or *aapN* directly in this study, it is obvious that there are many akinete-specific genes remaining to be discovered. With the availability of numerous sequenced genomes of closely related species, it is now possible to mine these resources using multiple sequence alignments of intergenic regions to form testable hypotheses regarding transcriptional regulation. Future efforts will be focused on determining if each of the conserved regions identified in this work are required for cell-type specific gene induction. The very limited amount of conservation found here between promoter regions of similarly regulated genes indicates that transcriptional regulation of akinete differentiation will likely turn out to be as complex as that found in heterocyst development.

Acknowledgments

This work was supported by NSF grant MCB93327, and NIH grants GM48680 and GM6378

References

Argueta, C., and Summers, M.L. (2005) Characterization of a model system for the study of *Nostoc punctiforme* akinetes. *Arch. Microbiol.* **183:** 338-346.

Argueta, C., Yuksek, K., and Summers, M. (2004) Construction and use of GFP reporter vectors for analysis of cell-type-specific gene expression in *Nostoc punctiforme*. *J. Microbiol. Meth.* **59:** 181-188.

Braun, H.-P., and Schmitz, U.K. (1995) Are the 'core' proteins of the mitochondrial bc_1 complex evolutionary relics of a processing protease? *Trends Biochem Sci* **20:** 171-175.

Campbell, E.L., Hagen, K.D., Cohen, M.F., Summers, M.L., and Meeks, J.C. (1996) The

devR gene product is characteristic of receivers of two-component regulatory systems and is essential for heterocyst development in the filamentous cyanobacterium *Nostoc* sp. strain ATCC 29133. *J. Bacteriol.* **178:** 2037-2043.

Cardemil, L., and Wolk, C.P. (1981) Polysaccharides from the envelopes of heterocysts and spores of the blue-green algae *Anabaena variabilis* and *Cylindrospermium licheniforme*. *J. Phycol.* **17:** 234-240.

Castenholz, R.W., and Waterbury, J.B. (1989) Oxygenic photosynthetic bacteria. Group I. Cyanobacteria. In Bergy's manual of systematic bacteriology. Vol. 3. Staley, J.T., Bryant, M.P., Pfenning, N. and Holt, J.G. (eds.) Baltimore: Williams and Wilkins, pp. 1710-1789.

Chauvat, F., Corre, B., Herdman, M., and Joset-Espardellier, F. (1982) Energetic and metabolic requirements for the germination of akinetes of the cyanobacterium *Nostoc* PCC 7524. *Arch. Microbiol.* **133:** 44-49.

Chen, C.J., Chin, J.E., Ueda, K., Clark, D.P., Pastan, I., Gottesman, M.M., and Roninson, I.B. (1986) Internal duplication and homology with bacterial transport proteins in the mdr1 (P-glycoprotein) gene from multidrug-resistant human cells. *Cell* **47:** 381-389.

Chevrier, B., D'Orchymont, H., Schalk, C., Tarnus, C., and Moras, D. (1996) The structure of the *Aeromonas proteolytica* aminopeptidase. *Eur. J. Biochem.* **237:** 393-398.

Dassa, E., and Bouige, P. (2001) The ABCs: a phylogenetic and functional classification of ABC systems in living organisms. *Res. Microbiol.* **152:** 211-229.

Devereux, J., Haeberli, P., and Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nuc. Acids Res.* **12:** 387-395.

Fundoiano-Hershcovitz, Y., Rabinovitch, L., Shulami, S., Reiland, V., Shoham, G., and Shoham, Y. (2005) The *ywad* gene from *Bacillus subtilis* encodes a double-zinc aminopeptidase. *FEMS Micrbiol Lett* **243:** 157-163.

Gonzales, T., and Robert-Baudouy, J. (1996) Bacterial aminopeptidases: Properties and functions. *FEMS Micrbiol Rev* **18:** 319-344.

Hawlitschek, G., Schneider, H., Schmidt, B., Tropschug, M., Hartl, F.U., and Neupert, W. (1988) Mitochondrial protein import: identification of processing peptidase and of PEP, a processing enhancing protein. *Cell* **53:** 795-806.

Herdman, M. (1988) Cellular Differentiation: Akinetes. *Meth. Enzym.* **167:** 222-232.

Kaneko, T., Nakamura, Y., Wolk, C.P., Kuritz, T., Sasamoto, S., Watanabe, A., Iriguchi, M., Ishikawa, A., Kawashima, K., Kimura, T., Kishida, Y., Kohara, M., Matsumoto, M., Matsuno, A., Muraki, A., Nakazaki, N., Shimpo, S., Sugimoto, M., Takazawa, M.,

Yamada, M., Yasuda, M., and Tabata, S. (2001) Complete genome sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120. *DNA Res.* **8:** 205-213.

Leganes, F. (1994) Genetic evidence that *hep*A gene is involved in the normal deposition of the envelope of both heterocysts and akinetes in *Anabaena variabilis* ATCC 29413. *FEMS Microbiol. Let.* **123:** 63-67.

Leganes, F., Fernandez-Pinas, F., and Wolk, C.P. (1994) Two mutations that block heterocyst differentiation have different effects on akinete differentiation in *Nostoc ellipsosporum*. *Mol. Microbiol.* **12:** 679-684.

Leganes, F., Fernandez-Pinas, F., and Wolk, C.P. (1998) A transposon-induced mutant of *Nostoc ellipsosporum* implicates an arginine-biosynthetic gene in the formation of cyanophycin granules and of functional heterocysts and akinetes. *Microbiol.* **144:** 1799-1805.

Livingstone, D., and Jaworski, G.H.M. (1980) The viability of akinetes of blue-green algae recovered from the sediments of Rostherne Mere. *Br. Phycol. J.* **15:** 357-364.

Marchler-Bauer, A., and Bryant, S.H. (2004) CD_Search: protein domain annotations on the fly. *Nucleic Acids Res.* **32:** 327-331.

Marck, C. (1988) 'DNA Strider': a 'C' program for fast analysis of DNA and protein sequences of the Apple Macintosh family of computers. *Nucleic Acids Research* **16:** 1829-1836.

Massar, J.P., Travers, M., Elhai, J., and Schrager, J. (2005) BioLingua: a programmable knowledge environment for biologists. *Bioinformatics* **21:** 199-207.

Meeks, J.C., Campbell, E.L., Summers, M.L., and Wong, F.C. (2002) Cellular differentiation in the cyanobacterium *Nostoc punctiforme*. *Arch. Microbiol.* **178:** 395-403.

Reuter, G.T., Janvilisri, T., Venter, H., Shahi, S., Balakrishnan, L., and van Veen, H.W. (2003) The ATP binding cassette multidrug transporter LmrA and lipid transporter MsbA have overlapping substrate specificities. *J. Biol. Chem.* **278:** 35193-35198.

Sili, C., A. Ena, R. Materassi, and M. Vincenzini (1994) Germination of desiccated aged akinetes of alkaliphilic cyanobacteria. *Arch. Microbiol.* **162:** 20-25.

Simon, R.D. (1980) DNA Content of Heterocysts and Spores of the Filamentous Cyanobacterium *Anabaena variabilis*. *FEMS Microbiology Letters* **8**.

Soriente, A., Gambacorta, A., Trincone, A., Sili, C., Vincenzini, M., and Sodano, G. (1993) Heterocyst glycolipids of the cyanobacterium *Cyanospira rippkae*.

Phytochemistry **33:** 393-396.

Summers, M.L., Wallis, J.G., Campbell, E.L., and Meeks, J.C. (1995) Genetic evidence of a major role for glucose-6-phosphate dehydrogenase in nitrogen fixation and dark growth of the cyanobacterium *Nostoc* sp. strain ATCC 29133. *J. Bacteriol.* **177:** 6184-6194.

Sutherland, J.M., Herdman, M., and Stewart, W.D.P. (1979) Akinetes of the cyanobacterium *Nostoc* PCC 7524: Macromolecular composition, structure and control of differentiation. *J. Gen. Microbiol.* **115:** 273-287.

Sutherland, J.M., J. Reaston, W.D.P. Stewart, and M. Herdman (1985) Akinetes of the cyanobacterium *Nostoc* PCC 7524: macromolecular and biochemical changes during synchronous germination. *J. Gen. Microbiol.* **131:** 2855-2863.

Taylor, A.B., Smith, B.S., Kitada, S., Kojima, K., Miyaura, H., Otwinowski, Z., Ito, A., and Deisinhofer, J. (2001) Crystal structures of mitochondrial processing peptidase reveal the mode for specific cleavage of import signal sequences. *Structure* **9:** 615-625.

van Dok, W., and B.T. Hart (1996) Akinete differentiation in *Anabaena circinalis* (Cyanophyta). *J. Phycol.* **32:** 557-565.

van Veen, H.W., Venema, K., Bolhuis, H., Oussenko, I., Kok, J., Poolman, B., Driessen, A.J., and Konings, W.N. (1996) Multidrug resistance mediated by a bacterial homolog of the human multidrug transporter MDR1. *Proc Natl Acad Sci U S A* **93:** 10668-10672.

Woebking, B., Reuter, G., Shilling, R.A., Velamakanni, S., Shahii, S., Veter, H., Balakrishnan, L., and van Veen, H.W. (2005) Drug-lipid A interaction on the *Escherichia coli* ABC Transporter MsbA. *J. Bacteriol.* **187:** 6363-6369.

Wolk, C.P., Ernst, A., and Elhai, J. (1994) Heterocyst metabolism and development. In The Molecular Biology of Cyanobacteria. Bryant, D.A. (ed.) Boston: Kluwer, pp. 769-823.

Wong, F.C., and Meeks, J.C. (2002) Establishment of a functional symbiosis between the cyanobacterium *Nostoc punctiforme* and the bryophyte *Anthoceros punctatus* requires genes involved in nitrogen control and initiation of heterocyst differentiation. *Microbiol.* **148:** 315-323.

Yamamoto, Y. (1975) Effect of desiccation on the germination of akinetes of *Anabaena cylindrica*. *Plant Cell Physiol.* **16:** 749-752.

Zhou, R., and Wolk, C.P. (2002) Identification of an akinete marker gene in *Anabaena variabilis*. *J. Bacteriol.* **184:** 2529-2532.

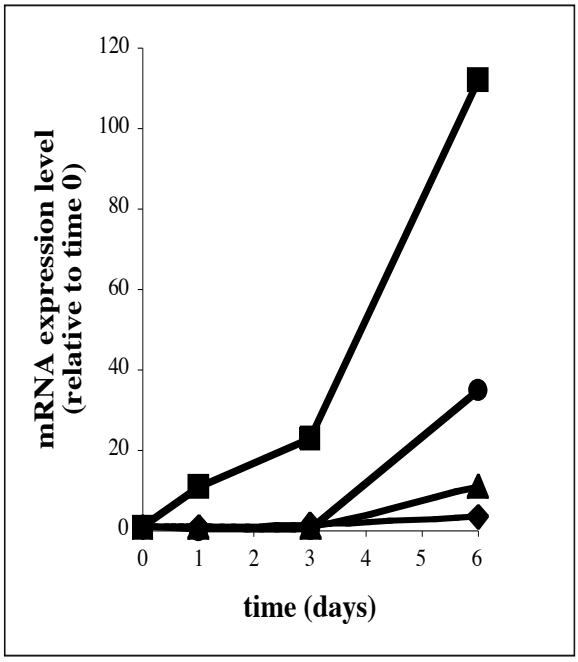

Figure 1. Kinetics of mRNA expression in UCD 466 after incubation in dark heterotrophic conditions as measured by RT Q-PCR and shown relative to expression at time 0 for each sample. Squares, NpR4070 (*hap*): Circles, NpF5999 (*aapN*); triangles, NpF0062 (*aet*); Diamonds,NpF6000 (nitroreductase) totaling 112-, 35-, 11-, and 3.5-fold increased expression by day 6, respectively. No corresponding increases were observed for these genes in the wild-type strain.

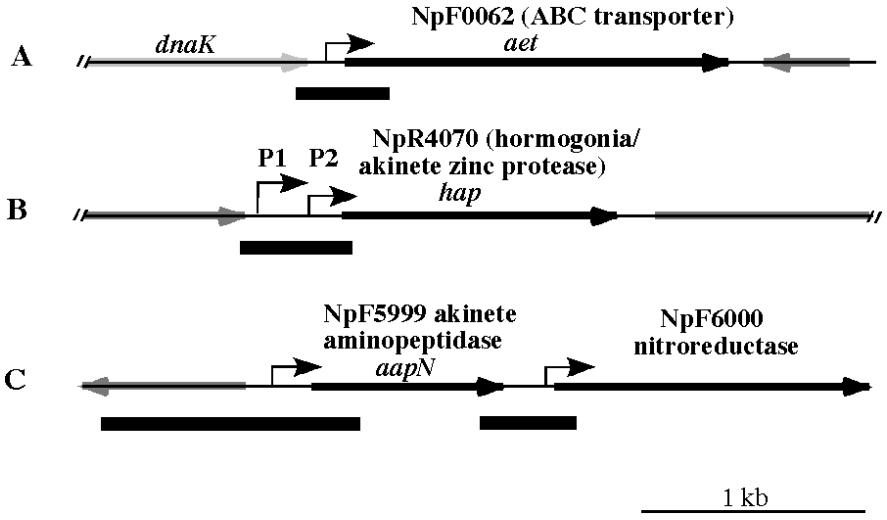

Figure 2. Genomic regions containing genes of interest. Heavy arrows depict extent and orientation of genes. Dark bar below each section of DNA denotes extent of promoter region cloned into pSUN202. Approximate locations of transcriptional start sites determined by RACE analysis are shown with smaller arrows.

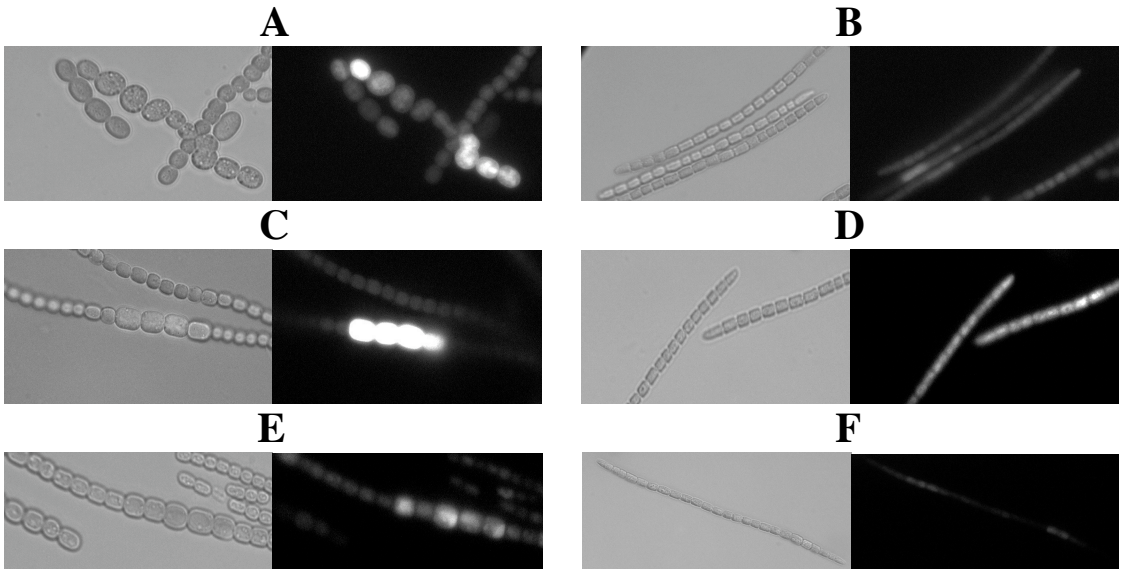

Figure 3. *N. punctiforme* strains bearing GFP reporter plasmids visualized under brightfield (left panel) or epifluorescence (right panel) microscopy. A and B) *aet* reporter; C and D) *hap* reporter; E and F) *aapN* reporter. Larger cells shown in A, C, and E are akinetes, filaments shown in B, D and F are hormogonia.

Figure 4. Alignment of promoter regions with upstream regions of orthologous genes from other filamentous cyanobacteria. The +1 transcriptional start is capitalized, and –10 regions are underlined. Bold letters denote identity in all three strains. Subscripts denote nucleotide length of the un-translated mRNA leader sequence. Double underlines denote inverted repeat sequences. Capital letters denote start and stop codons. Np, *Nostoc punctiforme*; Av, *A. variabilis*; An, *A.* PCC 7120.

Primer	Sequence (5'- 3')
Arbitrary 10 base primers (Differential Display)	
50.1	GTG CAA TGA G
50.6	AGG TTC TAG C
50.8	GGA AGA CAA C
50.9	AGA AGC GAT G
Sequencing	
M13 Forward	CGC CAG GGT TTT CCC AGT CAC GAC
M13 Reverse	TCA CAC AGG AAA CAG CTA TGA C
Reporter Construction	
aet P1	CCA CTG CAG AAA TGG GAA AGC AGG TCG GA
aet P2	GTC GAA TTC CTG CCT AAG ATC CAG TGT CA
hap P1	GCA CTG CAG TGC TTG GTG AAG GAA CAG A
hap P2	TCG GAA TTC TGC TTA CGC CAC CCC AGT TT
aap P1	AAT CTG CAG GGA CTG TAA TCA CTG CCC
aap P2	TTT GAA TTC CCT CTG GCT TGG CAG GGC AA
NR P1	TAT CCG CTG CAG ATG GTG ACA GAT ACA GCA T
NR P2	ACC AAT TGA ATT CCT GGG TTA TTG GCA AAT CCT
GFP Forward	TAT AGC GCT AGA GTC GAC CT
GFP Reverse	GAG TCT CCA GTT TGT TTC AGT
Real Time Q-PCR	
aet QP1	CCA GTT GGG AGA GAT GCT ACT A
aet QP 2	CGG TGA GAT ACA TTG TCA TAT C
hap QP 1	CTC GCA AAT CAC GCA CTA AAC G
hap QP 2	GCC CAC TAT CAG CCG GAA AAT A
aap QP 1	CGT GCT CGA GCG AGG TTT AAT AGT TCC CCA
aap QP 2	ACC CTT CAC ATA CAC CAC TA
NR QP 1	TGG CGA TCG CGT TTA CCG TT
NR QP 2	GTT TGC ATG CGG GTG GCC CAA TGT AGT GAT
RACE	
aet RACE P1	ATT CCC CGT TGA CT ACCT GC
aet RACE P2	CGG TAA AAG ACC AGC CAC TA
hap RACE P1	GGA AGC CTT GCC GTA CCT TT
hap RACE P2	TCG GAA GAC TGA GGC TGG AA
aap RACE P1	GGC AGG GCA AAT TTA ATA TC
aapRACE P2	AGG TGT GGA TTT CCA CAC TC
NR RACE P1	GCA CTG GCT GT TAG TCC AG
NR RACE P2	GCT GGG CAA TTG ATT GGT GT
avaK RACE P1	CTT CTA ATT TGT AGA GTA CC
avaK RACE P2	CGC TTG ACG TTC ATT GAA CTC A
DT 88	GAA GAG AAG GTG GAA ATG GCG TTT TGG
DT 89	CGC CAT TTC CAC CTT CTC TTC

Table1. Oligonucleotides used as primers for this work. Introduced restriction enzyme sites used for cloning are underlined.