Motif Finding in Upstream Regions of Chaperone Genes Downregulated in Response to Environmental Changes

Stephen J. Tam
Department of Biophysics
Stanford University
stephent@stanford.edu

Many methods exist for discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. Among them are Motif Discovery Scan, Bioprospector, Consensus, AlignACE, and MEME, which were used to discover the important sites in the 800 bp upstream sequences of 25 yeast chaperone genes downregulated in response to temperature shocks, hydrogen peroxide, the sulfhydryl-oxidizing agent diamide, the disfulfide-reducing agent dithiothreitol, amino acid starvation, nitrogen source depletion, osmotic shock, and progression into stationary phase. The purpose was to find motif(s) which might point toward a putative transcriptional promoter binding site responsible for the coordinated downregulation of these genes under the aforementioned conditions of cellular stress, as well as to compare which method(s) works best. Bioprospector came up with the most interesting motifs (TACTGACG and CATATTTA) while on the other hand, Consensus failed to find any conserved sequences from these coregulated genes, including the motif in a positive control gene set where the transcription factor binding site is known.

Introduction

To become functionally active, newly synthesized protein chains must fold to unique three-dimensional structures. Evidence accumulated over the last decade indicates that many newly synthesized proteins require a complex cellular machinery of molecular chaperones, protecting nonnative proteins chains from misfolding and aggregation, to reach their native states efficiently (F. U. Hartl 1996). Some native proteins unfold under conditions of stress, and thus are susceptible to aggregation. Therefore at those times many chaperones are synthesized at greatly increased levels. Yet quite interestingly, in DNA microarray data that measured changes in transcript levels over time for almost every yeast gene under the aforementioned conditions of cell stress, a certain subset of yeast chaperones were actually downregulated (Gasch *et al.* 2000).

In some cases, analyzing promoters of genes that are coregulated under the conditions examined often identifies common sequence motifs, suggesting regulation by known transcription factors and in others identifying novel promoter elements. Most of the early algorithms do not incorporate microarray information in motif finding such as Consensus and MEME, employing a progressive comparison method based on position-specific probability matrix and an expectation-maximization based iterative optimization algorithm, respectively. Improvements only began with a better understanding of the dataset to be investigated, namely the upstream sequences of genes with similar expression patterns (Conlon *et al.* 2002). This new breed includes Motif Discovery Scan, Bioprospector, and AlignACE. This project entails the use of motif discovery algorithms to find candidate motifs for these repressed chaperone genes in yeast, as well as to compare which methods works best.

Results

Each motif finding algorithm tested can take on various input parameters which may drastically affect the output consensus sequences. In order to 'standardize' the computational methods, a positive control dataset was put together consisting of 25 yeast genes (Trr2, Gtt2, Aad6, Aad4, Isu2, Gpx2, Bet3, Trx2, Lap4, Flr1, Ttr1, Ydl124w, Ylr108c, Yll055w, Yil167w, Ylr460c, Ycr102c, Ynl260c, Ykl071w, Ydr132c, Aad15, Yol118c, Ykr071c, Yml131w, and Ynl134c) shown to contain the Yap1p consensus binding site within 600 bp of the gene start site previously identified by an experiment where wild type and *yap1* strains were exposed to hydrogen peroxide treatment in duplicate, and approximately 70 genes whose expression was affected by deletion of Yap1 were identified as dependent on the factor. Those very genes were screened further for the known Yap1p consensus site 5'-TTAGTAA-3' using the MEME algorithm (Gasch 2001 and appendix).

Next, the upstream sequence retrieval of the Yap1p gene list was carried out by the 'Regulatory Sequence Analysis Tools' (http://embnet.cifn.unam.mx/rsa-tools/) website for the 800 bp before the gene start site. These sequences were queried through all five of the computational algorithms: Motif Discovery Scan (http://bioprospector.stanford.edu/MDscan/), Bioprospector (http://bioprospector.stanford.edu/), Consensus (http://ural.wustl.edu/~jhc1/consensus/), AlignACE (http://atlas.med.harvard.edu/), and MEME (http://meme.sdsc.edu/meme/website/). Input parameters were adjusted so the Yap1p consensus sequence was found, to maximize motif score, and to maximize the percentage of sequences which contribute to the motif (Table 1, 2).

As expected, the MEME computational method found the Yap1p motif as GCTTACTAAT in *all* of the input sequences. MD Scan found three overlapping motifs: TTACTAAT, TTACTAAT, and CTTACTAA with around 60 percent of the input sequences contributing to each consensus motif. Interestingly, most of those input sequences contained multiple segments with the Yap1p site, thus allowing the same sequence to increase the motif score in the dataset. The iterative stochastic sampling approach of Bioprospector obtained similar results (TTAGTAAT, TTAGTAAT, ATTACTAA) but with even greater contribution by the input sequences of 84 percent. AlignACE successfully found the consensus site ATTAGTAA as its top scoring motif while also digging up some noise in the motifs GAGGAAGAGGACGATGA and AACAAGAAAATG, with around 70 percent of the input sequences contributing. Unfortunately the Consensus algorithm failed to find the Yap1p consensus site and instead came up with the result of GGCCGGCC (Table 1).

To strengthen the putative regulatory consensus sequences that might be found within the upstream sequences of the downregulated chaperone yeast genes, a gene set of 20 *upregulated* chaperone yeast genes (Ssa2, Hsc82, Sba1, Ssa1, Yro2, Sis1, Hsp26, Hsp104, Ssa3, Sti1, Cpr6, Hsp82, Ynl007W, Cph1, Cpr8, Yfr041c, Caj1, Xdj1, Afg1, and Ypr061C) were put together from DNA microarray data that measured changes in transcript levels over time for almost every yeast gene under the various cell stress conditions (Gasch 2001, personal communication with J. Frydman). This would serve as a negative control to make sure that the motifs discovered would not be general yeast chaperone secondary regulatory sequences without specificity to downregulated genes.

As before, the upstream sequence retrieval of the Upregulated chaperone gene list was carried out by the 'Regulatory Sequence Analysis Tools' website for the 800 bp before the gene start site. These sequences were queried through all five of the computational algorithms again using the same input parameters that maximized results for Yap1p transcription factor binding site motif, motif score, and percentage of sequences contributing to the motif. The results are summarized in Figure 1, with MDScan strongly exhibiting the motif TTCTAGAA while most all the other computational methods came up with a lot of simple non-functional repeats such as GAAAAAAA and TCCCCCT. Consensus failed to come up with any common sequence segments enriched in the dataset.

To find motif(s) which might point toward a putative transcriptional promoter binding site responsible for the coordinated downregulation of these genes in response to environmental changes, a set of 25 *downregulated* chaperone yeast genes (Gim4, Fpr1, Ssb1, Gim5, Gim6, Gim1, Gim3, Gim2, Egd1, Egd2, Pdr13, Ssb2, zuo1, Fpr4, Npi46, Cct2, Cct8, Cct3, Cct5, Cct6, Cct4, Cct7, Cpr8, Cct1 , and YnL227) were run through the

‘Regulatory Sequence Analysis Tools’ website for the upstream sequences which were then queried as the *upregulated* and Yap1p genes were. Even equip with 3rd order Markov of the whole yeast genome as the background, MD Scan only found background noise such as a string of T’s and A’s. AlignACE, with a 0th order Markov background model, failed just as badly while Consensus once again did not work. Bioprospector was the only motif finding algorithm which come up with significant motifs: TACTGACG and CATATTTA (as shown in Table 1).

Discussion

The motif discovery problem has long been an interesting area of study for biologists and statisticians alike. Only recently has the microarray data of many organisms in many experimental conditions at many time points been available for the use in the discovery of known, as well as novel, transcription regulatory signals. Therefore, the integration of expression data into motif finding can improve specificity of motif finding programs, such as iterative masking to find multiple motifs and the use of Markov background models to reduce the noise found in upstream sequences (Conlon *et al.* 2002). Bioprospector, a C program using a Gibbs sampling strategy, was designed specifically to examine the upstream region of genes in the same gene expression pattern group and look for regulatory sequence motifs (Liu *et al* 2001). It is a modified algorithm from earlier samplers in that it allows for the modeling of gapped motifs and motifs with palindromic patterns, as well as providing 3rd order Markov background. This computational method was the only one that found putative motif(s) from the *downregulated* chaperone yeast genes: TACTGACG and CATATTTA. Most importantly, none of the putative regulatory motifs found in the yeast genes upregulated under the aforementioned conditions of cellular stress resembled these two sequences, suggesting a specific transcription regulatory site.

The model with the largest shortcomings was Consensus. While it had the advantage over methods that enumerate regular expressions such as Moby Dick in that it allowed for base substitutions and quicker results, it failed to come up with any common sequence segments enriched in the dataset excluding the one false positive for the Yap1p regulated genes. Possibly the cause of this failure was due to the fact that because in a progressive comparison method, the first sequence is the seed and the search proceeds linearly without iteration. Therefore if the first two sequences did not contain the putative transcriptional binding motif, then the algorithm would not pick it up. In effect, it loses accuracy and efficiency when there is variable number of motif instances in each sequence (Liu *et al.* 2002). Further tests can be done, such as order rearrangement of the source data (i.e. Yap1p gene set).

AlignACE, an updated Gibbs sampling update motif matrix that allows for base substitutions as well as allowing for variable motif width, found only strings of A's and T's from the *downregulated* gene set, which was the same thing it got from searching the *upregulated* gene set. The mostly likely explanation for its failure seems to be from the use of a 0^{th} order Markov as a background model instead of a 3^{rd} order, where the distributions of true positive and false positive have been show to separate very well (Liu *et al.* 2001).

MEME found only a string of A's in the *downregulated* data set on 92 percent of the input sequences because MEME is a tool for discovering motifs in a group of related DNA or protein sequences, but not for use specifically to search out genes selected for through clustered DNA microarray expression patterns. Furthermore, the algorithm does not allow for gaps and much of it is automated. MD Scan's false positives can be explained by the fact that this method was originally optimized for examination of the ChIP-array selected sequences and searches for DNA sequence motifs representing the protein-DNA interaction sites. It makes no reference or link in the algorithm to expression data, only co-immunoprecipitation by the predetermined protein crosslinked to DNA fragments (Liu *et al.* 2002).

Despite the failure of MD Scan to discover candidate motifs beyond simple background repeats, the program managed a mean search time around one minute by enumerating only existing *w*-mers in the top sequences so that the time increases only quadratically to the total number of bases in the top sequences for *all* motif sizes. For Consensus, the progressive comparison strategy is quite fast at a two minute mean search time because it seeds with the first sequence and proceeds linearly without further iterations (at the expense of its search). As iterative stochastic sampling methods, both Bioprospector and AlignACE where three place in the search race with mean search times of six and five minutes, respectively. These methods are time-consuming because of the multiple initializations required in order to avoid being trapped in a local maximum. MEME was last place at 15 minutes because it uses every word as a seed for the expectation maximization based iterative optimization algorithm.

A possible problem in the approach used to search out putative transcription regulatory motifs exists within the assumption that the response element in the *downregulated* chaperone yeast genes is characteristically similar to the Yap1p binding site (i.e. one block, around 7 base pairs wide, etc). All input parameters were adjusted so the Yap1p consensus sequence was found so that the response element would be missed completely if a two block motif. Further testing requires the use of several more positive control data sets which cover a majority of all motif-classes.

In conclusion, the Bioprospector method developed by Xiaole Liu turned out to work the best on these sets of clustered DNA microarray data under several conditions of cellular stress. The biological significance of these motifs must be established in order to complement their statistical significance. So the next step would be to follow up on the putative regulatory regions of these co-regulated genes with biochemical binding assay with proteins known to interact with DNA sequences similar to TACTGACG and CATATTTA.

REFERENCES

Timothy L. Bailey and Charles Elkan, "Fitting a mixture model by expectation maximization to discover motifs in biopolymers", Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, pp. 28-36, AAAI Press, Menlo Park, California, 1994.

Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO. (2001). Genomic expression programs in the response of yeast cells to environmental changes. Mol. Cell. Biol. *11*, 4241-4257.

Conlon E, Liu X, Brutlag DL, Liu JS. Motif Regressor: A Computational Approach to Integrate Sequence Motif Discovery and Microarray Analysis. Manuscript in preparation.

Liu X, Brutlag DL, Liu JS. (2002) A Fast and Accurate Motif Finding Algorithm With Applications To Chromatin Immunoprecipitation Microarray Experiments. Manuscript in preparation.

Liu X, Brutlag DL, Liu JS. (2001) BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. Pac Symp Biocomput. 127-8.

Minami Y, Hohfeld J, Ohtsuka K, Hartl FU. (1996) Regulation of the heat-shock protein 70 reaction cycle by the mammalian DnaJ homolog, Hsp 40. J Biol Chem. *271*, 19617-24.

able 1

ene Set	MDScan	Bioprospector	Consensus	AlignACE	MEME
p1p	TTACTAAT 56% TTACTAAT 56% CTTACTAA 60%	TTAGTAAT 84% TTAGTAAT 84% ATTACTAA 84%	GGCCGGCC	ATTAGTAA 72% GAGGAAGAGGACGATGA 72% AACAAGAAAATG 68%	GCTTACTAAT 100%
regulated	TTCTAGAA 75% TTCTAGAA 80% TTCTAGAA 85%	GAAAAAAA 80% TCCCCCCT 65% TAAATAAG 75%		AGAAAAAGAA 85% AAAAGCAAAAGAGA 80% AGAACTTCTCGGAAAA 75%	AAAAGAAAAG 95% CCGCCCCTGC 45%
wnregulated	TTTTTTTT 72% TTTTTTTT 72% AAAAAAAA 68%	CATATTTA 60% TTTTTTTT 60% TACTGACG 40%		GAAAAAAAAAA 76% AAAAAAAAAAAAAAAAG 84% GAAAATTTTTC 88%	AAAAAAAAAA 92%
ean result time	1 min	6 min	3 min	5 min	15 min

percentage of sequences which contribute to motif

able 2

	Parameters (others set as default)
DScan	used 3rd order Markov of whole yeast genome as the background model motif width 8 report top 3 motifs
oprospector	one block motif motif width 8 motif occurs not in all input sequences motif occurs on both stands of input sequences used 3rd order Markov of whole yeast genome as the background model
nsensus	basic version 6c ignore complementary strand desired pattern width 8
gnACE	background GC 0.38
EME	motif width 6 to 10

APPENDIX:

UPSTREAM NUCLEIC ACID SEQUENCES

Downregulated genes:

Gim4
Fpr1
Ssb1
Gim5
Gim6
Gim1
Gim3
Gim2
Egd1
Egd2
Pdr13
Ssb2
zuo1
Fpr4
Npi46
Cct2
Cct8
Cct3
Cct5
Cct6
Cct4
Cct7
Cpr8
Cct1
YnL227

>Gim4
AAAATTGGAAGTGACCAATCTGTAGATCTCACTTTTATTTTTGGCATTTGTATACTAGTA
CTTTCATCATTCACATCATCTTTACTTTCAGCATACAATGAACGTACGTATCAAAAATAT
GGAAAGCATTGGAAGGAGAATATCTTTTACAGTCATTTTTTATCATTACCTCTCTTTCTA
TTTAGTCGAAAGCAGTTAATTCACGAATATCGAGTGATGAGAAAGTCTGAAAGAATATTG
TGTTCAAATTTTGGAGGGAAGATTCTAGTCCCGCGAGAAGAAACTCTCCTTCTTTTCAAT
GTGTTAACACAATATTTTTGCGTCAAGGGTGTTAATATCCTCGCCAGTAAAACAAACGCT
CTGACACTTTCCATAACATTACTTGTAAGAAAATTTATAAGTCTTTTACTGAGCGTCCGA
CTTTTCGATAATAACTTGTCATATACTGGCTACATCGGGGTCTACCTAGTCTTTTTTGGT
GCCTTTATATACTCGCTGGGATCAATTCACCCCAGGCAAAATGACAAGGGGGCGATAAAG
AAAAGTAAATGAACACACATTATAAACTAACCTTAAATCCTACTTGTATGTTCCTTAACT
CTATATTAGTGGTGTTGACGTAAGTTTAGATAAAATCGCGAACGCTGCCGCCTGGAAGTA
AATGAAAAAAAAAAAAAAAAGAATTCATCACCTAAACTCACTAATCTTTAAAACTCATTC
ATGAGAAGTGGAGAACTATATTACAGCTATGGAACAAAGGAACAACGGTATGCAAAATTT
CGATCTCTGGGTTTTAATAA
>Fpr1
GCTGTTGGCCAGATGCTTATTCAATTGGCAAAAAAACTAAACGGTTTCAGCAAGATCATC
GTCGTTGCTTCTCGTAAACATGAAAAATTGTTGAAAGAGTACGGTGCAGATGAACTTTTT

```
GACTACCACGATGCTGACGTTATCGAACAGATAAAAAAGAAGTACAACAACATTCCTTAC
TTGGTGGACTGTGTCTCCAACACAGAAACTATTCAACAGGTGTACAAATGTGCCGCTGAT
GACTTAGACGCTACGGTCGTTCAATTGACCGTTTTAACCGAAAAAGATATCAAGGAGGAA
GACAGGAGGCAAAACGTCAGTATTGAAGGAACCCTTCTATATTTGATAGGAGGTAACGAC
GTCCCATTTGGCACGTTTACTTTGCCAGCAGACCCTGAATACAAGGAAGCCGCCATAAAA
TTTATTAAGTTCATCAATCCAAAAATCAATGATGGTGAAATCCACCACATCCCAGTGAAA
GTTTACAAGAACGGGTTAGATGATATCCCACAGTTACTTGATGATATTAAGCACGGGAGG
AATTCTGGCGAAAAGTTGGTTGCCGTCTTGAAATAAATAATCTATAGCTACTTATGTATA
TAGCGCTTGATCATTCTGTTTACCTCCACATTATAGTGCCGTTCAATTCCCCTCACCCGC
CTTAGTCTTTTTCGGGCAAAAAAAATTGCCGGAAGAGAGACGTCGTTGAAAGAATTTGGT
GGATAAACTCGTGAAAGCTTAAAGTAAGGCCTTTCACCTAAACTCGAGTATAAGCAAAAA
ATCAATCAAAACAAGTAATA
>SSB1
ACTATTATACGGTCACCGATATCAAACTCACACCCACGGACCCTCTTGTAGGACCTGTCA
AAACTGTTCACCACTTTTACTTCGACCTTTGGAAAGATATGAACAAGCCAGAGGAAGTAG
TACCCATAATGGAATTATGCGCTCACTCCCACAGCTTGAATTCCCGCGGGAACCCCATTA
TCGTACACTGTTCCGCAGGCGTGGGCAGAACGGGGACGTTCATTGCCCTAGATCATCTCA
TGCATGATACCTTGGATTTTAAAAATATTACAGAACGGTCAAGGCATTCAGACAGGGCTA
CAGAGGAGTACACACGGGACTTGATCGAACAGATCGTGTTACAGTTGCGCTCGCAAAGAA
TGAAAATGGTTCAGACAAAGGATCAGTTCCTATTTATCTACCATGCTGCCAAGTATCTTA
ACAGTCTTTCCGTGAACCAATAGACAGCTATATAAAAGTTCCTAATTGTGCATTTTTTCA
ATAACAATACTTATTCATCCTTATAATTATATTCTAGCTTCGTTGTCATGGGAACATAGC
CCATACACCGCAGTTATTTATGATCATTTCGAACGGGAAGTATGGATGAATCTTTTTTTT
TTTTTTTTATAGCACGCAACTGAAAAAAAAAAAAAGAAAAATTTTTCATCTTCGCTCGA
CGTTTCTTTTGTAGTACTCATCTCTTTTTATATAAAGATTAATTAGTTATTGTCGCTTTG
CTTTTCCTTCTTTAAAAAATGTTTCTTGCTTTTGGATTTTCAGATGTCCCAAGATCATTA
CAGTATTTTAATTGAACAAA
>Gim5
AATTGTTCATAAATTCAAATATTCAATATATTTTGCAGCTTGTTCAAAATTAAATGCTAA
AAGTAGAGTAAATCCGTTGAACATACATAGTTTTGTAATGTCATTAAGGGTATCTTCAGA
GTTGTTATCGTCTACTATGAATATCAGTATATTATTTTCGTTTCTAGATAAGGGCTGCAT
TCTCCTAGAGATATAATCGACATATAATTTGTGATAAGTCAAAGTTAAGAACAGTACACT
TCTTCCACGAACTAGATAGTCGTAGTATATCATATTAATTCCTGTTGAAGATACATATCT
CCAATTGGTGCTCTTTAAGTGGTTCAGGAGTGGATTTTCTTTTTGGGTAGTGTTCACTAG
TACAGTTTTTCCCGGTCGAGTACTCCTGAAGGGTCTCTTTCGATTATAGTCATCTGTAGC
CTTTGAGTCGGTCCATTCTTCCGGTTTTTGCTGATTGAAGGCGTTTATGACCTGGTTTGA
ATTTATGCGTCTTGATGTTTGCGGTTCCTGCTGCTGGAGTTTTGATGCATCTATTTCTAA
CGATTGAGATCCAGTAGTATCTGCACCACTCTTTTCCTTTCTCAATTTGGCCACACCAGC
CAATATACTTTCAAATGAAGTAGGATCAGTATTGTTCATCTTCTAGGATAACCAACTTTG
TTTATGCCCTTTTTTGTGTTGCGTGGCTGTCTCATAAGTTACAATTTTCTTTTTTGTTCT
AACGATTTTAAATAGAAAAGTGTCTTAGATGAAAATCTTGCTGTGACCCAAACAAGAAAG
AGCACTGTAAAAATCAAGCC
>Gim1
ACTGTGATATGATAGATCTACTTATCATGTTACCAAAAATCGGCGATATACTAATGAAAT
TCTCCTCCATAGAAAATAATAACGTCTTTCCAAATGAATAAATCGGGAAAGAAAAACCAT
ATACTGTTTCATCGGCGGTGTCATTTTTTTGATCAGGATCAACCTCGTTTGGAAAGTTTT
CGTCGTAATTCCACGAGTGTTTACCTGAAGATTGGGGAAGACTTGGACTTTTCACCGTCA
ATGTAGTTTTAATTTCACCAATAGAGGTTAATTGAAGCAGATTTTTGGGGAATAATGGGT
TCATTATCTTAGTAGTAATGACGGCTGTCGAATATTGAGGAACATCCAAGTCCTGAGGAA
ACTCAACTTTTGGAACCGGGCAGACTTCTAAAGATTGCAGGTTTTTTGAAATCTCTGGGA
GATCTAGCAGATGTTTTGGTTCTGGTAAGTCATTCCATTGTTTAAACTTCAAAAAGAATC
AAGTAACATAAGTTAGTCAAACCACTCTACTCTAGCAAAATATGATGGGGAAAAAAGATT
ATTGTAACATACAAGCATCGCTATAAATGAAAATTTTACTGTAGTGCGATGATATCTTGG
TCGCTGTTTCTTATTATTTCGCCTGTTGCCACCTATTGCCACGCAACTTCTTAAATTGAT
TTCAATGAAGTCTGAAATTTTGATGCGAAAGAAAAAAAAAAAGTAAAGGCAATAAAACTTC
ATCAAGTCAAACGTCATCTATAAATAATCCCACTTCAGAATTGAAAAAGACTACATTTGA
GTGGAGGAACAAAAGCAAAA
```

```
>Gim3
AGTAATCTGTCTCATCGTAGAACCATGAACATAGTCATACTCTGATTCTGTTGGCAATGG
TGGTATCGGCCTCATAGCCATAGATGAATCCAACCTGTCCATTTCTTTGTTAAGATCGTC
CCTGTTTAAACGAGGTACTTTGGGAGTGAAAGTCAACTCAATGAATATAGTGCCAGCAAA
TTCATCCCCACTTCTTTTCAGTTCGTACCAGGTACAATAACCCTCCTTCGGGTCTGCTCT
AATAGCATTTAAAAGGTCTATTTCACATCTTCCAATTGGCAGCGGAGATTTTTTCCTTCT
ATCGCAGTAAACTTCGACGTACATTAAAGGCTTTATCTCTGGAGTAATATCAAATTTTTC
TAAATAATGGAAAACAGGGTTTTGACCTGCTCGATGTAGAGTGTTGCTGGCTCTTGTCAT
ATGAGCTATTCGCAACCTAAGCATTACATTTTGCTTGTCCAGTTTATTCAAATTAGGCAA
GTCTCTAGCTTTGCTCACATATACGGAGAGGATTCCTTGGTTTCCATTCCATACTTCTTC
GGACATTCGACAGTGACTCAACGAGTTAACCCAATTCCAGTTGTGAAACCTAAGGTCTCC
GAATTCTCTCGTACTATTAACATACCTTTCTATGTTTTTATCTAAATCATATGAGGAACG
CTAGGCCGTCATTCATGATCAGCAGAGGCGGCTTTTCATACAATTTTACAGGAAGCTAAG
TACTTAGAGGACTTATGCTCTGTTTGGTCCTGCGGAAACTACCAGAATACACAACGCCAA
GAAACTAACAGAGGACCATT
>Gim2
AGCGTACTTTCGAGGTCGTCATCATCGTCATCATCGTGTAAATCTCCGTTTTGGGTGCCA
TCGTCCTCGGTGGCGTTCTCAGATAGTGAACCGGCAGTATCTATCGAACGAGCAGTGGTG
ATGATGATTGGGGTGGCCTCGACTGCACTGGCTGTGGTGTTTTTACCATCTGCTGTCCCG
CGAGTGTCCGCTTTCTTTGTGTGCGATCTGGACAAGGGAGTAATATCCTCCTTGGTGATG
ACCTTATTCGGTACCTGAAATGCCTTGCTGCTTTTATTCATTGCTACGTAAACTATGCTA
TATGATACAAGTGAACGCTCGTCGAGGGAATAACTGCTGCGTGTGCGTGAACGAGTAAAT
GAATGAATGAATGAGTTAACACATGAAAACGAAGATAATCGCAAAACGACAATGAACAAA
TAAACACATCCCTCATTTGCGGCACTGCTCATCTCATCTCATCCTTTATATTAGGAGGCA
ACGAGCTTTCTCGCTCACCAAAAAGAAACGTGCGAAAAAAACCGAAAAATAAAAAAAAAA
AAAAAAAAAAAAAAGAAACGCGCTGCAGGAACGTGACTTCCAAAAAAGGAAGGTGCCGTT
CCAATATGGAAACCGCAGGCACTCATGATTGTGCAGGACGGGTAACACATGATTTGTTTC
GTTTCTTCGATCATTTTGCAGAACTGGTCAGAAAAGAAAGTGAAAAAATCTAGTTTAACA
CTATTGTAGAGGCTATAAATTGTAGACAACATTTACTGCAGAAAACTGTACGCACATATA
TAACAGATACTACGGTTATT
>Egd1
ATAGATGGTAAAAGAGGATATTACACACCTTTATATAACGAAGTACCATCACACCGCCAA
TTACGTTTAGGGGACGCCGACAAGAAGTAAATTCACTATTATGTACTTACAAATATGTTT
CCACCGATAACCTTTCCCCCTTCGACGGTTATTTGAGCAGCGTCTGTCCGCCATTTTTCC
AGAGAAGCTACTTCATCAATCAATTGACTTCGCAATTCCCTTAAACAAAAAAAATTTCTT
TTGCCCTCCCTCGACGATTTTGGAGTAGAGAATACCTTAGGCCTTTTTTTTTCTTTTTTC
TCACGGCTGCTCTTCCTCTTTTTCGCATATTCTATTTTATCATCGACTTCCCTAATTCGC
ACTCGTACCAAAATGTTAAGCAGTATGGCGAAGAACGTGGCGCGCTGGAGTCGTGAATGT
TTGGGTCCTTGATGATGGACTACGGTAGTAAGTATGTAGTAGTTGCAACTTCATATGTTC
ACTTCTGATCCAAGGAAGAGCGGTTATGAATTAATCTCTTGGCATGAGCGGACGGGTAAG
GGGACACCGCCTTTTCTTCGATGGGAATCAGGGTAATGGTATATGATGGATTATTGTGGA
ATCATTTAGTACGGCAGATGTTGAAAAAAAAAGCAGAAAATTTTTGAATTTTTTTCGTTG
ACATTGGAAGATTTCGTAGTGGAAACAGCTGCAATTGCTTGTTAAGTAGTAACCCCTCCT
TTGTCACAAGAGAGCGAATATTCTTTCTAGGGAGGTTTAAGAATAGAACATCTCACACCA
GACGCGACTCATAATTCATA
>Egd2
GTCCAGAAGATGAATATGAAAATGTTGCTCACGAGGATCCATGAGATGAACGCACTTATA
TCGTCTGAATTGAACGGCCGGGTGGACTTTTTCAAGAACCATCGTATGTTTATTTTGAGC
TTAGTAAAAATATTGTTGGTCTGCACCAGCAGGGAATCCCTTTCAGTAATGTATTTCAGG
TTTGATAGCTTTTTTGATTTTTTGTCGGGAGGAGCATCCCTTCCAGTTTTAGACTCATTA
GAATATGCTCTTAGCGATGCCGCTGCTATTGACGGTATTCTCAAGTTACTTCGAAGGGCA
GGGCTTCTAGCCCAATACACATACCTTGCTACAGAAAACTGCCGGGGGCTCCTTAGAAAA
GGCCTGGTAAAAAGGGACATATAGTAGACTGCCAGCTTTCCCGTCCGAATTTACATATAT
ATGCCTACGTGATAATGCGTAGTCATGACACATTTCCGATCAACTCTTACTGAGGATAGT
CAAGTGGATATCCTTTAAGATAAAAAGGGAAAGTTGCAGTGAAATTTCAATTATTGTGAC
CTAATATTAGAAGGGTTTTTAGTTTATCTTCTTTAGGTTTCTTCTTTATTTAATAGGGCG
TAAGTTTTTTCGGTCATATAGTATAATAGGTCTTTTGCCCTTTTGTTAGGTCTGTTTTTT
```

```
TTTTTGAAAATTTTTCATCGCCGAAGAAAAAAATGACAAAAAGTTGAATGAAAAGGATTA
TTTTAAACAAGAACTAAGGTAAAGTCTCCTTCTGGGAGCAACCCTTACAAGCAACCACAC
ACCCACATACTCAATTGATA
>Pdr13
CCAACGCGGAAAGGTTGTCAGGATTATGTAATTTGCTCGAGTTTAGTGCCACTGCGCTAC
ATGGTGATTTGAATCAAAACCAGAGAATGGGATCTTTGGATCTTTTCAAGGCAGGTAAGA
GGTCTATTCTCGTGGCTACAGATGTCGCTGCGAGAGGGTTGGATATTCCATCAGTGGACA
TAGTTGTAAACTACGATATTCCTGTGGATTCCAAGTCCTACATTCATCGTGTGGGAAGAA
CCGCAAGAGCTGGTAGATCAGGAAAATCTATCTCACTAGTATCACAATACGATCTGGAAT
TGATCTTAAGAATTGAAGAAGTTTTAGGTAAAAAACTACCGAAGGAAAGTGTGGACAAGA
ACATTATATTGACCCTAAGAGATTCCGTCGATAAGGCTAATGGGGAAGTTGTAATGGAAA
TGAACAGAAGAAACAAGGAGAAGATTGCCAGAGGTAAGGGAAGAAGAGGAAGAATGATGA
CAAGAGAGAATATGGACATGGGAGAAAGGTAGGCTTTGTAAGCATGGTTTACCTTTTAAC
GCCCAAAAGTTCAATTTAACTTAGAAATATTTATCTATATGTTCATATATACATCTATCA
ATATCACATAAAATTGCCTATCATTTCTTATTCGTTTCCTATTTGTTTCTCGAGCCATTC
ATAACTCATTCCGCGTAATGGGAGCGGTAAAAGATGTGCAAAATAAGTAACTATTTACGA
AGTATTCATATTTTAGCTCTTTTTCTTCGTCGAATGTGATGGTGAAATTTTTTCCAAGAG
ATGATGAGGTCGGATAATAT
>SSB2
GGGTTTTTAATAATACTTTTTACACATATCACGATAACCGCCTACGTATTCTCCAAGAAG
ACTTTTCTCAATTGTTCAAAAAAATCAAAACTAAGGCTTCTGTACTATGTTTTACAGTTG
AGGAAATTTTTCTGACAAACCAAGAAATTTTACCTCAAAACTCAACAGTGGCAGAACTGC
AAAAGAGCACTAATAAAGTACAGACAAATGGGCCGCAACGGCACGATTTCATAGTCACTC
TAGAAATAAAACTGAACAAAACACAAATCACTTTCCTCATTGGAGCTAAAGGAACGAGAA
TTGAAAGCTTGAGGGAAAAATCAGGCGCCAGCATAAAAATAATACCTATTAGTGATAAAA
TGACTGCACATGAAAGGAACCACCCTGAATCTGTTCAACAAACAATACTAATTTCGGGTG
ACTTATACTCAATTGCATTAGCCGTCACCAGTATAGAGTCTGCATTAATTACTTTGGATT
TATAGATTATGCGGAAATGGTGTTGGAAGATACCAGATTGTAAGCTTATTTGATCGTTTC
AATGATGGCTTATACAAGATCACACTTAATTCGTATCATTCGAAACGGAAAAAGAAATGT
GGATTTTCTTTTTCAATTTTTTTTTTTTTCATGATTTTTTTGGGCGACGCAGAGATGAGA
TAAAAAAATTTTTCAACTGCTATTCTCAATCGAATTTTTTTTGTTATCCTCACCTATTTG
TTACATAAATTTCGTTATTTTGCACTTGTTACCCATCGTTTTTTCTTTCAAGAAACCAAG
AACCAATATCCTCATTAACA
>Zuo1
TGTTTCATGGCAGGTTGTAACAGTATCTTCACCGGTAAGAAAATGCTGACGACAATGTGT
AACGGTTGGGACGAAGACAAGGCAATGTTGGCTAAATGGGGATTGCAACCTATGGAGGCA
TTTAAGTACGACAGATCTTGAAGATAGGGATATGTGGATAATTCTACGATTCTAACTGTA
CATTTCTCCCTTATTTATTAAGAAAACCTATATATATATATATTTACCTATTTATTCTGC
CATCGTTAGCTGGCGTTTTATCTTTTATGCATCCAATATCTAATATTACTTCCGATCACG
CATTTAGTTCTGATTACAGCAGAAATCGTAGCGCGATGAGACATTTCATCAAATGGCCTT
TTTTTTTTGGGCAATTTTTTTATATCTTGAAATGATAGTTGCCTTGTACTTTCAACCGTT
CATTTCATTAAGAACTTGACTAAATATGAACATTTCTTAAAAAAAAGGTTGACATATAAA
AATAATCGAATATAAACGATGGAATTTTTATAAAATTAAACACATATATATATATATATT
AACTATAAATATGTCAAAGAAACCATACAATCATAGATTTATAACTATCTTTTGGATGAC
ATTAATGAACATAACGCTCCTAATACAAATGTCCAAAAAATATTACCCGCAAATACGAAT
CTTTTTTTTTTCTCGATGAAATTTTGCAAAGAGTTCGAAATTTTTATTTCAAGAGCTGGT
AGAGAAAATTTCATAAGGTTTTCCTACCGATGCTTTTATAAAATCTTCGTTTTGTCTCAC
ATATACCAACAAGAGTAACG
>Fpr4
ATCAGGTCATCGCCGTTCTGGGCCACTAGGCTATATCTATTCATTTACATTGGAAGCTGC
ATGGTCTACTCCAACCGCACTGAGGCAAAGTCTCTGTGGAGGAAATGCTGTCCTAGCGAG
CCATTGCAATAGGTATAATCAGAGTCGTCCTCCCAATGTGTTAACTTCTCTCAAAACCAC
CATATAAATTCTGTAACCTGCCAGCGATAATGTTATATATGTAAAAGGTAAACATACAAC
CCAATTTGTCCATACTAAGTGGATGCTGGGGAAGCAACTACAGTGCAGATGCTAAAGTTT
GCTGCGGCAAATAATCTCAACATTTTTGTGATTAAACCTGCCTCAGTGGGGTGTATGATT
AGTGCGTTGATATTTAATACACGCGTAAAAGTTAAGTCTGTTGTACGTTTTATTTCGAGT
GACGAAAGAACTACGAATAACTTCCTGCATACTGTTGCCTTTTTCTGTATTTACGTATAC
```

```
TTTTATATCAAATAGCTTGGCCAAAAGAAAGTAGCATTGTGTATCAAGTCATAATTCTAA
TTCTTAGTTTAACTTCAGCAAAAATAATATTCTTCGTCAGACATTCTCATACAATCATTA
ACCCATATGAACGTTAAATTTCTCAATTTCTTGCCTTCGATGCCCTATGTATATTTTTAT
CTCAAACCCGAGGCAGTGAAAATTTTAGCTGGTGCATCGAAAAGAGAAGGGCATTGAAAA
TTTTACAGTGTTAAAAGAAGTGCTTTTCATGTTAGAATGATAAGAAAAAAGAAAAAGACA
ACAAAATATACGTCTCTACC
>Npi46
GAAAAGCGCCATAGCTATTATTTAACGTAACTTACCAGAAGTCCTTGAGCAGACGATTGA
GCATTGAATAAAAAATAAAATTCTTGTAATTCTTGGAAGGAAGAACGACACAAATAGATA
AATGGCGTATTTTCTGAGCAATGAACTGTGTTCCATCAAAAACATTTTATCGCACACCAT
AGGCACTTAGAAAAACAAACCTTTGTGCCAGTATTTGAATGAAAACACTTTGGTTGAAAT
GTATTGTGAAATTAATCATCGAGCACTTTACGGTTGGCAACGACACATTACTGACGTTCT
TTCGCCGATTTGTTACTTTAATATATTGTACAAAAATGCATCAAAGAACATAAAAGAGAG
CGATCGCACAATCAAACTTCTTTTGGAAATAATCTGTCATCAATGCTCCAAACATTCTAC
TATCACCTGTTCATGATATTATGGGGAGAATATGCCAAATGTATATGACACATCGCATAC
TTTCTACCTTTTTCAACTGGCTTGCTTCCTCTACCTTTCACTAAAATCTGGATTGCATCT
ACATCTATATAGTCAAAACCGGCAAGAACAGAGATTATGATACATGCTTCATGTTCCATT
GATTAGAATCCACTAGTATTAGCTCGCTGTTCAATTTTCAGCCCTCTACGGACTGTCGTC
ATTCCTAAAGGACGGCAAAATTTAAAAGTGAAGGTAAAATTTTTAAATTCGAAAAAAGAG
CCTACTAACACGTTTCTATATAATACATAATTGTGTGAAAGTTCATACATAATTGAAAGC
AAGCATCCAACCAGCCCAAT
>Cct2
CAAAGAGTATGCTGTGATCTTGTATTCATGAATATGGGAGGGTCTACTTATAGGCTCTGC
AATGGTGACCAGAAAGTCCTGAGCCTGTTCTGCTAATGGAGAGAAACTCTCCAGAATGAT
CCTACCGTCACTTGGCGAGATCCAAAGGGGCCTGGAAGCATGGTCTGGTCTTAACCGCAA
ATAACTAAAGTCATGGCTTCTGAACATTCCAGAGACGGAATCTGGAACGAAATCTGCTGG
TATATCGCTATTAGTGTCCTGCAAAAGCGCTTTATCCTTAGCGTCCATTTGATTCATAGA
GGAATCCGCCGCTGTATGCCTGCTACTCTTGGTCTTTTTCCGAGGCTTCTTCAACCCTGT
TACCATGGCGCCGGTATCCCTTTCTCCACGTCCTTCTGACGTTTCCCTGTTATCATCATA
GTTCTCATCGATCTCTGCATCAGTAGCTGGTGAATCCTCATCTGATGAGAAAAAGGACTC
ACCCATATCTGGAAAGATTTTCCCCTTGCTCTTAGGTTGGTAGCCTTCAACGTCCGTCAT
GATTTATCTTCCACTGATTCCAACGTCAGTAGATGCTACTTTTACTAAATAATGTAGATA
TCATTACTATTATCTATCCTTACTAAGGACCTTTTTCCTTATACTTCTTAGAAGAAAAAC
CTCGTTTCTAACGATCTGAAATTTTTTTCTCTTCGCGGTGGCGAGTTTAAAAGACAAGAG
CAGAGCAATAAAAAAGAGGATGAAAGGTGGCCATTAAATAGGCAAGAAGAGTCAGTCGTA
CAAAGCTCAAAAGATTCAGG
>Cct8
ATACTAGCCCGTAGTATTTTACTGATGTGTTTCCAAGCTAGGGCATACTGAATATGTTCT
CTGGGCGCGTGTTTATATGCTTTTCTTATATACAATAAGAAAAGTTTACTTCTATTCTCG
GTAAAAACATTCTTACGGCTTGACCACTATTGATGGATATGGTTTATTGCTCATAGGTAG
AGGTCAAAAAGGCTTTATTAATGAAGATTGAATATCGATCTTCTTAGACTTATATAAGGG
GGATATAAAACACACGCCGATTGGTTCATTCCTAATTGTGACCAATGTAATAGTGGTTCC
GGCAGTTCTCATACATTTAATGTGAAGACCTGTATTGCTCATATATATGACTATCAGATC
ATATTCAAATAATGGATTTTTAATTCTGCTTCGGTACTGTCTGAACTGCTTCTTGAACAT
TTCCTTCTGTAATAGGTACATTGACAGACATATGCGTGGTGTATACTAGCATAAAATGTA
AGCATCGAAATTGTTACCAATTGAATTCATGTTGATATCTAATGATCACTTGTCCGAACA
TTTTGTATTATTAATCAATAAAGAAGCTGCCGGATTTAGATAATTAAAGAGTTGCGAGCA
CATTTACTATTTCAATAGCTTATTTTATATAAGCCAAGTTAATGGTGTTTTTGTCATTAT
TAACGGTATTCGAAAATGTTCATCACCTCCGCTAGTGGCAAAAAAAAGTGAAGAAAATAG
TATAGCATGATGATGTTCTCACTTTACAGATACCGACAAGACAACTTGAACCTCAAGGGC
AATAAAATCATCAATTAACC
>Cct3
AATGATCAGGTTCATTATTCCTTTGATGGATTGTGCAAGACTTTATGGGCGTCCAGTTCA
ATTCAAGACAGAAAGAACCGATATAAATTCGTTAATGTTTGGATTACCTAAATTGCCCTC
AGTGGATTATTTTGCTGAAGAGGAGATTGCCCATTTGATGACCCAGGAATTTAATCCTTT
GAAAGAAAAAGATACTGACGATACTATGGCTTTAACTATGTCAAGATTTTCTAGCTACCT
GCAAGAAAGAATGACAAAAGTCTCCAAGAGAAACACAGAACTCAACTTCAGGACCTTTAG
```

```
TGATGTAATGGGCATGTATAATACTACGAGGGATCACTCTAAACTTAATTGCATGAGTGA
CAAGGATAATTCTGTGAAGGAATTTTCATTGAGTGATAAAAGCAATGTTAGTTCTGAAAC
GACTAATATGATGAATCAATTGCAAGTAAACGCCCATGAATACAAAAGTTCTATGTGCGA
ACGTCCTATCGTCGCAAGTACTTCGCCCATTGCATAGTTCTTCGCATATATAGCCACATC
TCAAAAACTTACATACAGTGTATATATTTGTGGTACAGTTTCTGTATAGCCAAATATACC
CTGATATAATCTTATGTGATATTATATACTTCCATCCTAATACTAAACTTTTTCGTTTAA
CGTGACGCATGGTGCGAAGAAAAAAAATAGCAAATCGCCACTCAATAAAGTGAGTAGAGA
GCATTATATCAAAAAAGCATCGATAATATCGAGTGATCATACGTTGAAAATCATCAAGTA
CAATAGTGATATAAGCCACA
>Cct5
GTGAAGGGCATTGCATGGCAAGTGTAAAGAATCCAAGAAGAAAAGAAAGGCTTAAAGCAA
TCCAGGTTAGTACTTATAAGAATGTCTGTTGTAGGATCGTTAATTTTTTGCTTGGACTGT
GGTGATCTCCTGGAAAATCCTAATGCCGTATTAGGCTCTAACGTTGAATGCAGCCAATGT
AAAGCCATATATCCCAAGTCACAATTCTCCAATTTAAAAGTCGTCACCACGACGGCAGAC
GATGCGTTTCCATCTTCTCTTAGAGCCAAGAAATCCGTGGTTAAAACTTCTTTGAAGAAG
AACGAACTGAAAGACGGCGCTACTATCAAGGAAAAGTGTCCTCAGTGTGGAAATGAAGAG
ATGAACTATCATACTTTACAGTTAAGATCTGCAGATGAAGGTGCTACTGTCTTCTATACA
TGCACTTCCTGTGGTTACAAGTTCCGTACCAACAATTGATTGGGTTGTGTTCGCACATAT
TACAATTTGTACATTAATATCATCCCCGCTTTCATCATGTGAACATCTATAGAACATCAT
TTATAAAAACAAAAAATCTATGTACTGACGATATAATAGCGTACATTATCATATTATTCT
TACTATTTATCATACTCATCGCATCTACTGTTCTCGAGAACGTGTAAAGTGATTTATGCC
ATTGAATGAAAAATTTTGTTAAGGAACAACTAGAAATATAATAGTGTCTTGGTGGATGTG
GCAAATGAGGTGTGCGTGATAAAGAGCTACAAACTCAAAGTAATAACAGCTATCATCCAT
TTTTATAGCAATATTTGAAA
>Cct6
AAGAAGAAAAAGGAAACGAAAAAAAAAAAAAAGTAGGGAAAAGCTGAGTGATGAAGATCCG
GGCTGCCAAGATTGGGTTTGAAACGTGTAGTCAGCTACAAACCAAAAATTAAAAGAACAA
TAGCAAAGAACCTTAGTAGGAAAACAATAAAATAATAGACACGACACCCGTAGAAAGATA
GAACGGGACGAGGGGCCGGTTCAACAACTATCATACTTTCTTGTTCAACAACGCGATTAA
CCCAAATAATAAAAAAGCTACGAGACAAGAACAGACGAGGCAATACGAGCAACGAGCGCA
ACGAGAACGGTAAGCATGGCAACGGCGCGCACACCCACGCCCCGCAACCGCACACCTCAG
CTCAAAGTAACCCACTTTAACGTGACTGGACCTGAGCTGTCTGACTCGGTCGTTTGCAAA
TAGGTTAACGCCGTCGCTTCTCCCTTTCGTGTACCGTATATCTTTCTTTAATTGCATTCC
CTTGTTGTGATTATCGTTTCCCATTGCAGTGAGCTTTCATTTTGGGGACCCTGGGAGTAG
GCTGTTTTTCGATAGAGTGAGAGGTGTGGGTTCTGCGGCAGAATGGATACGGGCCGTGGA
TCACTGTTTACGGGAGTTTTGCTTTGGTGGCAAAATTTTCGCTCGGTGATGTGTTAGTAG
TTGTTATTCTGGTAAGAAACGAACGAAAGACGGAAAGGATTATGTATAACAGTGTATGTG
TGCAGTGTGTGGAAAAGGTAAAGGTTAAGGGTGTGGAAGCAGTGAGAAGCAGAAGATTAT
AATATAAGTAGAGTAGAAGC
>Cct4
CTGTGGATCAAAAGCTCCACACGATAAGTGGTCCTGGCCCACTTGACTGATATGGCCTGA
CCCGATCTTAGTGGATCCGATTAACTGGATGCCAGAAAGCACCGTGTAACGATACTGCTC
TCTGGGAAATTCCGCCCATATCTCCTTCTCAATATCGATTCCGTTCTGTACCAGTAAAAT
GTGTGTCGTCAATTGCGGCCCATGTAGCTCTTTATTCTTCTCAAGCACAGGTCTGATGAT
GTTGGACACCCTTGAATTAACGGGCCCATCTATGATGTTCTTGGTAGTCACTACAATATA
GTTGTATCCTTGGTTGTCTGCGGCAGAAGCAGCGTCCTCGACGCTGCTGTAAATATGGTG
GGGTCTCCAACCTTCTAACCGGCCATAGTCACATGATTCGATTGTGTAGCCATGCTTCAA
TACCCTATCGTAGTCAGATCTGACTACCAAGGAAACGTCGCTCTCCTTCTTAAGCCACAA
CGAAAGCGCAGTAATGACCCCAACGCCACCTGCCCCGATAACTAGTATTCTAGACATTTT
CGACATTAGCTTATGCCAACTTGTATCCTGGCACCGGGACTGTAGTTTATGAGGTATATC
AATATGTCCTAATTAAGGCTATATAATTTGGTATTACTATCACTTGCCACGGGGAAACTG
AAATTTTTTTCGTTTCGCCACCAAGAAAGGCTAATTGAATACAGTTCAAGAGGTTGATAT
AGCTAAGTGGGCATGAAGGTTATGAATTGTAATTGGAGGTGCGGCAAGTGTGCAAGACGG
GTAAAGGAGAGTATAGAACG
>Cct7
GGATTTAAGATCCGGAAAGGTGATTCGTACTTTGAAAGGGCATACAGATGCCATAACGTC
ATTAAAATTTGATTCCGCGTGCCTGGTCACAGGTTCATACGACAGAACAGTTAGAATCTG
```

```
GGACTTAAGAACTGGGTTGTTGAATAAGTTTCATGCGTACAGTGCGCCAGTTCTTTCGTT
AGATCTCTTCCAAGAGAACGCTGCAGTCGTTGTCGCAGACGAACCAAGTGTCCAGATATA
CGATAGCGAAAAGGATGAAAGCTGGTCTTGCGTCGAACAAGGTAACGAGACTAGCGTTAG
TACCGTTAAGTATAAAGAAAATTACATGGTTGAGGGTCGTGAAAATGGGGACGTAAATAT
TTGGGCCGTATGATAGGCATTGCATTGCAATCAAGTCAACCCATTTGAAGAAAAATAAAT
ATCGAGTGCATATGTGAATAGCTGCTCGCTCAGATATATATAAATAAATATTCGTTCTGG
AATGTAAATAAAGTATCGTCAATACCTTGGATGGTTTATCAGCGAAATCGGTGCTGTAGA
TGTTGCCCTGTCATATTACTACAGATGTCTTGGAAGCAATCACCATACCTGCATTATCCT
TACCCGGAAAAAGAAAACTTTCGTGAATAATGATTAATATTTAAGTGGCGACTTTTTCAT
CCTTGCTATTGAAAAAAAAAAAAAATGAATAATTGTTATTAAGCGTTAGCAGATGCCAAAT
TGTCGAAATGATTTCAAATACAGAGAATTAGCCGCAAATATTAAGTAGTACCTGAATCAC
GAGACATACAACTCGGAAAA
>Cpr8
ATCAGTTTGAAATGGAACTGTTGGTGGGGATGTCAATTCGTACCTTAGACGACGCTAAGC
AGGCATTTAAACTATTTCATAAAAAGTACCCTCGAGTAAGCCGAATAGTTGTCACTAGTC
TGGAATTATCTGAATTCCTGAGCAACGACACTTACGTAGTAGCAGGGTTTGATTGCTCGG
CCAGTGAAGAGATATTCTTCTATGAAATCCCCAAGATCAACGCAAAATTTAGTGGAAGTG
GTGATTTAATCAGCGCCATGCTCACTGACTCGTTACTAGGCGACCGCCGCTGCACACAGC
TGTCATTAAGCGCATCCTTGGGCCAAGTGCTGTGGCTCGTCACCAGTATCCTACAAAAAA
CGTATGATTTAAACATAGCTGAGCGGGGCCCACAAGATTCCACGATAGATATCAAGGATT
TAAAGCTGATTCAGTGTAGAGACATCCTAAAGCAAGATCTCATACCCTCCATTGGCAAGC
CAAAAACCATTAAAATATGATTCTTACGAAACGGTATGCTCCTTCCTCATTGGGGTGCGG
CTGCTGCAGGAACCTGCCTAAATAGGAAAATACATTCCTGATTTGGAAACTGTAGGCACA
CGACGCGTCCCTTTGGAGAAGATCGATCCTTCTTTGGGTCTTTCAATTTGATTGGAAAAA
AAATATTTAAAAAAATAGAAGAAAAATATAGAATATATAGAATATATAGAATCAACTACT
GCAGTAGATAAATAAAGAAGACTATTCATATTCAGTATCACAATTGAAAGATTATTACAA
TTTGTATACATTGGTTCCCA
>Cct1
AATTGAACACCTTGAGAATGAGTATGAACGTGACATATCATGAGGTGAGAATAGCAACAA
TAACTGCATTATTGAGAAGAGTTTACCACTTTATTGCAACTCAAACATTAGGTCCCAAGG
ACGCTGTGGTGAAGGTTTTTAATCAGTGGGGACTGTTGTTCAAGAGACAAGCCTTTGATG
AAGAAGAGTATATTGATTTGATGAACATCATAATGGAAAAAATTGTAGAACAGAGTTTTG
ACAAACCGGATTTGATTCTATTTAGTGCATTGGTTTCTCTATACGATAATGACATAATTG
AGGAGGATGTCATTTATAAATGGTGGGATAATGTTTCTACTGACCCTCGCTATGATGAAG
TCAAGAAATTAACTGTAAAGTGGGTTGAGTGGTTACAGAATGCTGACGAAGAATCTTCCT
CAGAAGAGGAATAAAAGTCCACGCAGTACCCTGCATAATCATGTATTACCATAAAAAGGC
TTTAATATGAGAAAAAGGTGAGCGCTATATAGATTTAGTAATTTACATGTTACAGAAGCA
AAAAGCAAATATATATATATTTTTTATCATATTAGTTTTTATCCTCAGTAGTACTACTGA
AATAATAATCTCAGATTTTTTTTTTTCACTTGCCGAGGGACCTTGTCCTAAGTGGCAAAG
AACTGAAAAAAAAAAAAAATTGAAATGCGGTGGTATCACATGAAGAGTAAATAACTTAAGC
AAAAGAGAGGCCACTGAGAGAGTACAAGAAAGGGACAATAACAATAGACAGAAGACTTTT
ATATTTGTAATAACTTAGCG
```

Upregulated genes:

Ssa2
Hsc82
Sba1
Ssa1
Yro2
Sis1
Hsp26
Hsp104
Ssa3
Sti1
Cpr6
Hsp82
Ynl007W
Cph1
Cpr8
Yfr041c
Caj1
Xdj1
Afg1
Ypr061C

```
>SSA2
TGCTGAGCTGCTTATTTTCATTTTGGTCCTTTTCAGAGCCATTCTGTGCTTCCAAAGGTC
TTGGATCATACTTGTTGTGTACGCCATCTTTATTAAGTTAAGATACGAAAACTCCAAGTA
CATGAAAGCCGCCTTTGCTCAATGGAGAGTTAGAATGGACGGGATCATAAGCCACCCATC
TATCCCACCATTCGTCAAAAGAGCTTATAATGCAATTAAAATGAGCTTGATCCGCTTGTC
TGAGTACCGCTTAAGCGGTGCCCCACAAGTGACCAAGAAGCAGAATTGAGTTTTATTATT
ACTTTTTACAAGTTCAATACCTTCTAATTACTTTGTATAACATTCCACTCCTCTCTTTTG
CTTTTTGTTACAATTTGTTATATATTTTTTATCAATTTTATATCTTTATTCCTTGAATTG
AGTTACTCAACCAACCCTTTTACGGTCTCACAGTGGTGAATTGACTGCATGGCACTGCAT
TCTTACTCTCTCTTGGGCCCTTATTCGTAGTATTTCCGAAGCTTTTTTTTCCGAGAAGTT
CTTCCGATTCACTTTCTCGAAGTTTTTTTCTAGAACATTCCATCGAAGATTTCGAAAAAA
AAAGCTTCGAGGTTTTGACGAAATTTCATTCTTAAGCACTATGGGAAGGTGTGGACGAAT
TGTTATATATAAGCCGCAATTGGGCTGGGTTTTCTCCAAAAAATGTTGAAAATATAATAC
TCTCTTATTTAAGTTACTTCTATTCTTCAATTGATTAATTCCAACAGATCAAGCAGATTT
TATACAGAAATATTTATACA
>HSC82
GTAAATTGGCTTATCTGATTATAGATACCTGTCTTAGTATTGTGAGGCCAAATGATAGTA
AACCATTGGACAATAGATGGCGTATGTCATCCATGTCAATCTTAGGTATGTGCCTTCAAA
TCAACGCCAGGGGTGTCTCGGATAGAATTCGCGATATGCTGGATTGTGTATTTGGAATAT
TGCAATTAGAACAACCACAAAATCATTTGAAGGACAAGGATGATTCTTTTTTGATGAGAA
GAAGTGCCGTACACTTGATACATGATTTGTTGTACAGTACAGGATTTGATTTACTTCCCT
TTGAGTACAACTACGATAAGCTAAAAACACTATTATCTTACGTTCGTGATCAAGATGAAG
ATTACATGGTTTGTGAACAGATTGACAAGCTTTTAACCGTACTAGATAGTTTATAACCCA
TTACGCATTTGATTATAATTTGCTTCTTAGGCAAAATTAATATTTACGTTCTTTTATATT
CTTTCTTTTTGTATTCATAGAACAGCAGCCATTACCAATAGAAAGTTAAAATAGCCGCCG
ATGCATTTTATTACCCGCCTTTCTGTTTTCTGGGCACTTTTCTTTCTAGAAGGTGAAAGA
ACAATTTTTCTCGTTTTCTCGAACTTCCACCAAGCGTTGGGTAATGAGGGAGGAGATTTG
TATAAAAAGAGTGGCATGTGAACTGCCTACCGTAAGTGACATGAACACATGCATTATATT
TTTTGTGATATATTCTTTCTCTTGTTTTCTTTTTCTTGAAACGCTACAGAACCAATAGAA
```

```
AAATAGAATCATTCTGAAAT
>SBA1
ACATGATAATATTACTCTAACTTCTTTTCTTCTATACTTTGCCATTCATTACCCTATCAC
GTTATCAATCCTTGCATTTCAGCTTCCATTAAGACCGATAACTGTTTCCCAATCTCTATG
TCATCATCTTACACCGCTGTGATAATATTCTAGGAAGACGAATACTAGCCAATAGATGAT
ACAAGGATTTCATTTCGACAGAAACAAAAAAAGCATGGAAAAACGAATGGAGACAAAGAG
GAAGTTAAAGAAAGTTCATTTGTGACCTCATTTACTCAGAAGAATTTCGTAAATCGGTTG
TTATAGAACTGTTTTATTGTTTAAAAGAGCTTGTTATAGTAATCTAAGTGGAAATACACT
AACAGTAAATAGGGCGTGTGGCGTAGTCGGTAGCGCGCTCCCTTAGCATGGGAGAGGTCT
CCGGTTCGATTCCGGACTCGTCCAATCTTTTTATACTTATTAATAATTTTTTTCCTGCCG
TTACTTGCTTTTAAAATAACTGCCTTTTATGAATACAGAGTATAATTTTTGATATACAAA
GAGGTTGACTGTGATAATCAATACTTAATTTGTGGTTATTGGTACACATATACCTACAAA
AGTTACCAACAAACTGTTCGACTTTTAATGCTACCCGCCTTCCGAGTGTTTTTGAAGGGG
CGGAGAGGAGCGGCAAGAATTAGCATGGAAAAAAGCATAAAAAGACGAAATGGGTGGCAA
TGTATTAACTTGTTCGAGAAACCTAGTGGACTCAATTCATTACAACAACAAGTTCCCAAG
ATCATCGATTCATAATAGTC
>SSA1
CCAAAGAAGCGTTTAATTGTTTCAAAGTTTCAATCTTGGAGAAATCGGTGGATGCCATTA
TGTGTGTATATATTCGGTTGGACTCTTCAGTGCTGAAAGGTAGAAAGCATGAGGTAAAAA
GATAAAGAGTAGAAGGAAAAAAAAAAGAACCAACTAAATATAGTGGGTAGATAACCTTAA
ACTATTTATTTAGAGGCGACAATGTGAAAAATTTTTTCTTCAAAGGCTCGGTTGTCGACA
AATTGTTACGTTGTGCTTTGATTTCTAAAGCGCTTCTTCACCTGCAGGTTCTGAGCCCTA
AGAAAAAAATTTCCTTGGTTGAAAATGGCGGAAAAAAAAAAATTCAGAAAAAGAAATAAA
GCACGTGTGCGCGGTGTGTGGATGATGGTTTCATCATTGTCAACGGCATTTTCGTTCTTG
TGGATTGTTGTAAACTTTCCAGAACATTCTAGAAAGAAAGCACACGGAACGTTTAGAAGC
TGTCATTTGCGTTTTTTCTCCAGATTTTAGTTGAGAAAGTAATTAAATTATTCTTCTTTT
TCCAGAACGTTCCATCGGCGGCAAAAGGGAGAGAAAGAACCCAAAAAGAAGGGGGGCCAT
TTAGATTAGCTGATCGTTTCGAGGACTTCAAGGTTATATAAGGGGTGGATTGATGTATCT
TCGAGAAGGGATTGAGTTGTAGTTTCGTTTCCCAATTCTTACTTAAGTTGTTTTATTTTC
TCTATTTGTAAGATAAGCACATCAAAAGAAAAGTAATCAAGTATTACAAGAAACAAAAAT
TCAAGTAAATAACAGATAAT
>Yro2
CTTCCGTGATTTTCGATTCTTCAATTTCCGCATAATTTGCCTCAAATAGGAAATCTGAGC
TGTTTTTTTCTTACCCTCGAGTGGTCGTTCTGCGGACTTTTCTCCGGTCCAATGCATGAA
TCCCTTTAAGAAGCATGAAGAAAAAAAAAATAGCCTCTTGTAAAAGCAAAAAATGCCTTT
CACTATTTCTGTGTAAACGAATAATTTCGTCATCTTAGTAGTTCACTGGATAATTGTGCT
TTCCAATCAGATTAGTTAGTAAAACCCCTAGAAAATTTCAATGTTAGATCGGGATGATTC
CTCATCGGGTTGTTGCTTATTCATGATCCTTTCAATTACTGGAAGCTTACCGAAAAGATA
TATAAAGAACCGAACTTATTCGGTTTACATGTAAGACAGTGTGTTAGTTATGGTTTTCCG
CCCTTTTCTTCTCTTGTTTTCTTCGATATTTGTTATTAACCTTTTCTCATTTACGATTAT
ATAATCAAGTGTCAAGATTGTGTCAGCAAATAACAGTAATATTCCAACCCGAAAAAAACT
ATCTATTTATCTTTGGATACTTTAACATAGACAAGAAGTCCTAAGAACTTACCATAATTT
TTTTTCTAAAAACCCTTTTTTACTCAAAACTAATAGCTTATTCTGGTTATTTTTAATCAA
TCTAGAATTTTTTTCTTTTAATTCCAAACAATCTAATAACGTTTAACCACTAATTCTTTA
ATTATTAATATTTAATTTTAAAAATCTTCTAATTTTCCTGATAACTTAAAGTGACACTAT
TTTTTAAAAAAAGCATCAAA
>SIS1
TTAATGGTCCCATCTTCGCTAGAGGTCACCATCCATCTGTTGTCCTGTTGAAAAGAAACG
GAGGTAACGTTTCCCCGATGACCCTCAAAGGATGCTACTGGATTGGGATTTGTCGTCCGA
ATGTCATAAAGCCGTACGTTTTGATGGCCTGCTGTTGCCAACAACTTTTTATCGTTGGTG
ATTTCTAAGCGATTAACCTGTGAGTCGGAATGCTGAATTGTTCTTGAACAAACCCCAGTA
AGAGCCTCCCAAAATCTTATTGTGTGGTCATAGCCAGCAGATACTAAGATAACAGACATC
AATCAGAACTCCTTTTTTATATGGATGACAAGACTGCCACGTTGATCGCTGCGCTAATAA
TATATGGTTGTCGCACTCTTCTATAACATGATCAGTAATAGCTAGTCGTTACTTTTCTGA
ATCTTAACGTATTGTGTTACCCTCTTTGGTTTATATGAACGTTCCAGAAACTTCTGGAAA
AAGAATGGGATATCCTTTTACGGGTTTTTCAACCACTGTCAGCACAAACTCCTTGCCAAG
GTAGTTCGAGATTTAATATATATTGTATAGATTTCTCGTCCCTTTACCGTATTGTTAATC
```

```
CCTTTTTAAACTTATGCGCTTTCTTGTCAAAAATTCTAATTAATATTATTTAAGCTCTTA
CACGAATTCTTTTAACGACTTAGAATTGTATAACCATCCAGAAACTCTTCAGAGGATAAA
GGATAAGTTGTTTGCATTTTAAGATTTTTTTTTTAATACATTCACATCAATATAATAGAG
TATAGTATACAGAACTAATA
>HSP26
AATATTCTGCGCACATCAATCATTTTCTTACTACATACACTAACATTACTCCTAGTTTAA
TTTAATTGAATTTTTAACTTTCTTTTCTTTTCATTTGGCAATTTGGCTCCTTGAAAACAA
GACTATGGGTCTCTCTCATAAGCCTCAGGGGGGGACCCCAAAAAAATAACGCGGCCATCT
TGCATGCACCGTTGAACCTGTAGCTTACAGTAAGCCACAATTCTCTTACCTTCTTGGCAA
TGTGGCACAAAATAATCTGGTTATGTGTCTTCATTTGGTAATCACTGGGATGTTACTGGG
GCAGCAGCAACTCCGTGTGTACCCCTAACTCCGTGTGTACCCCTAAAGAACCTTGCCTGT
CAAGGTGCATTGTTGGATCGGAATAGTAACCGTCTTTACATGAACATCCACAACCAACGA
AAGTGCTTTTTCAAGCATTGCTTGATTTCTAGAAAGATCGATGGTTATTCCCTCCCCCTT
ATGCGTCCAAAAATATAGGGTGCTCGTAACAGTAAGGTATTCGCACTTAGCGTGCTCGCA
ACACAAAATTAAGTAATATGCGAGTTTTAGATGTCCTTGCGGATCTATGCACGTTCTTGA
GTGGTATTTCATAACAACGGTTCTTTTTCACCCTTATTCCTAAACATATAAATAGGACCT
CCATTAGTTAGAGATCTGTTTTTAATCCATTCACCTTTCATTCTACTCTCTTATACTAAT
AAAACCACCGATAAAGATATATCAGATCTCTATTAAAACAGGTATCCAAAAAAGCAAACA
AACAAACTAAACAAATTAAC
>HSP104
AGACCAAGAAAGGCGCTGATCACCCTGAGTCCGAAGGCAATCAAGCATTTAAGGGCACTG
CTAGCTCAGCCGGAACCTAAATTGATTAGAGTTAGCGCTAGAAACCGTGGATGTTCAGGA
CTAACGTACGATCTACAATATATCACCGAGCCGGGGAAATTCGATGAGGTAGTAGAACAA
GATGGCGTTAAAATTGTCATCGATTCAAAGGCGTTATTCAGCATCATTGGAAGTGAAATG
GACTGGATCGACGACAAGTTGGCCTCTAAGTTTGTCTTCAAGAATCCAAACTCCAAGGGC
ACATGCGGTTGTGGCGAGAGTTTCATGGTTTAAAAACCTTCTGCACCATTTTTAGAAAAA
AAGAATCTACCTATTCACTTATTTATTCATTTACTTATTTATTTACATATTTATCATACA
TATTAACATTGAACCCTCCATCGTGGTAGTGTTTGCTGTTCCTAACTTTTCTTTCGTTGT
TCTTGTAGATATATATTTTTCCAGAATTTTCTAGAAGGGTTATTAATTACAATCTTAAAC
GTTCCATAAGGGGCCGCGATTTTTTTGTTCAATTTTCAACAGGGGGCCCATCTCAAAGAA
CTGCAAATTATATCACAGTAAAAGGCAAAGGGGCGCAAACTTATGCAACCTGCCAGATTA
TTATATAAGGCATTGTAATCTTGCCTCAATTCCTTCATAATTCGTTCCTTTGTCACTTGT
TCCTTTTTACCCTTGAATCGAATCAGCAATAACAAAGAAAAAAGAAATCAACTACACGTA
CCATAAAATATACAGAATAT
>SSA3
TAAACTTCAAATCTAGAGAAGCCATAAGGCCTGGCCATGAAATGGAGGATTTTTTAGACA
AGTCTTACTACTTGAACACTGTAATGCTACAAGGAATTTTTAAAAATTCAAGTAATTATT
TTGGGGAGTTGCAGTTTGCGTTCTTAAATGCCATGTTTTTTGGTAACTACGGGTCGAGTT
TGCAATGGCATGCTATGATCGAACTGATATGTTCAAGCGCTACGGTGCCTAAACATATGC
TCGATAAATTAGACGAAATCTTATATTATCAGATAAAGACATTGCCTGAACAATACTCAG
ACATCTTGTTGAATGAACGAGTTTGGAATATTTGTCTGTATTCGTCATTTCAAAAAAACT
CCCTACACAACACAGAAAAGATAATGGAAAACAAATATCCAGAATTGCTTGGTAAAGACA
ATGAAGACGACGCTCTTATTTACGGTATCAGTGATGAAGAAAGGGATGACGAGGATGATG
AGCACAACCCTACCATTGTTGGCGGTCTCTATTACCAAAGGCCATAACGATCATCGTGCG
GCGCTATCATCAAACGTATTTGACTTGATGCCTATGGAGGTTATGGGTGCCCTTAATTAG
GGATCGCTGTGGAAAGTTATAGAATATTACAGAAGCAGCCACAAGGGTGACCAGAAGATG
GTTAAGGGATGTATCATATTGCACAATTGGAAACGAATGGAAGGGTATATAAAGTGACTG
AAATTGGTAGCATAAACATTCTTGTATGTCAATGTTTGTCACTAAACGGATAGAATAGGT
ACTAAACGCTACAAAGAAAA
>STI1
CAAACTCCTCTCGATTGATCGTTGGTATGAACAATAGTCAGGTTCAATGGTTTCGCCTGC
CACTCTGTGAGGATGATAACGGAACAATTGAAGAATCAGGACTGAAGTACCAAATAAGAG
ATGTCGCTCTTTTACCGAAAGAACAAGAAGGTTATGCATGTAGCAGCATTGACGGGCGAG
TTGCTGTGGAGTTTTTCGATGATCAGGGCGATGATTACAACTCAAGCAAAAGATTTGCAT
TTAGATGCCACCGTTTGAATTTAAAAGATACAAACTTAGCGTATCCAGTAAATTCTATTG
AATTTTCCCCCCGTCATAAGTTCCTATACACGGCTGGCTCTGATGGCATAATTTCATGCT
GGAACCTACAAACCCGCAAGAAAATAAAAAATTTCGCCAAATTTAACGAAGACAGCGTGG
```

```
TTAAAATTGCTTGTTCGGACAATATTCTATGTCTGGCAACTTCTGATGATACTTTCAAGA
CAAACGCCGCAATTGACCAAACTATTGAACTAAACGCAAGTTCAATATACATAATATTTG
ACTATGAGAACTGATATCTTCGTGAAGATTCGTGTAGTATGATAGAACATTCCAGAAAAA
AAATTCAGATTCATCGCTCTCTCTTCGCTTCTCCTCCTTTAAGGAATAAAGAAAAAATCA
CATACATAGATTAAGTAAATAGGATCTGCTAGAAAAATTATATATAGATCAATCATCTTA
TTAAGGTATCTTGTTTAAGCCCAAAAGTCTGCTCCCAAATTCCTCACTGTAGCTACTAAA
ACAACCTATACGCAAGAAAG
>Cpr6
CCAAATCAAGCAATTTATTGAAGTATGGGTCCCTGATGTTTGTCCTTCTCGTTCTTGCTT
TCCCTCCCCTCTTCTCATCCATTCGCTTCATTTATGCTTTTATTTTCTTCTTTTTTTTCT
CACAACAAATGTCCGAAACAAAAAAAGGGGCCTCAGCCATTAAAGGTAGTGAAAGGCACT
TCTAAGAGTTGCGCCAAACAAGGGAAGATTAAAGGGAAAGAAAAGCTCTGGCAGGCTAAA
ACTGGCAGACTGGTCATGAGTGAAACTGGCGAAACTTCAGAGTATTATAAGCAGGCGCTG
GAAGAATACAAAGAGGTCCAAGAAGACGAAGACCCGGATGTGTGGGACACGAGAATATCC
AAGACCGGATGCTACGTAGAGAACCTCGCATTACAGCTGTGCCATGCCGAAACAGGTGAC
TGGAGGCAGTGCTTCAACGAGATGGCGTTATTTAGGAAGTGTTGGGAAAAGAATGGTAAT
AGAGAGCGCGTAAGCACAGTGGACGTGGATGGGACGACCAGTAAGGATTCGGAAAAGAAG
AAATGAAAATCTAAATGTCGTGATGTATAACTTGTATATAATAGACAGCTGCAGTGATCG
AAACACATTGTTTCCCTTTATAGAACATAACTGTTACGCTTTTGAACGGCATTTCTATGA
GCTTCTAGAATATTTTTCCGCCCTAGCTGGAGAAAGTTCAGACAGAAAATTATTTAAATA
AGTCGAATATCAGAGGTGCTGATGCGCTCACATCACATAGAAAACTGGTAAGACAATATT
CAGGCGATCAAGGAGTAAAA
>HSP82
GAGCGCGAACAATTTTGCCTATCGTATTACAATTCGTTTTTGTTTTACACAAATGATAAT
TAACTCATCGATTTTTTCTCAGATTCTCTTTTTCTTTTATTACTTTTCCTTCTTATTCTA
TGAAAATTTCATATCATCGTTCATGGTGATATTACCACGGTGACATCAAATGGAATTGTC
CTATTTTTCTTTTCCTGTAATTAGTGGCGCGGGACTTTTTATTCCTCGTGGCAAAGAATG
GCTACTAAGAGGAGAATAAGAGATGGTTAGAAGTATTTGGAAGCTTGATGACAGAATATT
TTAAAATATAACTTAGCTTGCGTGTGCGTATTTGTATTCTTCGTGGTTACGATATGAACC
AACGTGCAAGCGTGTGATATATCACATTCCGGAGGGTGTCACCCCCCCTCTCTCAACACA
GTAATCCATAAACCAGTTTTACATACACGTAAAAAAGAACAGGAATAAAGCTTAATCGGA
TTATTAACTCATACGCTTGTCACATATTGTTCGAACAATTCTGGTTCTTTCGAGTTTCGC
AGAACTTTTTGAATTTTTCTTTTTTTTTCTAGAACGCCGTGGAAGAAAAACACGCGCATGG
TTTTATGAGCGGTTAATTCTCATCTTAATACCAACCAGGTCCTTCCGCCACCCCCTAAAA
CATATAAATATGCAGCTTATCCCTTCAATTCTTAACATCTGTGACCTCCTCATTTCTTCC
CGCTGTATTAGAGTTCAAGAAATCATACCTGATAGAAAATAGAGTCCTATAAACAAAAGC
ACAAACAAACACGCAAAGAT
>Cph1
AATCCTCTTTAATTGCGATATGGAGCTGGAAAGGTTGGTGTCGATGTTGAGGCTGGTGTT
AGCCACAGCGGCAGCGTCCGCGTCAGCGGCGTTGCCGCTGATGGTATTCATGGTAGTACT
ATCAATAATGTTCTCCTTCTCATCGATAAAGCTCTCCAGAAATTGAAGAACCTCGTCAGT
AGAGACATGTTGGGGCAGCTGCGTGGCATGAATGACCACAGGCGTGTTTAGCGTGGTAGC
AGTGTTATTACCTGTTCTTCTAGAACCTTTCATCATCTTGTATTATCGAAATCTTACCTC
AGACGTCTCCAGTTCTGCTCTGCCTTACCGATGAGACCTCTATAATCAATGTGATTTAAA
CATATGCATCATCTTGAAGTAAGCTCGAAATGAAAAATTTTACACCCGCAATATGCTTAC
CCGTCGCTGTTAGATGTTACCCGCCTTTCGCCCCTTAGGGGTTTCCACCGCCCCCCTCGT
CCTCTACGGTATTTTCTGGGGCGCCCAGCCCGCCACGTCCCGCCTCGTCTCGAATTTCTT
CCTCTGCTAGACGGGCTTCCACGCGCTTCCACTCATTTCTGTCTCTGGTAATGGCCGTGG
CCCTTCTCACTTTGGTTGGGCTTACGCTGACAAGTGTCTGTTCGATTCCCTGTATAAATA
TAAACGTATTCTCTTGAGCCTTCTATCCTTTTGCCACTGTCGTCATCATTTGTTCCTCCT
TTTTCGCTAGATAGGTTATATTAAGATTTGTCTTGAATTTAATATCTCAACTCAATCCAA
ACTCAACCGCTAATACTACC
>Cpr8
ATCAGTTTGAAATGGAACTGTTGGTGGGGATGTCAATTCGTACCTTAGACGACGCTAAGC
AGGCATTTAAACTATTTCATAAAAAGTACCCTCGAGTAAGCCGAATAGTTGTCACTAGTC
TGGAATTATCTGAATTCCTGAGCAACGACACTTACGTAGTAGCAGGGTTTGATTGCTCGG
CCAGTGAAGAGATATTCTTCTATGAAATCCCCAAGATCAACGCAAAATTTAGTGGAAGTG
```

```
GTGATTTAATCAGCGCCATGCTCACTGACTCGTTACTAGGCGACCGCCGCTGCACACAGC
TGTCATTAAGCGCATCCTTGGGCCAAGTGCTGTGGCTCGTCACCAGTATCCTACAAAAAA
CGTATGATTTAAACATAGCTGAGCGGGGCCCACAAGATTCCACGATAGATATCAAGGATT
TAAAGCTGATTCAGTGTAGAGACATCCTAAAGCAAGATCTCATACCCTCCATTGGCAAGC
CAAAAACCATTAAAATATGATTCTTACGAAACGGTATGCTCCTTCCTCATTGGGGTGCGG
CTGCTGCAGGAACCTGCCTAAATAGGAAAATACATTCCTGATTTGGAAACTGTAGGCACA
CGACGCGTCCCTTTGGAGAAGATCGATCCTTCTTTGGGTCTTTCAATTTGATTGGAAAAA
AAATATTTAAAAAAATAGAAGAAAAATATAGAATATATAGAATATATAGAATCAACTACT
GCAGTAGATAAATAAAGAAGACTATTCATATTCAGTATCACAATTGAAAGATTATTACAA
TTTGTATACATTGGTTCCCA
>YFR041c
TTGTTAATAGTTTTTTGTATGCTGTTAGTAGTACAGGGAAGTATGCTAGTAAAACCACCC
TGTTGATTATGTCCAAAGTCGGGGTAGAATGATAGTTCTTAACCTGGATATTTAATAGTA
GCAATGCAAATTCACTCAGGTAGAGAATTGGCAATGTAACAAATGTGTTAACCTTCCATA
CAAGTGTTCTTTCCTTTACGCCCTTACGTTTAAAAAGTTTGCACCAATGGCAATATGCCT
GTGATAAAGATAAGGAAATAATGAAACTAGAATATGATGCATTCTTGAGGATCAATGCAT
GCTCATAATCGTCATAAAAATGGATCACGCTGACTAAATAGATAAAATTTAGGTCTTTCA
AAAACGTTCTTGAAAGCAAACCATATTTGAACCCGTGGAATCTAAGCAACCAAATAACTT
CCATGATGCTGGAATAAAACATCAAATATATCAAGAATTCGTGGATTCCTCCCGGTAAAA
ACCGAGACCCTACCAGTGGCATTAATATCAGCCACCGTATCAATAATGCAGCATACAAAA
AGCTTGTGGCTAATTGATAATACTGATAAAGCTTATGCGTCAACTTGATACCTGCCATCC
TCAATGCTAGTTAAATATATGTGACCGAACAAATACTTTTATTGACGTTGTATTTACATT
ATTTCCCAATTTGAAGGTAAATCATATTTAATTTCCATATTCAAAATTAATAATCACGAA
AAAAATGAGATTTTCAAAATAACACGTACCAAAAGAATAGAAAACGAACAAAAAATCAAG
GTTACAAGATCACTAAAAGA
>Caj1
TATTCCTTCACCAAATTTGAATACGAATCAGACAATTCTAAAATTACGTCATATTTTAAC
AACACATATCATAATATTCCACCAAAAGTCGCTTTTGATTAATAAATGATACAGTATTTT
TTTCCCAACGAACGGTCAGTAACATATCAGTAGACTCTAAACACATTTTGTATCAGTACA
CCTTCTTACACTGCTAAAATGTTTTGGTCTTTTTTTGTTTACATTAATATTTTCATTTAC
GCAACAGAGAATTGCCGCTAAATATGAACCAGCTACAAACAACTGGGTATGTATTAAATA
TAAGAATGCGAATTGATACCCTTATTACATAATTTTTTGATGCTCATCAGTATCAACAAA
TAGAATATAGCAAACGGAATTACCGGACATTCTAACATTTCCTGGAGAATATCTGCTAGA
GTCAAATCAAATGATGAAGACTTTCGGTTAGCGTTTCAGAAATAAACAATGATGATTAAT
CCAAAAAATTTCGTATTTGAGAAGTAAGACAGTGAATCGAAGTGGTTCTCCTATGGATTG
TTTTAGACCCTGGGAACGTCACTGCGGTGTTTTCGCCCGTCATACCAAGATGGGGCGGAG
AACGGAAAATTTGATGGCGAATGTACACTGCCAAGGACAAACAGGGACGATAAAGTGTTA
AAGTGGTGATATAGTCGGATCCGCGGTTAGCCGTAACCATTGCTTTCCAATGGTTTACGC
CTGTAAAAAGAGCAACTTAGATAATTGTCGGCTTCACGTAATTAAGACGTATTAACTCAA
AGGAAAAGAAAAGAGGGAAT
>Xdj1
TGATATGTGGCATTTTTTCTGCGCTTGAAGTGAGAGTCCTTATATGATATGACTTAATAA
CGTGAGTTATAGAATTACTCACTGTGAAGTGATTTTTGGCAGACAGTGATAACATCGTGG
AGAGAGCAAAATAAAGACTGGGGAAATATATTTATATGCGTTCAATTGAGAAGCAGAAAC
AATTAAATGCGGGGGAGAAACAAAGAAGCTGGTCCTGTTCTCGAGCAACGTGTGACTAGT
ATGAACAACTAGAGTAAAATAGCGAGAAAGCAGTAGTAGTAACGGAAGAGCTTTTCCTAG
TATGGTCCGCGGGTCACGATCGTTTACAAAGGGCGCAGCAAACTCTAAATACGTCAACCA
CAGCAGGTCAGAAGAGTACGTGAAGTTGTCTCAACGAAAGGTCGAAAATCGTTAGGCTTG
GACCCCTTCAGAATGAAAAAAAAAAAAAGACGCAAGAGTGAAAAAAAAAACCAGTACTGTAA
AAGCAGAAAACGTCTAATGGGAAGTGCTACTATGTAAGATTATATCATGAATGTGCTGTG
TATTATATATGGATTATATAGCCACGAATATGCGGTAAATAAACATATGCATATATGTGT
GTGGGGCAGCGCTCTGCTGTGCAATCCTTGTTTCTTTGTTTCTTGGGTCGGGATTTGAGT
AGCGACGAGCCATCCTCTGCAGAAGGGAAGAAGAAAGAAAGGCAGAGAGGCAAAGAAAGA
GTCAGGCTTATCTTATCTTAAGAGATATGCCGCTTGGACCCAAAAGAAAGGGAAAAGTAA
GTACGCAGGTGTAGTTTTGA
>Afg1
GTAGTGGGTAAAGTCTGGTAGTCATTCAGCCAGGAGATAACACATCTAAGTAATCTATCC
```

```
GCTATGGCTGCTACACTTGCCAACGGTGGATCGTAGGCCACGTTCACATCAAATTCTATT
TCTTCTGCTGTCAATTCTATCAGACTGGAGTCTAATTTGGAGTTGTTGACCTCTAAAGAA
TGTGTACCCTCGAACAAATCAAACCTGGGATCTTTCACGATAGCCTCTGGTTTTAGTTTT
GAACACAATTCATCAAACAGACTTGTGACGTCAACCAATTGGTCAAAATCATCAGTTATG
CTCAAGGAACCTAATATACTGTCTACTTCCATTCTCGGAAACGTACCAGCTATTATCAAA
GTTGCTCTACACTAGCACAGTTGTTTTCAATCTCAGCGCCACTCTTCTTGCTATTGAAGA
GCTGCTATTTTACTTGTATCTCTTCGTGAAATTTCACTCCATCTCATAATTATTACCCGA
AAACATAACAATGTGAAATAATCGTATTATGTAAAGGTTAGATAATATAACAACTACTAG
ATAATACAGTACTCTTATTCATTTATTTAGGCACCTGATTCCACTATTACCTATTCAATA
GGTGTGCCGAGTTCCTTTGTCTTACGTAATTTTACACATCTTTTATGGGTCATATTTCGA
TATGTAAGGTAATAACGGGAATCTCCGGATTTACTTTAAAGTGGGAGGGAGAAGTTCCTT
TCTTCTGATTTGCGGGTGCTCCGAAAGAACAAAGAATAAATAGGTAAAGCAGTGAAACAT
AGTATTTGCTGGTTAAAATC
>YPR061c
ATATATAAAATTAAATACGTAAATACAGCGTGCTGCGTGCTCTACTCACCAATATCTTCA
AACCAATCCTGAGGTCTTTCATCGATAAATTGTTTCATGATCTTTTTACACCTCTCATCG
TCAACAACAACAACCTCGTGACCTCTAGTTTGTAAATATTTCTCGCCCTTACTTTTGAAA
TTAACGTTCTCACCGACAACACAGCGTGGAATACCATACATGATGATGGCACCTGTACAC
ATGTCGCATGGAGACAGCGTCGTATACAAAGTGGTATCTTTGTACACTTTGCCCTCTAAT
CTCCCACAGTTTTCCAAAGTGGAGATCTCACCATGTAGTGTGGCGGATCCCTTTTGAAAT
CTCATGTTGTGACCACGACCGAGAACACTTCCGTCTTTGTTATTGATAAGACATCCGCCA
ATAGGAACACCACCCTCTTTGTAACCTAAGGCCGCCTCCTCATAGGCAATGTCCATACCC
TTCTGATCCCACTTGCTTGCCATTCCCCCTGTCACCATTAGCTATGAAATTTTTAACTCT
TTAAGCTGGCTCTCATCAACATAAAACAATGGCTTGTAGATCTCATTGTCGAATTGAAAA
TGAACTTTTTAATTTCGCAAAAAAAAAAAGAAAAGTTCAGGGTCTCCGGGTACTACCGCC
CCTGCAAGGAATGACCGGTGATAATGGAAGAGAAGATATTACAGCGACAGTATAATTAAC
AAACACGTTGTGTCTTTAGGCTCGTAACTAACTGTACGTATATCTACACACCATAGGTCA
CTACTTGCAATATATTTGTT
```

Genes with Yap1p response element (5’-TTAGTAA-3’)

Trr2
Gtt2
Aad6
Aad4
Isu2
Gpx2
Bet3
Trx2
Lap4
Flr1
Ttr1
Ydl124w
Ylr108c
Yll055w
Yil167w
Ylr460c
Ycr102c
Ynl260c
Ykl071w
Ydr132c
Aad15
Yol118c
Ykr071c
Yml131w
Ynl134c

```
>Trr2
ACCCATCCAACTACTTGACGAAGACTCCACGGAGCCTGAACTCGACATTGACTCACAACA
AGAAAATGAGGGACCCATCAGTGCGTCAAACAGCAATGATAGCACTAGCCATAGTAATGA
TTGCGGTGCCACAATTACCAGAACAAGACCTAGACGAAGCAGTTCTATCAATGCAAACTT
TAGTTTTCAAAAGGCTCATGTCAGCGATTGCACCATAGTCAATGGCGACCATGGAACAAA
GTTTGCTGTCTGGAGAATTACCGTATTTCTTGAACCCAACTTGAAGGCTTTTGCGGCCAA
GAGGGAAAGCTATAAAATCCAAACCTATAAACGATACTCCGATTTCGTCAGATTACGAGA
GAATTTGCTCACAAGAATCAAGACAGCGAAACCTGAGAAACTTAACTGTTTGCAGATTCC
ACACCTTCCCCCTTCAGTGCAGTGGTACAGTTCTTGGAAATATCAAGAAGTGAATCTGAA
CAAGGACTGGCTGGCAAAAAGACAGAGAGGGCTCGAGTACTTCCTCAATCACATCATCCT
TAACAGCAGCCTCGTAGAAATGACCAAAGATATACTCATACAGTTTCTAGAGCCTTCAAA
ACGAGTTGCATAGCTCACCATCCCTATCCAACCGACTATTCTTCTCATCGACTACTACTA
TCCCATTTAACTCGGGCGCGTTGTTAATTAATCACTCGATGGGGAATGCCTTGAGCTGAC
CGCAATGAAAACTTTTAGGGGATCGTCCAACATTAAAGGAAGAACGAAACGGACTCCACA
GTTTCTAATATAAATAAACA
>Gtt2
CACTCGCGAATCCACATTTTCACGTGAGCTTGCAACAGAGAATAATAGCGGAAACTTTCC
CGTATTACTTAGATATCATATCTTCAGACAGCTAGATATCGAAAATGGAATTTTATTATG
TCAAGTTCTGCAGGCGCTAATTATTGTTCAAGTAATGTTTTCGTGAGTCTAAAAGCTCGT
ACAATAGTTCTTTTTTTGGAACCCATTATGAAGCGGAGTGGATGAATAATACGGAATGAA
AAGGGCGATAATGGCAAACAATTACGAGCAAAATTCGATGGATGCCGTATCACATTTGCG
```

```
ACAAAAAAATGATCAAAGATGGTAATAGAGTAAAATTTCACGAATGGAGGGCAGATACAA
GAGCGAGATGCCAATAGATCGCAGATTAATGCAGCTGCATTTTCATCAGAAAGAGGTGCG
CGTGAACTGTAGGAACAGTAGTGCAATTGGGAGTACGTTAACCACACATTTTGAAATCAA
GAAGGAAAGTAGGCGCTGGAAATGACCCGAATCCTATAATCGGACAAGAAGGACATAGAA
ATCGGTATTTACAGTAAATATTCAAAAGACCTTCTACTACCGTGTGCAAAACAGGGAGTG
GATACAAGTGGTTTGATCAAATGGAAGTATTTGATACTTCAGTTTCTCCGTTCCGTGTCC
GTTCCCTTAGGTCGCGCTTAGTGCTCGAATTTCTTTCACTGCTTACTAATTCATTCGTTT
ACGTCTCAAATGATTAATAATGTGCTTCCTGCTAAAGGATAGAAACTTGGTGCTCTATAT
AAAGTACCTACAAAGGATAC
>Aad6
AAGGATTCTATGTGGGAATCAATAAAAAATAATAAAAAAAAGATGGCAGAAAAGAAGTGC
GCAGGAGCACATATTTGGTGAAAGCGTGACTTTAATCAAGGGCTAAGTAATTGATAATCG
CAATCATCCGGCATTTAATTTTCTATGTACAAAATTGTAGAAAATGATACACCGACCTGT
TCTCAAATACAGTGCTAATAATCACATGGTTCGTAGAGGGCGTGAACAGGAGAAGGAGCT
GGCTTGTCGAGTTTAAAATGTGCGTACAGGAAGCCGTAAAGATCTATATTGCATCTGCTT
GCCTTGCTGGTTTTAAAAAGCAAAACAACGCACGGATTGATAACAATATTTAGACTTTCT
GGTTTGCCACCGCTCCTATAGAATCTTTAGCCTTTTCAAATGGAATCTGAAAGAATTACT
AGGATGGAGTAACGCAAAATGAACTTTGATATTACGAAGAGATAGTACCGTTCCGAATCA
ACACAGGTTCTTGCAGTTTTGTTGTGGAAACCGTAGAACATGCAGTGGCTAATTCGTAAG
CCAATTTTCACATTTCGCTTACTAATTTTGCTCTCAATTAGTAAGCAAATCCGTATCGAA
GGTATACAACCACGTTCACCGTTTTCACGTGCACAAATTCGAGAAAATAAGCACTCAATC
TTCAGACATATAAAAGGAGCAATACCGTCGCAATGAGAAACAATTTCCTCTCTGTTCAAC
CTAAATCAGTCAATCAATCGATATAAATTCAGTACTAACTGAATATCTCTTCTATATCAT
TTTCAATAGCGATTTTTAAT
>Aad4
AAGGTTTTTTTAAATAAGGCCGAACCTAAATGTTGCCTTCAAAATTAACCTGTATCAAAT
TTTGAAGCCGAAGTTGGTATCCTCACATTGTCTAACCTCATGGTCTCCTAAAGTCCGGCG
ATCACTTTCCGTTGCGTGACTATTGATGATCAGACTGTCAACAATGTCGTTCCGGGCCTC
CATAGCTTGAAGAATTTTCGTTGACGGAACTAGAGGAATGGCAGATGGAGATGACCTTAG
TTTCATGTCGCCTTGATAGTTATCAAAGAGTACAGAAACAAGACGTCTCTTGCTCAATAA
TAAGGATTGGATCAGGTGCTTACAAATTTTATTCGTCATACAAACAATTACTAATATTTT
TTTAATATTAAGCGTACCCCTATTTGATTGGCTCTCTTCTAAATTCATACGTTTTAAAAT
CCGCTTTCAACCGTCAACATAAAATTAGTAAGCAGACCTCGTTCCAAGACCCAACTAAGT
TGCACTAAATATCCTCGAACAAATTGGGGGCAGATTTTTGAACCATATAAAAGTCATGAC
AATATTTCCAGATCAAGGACAGTTGCAGTTCTTCCTTTCATCGGCCAGCTGACGAATTGA
CTGAGCGGCTTTGTATTTCACTAGAATACACTTCCTATAGCTACAAAATATTATTCAATA
ATATTAGTGAATTATTTAAACCTCTACCTGAACCACCTACCGAGTTGGGACGTCTCAGGG
TTCTTTCTAAAACTGCCGGCATAAGAGTTTCACCGCTAATTCTGGGAGGAGCTTCAATCG
GCGATGCATGGTCAGGCTTT
>ISU2
CCCCTTATACAACCAGTCCAATGGCACATCTATATCAGTGTACTTCAACTTGGAGCTCAT
GTCAGACTAGATTCAACCAGATACAAATTAATACATATACAATGGCCAAAGAGAATGAAG
AAAAAAGGAAGGGACTCAGTCAAATATTAGACGACAGTATCACTGAAGGAAGGGAACTGG
TGCGAGGAGGAAAGGAGAAAATACGTAATGCAAGTGGTAATAAAACGGGAACAGGCTAGT
TCTCGAGAAATGTGGCAACTGCAACCAAATGGAAGAGTCAAATGTAAGTTATAGAGGGTT
CGCTTTCCACCTTAACCCATCGAACTGTCAATCATGTCGCTTCTTTAACCTGCCCCTTAA
CGTTACTAAGCAAAAACAAAAGGCAGACAAACTCAAGAAATAATTAAGCATAACCCGGAA
AAGGAAACGCCAAACCACTATGGAATTATTCTTATTTCATTACAGGGTCCGCACCTTCCC
AGTCTAGAAAGGCATGGTATGGAATTATGACTGCAACTTTATAGTGGTCAGATGTTTATA
TATGTACACCCATAATACAGATACACAGCATACTGCATATATCCGAGTCACCATGATTCT
ATTGTTTCAGAGTCCGGTTCAGGCCCAACGTTTTTTGGCGCCCTCGTTTTTTCCCTATGC
GGAACCTACTATTAATAAGCACCTGGGGCCGGCCGGTGCAGCAGTATAGCAGTATATAAC
AACAGCAAAAGAAGCATATGTAGACAACGCTGAGGTTTCATTACAGCTCATACATCACCA
ACAAACATAGCACAGCATAG
>Gpx2
TGGGCCCTGATTTTTGGAATAGTAGATTAAGCATGTGATAAGTGCCAGTGAAAAAAGTCG
CAACATATAGTCTACTTTTTGTTACCGTTCTCCGCCTTGCTCTACTTGATAAAGTTTTTA
```

```
TTTACATATTTCACAGTTATATACACATGTTAACAATTTAATTTTCGCGCTGTTTAGTAA
AGCATTTTTCGAGATTGAGGCAATATAAAAGGAAAGGGCCATATACAGGAGATGAAGGAA
ACCTGTACTATTATTGTATTAAACACATGGTAGTAGCTAGAAAAAATTCATGCATCAACC
TCTGTGTTTGTATTCGTAACATATAAATCCACAAGAAATTTTTTACTCATCATCCGCTAA
AAAAAGGAAAATACGGACGTTACTTACTAAATGTACGACGAACATGTACACAAATCTTGA
CTATTTTTGATAGATGCAGTCTCTTGAAATGACCAACGACTCAGTAGTGTTTTCCAAGCA
ATTTTACTTTCAACTGTTTAATATTTTTTAGAGGCTACTCGGCCGGCCATATCTAGACAA
TTAGTAATTCCGATGACAAAGAGTTTAAAACCACAATATCTGCCACACATATCGTGAAAG
GCATTAGTTCCATTGTCTCCGTATTAGTGCAAAGCCCCTAAGCCAGCCTGCTTTAAAAAA
TGTGCAGGTACTTTTGTTATTGTTTTGATATATAAAGAACTACTATTTGCTAATATGTGA
TGTAGCGCAGAGAACTTGCATTTCCCTGCATGGCTGACCTTGACAAGAATACAATTAGCA
ACAAAAAGTAATAATTCAAA
>Bet3
GTGGACCCTGTACCAATAAGCAAATGAACCTGATCAGTAGCAAAACCAGCAGTAATCAAG
GGCAGTGATGAAGGCAATGTCACTAAGTCTCGTACCATTCTTTTGAAATTTTCTTAGTTT
ATTTATCACTCTATCCGTGCTTTTTTTAATTGGGATGGAATATCTGAGCAATTCTAAATG
AAGCCTACAACTTTGCAGAAACACATAAGTAAATTATATATATATATCATTTTTTTGACT
TAAGCATACTCGACTTTTGTAGCGAGTTTATTAGCGATGATTCACATAGTGAAGCAAATT
TCATTATGAGTCCGTTCACAGTTTATTATGAGAGCCTTTTTTCACCACAATTTGTAATGT
CACGTGCATAATGGCATTTCTAGAAAGAGAAAAACGTATAAAATGATTCAAGTTTGACGT
ATCGTATATAAAAAGGTCGAGAAACCAATAACGAAAATAGAAAATGATATCTTTACTGTT
GCTGTAGTTGTGATTTCTACAGTTGCGAGCAAATATGTTGGACTCAGAGAACGAGGATTG
TACATCTTGGATATAAATAGATAAATAAGCGCTATTCTATGCACCTTGACAGCAGGAGCC
AAAAATTTGAAAACTTGCGAAACTCCAAGCCAAAGGTAAGGCCAACCTGAAAGTGTTACA
TGTTACCCGGTTTGGCCAACATTGAACTTAATATAATAATAACAGAGTAAGTGATCGTAG
CTGTGATTACCCATAGCTATAGAGATGGGAATGCAGAGTAAAGCTTCAAGTGTTCATTGA
TAACCAAAACTGGGTCAAAA
>Trx2
AATGTTTCTAAAATCATGCTAAAGTATCCTGACACCAAAGCTGTACTGACCAGTGCAGTG
ACGGAAGTTTTTTTGAGGAAATTGAAAAAGCAAGGTTTTATATTATACGAGGAGGGCACG
GAATTTGTTGACACATTTGAGGACGAGCCAGACAAACTAAAGCCACTGGACGACAAGATC
CTTAACAGAATTGAAAATATCAGAGATTTGGTCCTGTATAGATTCGGATCTACGGGTGTT
GTTCAGGTTCTGCAAGCAGCAACGGACATCCTTGGTCTAATACCGGTGTACACTGTGAAG
AACATCCAGACTTTTACGGGTGGCAACGGAACCAACGTATTTAGAGATTGTTTTTTGGTC
AAGCGAGGAACCCCTGTTGGCAAAGTTGCCAGGTATATCATGGGTGGCGAGGTCACCATT
GCAAGCATTGAAACCGTTGGCGGCGTGAGAGTCAGTGAAGAAAGTCTTGTTGAGCCCGGT
AAGAATGACATACTCGGCTTCAAGATCGCTCCAAGATCAGCATAACTTGAGTGCCAGTGA
ATATTAAGTAATCATCAAAGTATATGTGTAATTGTTTATACTCTTAGTAAAGGATGCTCC
CTACAAGGTGGCTCTTTTCTTACTAAGCGCGTTCAGTTTCCAGCCAGCCGAAAGAGGGAT
ATCAGTATATAAGAAAGCCATTCGGGGGATGAAAAGCTGACAAGAGAATAACGAGGACCA
GTTTTTATTTGTTGTCTAGCAAGAATTATACACGCACACATACACGAGAGTCTACGATAT
CTTTAAATAACACATCAATA
>Lap4
AACCCAACGCGAAAATGACCACGCGAAAAGAAAAATTAGTAAAAATATTTCACAACGCGA
AAAATAAGGACGAAGGCAACACATTCCACCTTCCCTGTTATTGTAAAACGCACAAATGTG
TGCTTTATTCGCTGAGGTTTCATCGACAGCTGCAGCCAGTTGTTCATTCGGTGTTCTCTT
GTGCAGCGGCTCAAATTGTGGCAGGCTTGTAGGCTCCGAACGCCGTAAGGTACTTTATGT
TTTTAATGTCAACGTTTGCAAGAAAATGTCATCCTATTACTTTCTGAGGCATGATAACAT
TGTAATCCCTTATCTCTTGCGCTTGTTGGTATCAGATAAAGAAGCATCAAACAAGAACCC
ACTCCTTCCTTTTCTCATGGATAAAGAGAGGTCGCACCATTTCGTGAGGAATATGAATTA
GTCAGCTTCTACTTTAGGGTATAGGTTGCAAACTTCCGTTTTTCTCAAGGTTTAATCAGA
AGCATAATAGGTCTTCGGATGATTTCTTGTCTAATTCAAAAAAAAAATAATCTGGGATAG
GTTTTTCGGGAAAAACAACTAGGGGTGATGCTGTTTTGCACCTCTATTACTGTAAGATAG
GCTATGCAAGCAAACGACAGTAAGAAAGAATAGGGTTAAGTGAAATAGTTGACACACCCC
TTAATCGTTCCATTAGATAATAAAAATTCATTTATAAGCAGTTGACTTCACAGAACAAGA
CATACACTGTATAGAGCCTGATCAGTAATCTTAGTGCAATTGTAGAAACCTGCACAACCA
ACAAAATTAAGAAAAAAAGA
```

```
>Flr1
AAAATAAACAAAGGTTTAAAAAATTTCAACAAAAAGCGTACTTTACATATATATTTATTA
GACAAGAAAAGCAGATTAAATAGATATACATTCGATTAATGATAAGTAAAATGAAAAATC
ACAGGATTTTCGTGTGTGGTCTTCTACACAGACAAGATGAAACAATTCGGCATTAATATC
TGAGAGCAGGAAGAGCAGGATAAAAGGTAGTATTTGTTGGCGATCAAAAGGAACATTGAT
ACTCCATATACTGATGGCCGGCAATTGTAAGCCCATTGACAACAGAGAGACATATTACTA
GCACATTTTCTGCAATATATTATGCTAAACCTCTGATGCTGACACACGCTACCGCGTATA
TGCGCGGAAAGGTCTCTTCTGTATAGAAAATTCTATGAGGTTCGCTTCTTCATTTTAAAA
TACGTTATGATGGTGATTACTAAGTATAGGAATGCCCTGCGCTTTTCTCTACAATATTGC
GCATCACTAGTTTTTCCGGAAGATTATGAAGATTTCAAATGAAAGAAGACAAATGCAAAG
AAAAAAAAAAGGATATTTGCTTAGGTAAGGAGCAATAACAGTGCGAAAAGGGACATGATA
GTACACATTTGAGTAAAATTGGTATCAATCATCTGACTAATGGGCGGGATAATTAGTCAG
GTAAAAGGGGAACAGGAAAGAAGACGAGAATTGCATTTAAACAGCTCTTTACGAGGCTAA
AATATCTACATCCTTATGCCGTCTTTGTCATTATGGTGAAGCTATTATAAAAGAGTATCC
GTTGAACGACAATCTCCACT
>Ttr1
CTTTTTAGCTTCCTGGCAAGATTATCTCCACATGATTTCCAACGGTAAATGGAAAGATGC
CACTTTATCTTCGGAAACCCTGGAGAAACTATCCCCTGAGCAAACAGTTCAACTGTATGA
ATTGATGAAGGAAACACAGAAACTTCACCAAGACAATGAAATAGAATCAAGTAAGGACGT
GAAGCGGAACAACAAGGATTAGTGTTCGTCTTTTTATTGCGCTTGTAAATAAGAAATCCA
TATACTATATAGTACACATGCTAAAAAAAAAAAGCTAAAAATCTCACTTATTTGCAGTGG
TCTCCCTCTCCTGCCATATAACCCCACTGGTATTTTCCAATGCCTTATTGTTGGAAACCT
GATCTTTATACCATTCCTGCACTTTCACAGGGTCATTGCCGTGGATAATACAAAACTTGA
ATTTGGACACCTGCTTGTCACATGATGTAAAATCTCCATATCTGTAATAGCTTCTAAATT
GCCCTCCAATCGAATAGCAACTCGTCAGTTGATCAAATGCCTCTCGGCAACTCATCGTTG
TCGGATATTTTGTACTCATCCTTTCCTGTTTCTTCCTCAAGCTGCTCTCTTTTACCCTAA
TAGAACCATCGCCTCCCTCTTGATTTATGCTAATACCACATCCAATAGCAGAACTATTAC
TAAGATCCGATATTTCGGCCCCCTTCGCAAAGGGGCCCGCCGCACTTTCTTCATGAATTT
TCATATAAAAAGTCCCAGGACGCCAAGAAAAGGTGCCCTCTTGATTAACGGACACTCCAA
CTACTGTTATATATTGTTTC
>Ydl124w
TATCGCCATCATTGGTGTGAGTGTAAGCTCTTTATTAGTAATGTGACATACTTTTAATAA
TCTTGCAAAAAATACAGGCAGCATCAAGCGTAGCAGGAGCAGACATTTTTGTTGTTGTTA
GGTACGGTGCCGATATCAATGGCTCCATCAATGTTTCTTAAAAGTCTGGAAGGCGTGCTG
ATTGTCCTTTTACTTCATAATGTTCTCCGCTTTCCAGTCGAACACACATCACATTACACG
ATCGTACAGCAATGCTTCAGTAGGGCGGTAACTTCTTCAGAAGAGAACGGACCCGCCATT
ATTTCTGCAAGGAATATTCTGTAACGATTCGCACCATATCTTAAAGCTTCTCAAGCTTAC
TGTTTTATGTGACCTCCTTTTGTACTTTTCAAGTTCTGTAGACAAGTGTCAATTCAACCT
TTGAGGCTACGTGGCCAACAAGGTCTTGCGCGATTTACCGGTGTTTCCTTTGCGCCCCCT
GCTCATGCCATCAAAACGGCACGCTGTTTTGTAGAAGCCTCCATGCGTCTCCTATTTGGG
TCCTTGCAAAATAGTCTAACGAAAGCGCGGAGAATTAACCGGAAAAAGGAAGTTGCCACA
CTCCCCTTACTAATCCACACCTAGGATCTCCTTTACTCGTATTTTGAATATCTACGAAAA
CACTCACATACGCTCATCCCATATATATATATATATATATACGGATATATTGATAGTCGG
ATACTACGTGGATATGGATACACATGTAAACGGTTGTGTTACCCTAAAGAAACGGAGGTC
AGTTAACAACACTTTTACAG
>Ylr108c
CTTCTTGTTTTCAGAAATTAATTTGGACCATTCAACTGTTTGTGGCATCTTACAAGATTC
ACTGTCAGCATCACTTTGGCTGATAGCAATGTATTGGAATTTGTAGTCGCCAGCTGGGAA
TTTCTTGTTAACTAAGTCAGACATGTTTTGTTGTGGTTATTGGTAGTACTTGTGTTGTTC
TGGTTTGTTGAAATTTCGTTATAAAGAAACGGTATATGCACAATTGGCAAGGGAGAGGAA
ATTGTGCTCCTTTATATAATGTTTCGTTTATGCTTTTCCCTTACATCTCGTCCTTAATTT
TCCACCCCTTTATTTTTCATTTAACCTCTCAACAATATGTGAATACCGTCCGGAAGGGGT
GTGGAAAAAGAAAAAGAATGATACAGAAGAAAATCAAAAGATGATATTCAACTTCATCAC
CAAATACTGACCTTTTTGGCCAGAGCTCCCCGTATAATTACAAATCAGAAAAAATCCGCG
CTCACTACCTTAACGGAAGGAAATATTCCACATGTCGGATTAATTTTTCTCTGGTCCCAT
GACCATGTGCCGAAATCTTTTCAGTGACGTCAGGACCGGTTTTTCCGAGAGCCCCACAGG
GGACGCCGATGATTTAGGGTTCTTTTCTTGATTACTAATAGACAACATAAGCGCAATAGA
```

```
TCGAGGCCAAGGTGGTGGTTTATAAATTGTATTACCTTAAAGGCTCATTATTAGCTTGAA
CACGAATATCCAAAGAACACGAACAACTTCAGGACGTGGATAGACAAGTGTAAAGTCACT
CACACGCACGCTCATATCAG
>YLL055w
TTGTCTACTTGCGTGTGCTTTCAGTTCTTTACCAGTACGATTTGGTCACGTGACACGCGA
TCAACTACATTCAATCCATCTGCCCCTTGCTGGTTCCGAGTATCAATTCTAAATTGTGGC
TACCACTATAGCAATAGTACGCCAAGCTCGGATGCAAATCACGGATATCTTCTTAGCATC
CGAGTTCAACATCCGAGTTCAGCATCCGAGCTTTGCGGCGAAAACTTAACAAGATAACGG
ATAATCAAAGTTCTCTGTGATGTTACGGAGTCTTTGGTCTGCGCTGCTGACTAAAATAGT
GTGACACAACGTACATAGCGCCACCCATATTGGCCGATAAGGACCAGACTTGTGGCGATG
ACTGTTAACTCTCCGCGAATCTTTTTCGAAGTCGCAGGTGGCATTGCAGAAACAAAAATG
TGTGGCTTCTGAAAAAAAAAATAGGCACCCCATTTGATGTAAGACATCTGGCTACCCCTT
GTACATTTCAGTCACTGAGCATCGAAGGCACGTCTTTGCAGCATGCAATAGTGTTCGAAG
TTCTTCCATTTACGGTGGCTCCCCTCACTGGAACGGTGATGTAATATGTAGATAGCTAAA
ACATTTCGCAGGTCCCTAGGAACTGGACGTCAAGATGAATAGGTATTGTTAAAAATGCGT
ACCTGTTGAATTGAAGAGGTCGGAAAAATCAAAGTACTTAAACAGACGATTGATTCGACA
TTTAGCTATTCAGGTTGCACTTCTTGAAGGTTGAACAAAACATCAAGAGAAAATAAAAAA
GAAAATATAACAAAAAAGAA
>SDL1
CAAATCAAATCGTCATCGTTTGGGCGGAATCATATCCTCAAATTCTTGACATTGCATAGC
GTAGTATATATGGTATTAAGAAAGCTCGTACAGCATAAGATATCAGGTTCTTTCTTTGAA
CATTATCGAACGCTGTTAGCCTCTTGTAGTATTTCCCCATTGGGCTAAACGGCAATGTCC
TTATCGTAGTTAGAATATGGAAAAATTACATAATTTCGAAACTGGAAATGATGCTAGATA
AAAGTGGAAGATTACTAAGCGTAGTATATTAGCCGGAAAGAATTCATATATAGTCATGAA
GACATTTTGAAAAAATTACATTTATAGAAGTTTATTATAACGTAACAGTTAACAACAAAG
GCTCATAGTCCGAGATCAATATAATAAATTATTCAAGGTTTAAGGAAGAAGTTACCATGG
AAATGACTTACTATGAAAAGACACCTTTGATTCGTCAATTTTTGAACAATGGTAAGACAA
ATTCGTGGTTTTACGTTAAGCATGAGATGTTACAACCAGGTGGAAGTTTCAAATCGAGAG
GAATCGGGCATTTGATAAGGAAGAGTAATGAAGAAGCGCTAAGCGAGGGTTCTGGGAAGC
TTGCTGTATTTTCTAGCTCTGGGGGAAATGCTGGTTTAGCAGCAGCAACTGCCTGCAGAT
CGATGGCACTTAATTGCAGTGTAGTGGTTCCTAAAACTACAAAACCTAGAATGGTAAAGA
AAATTCAAAGTGCAGGAGCCAAAGTCATTATCCATGGTGATCATTGGGGGGAAGCAGATG
AATACTTGAGGCACGAATGA
>YLR460c
TTGTAGTGCTTCTTCGATGAAAAGGGATAGAATCTTTAAAAGACTTGATCTTGGCATTTT
TATCCTCCTTCAAGCAGAATGGCTCCAGTAGAGTTGTTCTGAACGCTTCTGGGCAGTGAG
CAGCATTGAAGAAAAAGTACTCAGCATTCCAAGCTTTGTTTTCATATAATTATAAATTAA
AACTCATAGCCACCAATTTCCAGGTTTTGGTCGCAATATGCTGCAGAGTCAATGGCTTGT
TTCATTACTATACTAGTCCGACGGAGTTTACTTTACTCTAAATTTTCTGAAATAGAAAGG
CCACACGCCATACAATTAACAATAGTTCATTTAGTTGAAAGGTTTCCGAGATAGAAAGTG
CGGTAGTGGAGAAGAACAATAAACGTATCAAACTTTGCGCAGTCATGGAATTTCCAGGAA
ACATTTTTAGGTTTTCATATGACAGCAGTTCCTGTCCGAGAGCTGTGCCTTTCTTGTCTC
AGCTGACTAAGATGTCTCGGATCCGCCCGGCTGACGCGCGGCTCCGTCTAGTGGGATAGG
CTTTCAACACATAGAAAACCCTTCAGATGATGAGAACACATCGAATCCCAGAGAGTAATA
TCACTAAGAATAAAGCAACCATGATAGACGATATTACTAAGTGTTCCTCAACAGATATTG
AGAAAAGTTCTTATAAAGCGGGCAGGCTTTTTTCTATTTTTTTCTTTTTCTTTATCTAGC
TTAAGATCTTTCTAACCATATTCTATAAAATGGCAAGCTGAATACAGATTATCAACGACA
ATTACAGATAACCTACAAAC
>YCR102c
GGATAGAGTCTTTAAAAGACTTGATCTTGGCATTTTTATCCTCCTTCAAGCAGAATGGCT
CCAGTAGAGTTGTTCTGAACGCTTCTGGGCAGTGAGCAGCATTGAAGAAAAAGTACTCAG
CATTCCAAGCTTTGTCTTCATATAATTATAAATTAAAACTCATAGCCACCAATTTCCAGG
TTTTGGTCGCAATATGCTGCAGAATCAATGGCTTGTTTCGTTACTATACTGGTCCGACGG
AGTTTACTTTACTCTAAATTTGCTGAAATAGCAAGGCCACACGCCATACAATTAACAATA
GTTCATTCAGTTGAAAGGTTTCCGAGATAGAAAGTGCGGTAGTAGAGAAGAACAATAAAC
GTATCAAACTTTGCGCAGTCATGGAATTTCCAGGAAACATTTTTAGGTTTTCATATGACA
GCAGTTCCTGTCCGAGAGCTGTGCCTTTCTTGTCTCAGCTGACTAAGATGTCTCGGATCC
```

```
GCCCGGCTGGCGCGCGGCTCCGTCTAGTGGGATAGGCTTTCAACACATAGGAAACCCTTC
AGATGATGAGAACACATCGAATCCCAGAGAGTAATATCACCAAGAATAAAGCAACCATGA
TAGACAATATTACTAAGTGTTCCTCAACAGATATTGAGAAAAGTTCTTATAAAGCGGGCA
GGCTTTTTTCTATTTTTTTCTTTTTCTTTATCTAGCTTAAGATCCTTCTAACCATATTCT
ATAAAATGGCAAGCTGAATACAGATTATCAACGACAATTACATATAACCTACAAACAGGC
AAGTTGCAATTCCAGAAACC
>YNL260c
ATGGAGATGTTTTACAATGGGGATTTTTATCTTTTTGCGACAGACGAACTAATCCAATGG
ATAGAGTTATTGTTTGCCGACACTCCCCTGCGAAGGAATGCCATTGATGATATTTACGAA
ATTAGAGGCACTGCTCTAGATGATTAAAAAATAATTAATCGTAGGATAGGTTACATGATC
ATTGCAATAGATGTACATCTTGTATTACGTATGCGTAAGTTGTCTTTTTTTCAGTTTTCT
AATAAATTTCACGTGATGCGGTAGTTTTTTTTGGGGGTGCAAAGTTGGCACATCTTGTAG
TGCATTTTTGAGCTTCTTAAATGACTTGAATCATCTTAAAACATTGCAGGATGAGATTTC
CAACGACACAAGAGAGAACTAGCGCAAAAGATGGCAGAGATAAAACATTATCAATTCAAT
GTCGTCATGACCTGCTCGGGATGTTCTGGTGCAGTCAATAAAGTTTTGACGAAATTAGAA
CCAGACGTTTCCAAAATAGATATATCCCTTGAAAAGCAATTAGTAGATGTATATACTACA
CTTCCTTACGATTTTATTTTAGAAAAGATTAAGAAAACCGGAAAGGAAGTTAGATCTGGC
AAACAATTGTGAATGAAAGTGAAGAGAAATAGTTTTTTACTCTATTTTGTATTGTAGCTA
CTTAAGTAAATGATTAATATTCTCAATAGGATGCTTACGAAGAAAAAAGTCTTTTTTTAT
TGTTTTATTGTTTTATTCTTTGTTGGCTATGTAGCCATTTGCCTTTTCAGTATAATCTTA
CTAATATAAAAATCTCAATC
>YKL071w
ATAATTATTCCTGTTTCTTTAACCTGGTAAAAAAAAGTACAAACACTTAAGCTTTTTGAA
ACAGCTTTATTTTGCTTCATTAAATAGCTAGGATAAGAAATCCCTCATCCGAAAGGTTTT
GTATCTAACTACCCTAGAGAACATTTGTCCTGATCAGGTTCATTTGGAGTTTATATTTTT
TAGAAGCTCAAAGTTTGTTGGACTCATTACCATGGAAGAAAAAAAGAAGATACTACGAAA
TATTGGTTTCTCAGGTTAAATAAGGGACACCATTTTCCTATTAGGCTAGTCGAGCTTAGT
TCTTCTAATTTCTTCAGATCTTCTATAATTTCCTATCTTCTACCTGATGTGTGCATGATA
TATCTATGAGCTCCTGATATTGCTTGTTTTACTTTAGCTTGCATGACTTGCAATAATCTA
ATCATATATGTTCCCGATTAATATACTGTGCACAAATTGCAGGACATATAATTTTTCCGT
GGATTATATCTTCGATTAACGTCCGCGGGTCTCATAAAAAGCAAACCAACTTCGCAATTC
CCTAGAAATACCTCAATAGAAAGTTATTTGTAATGAGATTAGTAATGAGATTAGCAATGA
GATTAGTAATGAGATTAGTAATGAGATTAGTAATGTGATTAGTAATGCATAGCGGTATAA
ATGGTAGTACTAATAAGTAAGATAGTATACCAGTTATAATAAATAGGCGGCGATGCTTCA
AAACTAATTTTTGACGTTTTTAAGAATAAAGCCTTTACCAGTGGCATAAATCAGTAGAAT
TCTAAGCAAACAAAGTCGAT
>YDR132c
TTGTTGGCTGGTGCTGCCGCTTTATTGTTGTAATTTACTCAACCTTTTCTTTAATATATT
TTTAGAAAAATGGTTAAGTACTTTTCCGTCAATACAGCTTCCACAAAATCGTTTTATTTC
AATTAATAAGATATTCTGGTAATTTCTTATTCTAATTCTAGGAGGCAACACTCCATAACT
AAATGTTTATTATAATAATATTAATATAAAGTTAATATCTCTTAGGTTACACGCAACTAA
ACTTTTTTTATATATACTGGTTGCATTGTATTTTTATCATATATCTTAATATAGGTTGTT
TAAGGTTCGTTAATTCAAAAATAATGTTTTTTTGTTTGCCCAGTCGGTTTACGCTAATCA
AGAACATGGATTCAATATCTTACATAATAAGCGGACGATGAGTCTATCTCGGTTTCCTCT
TTTCTGAAATATAATAGTTTTATATTAGTGATTCATTATTACGATCTGGCCGGCTTATCA
TTGCTTCTTTTACAAACGTATCCTTTTAAGATTTTATCCGGACCACTGAGCTGGCATTTT
TCTTAGCTTGCTTTTGTCTTTGTTGCACGGGCGACTGAACTTCTTGGAAAATCATGTAAG
CCGCAATTGCTTACTAAGTCAAAATTCTAATGGATTCTCCTCTCGAATAAAATTGAATAG
TCTGAGCAGTGAAAAAAATGTATATCTTATGTGTGAATGTTATATACCATAATCTAACGA
AAGATAGCGAATTAAAGACCGGTCTGTGGCACTCGTAAATATTACCAACAAAGAAAGTTC
CCAATATAATCATATAGAAA
>Aad15
ATTTAATGGTTTGAAATAAACTGAAGAATAAAATTTACATAATAAAGCGTATATTCTTCA
ACTTTTTGAATTCAGAAGATTTTCTATTCAAGATCCGTGCAGATCTCTATTACACTCATT
ATTACAAATCATTGAGAATGAGATTTTGACACCATTACTAATAATCCTATTCACTAATAC
TTACTGTAAGCGGTAGAAAAACACATGTACAAGAAATTGGCTGCTGAAATGCTGCCATGG
ATTCAACCACAGTGAAAAAAGGTTAAGTGGAAACGTGATAAACTGCGGCCTGCTTGCTCG
```

```
CGATGCCAACAAAATTCATTACGGTGTGAATATGAAGAGAGAACAGATTTATCTGCCAAT
GTTGCAGCAAACGACTCTGATGGATTCAATTCCTCTCATAAGCTCAATTTCGAACAGCAA
CCTGTACTTGAAAGGACTGGGCTTAGATATTCCTTACAAGTGCCTGAAGGTGTCGTTAAT
ACTACGCTGTCGATATGGAACGCCGAAGATATGCTAGTTATAGTAGGATTAGTCACATTT
CTGGATTATCCTTTTGCTGCGCATAGTCTGGCGCAACATGACCACTATATCAGGGCACTT
TGTGCTTCGTTGTACGGCATGGCGCTTGTTGACTTTAGCAATTATGCTAATGGTATTCCT
TGTGAAGACACATCAAGAAGTATACTAGGACCATTGTCATTCATAGAAAAGGCCATTTTT
AGACGGATAGAACATAGTAAGCAATTTCGAGTTCAGCCTGGAACGTGTTGAACAGAGATT
TTGAGCGTGATATTATTCCA
>YOL118c
GCTTATGTCGTGCCGCTCCTCGATCTTGATCATCCTGACTATTGTCTTCATTTATTTCAA
TGTTTTGTTTGTCTTCATCATCGGTTTCTTCCTCCATTTCAATTGACTCCATGCCTTCGT
CATCCGATAACATGAAGTCGTCATAATTGTTATCTTCATCAGACATTTTGTGTATAGTGA
ACCTTTTTGCCTTTTACCGCGTCAAACGTTTCAATAGTTCATTTAGTCAAAGAGATCACA
AAACAATTGATCCAGCTCCCCTTATAATATTTTTCAGCGAGGACCAGATAAATTTCTTCG
AAAAGGAGCTGCACTTTGGGCAAACGAGAAAAACGTGCCAACCTGGCATTTAGCTACAGA
GTACTTTCGCCAGTGTTATCATATATGGGCCACTACTCTCATACATGCACAAATGTTATA
CTCATTATATACTTAGTGTGTATTAAGAGATCAAATTTGAAAAAGGTAAGAAAGTAAGTT
TGTGCAATCTTGTTGCCATCATATAAAACTCGAATCTCAAACTCGACATTCCCATAATGA
AATCCAGAAAGTGTAAGATCACTGTCACTGCTTTCGAGGAATTAGTTATCTAAAGCAAGT
AACCTAGAAGAAAACGTTCTATTGTGCGATAGATGAAATTCTAGTCGCAGTAAAAGCGGC
ATCTATAAAAGATTTTCTCAGTATATAAATGGATGTTTGACATCGAGGAATTGGCTAATT
TTACAGTCGTATTATGGTAACACTTGATGGACTATAACTACCGGAAACGGCTGAGTTGTC
TACCCCCCGTATGATCTTGA
>YKR071c
GGAAATGTCCTGGGTCACTGTTTTCAAGCCGTCTGAGAAAGTTATGGATATAAATGGTGA
TATTGGTCTCGGGGGTATGATGGACTGGAATGAGATTGTTAACAAAATCGTTATGAAATT
GGGTGGGTACCCAAAAAATAACGAGGAAGAGGACGATGATGAGGATGAGGAAGAAGACGA
CGACGAAGAAGAAGATACAGAAGATAAGAACGAAAACAACAACGATGATGATGACGATGA
CGACGATGACGACGACGACGACGACGACGACGACGACGACGACGACGATGATGACGAGGA
CGAGGACGAAGCAGAAACCCCAGGTATAATAGATAAGCATCAATAAATAGTCTATGAATT
ACAAATACACTAAACACATTCTACTTATTTGTGTTTTGTAATTTCTGGAAAAATACTCAT
TAGGCAAATTCACGCAGCATATTTACAAAATGTCACAAATTTTTTTCCTTTTTTACTTGC
AGTAGGTATGTAATGCCATACTAGCGTACATAGAGCTCTTCATATAAACCGCTATTTATT
ACTAATTTTAATAATTCAGTAATTCATTAAATACATTCATTACTTTCTGGGTAAATATCT
ATTCGTTTCTCATTTTTCTTTTTTTGACATATCCCGGTACCGGAATAAGTCCGATACTAT
TCAAAAGCTTCTGGAAAAAAAAAATAAAAAAAAGAAATGATTAGTAATCATTTAATTTCA
TGACATATTAAAAAGGTATTCTAAGCAGAAGTCTTTCATCACATTTGATCTAAGCATATA
CACTGTAAGTGAAGGTATCG
>YML131w
TTCTGAAAAAAGTTTTATATATTATTTCACAACATTACCAGTCACTTCCATGCGTGCCTT
CCCTCTTTGAAGAATTAAATTACTGGCATAAACGAAGAATATCATAATTTTACTTGTAAC
TTCTATACAATTTTTTTTTTTTTTGCCTTCTTCATACTTTTACTCCTGCTTTTATTACTC
TAAATTTCATTTTTATTTATTCTCTTCTACCGTGTCAATTGGACAAAAAATTCAAGTTCT
GATATCTATGCAGGTCACCCTTTATCAGAAATAAATCTTCAGTCCTCCTTCTTTTAGTCA
TATCTTAGTAGTTACATAACATACATGAAATGTAAGAAAAAATGATAACTGAAGGCCCCC
AATTTTCATATCAGGGGTTATGAAAAAGCTTGTTTTTATTAGTCGTTTGCATACTTGAAT
TTTTTTTGGTTTACACCAATTTGAAAATTTCAGTCATCATTTTAGCAAATTTATCATGAG
AGCGCTTCAGTGCCACCTCTTTAATACGGTGGTGATTAACACTTCGCTTTTCAGTTTGGA
GAAAATTAATTGTCTCCGCTTTTCGTACTTCTATTTTCCTCTTTCGCTTAATAATAATCT
TGGCGCACATTACGTAATAGCCAAACCCTATTTCTCTCCTCTTTGAACTCGAATTGAAGG
AGTTCTTTCGGAATGAAAGAATATAAAAATGAAGGGTTACCTACTTAACTTTTTGTTAGC
GGCACCTCTTTCATTAAAGGAATCTTCAAATAAAAAAAATAAAGGCATAGGTCTAAGACT
AGAAAAGATCGCATAACGAT
>YNL134c
GAATAACTACTTAGACGTCAAGGCAATCAATTTTGAAAAGAAAAAAAAAGCATTACTAAT
GGAGAACGATTCTACTACTAATACAGAAAAATATTTAGAGACCGGCACTAGTTTGGGCAG
```

TAAAACACAAATAAAGGGCGTAAAGACATCCAGCCCTAAACCAAAAATCAAGGTATCCAT
AAAAAAATTGGGTAGAAAACTGGAGAACTGATGTGATGGCCTACACACTTCGTCACAACT
AGAGCTTTATGCCAAGATTAAACTTCAAGTTATATAAATTATGTAAACATTGGACATACA
AAACTCGGAAACTCTTTTTCTATTCATCTTCCCTTCTCGTACGTGCCCACGGAAGCAATA
AAAAGAACCGAAATAACCAACACCCGTAACGTCAAAGCATTCATGCTTAGAATGGAAACC
ATTTCGTGGAATGAAATGGCAAATTGATCACATTGATTGCTCGTTCCACTACCTGTGTCC
GCAATTTTTTAATGGTCATCACAGCCCCTGCTGACTAAAGTTCCTCGGATCCGCTTACGG
TTGTCGCGCGGTTCCGCCCCTGCGTACTCTTAGTACCTAGCATATGGGCTCCCTCCGTTG
GATTGGCATCGATTAGTAAGGACAGATGTTAAGGATTTAAGACCGTTTTTAAGGTATTTC
GGCAATGCTTCGATTTAAAAGGAGAGAGTTTTTTTTTTGCCGTTTTCTTCCTCTCACTTC
TTGATTAGTACTGTAATTCTAGTTGAAAAAAAATCGTTAACTATACACAGCAAAAAGCAA
TATCATACTGCATATCAAGC

AlignACE 3.0

Yap1p

```
04/13/02
/home/jhughes/bin/AlignACE -i /home/alignace/Yap1p-
1.seq -numcols 10 -expect 10 -gcback 0.38
Parameter values:
 expect =         10
 gcback =         0.38
 minpass =        200
 seed =           1022780232
 numcols =        10
 undersample =    1
 oversample =     1

Input sequences:
#0   Trr2
#1   Gtt2
#2   Aad6
#3   Aad4
#4   ISU2
#5   Gpx2
#6   Bet3
#7   Trx2
#8   Lap4
#9   Flr1
#10  Ttr1
#11  Ydl124w
#12  Ylr108c
#13  YLL055w
#14  SDL1
#15  YLR460c
#16  YCR102c
#17  YNL260c
#18  YKL071w
#19  YDR132c
#20  Aad15
#21  YOL118c
#22  YKR071c
#23  YML131w
#24  YNL134c
```

```
Motif 1
ATTAGTAAGC   1   700  0
ATTAGTAAGC   2   556  0
ATTAGTAAGC   2   576  1
ATTAGTAAGC   3   443  1
CTTAGTAAGA   7   617  0
ATTAGTCAGC   8   416  1
CTTAGTAATC   9   434  0
ATTAGTCAGA   9   631  0
ATTAGTCAGG   9   651  1
CTTAGTAATA   10  654  0
ATTAGTAATG   11  33   1
ATTAGTAAGG   11  604  0
ATTAGTAATC   12  629  0
CTTAGTAATC   14  249  0
CTTAGTCAGC   15  481  0
CTTAGTAATA   15  631  0
CTTAGTCAGC   16  457  0
CTTAGTAATA   16  607  0
ATTAGTAAGA   17  775  0
ATTAGTAATG   18  577  1
ATTAGTAATG   18  601  1
ATTAGTAATG   18  613  1
ATTAGTAATG   18  625  1
ATTAGTAATG   18  637  1
CTTAGTAAGC   19  608  0
ATTAGTAATG   20  152  0
ATTAGTAATA   22  536  0
ATTAGTAATC   22  699  1
ATTAGTAATG   24  50   0
ATTAGTAAGG   24  611  1
**********
MAP Score: 39.3932

Motif 2
GAGGAAGAGGACGATGA 22   142  1
GATGACGATGACGACGA 22   229  1
GATGACGAGGACGAGGA 22   289  1
GACGACGACGACGACGA 22   247  1
GACGACGACGACGACGA 22   271  1
GAGGATGAGGAAGAAGA 22   160  1
GACGACGAAGAAGAAGA 22   178  1
GACTCAGAGAACGAGGA 6    520  1
```

```
GAAACCGATGATGAAGA 21   71   0
GAGTAAAAGTATGAAGA 23   148  0
GACTAAAAGAAGGAGGA 23   282  0
GATAACAATTATGACGA 21   137  0
GAAAACAACAACGATGA 22   211  1
GATAAAGAAAAAGAAAA 15   699  0
GATAAAGAAAAAGAAAA 16   675  0
GCCAAAGAGAATGAAGA 4    104  1
GAAAATGATATAGAAGA 2    767  0
GAACAAGACCTAGACGA 0    140  1
GAACCTGATCAGGACAA 18   144  0
GAGCAGGAAGAGCAGGA 9    183  1
GAAAATAAAAAAGAAAA 13   768  1
GCGAAAAATAAGGACGA 8    56   1
GAAACAGGAAAGGATGA 10   555  0
GACAATAGTCAGGATGA 21   29   0
GAAACCAATAACGAAAA 6    440  1
**  * **  * ** **
MAP Score: 35.4272

Motif 3
AACAAGAAAATG  0    56   1
AATAAAAAAAAG  2    30   1
AAAAGGAAAATA  5    361  1
AAGAGAAAAACG  6    384  1
AAAAAAAAAATA  8    517  1
AAGAAAAAAAGA  8    788  1
AAGAAAAAAAAA  9    537  1
AAAAAAAAAAAG  10   262  1
AAAAGAAAAAGA  12   365  1
AAAAAAAAAATA  13   431  1
AAAAAGAAAATA  13   775  1
AATAGAAAAAAG  15   686  0
AATAGAAAAAAG  16   662  0
AATAAAAAAAGA  17   709  0
AAGAAAAAAAGA  18   215  1
AATAGGAAAATG  18   270  0
AACAAAAAAACA  19   324  0
AAAAGGAAAAAA  22   460  0
AAAAGAAAAATG  22   610  0
AAAAAAAAAATA  22   674  1
AAAAAAAAAAAA  23   130  0
AAGAGGAAAATA  23   571  0
AAAAGAAAAAAA  24   35   1
```

```
AATAGAAAAAGA  24   312  0
** ******* *
MAP Score: 26.3439
```

Downregulated Genes

```
Input sequences:
#0 Gim4
#1 Fpr1
#2 SSB1
#3 Gim5
#4 Gim1
#5 Gim3
#6 Gim2
#7 Egd1
#8 Egd2
#9 Pdr13
#10 SSB2
#11 Zuo1
#12 Fpr4
#13 Npi46
#14 Cct2
#15 Cct8
#16 Cct3
#17 Cct5
#18 Cct6
#19 Cct4
#20 Cct7
#21 Cpr8
#22 Cct1

Motif 1
GAAAAAAAAA 0 663 1
GCAAAAAAAA 1 675 1
AAAAAAAAAA 2 592 0
GAAAAAAAAA 2 622 1
AAAAGAAAAA 2 632 1
ACAAAAGAAA 2 662 0
AAAAAAGAAA 3 705 0
GAAAGAAAAA 4 106 1
GAAAGAAAAA 4 686 1
AAAAAAAAAA 6 530 1
AAAAAAAAAA 6 540 1
AGAAAAGAAA 6 689 1
ACAAAAAAAA 7 224 1
GCAAAAGAAA 7 234 0
GAAAAAAAAA 7 283 0
GAAAAAAAAA 7 622 1
AAAAAAAAAA 8 654 0
GAAAAAGAAA 10 587 1
GAAAAAGAAA 10 604 0
AAAAAAAAAA 10 616 0
```

```
ACAAAAAAAA 10 694 0
GAAAGAAAAA 10 759 0
AAAAAAAAAA 11 358 0
GAAAAAAAAA 11 662 0
AAAAAAGAAA 12 764 1
AGAAGAAAAA 14 649 1
GCAAAAAAAA 15 697 1
GAAAGAAAAA 16 180 1
AGAAAAAAAA 16 677 1
AGAAAAGAAA 17 38 1
AGAAGAAAAA 18 1 1
GAAAAAAAAA 18 17 1
GAAAAAGAAA 20 606 1
GAAAAAAAAA 20 670 1
GGAAAAAAAA 21 653 1
AGAAGAAAAA 21 676 1
AAAAAAAAAA 22 615 0
GAAAAAAAAA 22 664 1
**********
MAP Score: 46.8942

Motif 2
AAAAAAAAAAAAAAAAG 0 664 1
AAAAATTGTTGAAAGAG 1 82 1
AAGAACGGGTTAGATGA 1 486 1
AAAAATTGCCGGAAGAG 1 680 1
AAGAATTTGGTGGATAA 1 709 1
AAAAAAAAAAAAAAAGA 2 590 0
AAAAAAAAAAAAGAAAA 2 624 1
AAAAACCGGAAGAATGG 3 432 0
AAAAAGGGCATAAACAA 3 657 0
AAGAAACTAACAGAGGA 5 778 1
AAGAAACGTGCGAAAAA 6 502 1
AAAAATAAAAAAAAAAA 6 524 1
AAAAAGAAAAAAAAAGG 7 281 0
AAAAAAAAAGCAGAAAA 7 623 1
AAGAAACCTAAAGAAGA 8 566 0
AAAAATTTTCAAAAAAA 8 658 0
AAAAAAATGACAAAAAG 8 686 1
AAAAAACTACCGAAGGA 9 329 1
AAGAAACAAGGAGAAGA 9 427 1
AAGAAGAGGAAGAATGA 9 460 1
AAAAAATTGAAAAAGAA 10 605 0
AAAAATCATGAAAAAAA 10 622 0
AAAAATTGCCCAAAAAA 11 362 0
AAGAAAAAAGAAAAAGA 12 761 1
AAGAACGACACAAATAG 13 100 1
AAAAATGCATCAAAGAA 13 331 1
AAGAAGCCTCGGAAAAA 14 334 0
AAGAAGTATAAGGAAAA 14 632 0
AAAAAGAGGATGAAAGG 14 730 1
AAAAAAAGTGAAGAAAA 15 700 1
AAGAACGAACTGAAAGA 17 297 1
AAAAAGGAAACGAAAAA 18 6 1
AAAAAAGTAGGGAAAAG 18 25 1
AAAAATTAAAAGAACAA 18 103 1
```

```
AAAAAGCTACGAGACAA 18 251 1
AAGAAACGAACGAAAGA 18 673 1
AAGAACATCATAGATGG 19 265 0
AAAAAAAAAATGAATAA 20 674 1
AAAAAATATTTAAAAAA 21 657 1
AAAAAATTGTAGAACAG 22 216 1
AAGAACTGAAAAAAAAA 22 657 1
*****      *** **
MAP Score: 31.9588

Motif 3
GAAAATTTTTC 8 665 1
GAAAATCTTTC 14 491 0
AAAAATTTTTC 10 663 1
AAAAATGTTTC 2 734 1
AAAAATTTTTC 2 636 0
GAAAATTTTAC 12 715 1
GAAAATTTTAC 4 568 1
GAAAATCTTGC 3 750 1
GAAAATGTTGC 8 17 1
GAAAATTTTGC 18 629 0
GAAAAAATTTC 9 763 0
AAAAATATTAC 11 635 1
AAAAATTTTAC 13 693 0
AAAAATATTAC 2 260 1
AAAAAAATTTC 19 659 0
AAAAAAATTTC 7 227 1
AAAAAAATTTC 14 677 0
CAAAATTTTTC 17 667 0
AAAAAACTTAC 8 599 0
AAAAAAATTGC 1 678 1
AAAAAAATTGC 11 370 0
CAAAAATTTTC 7 635 0
CAAAAACTTAC 16 541 1
****** ****
MAP Score: 14.7235
```

Upregulated Genes

```
Input sequences:
#0 SSA2
#1 HSC82
#2 SBA1
#3 SSA1
#4 Yro2
#5 SIS1
#6 HSP26
#7 HSP104
#8 SSA3
#9 STI1
#10 Cpr6
#11 HSP82
#12 Cph1
#13 Cpr8
#14 YFR041c
```

```
#15 Caj1
#16 Xdj1
#17 Afg1
#18 YPR061c

Motif 1
AGAAAAAGAA 1 746 0
AGAAAAAGAA 3 345 1
AGAAAAAAAA 9 593 1
AGAAAAAAAA 10 110 0
AGAAAAAAAA 11 559 0
AGAAAAAAAA 3 300 1
AGAAAAAAAA 4 138 1
AGAAAAAAAA 4 597 0
AGGAAAAGAA 11 186 0
AGGAAAAGAA 15 780 1
GGAAAAAGAA 3 533 0
GGAAAAAGAA 5 475 1
AAAAAAAGAA 7 355 1
AAAAAAAGAA 3 139 1
AAAAAAAGAA 18 622 1
GGAAAAAAAA 0 522 0
GGAAAAAAAA 3 330 1
GGAAAAAAAA 13 653 1
AGGAAAAAAA 2 466 0
AAAAAAAAAA 16 437 1
AAAAAAAAAA 5 744 0
GGGAAAAGAA 10 463 1
GAAAAAAGAA 7 756 1
AAGAAAAGAA 2 21 0
AAGAAAAGAA 6 79 0
AGAGAAAGAA 3 569 1
AGAGAAAGAA 1 732 0
GGGAAAAAAA 15 116 0
GAAAAAAAAA 0 593 1
AAGAAAAAAA 4 667 0
AAGAAAAAAA 4 62 0
GAAAAAAAAA 16 458 1
AGAAGAAAAA 13 676 1
AGGGAAAGAA 10 212 1
GGAAGAAGAA 16 685 1
AAAAGAAGAA 10 532 1
AAAGAAAGAA 1 478 0
GGAGAAAAAA 3 491 0
GGAAGAAAAA 11 579 1
AAAAGAAAAA 11 88 0
AAAGAAAAAA 9 647 1
AAGGAAAGAA 14 186 0
AAGAGAAGAA 4 425 0
AGAGGAAGAA 12 535 0
GAAGAAAAAA 3 211 0
**********
MAP Score: 47.6051

Motif 2
AAAAGCAAAAGAGA 0 351 0
```

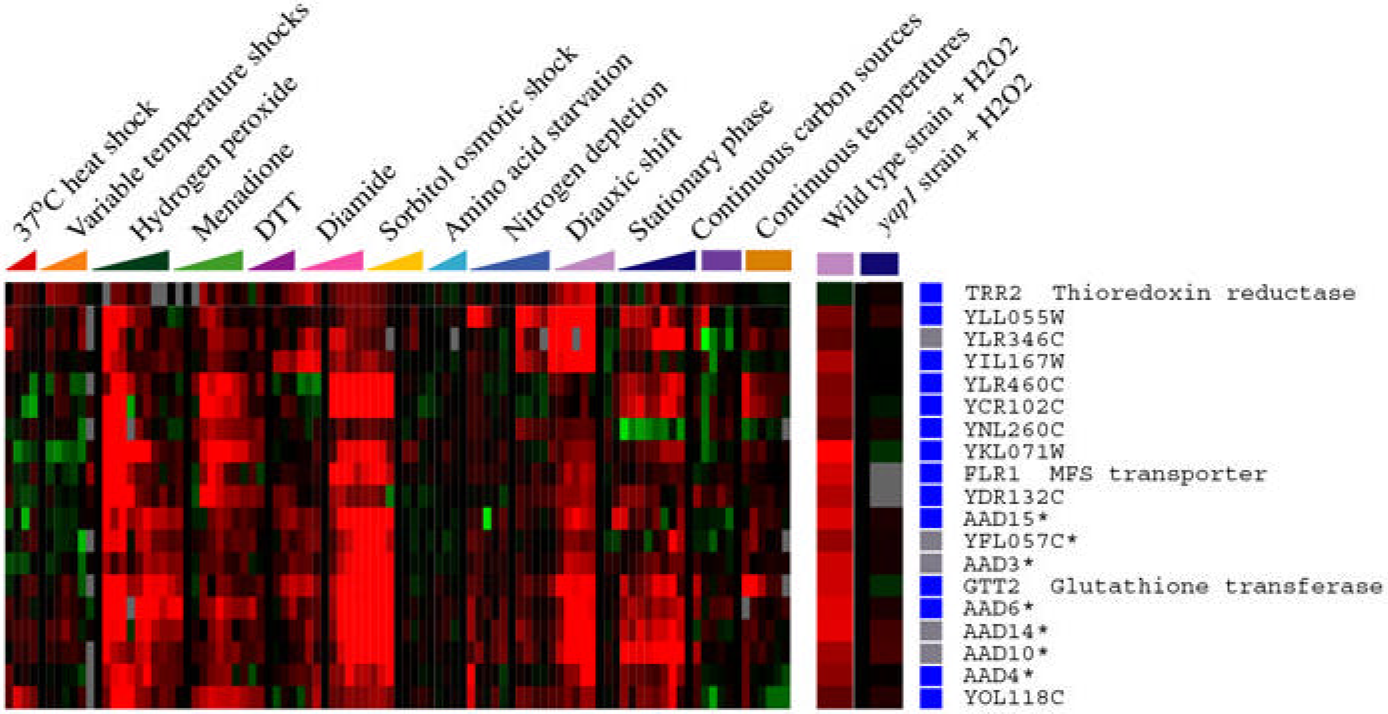
```
AAGAATAGAAGTAA 0 733 0
AAGAATATAAAAGA 1 469 0
AAGAAAACAAGAGA 1 737 0
AAGTATAGAAGAAA 2 25 0
AAGAGTAGAAGGAA 3 124 1
AGAAAAAGAAATAA 3 345 1
AGGGAGAGAAAGAA 3 565 1
AAGAAAAAAAAAAA 4 137 1
AGATATATAAAGAA 4 355 1
AAGAAAACAAGAGA 4 429 0
AAGCAAACAAACAA 6 771 1
AAATGAATAAATAA 7 378 0
AAGGGTAAAAAGGA 7 720 0
AAAGAAAAAAGAAA 7 753 1
AAGAAAATAAAAAA 9 377 1
AGGAGGAGAAGCGA 9 624 0
AAAAGCATAAATGA 10 87 0
AAAAAAAGAAGAAA 10 103 0
AAAAGAAGAAATGA 10 532 1
AAAAGAAAAAGAGA 11 84 0
AGAAGGAAAAGTAA 11 99 0
AGGAGAATAAGAGA 11 249 1
AGAAAAAAAAGAAA 11 555 0
AAAAGCACAAACAA 11 774 1
AGAGACAGAAATGA 12 572 0
AAAAATAGAAGAAA 13 670 1
AGATAAATAAAGAA 13 725 1
AAACGAACAAAAAA 14 761 1
AAGCGCAGAAAAAA 16 12 0
AGAAGCAGAAACAA 16 168 1
AAAAAAAAAAAAGA 16 435 1
AAACAAAGAAACAA 16 627 0
AAGGGAAGAAGAAA 16 682 1
AAAGGGAAAAGTAA 16 766 1
AGAGATACAAGTAA 17 429 0
AAAAAAAAAAAGAA 18 618 1
*** * * *** **
MAP Score: 32.1684

Motif 3
AGAACTTCTCGGAAAA 0 526 0
AGAATAAGAAGGAAAA 11 103 0
AGAAAAGGGCGGAAAA 4 413 0
AGAAACTTCTGGAAAA 5 465 1
AGAAAATTCTGGAAAA 7 496 0
AGAAGCATGAAGAAAA 4 128 1
AGAACATTCCAGAAAA 9 583 1
GGAAGTGTTGGGAAAA 10 454 1
GGAACGTTCTGGAAAA 3 537 0
AGAAAAAAAAGAAAAA 11 553 0
AGAACCAATAGAAAAA 1 767 1
AAAAAAGCATGGAAAA 2 206 1
AAAAGATAATGGAAAA 8 375 1
AGAATGTTCTGGAAAG 3 434 0
AGAAAAGTTAGGAACA 7 456 0
GGAATGTTCTAGAAAA 0 565 0
```

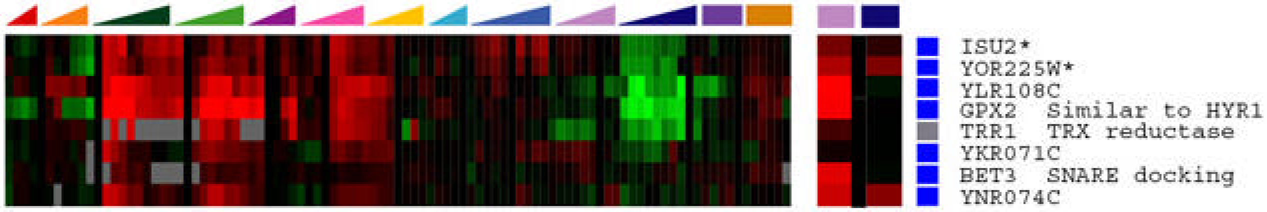
```
AGAATCTGAGAAAAAA 11 71 0
AGAAGGAAAAAAAAAA 3 130 1
AAAAAAATTCAGAAAA 3 335 1
AAAAAATAGAAGAAAA 13 669 1
AGAATGAAAAAAAAAA 16 429 1
AAAAGTGCCCAGAAAA 1 565 0
AAAAAAAAAAAGAAAA 18 618 1
AAAAAAAAGAAGAAAA 10 102 0
AGAAGGTGAAAGAACA 1 587 1
AGAAAACAAGAGAAAG 1 734 0
GGAATTAAAAGAAAAA 4 669 0
AAAATCTGGAGAAAAA 3 492 0
GGAAGAAGAAAGAAAG 16 685 1
AAAATTGAACAAAAAA 7 560 0
AGAATATAAAAGAACG 1 466 0
****      ******
MAP Score: 31.3674
```

Figure iv.

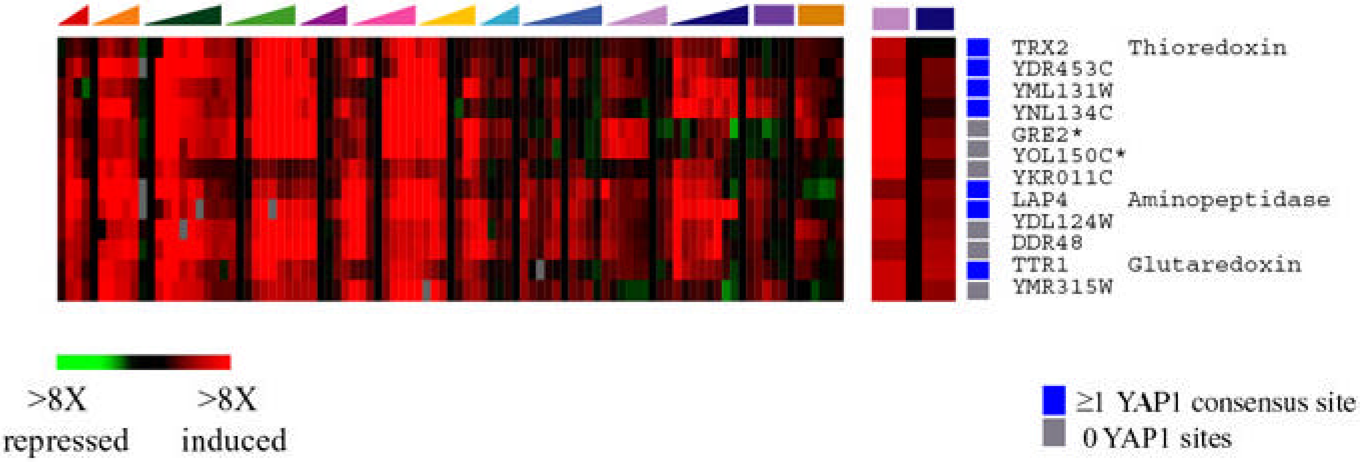
A. The "TRR2 cluster"



B. The "TRR1 cluster"



C. The "TRX2 cluster"



MEME RESULTS

Upregulated

Date: Sun, 2 Jun 2002 21:09:03 -0700
From: MEME <meme@SDSC.EDU>
X-Accept-Language: en
Subject: MEME job 52458 results: upreg (Use web browser to view results)
To: undisclosed-recipients:;

Content-Type: text/html; charset=us-ascii
 name="attach"
Content-Transfer-Encoding: 7bit
Content-Dispositon: inline;
 filename="attach"

ATTENTION: 0031-408 8 tasks allocated by LoadLeveler, continuing...
Command line Training Set First Motif Summary of Motifs Termination Explanation

Search sequence databases with these motifs using MAST.
Submit these motifs to BLOCKS multiple alignment processor.
Build and use a motif-based hidden Markov model (HMM) using Meta-MEME.

ld

MEME - Motif discovery tool

ld

MEME version 3.0 (Release date: 2001/03/05 14:24:28)

For further information on how to interpret these results or to get a copy of the MEME software please access http://meme.sdsc.edu.

This file may be used as input to the MAST algorithm for searching sequence databases for matches to groups of motifs. MAST is available for interactive use and downloading at http://meme.sdsc.edu.

ld

REFERENCE

ld

If you use this program in your research, please cite:

Timothy L. Bailey and Charles Elkan, "Fitting a mixture model by expectation maximization to discover motifs in biopolymers", Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, pp. 28-36, AAAI Press, Menlo Park, California, 1994.

ld

TRAINING SET

ld

```
DATAFILE= pasted sequences
ALPHABET= ACGT
Sequence name            Weight Length  Sequence name            Weight
Length
```

```
------------             ------ ------  ------------             ------
------
SSA2                     1.0000    800  HSC82                    1.0000
800
SBA1                     1.0000    800  SSA1                     1.0000
800
Yro2                     1.0000    800  SIS1                     1.0000
800
HSP26                    1.0000    800  HSP104                   1.0000
800
SSA3                     1.0000    800  STI1                     1.0000
800
Cpr6                     1.0000    800  HSP82                    1.0000
800
Cph1                     1.0000    800  Cpr8                     1.0000
800
YFR041c                  1.0000    800  Caj1                     1.0000
800
Xdj1                     1.0000    800  Afg1                     1.0000
800
YPR061c                  1.0000    800
```

ld

COMMAND LINE SUMMARY

ld

```
This information can also be useful in the event you wish to report a
problem with the MEME software.

command: meme meme.52458.data -dna -mod zoops -nmotifs 3 -minw 6 -maxw
10 -evt 10000 -revcomp -time 7200 -maxsize 60000 -nostatus -maxiter 20

model:  mod=         zoops    nmotifs=         3    evt=         10000
object function=  E-value of product of p-values
width:  minw=            6    maxw=           10    minic=        0.00
width:  wg=             11    ws=              1    endgaps=       yes
nsites: minsites=        2    maxsites=       19    wnsites=       0.8
theta:  prob=            1    spmap=         uni    spfuzz=        0.5
em:     prior=   dirichlet    b=            0.01    maxiter=        20
        distance=    1e-05
data:   n=           15200    N=              19
strands: + -
sample: seed=            0    seqfrac=         1
Letter frequencies in dataset:
A 0.312 C 0.188 G 0.188 T 0.312
Background letter frequencies (from dataset with add-one prior
applied):
A 0.312 C 0.188 G 0.188 T 0.312
```

ld

P N MOTIF 1 width = 10 sites = 19 llr = 188 E-value = 5.3e-002

```
Id
SimplifiedA77aa1aa881
pos.-specificC:::::::11:
probabilityG23::9:::29
matrixT1::::::1::
.
bits 2.4
2.2
1.9
1.7
Information 1.4
content 1.2
(14.3 bits)1.0
0.7
0.5
0.2
0.0
Id
.
Multilevel AAAAGAAAAG
consensus G
sequence
.
NAME STRAND START P-VALUE    SITES

YPR061c+6263.16e-06 CGCAAAAAAAAAAAGAAAAGTTCAGGGTCT
Caj1+7843.16e-06 AACTCAAAGGAAAAGAAAAGAGGGAAT
HSP26-823.16e-06 TTGCCAAATGAAAAGAAAAGAAAGTTAAAA
SSA1+7443.16e-06 TAAGCACATCAAAAGAAAAGTAATCAAGTA
Yro2-4235.07e-06 AGAAAACAAGAGAAGAAAAGGGCGGAAAAC
SBA1-245.07e-06 GGCAAAGTATAGAAGAAAAGAAGTTAGAGT
Cpr6+2166.98e-06 AGATTAAAGGGAAAGAAAAGCTCTGGCAGG
HSP104-4666.98e-06 CAAGAACAACGAAAGAAAAGTTAGGAACAG
HSC82-5776.98e-06 TCACCTTCTAGAAAGAAAAGTGCCCAGAAA
Xdj1+7628.89e-06 GCTTGGACCCAAAAGAAAGGGAAAAGTAAG
Afg1-7171.12e-05 CCCGCAAATCAGAAGAAAGGAACTTCTCCC
HSP82+4531.31e-05 ATACACGTAAAAAAGAACAGGAATAAAGCT
Cph1-2521.43e-05 GAAAGGTTCTAGAAGAACAGGTAATAACAC
SIS1+512.24e-05 TGTCCTGTTGAAAAGAAACGGAGGTAACGT
YFR041c+7513.50e-05 AACACGTACCAAAAGAATAGAAAACGAACA
Cpr8+6773.81e-05 TTAAAAAAATAGAAGAAAAATATAGAATAT
SSA3+4555.09e-05 GTATCAGTGATGAAGAAAGGGATGACGAGG
SSA2+5955.09e-05 GAAGATTTCGAAAAAAAAAGCTTCGAGGTT
STI1+6479.36e-05 CTTTAAGGAATAAAGAAAAAATCACATACA


.
Multilevel CCGCCCCTGC
consensus GAAT
sequence TT
.
NAME STRAND START P-VALUE    SITES

YPR061c+6569.24e-08 CCGGGTACTACCGCCCCTGCAAGGAATGAC
Cpr8+2841.85e-07 TACTAGGCGACCGCCGCTGCACACAGCTGT
SBA1-6545.83e-07 CCGCTCCTCTCCGCCCCTTCAAAAACACTC
HSC82+5351.55e-06 GTTAAAATAGCCGCCGATGCATTTTATTAC
```

```
Caj1-5891.95e-06 TTTCCGTTCTCCGCCCCATCTTGGTATGAC
Cpr6+6782.94e-06 GAATATTTTTCCGCCCTAGCTGGAGAAAGT
SSA2+1293.95e-06 TACATGAAAGCCGCCTTTGCTCAATGGAGA
Cph1+5206.70e-06 CCGCCACGTCCCGCCTCGTCTCGAATTTCT
SSA1-5528.86e-06 CTCCCTTTTGCCGCCGATGGAACGTTCTGG
```

Downregulated

Date: Sun, 2 Jun 2002 21:11:04 -0700
From: MEME <meme@SDSC.EDU>
X-Accept-Language: en
Subject: MEME job 42042 results: downreg (Use web browser to view results)
To: undisclosed-recipients:;

Content-Type: text/html; charset=us-ascii
name="attach"
Content-Transfer-Encoding: 7bit
Content-Dispositon: inline;
filename="attach"

ATTENTION: 0031-408 8 tasks allocated by LoadLeveler, continuing...
Command line Training Set First Motif Summary of Motifs Termination Explanation

Search sequence databases with these motifs using MAST.
Submit these motifs to BLOCKS multiple alignment processor.
Build and use a motif-based hidden Markov model (HMM) using Meta-MEME.

ld

MEME - Motif discovery tool

ld

MEME version 3.0 (Release date: 2001/03/05 14:24:28)

For further information on how to interpret these results or to get a copy of the MEME software please access http://meme.sdsc.edu.

This file may be used as input to the MAST algorithm for searching sequence databases for matches to groups of motifs. MAST is available for interactive use and downloading at http://meme.sdsc.edu.

ld

REFERENCE

ld

If you use this program in your research, please cite:

Timothy L. Bailey and Charles Elkan, "Fitting a mixture model by expectation maximization to discover motifs in biopolymers", Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, pp. 28-36, AAAI Press, Menlo Park, California, 1994.

ld

TRAINING SET

ld

```
DATAFILE= pasted sequences
ALPHABET= ACGT
Sequence name            Weight Length  Sequence name            Weight
Length
-------------            ------ ------  -------------            ------
------
Gim4                     1.0000    800  Fpr1                     1.0000
800
SSB1                     1.0000    800  Gim5                     1.0000
800
Gim1                     1.0000    800  Gim3                     1.0000
800
Gim2                     1.0000    800  Egd1                     1.0000
800
Egd2                     1.0000    800  Pdr13                    1.0000
800
SSB2                     1.0000    800  Zuo1                     1.0000
800
Fpr4                     1.0000    800  Npi46                    1.0000
800
Cct2                     1.0000    800  Cct8                     1.0000
800
Cct3                     1.0000    800  Cct5                     1.0000
800
Cct6                     1.0000    800  Cct4                     1.0000
800
Cct7                     1.0000    800  Cpr8                     1.0000
800
Cct1                     1.0000    800
```

ld

COMMAND LINE SUMMARY

ld

```
This information can also be useful in the event you wish to report a
problem with the MEME software.

command: meme meme.42042.data -dna -mod zoops -nmotifs 3 -minw 6 -maxw
10 -evt 10000 -revcomp -time 7200 -maxsize 60000 -nostatus -maxiter 20

model:  mod=         zoops    nmotifs=         3    evt=         10000
object function=  E-value of product of p-values
width:  minw=            6    maxw=           10    minic=        0.00
width:  wg=             11    ws=              1    endgaps=       yes
nsites: minsites=        2    maxsites=       23    wnsites=       0.8
theta:  prob=            1    spmap=         uni    spfuzz=        0.5
em:     prior=   dirichlet    b=            0.01    maxiter=        20
        distance=    1e-05
data:   n=           18400    N=              23
strands: + -
sample: seed=            0    seqfrac=         1
Letter frequencies in dataset:
```

```
A 0.311 C 0.189 G 0.189 T 0.311
Background letter frequencies (from dataset with add-one prior
applied):
A 0.311 C 0.189 G 0.189 T 0.311
```

Id

P N MOTIF 1 width = 10 sites = 23 llr = 220 E-value = 2.8e-002

Id

SimplifiedA787aa8aaaa
pos.-specificC:2:::::::
probabilityG3:3::2::::
matrixT::::::::::

.
bits 2.4
2.2
1.9
1.7
Information 1.4
content 1.2
(13.8 bits)1.0
0.7
0.5
0.2
0.0

Id

.
Multilevel AAAAAAAAAA
consensus GG
sequence

.
NAME STRAND START P-VALUE SITES

Cct1-6168.43e-06 CGGCAAGTGAAAAAAAAAAATCTGAGATTA
Cct7+6728.43e-06 CTTGCTATTGAAAAAAAAAAAAATGAATAA
Cct6+198.43e-06 AAAGGAAACGAAAAAAAAAAAAAGTAGGGA
Zuo1-3598.43e-06 AAAATTGCCCAAAAAAAAAAGGCCATTTGA
SSB2-6178.43e-06 AATCATGAAAAAAAAAAAAATTGAAAAAGA
Egd2-6558.43e-06 AAAATTTTCAAAAAAAAAAACAGACCTAAC
Gim2+5318.43e-06 ACCGAAAAATAAAAAAAAAAAAAAAAAAAA
Gim1+6928.43e-06 GATGCGAAAGAAAAAAAAAAGTAAAGGCAA
SSB1-5938.43e-06 TATAAAAAAAAAAAAAAAAAGATTCATCCA
Gim4+6658.43e-06 GAAGTAAATGAAAAAAAAAAAAAAAAGAAT
Cct3+6771.87e-05 GCATGGTGCGAAGAAAAAAAATAGCAAATC
Egd1-2841.87e-05 TGAGAAAAAAGAAAAAAAAAGGCCTAAGGT
Fpr4+7662.89e-05 AATGATAAGAAAAAAGAAAAAGACAACAAA
Gim5-3352.89e-05 AACACTACCCAAAAAGAAAATCCACTCCTG
Cct2-6823.21e-05 CCACCGCGAAGAGAAAAAAATTTCAGATCG
Cct4-6643.83e-05 TGGTGGCGAAACGAAAAAAATTTCAGTTTC
Cct8+6983.83e-05 TCCGCTAGTGGCAAAAAAAAGTGAAGAAAA
Fpr1+6763.83e-05 TCTTTTTCGGGCAAAAAAAATTGCCGGAAG
Cct5+354.45e-05 TAAAGAATCCAAGAAGAAAAGAAAGGCTTA
Pdr13-7417.06e-05 ATCACATTCGACGAAGAAAAAGAGCTAAAA
Cpr8+6547.69e-05 CAATTTGATTGGAAAAAAAATATTTAAAAA

Npi46+1909.05e-05 TAGGCACTTAGAAAAACAAACCTTTGTGCC
Gim3-2899.05e-05 ACTGCGATAGAAGGAAAAAATCTCCGCTGC

Yap1p

Date: Sun, 2 Jun 2002 21:11:04 -0700
From: MEME <meme@SDSC.EDU>
X-Accept-Language: en
Subject: MEME job 42042 results: downreg (Use web browser to view results)
To: undisclosed-recipients:;

Content-Type: text/html; charset=us-ascii
 name="attach"
Content-Transfer-Encoding: 7bit
Content-Dispositon: inline;
 filename="attach"

ATTENTION: 0031-408 8 tasks allocated by LoadLeveler, continuing...
Command line Training Set First Motif Summary of Motifs Termination Explanation

Search sequence databases with these motifs using MAST.
Submit these motifs to BLOCKS multiple alignment processor.
Build and use a motif-based hidden Markov model (HMM) using Meta-MEME.

ld

MEME - Motif discovery tool

ld

MEME version 3.0 (Release date: 2001/03/05 14:24:28)

For further information on how to interpret these results or to get a copy of the MEME software please access http://meme.sdsc.edu.

This file may be used as input to the MAST algorithm for searching sequence databases for matches to groups of motifs. MAST is available for interactive use and downloading at http://meme.sdsc.edu.

ld

REFERENCE

ld

If you use this program in your research, please cite:

Timothy L. Bailey and Charles Elkan, "Fitting a mixture model by expectation maximization to discover motifs in biopolymers", Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, pp. 28-36, AAAI Press, Menlo Park, California, 1994.

ld

TRAINING SET

ld

```
DATAFILE= pasted sequences
```

```
ALPHABET= ACGT
Sequence name            Weight Length  Sequence name            Weight
Length
-------------            ------ ------  -------------            ------
------
Gim4                     1.0000    800  Fpr1                     1.0000
800
SSB1                     1.0000    800  Gim5                     1.0000
800
Gim1                     1.0000    800  Gim3                     1.0000
800
Gim2                     1.0000    800  Egd1                     1.0000
800
Egd2                     1.0000    800  Pdr13                    1.0000
800
SSB2                     1.0000    800  Zuo1                     1.0000
800
Fpr4                     1.0000    800  Npi46                    1.0000
800
Cct2                     1.0000    800  Cct8                     1.0000
800
Cct3                     1.0000    800  Cct5                     1.0000
800
Cct6                     1.0000    800  Cct4                     1.0000
800
Cct7                     1.0000    800  Cpr8                     1.0000
800
Cct1                     1.0000    800
```

Id

COMMAND LINE SUMMARY

Id

```
This information can also be useful in the event you wish to report a
problem with the MEME software.

command: meme meme.42042.data -dna -mod zoops -nmotifs 3 -minw 6 -maxw
10 -evt 10000 -revcomp -time 7200 -maxsize 60000 -nostatus -maxiter 20

model:  mod=         zoops    nmotifs=         3    evt=         10000
object function=  E-value of product of p-values
width:  minw=            6    maxw=           10    minic=        0.00
width:  wg=             11    ws=              1    endgaps=       yes
nsites: minsites=        2    maxsites=       23    wnsites=       0.8
theta:  prob=            1    spmap=         uni    spfuzz=        0.5
em:     prior=   dirichlet    b=            0.01    maxiter=        20
        distance=    1e-05
data:   n=           18400    N=              23
strands: + -
sample: seed=            0    seqfrac=         1
Letter frequencies in dataset:
A 0.311 C 0.189 G 0.189 T 0.311
Background letter frequencies (from dataset with add-one prior
applied):
```

```
A 0.311 C 0.189 G 0.189 T 0.311
```

ld

P N MOTIF 1 width = 10 sites = 23 llr = 220 E-value = 2.8e-002

ld

SimplifiedA787aa8aaaa
pos.-specificC:2::::::::
probabilityG3:3::2::::
matrixT::::::::::

.
bits 2.4
2.2
1.9
1.7
Information 1.4
content 1.2
(13.8 bits)1.0
0.7
0.5
0.2
0.0

ld

.
Multilevel **AAAAAAAAAA**
consensus GG
sequence
.
NAME STRAND START P-VALUE SITES

Cct1-6168.43e-06 CGGCAAGTGA**AAAAAAAAAATCTGAGATTA**
Cct7+6728.43e-06 CTTGCTATTGAAAAAAAAAAAAAATGAATAA
Cct6+198.43e-06 AAAGGAAACGAAAAAAAAAAAAAAGTAGGGA
Zuo1-3598.43e-06 AAAATTGCCCAAAAAAAAAAGGCCATTTGA
SSB2-6178.43e-06 AATCATGAAAAAAAAAAAAAATTGAAAAAGA
Egd2-6558.43e-06 AAAATTTTCAAAAAAAAAAACAGACCTAAC
Gim2+5318.43e-06 ACCGAAAAATAAAAAAAAAAAAAAAAAAAA
Gim1+6928.43e-06 GATGCGAAAGAAAAAAAAAAGTAAAGGCAA
SSB1-5938.43e-06 TATAAAAAAAAAAAAAAAAAAGATTCATCCA
Gim4+6658.43e-06 GAAGTAAATGAAAAAAAAAAAAAAAAAGAAT
Cct3+6771.87e-05 GCATGGTGCGAAGAAAAAAAATAGCAAATC
Egd1-2841.87e-05 TGAGAAAAAAGAAAAAAAAAGGCCTAAGGT
Fpr4+7662.89e-05 AATGATAAGAAAAAAGAAAAAGACAACAAA
Gim5-3352.89e-05 AACACTACCCAAAAAGAAAATCCACTCCTG
Cct2-6823.21e-05 CCACCGCGAAGAGAAAAAAATTTCAGATCG
Cct4-6643.83e-05 TGGTGGCGAAACGAAAAAAATTTCAGTTTC
Cct8+6983.83e-05 TCCGCTAGTGGCAAAAAAAAGTGAAGAAAA
Fpr1+6763.83e-05 TCTTTTTCGGGCAAAAAAAATTGCCGGAAG
Cct5+354.45e-05 TAAAGAATCCAAGAAGAAAAGAAAGGCTTA
Pdr13-7417.06e-05 ATCACATTCGACGAAGAAAAAGAGCTAAAA
Cpr8+6547.69e-05 CAATTTGATTGGAAAAAAAATATTTAAAAA
Npi46+1909.05e-05 TAGGCACTTAGAAAAACAAACCTTTGTGCC
Gim3-2899.05e-05 ACTGCGATAGAAGGAAAAAATCTCCGCTGC

CONSENSUS RESULTS

Date: Sun, 2 Jun 2002 20:11:23 -0500 (CDT)
To: stephent@Stanford.EDU
From: stormosysadmin@genetics.wustl.edu
Subject: The consensus/wconsensus program output results

```
COMMAND LINE: ../../Program/consensus/consensus-v6c -L 7 -q 1000 -A a:t c:g -c0 -pr2 -pt 4 -
pf 0 -f /ural/d/choi/public_html/consensus/user/128.12.184.102/consensus/sequence

***** PID: 20915 *****
L-mer Width: 7
Minimum distance between starting points of words: not relevant
Save the top alignments derived from each intermediate alignment
    Maximum number of matrices to save between cycles: 1000
Status of complementary sequence: IGNORE.
Algorithim options: one match per sequence.
             Stop only when the maximum number of cycles is reached.
The number of matrices to print.
    Top Matrices saved from each cycle: 4
    Matrices Saved from the last cycle: NONE

***** Sequence information from file
"/ural/d/choi/public_html/consensus/user/128.12.184.102/consensus/sequence". *****
sequence 1: Trr2
    fragments: 1-800
sequence 2: Gtt2
    fragments: 1-800
sequence 3: Aad6
    fragments: 1-800
sequence 4: Aad4
    fragments: 1-800
sequence 5: ISU2
    fragments: 1-800
sequence 6: Gpx2
    fragments: 1-800
sequence 7: Bet3
    fragments: 1-800
sequence 8: Trx2
    fragments: 1-800
sequence 9: Lap4
    fragments: 1-800
sequence 10: Flr1
    fragments: 1-800
sequence 11: Ttr1
    fragments: 1-800
sequence 12: Ydl124w
    fragments: 1-800
sequence 13: Ylr108c
    fragments: 1-800
sequence 14: YLL055w
    fragments: 1-800
sequence 15: SDL1
    fragments: 1-800
sequence 16: YLR460c
```

```
     fragments: 1-800
sequence 17: YCR102c
     fragments: 1-800
sequence 18: YNL260c
     fragments: 1-800
sequence 19: YKL071w
     fragments: 1-800
sequence 20: YDR132c
     fragments: 1-800
sequence 21: Aad15
     fragments: 1-800
sequence 22: YOL118c
     fragments: 1-800
sequence 23: YKR071c
     fragments: 1-800
sequence 24: YML131w
     fragments: 1-800
sequence 25: YNL134c
     fragments: 1-800

Total number of sequences: 25.
Total number of sequence fragments: 25.

#**** Information on observed frequency and occurrence of each letter. ****#
#Total number of letters in the input sequences = 20000
A  0.323650; observed occurrence =  6473 (letter   1)
C  0.188250; observed occurrence =  3765 (letter   2)
G  0.188100; observed occurrence =  3762 (letter   3)
T  0.300000; observed occurrence =  6000 (letter   4)

PRIOR FREQUENCIES DETERMINED BY OBSERVED FREQUENCIES.
***** Information for the alphabet from the command line. *****
letter   1: A (complement: T)  prior frequency = 0.323650
letter   2: C (complement: G)  prior frequency = 0.188250
letter   3: G (complement: C)  prior frequency = 0.188100
letter   4: T (complement: A)  prior frequency = 0.300000


INFORMATION CONTENT IS CALCULATED USING NATURAL LOGARITHMS (i.e. BASE e).
DIVIDE BY ln(2) = 0.693 TO CONVERT TO BASE 2, WHICH WAS USED IN
PREVIOUS VERSIONS OF THIS PROGRAM.

      []          MATRICES SAVED FOR NEXT CYCLE            []
      []---------------------------------------------------[]
      []  total   | top adjusted |   ln top    [] ln expected []
CYCLE []  number  | information  |   p-value   []  frequency  []
------[]----------|--------------|-------------[]-------------[]
    1 []    19850 |       2.2114 |      0.0000 []      9.8960 []
    2 []      688 |       5.7728 |    -20.5323 []     -1.4743 []
    3 []      891 |       7.4711 |    -31.1067 []     -3.3347 []
    4 []      760 |       8.4692 |    -42.3972 []     -6.2435 []
    5 []      868 |       8.6191 |    -49.9219 []     -5.6559 []
    6 []      785 |       8.9341 |    -59.5587 []     -7.4117 []
    7 []      814 |       8.9844 |    -67.2030 []     -7.3805 []
    8 []      790 |       8.9109 |    -73.7638 []     -6.4532 []
    9 []      804 |       8.8804 |    -81.0224 []     -6.3987 []
```

```
   10 []      810 |       8.7827 |    -87.4588 []     -5.6881 []
   11 []      835 |       8.6544 |    -93.3971 []     -4.6391 []
   12 []      827 |       8.5388 |    -99.4070 []     -3.8177 []
   13 []      847 |       8.4083 |   -105.0357 []     -2.7694 []
   14 []      851 |       8.2334 |   -109.7098 []     -0.9206 []
   15 []      854 |       8.1033 |   -114.9378 []      0.2184 []
   16 []      869 |       7.9391 |   -119.3099 []      2.0533 []
   17 []      895 |       7.8006 |   -123.9385 []      3.4658 []
   18 []      902 |       7.6496 |   -128.0661 []      5.2044 []
   19 []      901 |       7.5035 |   -132.0469 []      6.9021 []
   20 []      916 |       7.3590 |   -135.8110 []      8.6112 []
   21 []      942 |       7.2356 |   -139.8248 []      9.8394 []
   22 []      948 |       7.1286 |   -144.0270 []     10.6095 []
   23 []      944 |       6.9983 |   -147.4440 []     11.8327 []
   24 []      957 |       6.8779 |   -150.8896 []     12.5793 []
   25 []      980 |       6.7325 |   -153.4355 []     13.4915 []

INFORMATION CONTENT IS CALCULATED USING NATURAL LOGARITHMS (i.e. BASE e).
DIVIDE BY ln(2) = 0.693 TO CONVERT TO BASE 2, WHICH WAS USED IN
PREVIOUS VERSIONS OF THIS PROGRAM.

THE LIST OF TOP MATRICES FROM EACH CYCLE--sorted by expected frequency (total of 34):

MATRIX 1
number of sequences = 6
unadjusted information = 11.0559
sample size adjusted information = 8.93408
ln(p-value) = -59.5587   p-value = 1.36146E-26
ln(expected frequency) = -7.41169   expected frequency = 0.00060415
A |   0   0   0   0   0   0   0
C |   0   2   6   6   0   0   6
G |   6   4   0   0   6   6   0
T |   0   0   0   0   0   0   0
  1|5   :   5/687   GGCCGGC
  2|3   :   6/521   GGCCGGC
  3|4   :  10/256   GGCCGGC
  4|2   :  16/505   GCCCGGC
  5|1   :  17/481   GCCCGGC
  6|6   :  20/468   GGCCGGC

MATRIX 2
number of sequences = 7
unadjusted information = 10.7779
sample size adjusted information = 8.98439
ln(p-value) = -67.203   p-value = 6.51761E-30
ln(expected frequency) = -7.38046   expected frequency = 0.000623313
A |   0   0   0   3   0   0   0
C |   0   0   7   4   0   0   7
G |   7   7   0   0   7   7   0
T |   0   0   0   0   0   0   0
  1|7   :   5/687   GGCCGGC
  2|4   :   6/521   GGCCGGC
  3|1   :   9/200   GGCAGGC
  4|5   :  10/256   GGCCGGC
  5|2   :  16/681   GGCAGGC
  6|3   :  17/657   GGCAGGC
```

```
  7|6   :  20/468   GGCCGGC

MATRIX 3
number of sequences = 8
unadjusted information = 10.4573
sample size adjusted information = 8.9109
ln(p-value) = -73.7638   p-value = 9.22154E-33
ln(expected frequency) = -6.45317   expected frequency = 0.00157552
A |   0   0   0   3   0   0   0
C |   0   0   8   4   0   0   8
G |   8   8   0   0   8   8   0
T |   0   0   0   1   0   0   0
  1|7   :   1/489   GGCTGGC
  2|3   :   5/687   GGCCGGC
  3|8   :   6/521   GGCCGGC
  4|1   :   9/200   GGCAGGC
  5|4   :  10/256   GGCCGGC
  6|2   :  16/681   GGCAGGC
  7|6   :  17/657   GGCAGGC
  8|5   :  20/468   GGCCGGC

MATRIX 4
number of sequences = 9
unadjusted information = 10.2354
sample size adjusted information = 8.88044
ln(p-value) = -81.0224   p-value = 6.49283E-36
ln(expected frequency) = -6.3987   expected frequency = 0.00166372
A |   0   0   0   4   0   0   0
C |   0   0   9   5   2   0   9
G |   9   9   0   0   7   9   0
T |   0   0   0   0   0   0   0
  1|7   :   5/687   GGCCGGC
  2|4   :   6/521   GGCCGGC
  3|1   :   9/200   GGCAGGC
  4|5   :  10/256   GGCCGGC
  5|8   :  11/694   GGCCCGC
  6|9   :  12/498   GGCACGC
  7|2   :  16/681   GGCAGGC
  8|3   :  17/657   GGCAGGC
  9|6   :  20/468   GGCCGGC
```

BIOPROSPECTOR RESULTS

Downregulated

Date: Thu, 30 May 2002 11:40:26 -0700 (PDT)
From: Web Administrator <web@SMI.Stanford.EDU>
Apparently-To: stephent@Stanford.EDU

```
****************************************
*                                *
*      BioProspector Search Result     *
*                                *
****************************************
```

```
The highest scoring 3 motifs are:

Motif #1:
******************************
Width (8, 0); Gap [0, 0]; MotifScore 3.306; Segments 22

Blk1    A     C     G     T      Con  rCon Deg  rDeg
1     0.00  0.99  0.00  0.00      C    G    C    G
2     0.99  0.00  0.00  0.00      A    T    A    T
3     0.00  0.00  0.00  0.99      T    A    T    A
4     0.99  0.00  0.00  0.00      A    T    A    T
5     0.00  0.27  0.00  0.72      T    A    Y    R
6     0.45  0.00  0.00  0.54      T    A    W    W
7     0.00  0.41  0.00  0.59      T    A    Y    R
8     0.99  0.00  0.00  0.00      A    T    A    T

>Gim5   seg 1    f82       CATACATA
>Gim2   seg 1    f775      CATATATA
>Egd1   seg 1    r345      CATACTTA
>Egd2   seg 1    f414      CATATATA
>Egd2   seg 2    f786      CATACTCA
>Egd2   seg 3    r783      CATATTCA
>Pdr13 seg 1     f584      CATATATA
>Zuo1   seg 1    f523      CATATATA
>Zuo1   seg 2    r248      CATATTTA
>Zuo1   seg 3    r353      CATATTTA
>Fpr4   seg 1    r580      CATATATA
>Npi46 seg 1     f764      CATACATA
>Npi46 seg 2     r334      CATATACA
>Cct2   seg 1    r765      CATATTCA
>Cct8   seg 1    f340      CATATATA
>Cct8   seg 2    r745      CATATTCA
>Cct8   seg 3    f360      CATATTCA
>Cct3   seg 1    f525      CATATATA
>Cct5   seg 1    f611      CATACTCA
>Cct4   seg 1    r444      CATATTTA
>Cpr8   seg 1    f747      CATATTCA
>Cct1   seg 1    r775      CATACTCA
******************************

Motif #2:
******************************
Width (8, 0); Gap [0, 0]; MotifScore 3.295; Segments 26

Blk1    A     C     G     T      Con  rCon Deg  rDeg
1     0.00  0.00  0.00  0.99      T    A    T    A
2     0.00  0.00  0.00  0.99      T    A    T    A
3     0.00  0.00  0.00  0.99      T    A    T    A
4     0.00  0.00  0.00  0.99      T    A    T    A
5     0.00  0.00  0.00  0.99      T    A    T    A
6     0.00  0.00  0.00  0.99      T    A    T    A
7     0.00  0.00  0.00  0.99      T    A    T    A
8     0.00  0.00  0.00  0.99      T    A    T    A

>Gim4   seg 1    r121      TTTTTTTT
```

```
>Gim4   seg 2    r129     TTTTTTTT
>Fpr1   seg 1    r116     TTTTTTTT
>SSB1  seg 1    f593     TTTTTTTT
>SSB1  seg 2    r165     TTTTTTTT
>SSB1  seg 3    f601     TTTTTTTT
>Gim1   seg 1    r100     TTTTTTTT
>Gim2   seg 1    r247     TTTTTTTT
>Gim2   seg 2    r255     TTTTTTTT
>Gim2   seg 3    r263     TTTTTTTT
>Egd1   seg 1    r567     TTTTTTTT
>Egd1   seg 2    f284     TTTTTTTT
>Egd1   seg 3    r169     TTTTTTTT
>Egd2   seg 1    f655     TTTTTTTT
>SSB2  seg 1    f695     TTTTTTTT
>SSB2  seg 2    f617     TTTTTTTT
>Zuo1   seg 1    r335     TTTTTTTT
>Zuo1   seg 2    f359     TTTTTTTT
>Zuo1   seg 3    f662     TTTTTTTT
>Cct8   seg 1    r94      TTTTTTTT
>Cct3   seg 1    r114     TTTTTTTT
>Cct6   seg 1    r770     TTTTTTTT
>Cct7   seg 1    r117     TTTTTTTT
>Cpr8   seg 1    r138     TTTTTTTT
>Cct1   seg 1    f616     TTTTTTTT
>Cct1   seg 2    r123     TTTTTTTT
******************************

Motif #3:
******************************
Width (8, 0); Gap [0, 0]; MotifScore 3.246; Segments 16

Blk1    A     C     G     T      Con  rCon Deg  rDeg
1     0.00  0.00  0.00  0.99      T    A    T    A
2     0.37  0.00  0.31  0.31      A    T    D    H
3     0.00  0.99  0.00  0.00      C    G    C    G
4     0.00  0.00  0.13  0.87      T    A    T    A
5     0.37  0.00  0.62  0.00      G    C    R    Y
6     0.99  0.00  0.00  0.00      A    T    A    T
7     0.00  0.99  0.00  0.00      C    G    C    G
8     0.00  0.00  0.99  0.00      G    C    G    C

>Fpr1   seg 1    r479     TACTGACG
>Fpr1   seg 2    f132     TGCTGACG
>Gim5   seg 1    f717     TTCTAACG
>Gim5   seg 2    f535     TTCTAACG
>Egd2   seg 1    r392     TTCGGACG
>Zuo1   seg 1    f44      TGCTGACG
>Npi46 seg 1    f289     TACTGACG
>Cct2   seg 1    r230     TACTGACG
>Cct2   seg 2    f394     TTCTGACG
>Cct2   seg 3    f666     TTCTAACG
>Cct3   seg 1    f193     TACTGACG
>Cct3   seg 2    f476     TGCGAACG
>Cct5   seg 1    f563     TACTGACG
>Cct7   seg 1    r91      TGCTAACG
>Cct7   seg 2    r499     TACTAACG
```

```
>Cct1   seg 1   f401    TGCTGACG
*****************************

Thanks for using BioProspector.
For questions, please contact Xiaole Liu at xliu@smi.stanford.edu.
```

Upregulated

Date: Thu, 30 May 2002 11:41:50 -0700 (PDT)
From: Web Administrator <web@SMI.Stanford.EDU>
Apparently-To: stephent@Stanford.EDU

```
***************************************
*                             *
*     BioProspector Search Result     *
*                             *
***************************************

The highest scoring 3 motifs are:

Motif #1:
*****************************
Width (8, 0); Gap [0, 0]; MotifScore 3.378; Segments 26

Blk1    A     C     G     T      Con  rCon Deg  rDeg
1     0.00  0.00  0.99  0.00      G    C    G    C
2     0.99  0.00  0.00  0.00      A    T    A    T
3     0.99  0.00  0.00  0.00      A    T    A    T
4     0.99  0.00  0.00  0.00      A    T    A    T
5     0.99  0.00  0.00  0.00      A    T    A    T
6     0.80  0.19  0.00  0.00      A    T    A    T
7     0.99  0.00  0.00  0.00      A    T    A    T
8     0.99  0.00  0.00  0.00      A    T    A    T

>SSA2  seg 1   f594    GAAAAAAA
>SSA2  seg 2   r270    GAAAAAAA
>SSA2  seg 3   r231    GAAAAAAA
>HSC82         seg 1   r52     GAAAACAA
>SBA1  seg 1   r327    GAAAAAAA
>SSA1  seg 1   f302    GAAAAAAA
>SSA1  seg 2   f332    GAAAAAAA
>SSA1  seg 3   f136    GAAAAAAA
>Yro2  seg 1   f531    GAAAAAAA
>Yro2  seg 2   r195    GAAAAAAA
>Yro2  seg 3   r731    GAAAAAAA
>Yro2  seg 4   r126    GAAAAAAA
>Yro2  seg 5   f140    GAAAAAAA
>Yro2  seg 6   r360    GAAAACAA
>HSP26         seg 1   f113    GAAAACAA
>HSP104        seg 1   f355    GAAAAAAA
>SSA3  seg 1   f387    GAAAACAA
>STI1  seg 1   f595    GAAAAAAA
>Cpr6  seg 1   r682    GAAAAAAA
>HSP82         seg 1   r233    GAAAAAAA
>Cpr8  seg 1   f655    GAAAAAAA
```

```
>YFR041c        seg 1   f718    GAAAAAAA
>Caj1   seg 1   r677    GAAAAAAA
>Xdj1   seg 1   f459    GAAAAAAA
>Xdj1   seg 2   f435    GAAAAAAA
>Afg1   seg 1   r413    GAAAACAA
******************************

Motif #2:
******************************
Width (8, 0); Gap [0, 0]; MotifScore 3.188; Segments 15

Blk1   A     C     G     T     Con  rCon Deg  rDeg
1     0.01  0.33  0.27  0.40    T    A    B    V
2     0.01  0.99  0.00  0.01    C    G    C    G
3     0.07  0.46  0.46  0.01    C    G    S    S
4     0.01  0.59  0.13  0.27    C    G    Y    R
5     0.01  0.99  0.00  0.01    C    G    C    G
6     0.01  0.99  0.00  0.01    C    G    C    G
7     0.01  0.99  0.00  0.01    C    G    C    G
8     0.01  0.00  0.00  0.99    T    A    T    A

>HSC82          seg 1   r147    TCCTCCCT
>SBA1   seg 1   r138    CCGCCCCT
>SSA1   seg 1   r204    GCCCCCCT
>SSA1   seg 2   r153    CCACCCCT
>Yro2   seg 1   f417    TCCGCCCT
>SIS1   seg 1   f576    TCGTCCCT
>HSP26          seg 1   r647    CCCCCCCT
>HSP104         seg 1   r165    GCGCCCCT
>Cpr6   seg 1   f677    TCCGCCCT
>HSP82          seg 1   f402    CCCCCCCT
>Cph1   seg 1   f470    GCCCCCCT
>Cph1   seg 2   f448    TCGCCCCT
>Cpr8   seg 1   f605    GCGTCCCT
>Caj1   seg 1   r151    TCGTCCCT
>YPR061c        seg 1   f656    CCGCCCCT
******************************

Motif #3:
******************************
Width (8, 0); Gap [0, 0]; MotifScore 3.160; Segments 19

Blk1   A     C     G     T     Con  rCon Deg  rDeg
1     0.00  0.00  0.00  0.99    T    A    T    A
2     0.99  0.00  0.00  0.00    A    T    A    T
3     0.99  0.00  0.00  0.00    A    T    A    T
4     0.99  0.00  0.00  0.00    A    T    A    T
5     0.00  0.00  0.00  0.99    T    A    T    A
6     0.99  0.00  0.00  0.00    A    T    A    T
7     0.26  0.00  0.31  0.42    T    A    D    H
8     0.00  0.00  0.73  0.26    G    C    K    M

>SSA2   seg 1   r70     TAAATAAG
>SSA2   seg 2   r5      TAAATATT
>HSC82          seg 1   r335    TAAATATT
>SBA1   seg 1   f366    TAAATAGG
```

```
>SSA1  seg 1   r613    TAAATAGT
>Yro2  seg 1   r67     TAAATATT
>HSP26         seg 1   f649    TAAATAGG
>HSP104        seg 1   r388    TAAATATG
>HSP104        seg 2   r416    TAAATAAG
>HSP104        seg 3   r400    TAAATAAG
>SSA3  seg 1   r359    TAAATAAG
>STI1  seg 1   f676    TAAATAGG
>Cpr6  seg 1   f715    TAAATAAG
>HSP82         seg 1   f665    TAAATATG
>Cpr8  seg 1   f559    TAAATAGG
>Cpr8  seg 2   r132    TAAATATT
>YFR041c       seg 1   r111    TAAATATG
>Caj1  seg 1   f259    TAAATATG
>Afg1  seg 1   f757    TAAATAGG
******************************

Thanks for using BioProspector.
For questions, please contact Xiaole Liu at xliu@smi.stanford.edu.
```

Yap1p

```
Date: Thu, 30 May 2002 19:26:38 -0700 (PDT)
From: Web Administrator <web@SMI.Stanford.EDU>
Apparently-To: stephent@Stanford.EDU

Pm 0.2638      Minimum match (6/8)

Top 3 motifs   Wid     Score1  Segment        Con     Deg
Mtf 1    8     3.831   57      TTACTAAT       TTASTAAK
Final Motif 1: Wid 8 Score1 3.831 Segment 57
        A    C    G    T     Con  rCon Deg  rDeg
1       2    2    2   94      T    A    T    A
2       2    6    8   83      T    A    T    A
3      94    2    2    2      A    T    A    T
4       2   56   39    2      C    G    S    S
5       2    2    6   90      T    A    T    A
6      78   11    2    9      A    T    A    T
7      94    2    2    2      A    T    A    T
8      11    2   42   46      T    A    K    M
Seq 2   St f65   TTACTTAG
Seq 2   St b92   TTAGTAAG
Seq 2   St f703  TTACTAAT
Seq 3   St b217TTACTAAT
Seq 3   St b236TTAGTAAG
Seq 3   St b262TTACGAAT
Seq 3   St f452  TTACGAAG
Seq 3   St f559  TTACTAAT
Seq 3   St f578  TTAGTAAG
Seq 3   St b691TTACTTAG
Seq 4   St b131TCACTAAT
Seq 4   St f348  TTACTAAT
Seq 4   St b350TTACTAAT
Seq 4   St f445  TTAGTAAG
Seq 4   St b447TTAGTAAT
```

```
Seq 5   St f363  TTACTAAG
Seq 6   St f174  TTAGTAAA
Seq 6   St b254TTACTAAT
Seq 6   St f343  TTACTCAT
Seq 6   St f384  TTACTAAA
Seq 6   St b411TTAGTAAG
Seq 6   St b482TTACGAAT
Seq 6   St f541  TTAGTAAT
Seq 6   St b621TTACTAAA
Seq 7   St f80    TCACTAAG
Seq 7   St b589TTACTTAT
Seq 8   St b175TTAGTAAG
Seq 8   St b211TTACTAAG
Seq 8   St f584  TTAGTAAA
Seq 8   St f620  TTACTAAG
Seq 9   St f36    TTAGTAAA
Seq 9   St b377TGACTAAT
Seq 9   St f418  TTAGTCAG
Seq 9   St b759TTACTAAT
Seq 10  St b142TGACTAAT
Seq 10  St b161TTAGTCAG
Seq 10  St b358TTAGTAAT
Seq 10  St f437  TTACTAAG
Seq 10  St f634  TGACTAAT
Seq 10  St f653  TTAGTCAG
Seq 10  St b692TTACTTAT
Seq 11  St b138TTAGTAAT
Seq 11  St f657  TTACTAAG
Seq 12  St f35    TTAGTAAT
Seq 12  St b188TTAGTAAG
Seq 12  St f607  TTACTAAT
Seq 12  St b760TTACTAAT
Seq 13  St b163TTAGTAAT
Seq 13  St f632  TTACTAAT
Seq 14  St b507TTAGTCAG
Seq 15  St f252  TTACTAAG
Seq 15  St b543TTAGTAAT
Seq 16  St b161TTAGTAAT
Seq 16  St b311TTAGTCAG
Seq 16  St f484  TGACTAAG
Seq 16  St f601  TCACTAAG
Seq 16  St f634  TTACTAAG
Mtf 2     8          3.801     54         TTACTAAT        TTASTAAK
Final Motif 2: Wid 8 Score1 3.801 Segment 54
         A     C     G     T     Con  rCon Deg  rDeg
1        3     2     2    94      T    A    T    A
2        8     8     8    75      T    A    T    A
3       94     2     2     3      A    T    A    T
4        3    56    39     3      C    G    S    S
5        3     2     5    91      T    A    T    A
6       87     2     2     9      A    T    A    T
7       94     2     2     3      A    T    A    T
8       11     2    34    53      T    A    K    M
Seq 2   St f65    TTACTTAG
Seq 2   St b92   TTAGTAAG
Seq 2   St f703  TTACTAAT
```

```
Seq 2    St b730TAAGTAAT
Seq 3    St f104  TAAGTAAT
Seq 3    St b217TTACTAAT
Seq 3    St b236TTAGTAAG
Seq 3    St b262TTACGAAT
Seq 3    St f559  TTACTAAT
Seq 3    St f578  TTAGTAAG
Seq 3    St b691TTACTTAG
Seq 4    St b131TCACTAAT
Seq 4    St f348  TTACTAAT
Seq 4    St b350TTACTAAT
Seq 4    St f445  TTAGTAAG
Seq 4    St b447TTAGTAAT
Seq 5    St f363  TTACTAAG
Seq 6    St f174  TTAGTAAA
Seq 6    St b254TTACTAAT
Seq 6    St f384  TTACTAAA
Seq 6    St b411TTAGTAAG
Seq 6    St b482TTACGAAT
Seq 6    St f541  TTAGTAAT
Seq 6    St b621TTACTAAA
Seq 7    St f80    TCACTAAG
Seq 7    St b589TTACTTAT
Seq 8    St b175TTAGTAAG
Seq 8    St b211TTACTAAG
Seq 8    St f545  TAAGTAAT
Seq 8    St f584  TTAGTAAA
Seq 8    St f620  TTACTAAG
Seq 9    St f36    TTAGTAAA
Seq 9    St b377TGACTAAT
Seq 9    St f742  TCAGTAAT
Seq 9    St b759TTACTAAT
Seq 10  St b142TGACTAAT
Seq 10  St b358TTAGTAAT
Seq 10  St f437  TTACTAAG
Seq 10  St f634  TGACTAAT
Seq 10  St b692TTACTTAT
Seq 11  St b138TTAGTAAT
Seq 11  St f657  TTACTAAG
Seq 12  St f35    TTAGTAAT
Seq 12  St b188TTAGTAAG
Seq 12  St f607  TTACTAAT
Seq 12  St b760TTACTAAT
Seq 13  St b163TTAGTAAT
Seq 13  St f632  TTACTAAT
Seq 15  St f252  TTACTAAG
Seq 15  St b543TTAGTAAT
Seq 16  St b161TTAGTAAT
Seq 16  St f484  TGACTAAG
Seq 16  St f601  TCACTAAG
Seq 16  St f634  TTACTAAG
Mtf 3     8           3.792    52         CTTACTAA       MTTASTAA
Final Motif 3: Wid 8 Score1 3.792 Segment 52
        A     C     G     T     Con  rCon Deg  rDeg
1       41    44     3    11      C    G    M    K
2        3     2     2    94      T    A    T    A
```

3	6	2	10	82	T	A	T	A
4	94	2	2	3	A	T	A	T
5	3	51	44	3	C	G	S	S
6	3	2	2	94	T	A	T	A
7	82	2	7	10	A	T	A	T
8	94	2	2	3	A	T	A	T

Seq 2	St f64	ATTACTTA
Seq 2	St b91	ATTAGTAA
Seq 2	St f702	CTTACTAA
Seq 2	St b729	CTAAGTAA
Seq 3	St f103	CTAAGTAA
Seq 3	St b216	CTTACTAA
Seq 3	St b235	ATTAGTAA
Seq 3	St f558	CTTACTAA
Seq 3	St f577	ATTAGTAA
Seq 3	St b690	ATTACTTA
Seq 4	St f347	ATTACTAA
Seq 4	St b349	CTTACTAA
Seq 4	St f444	ATTAGTAA
Seq 4	St b446	ATTAGTAA
Seq 5	St f362	GTTACTAA
Seq 5	St b431	CTTAGTAA
Seq 6	St f173	TTTAGTAA
Seq 6	St b253	ATTACTAA
Seq 6	St f383	CTTACTAA
Seq 6	St b410	TTTAGTAA
Seq 6	St f540	ATTAGTAA
Seq 6	St b620	TTTACTAA
Seq 7	St b714	CTTAGTGA
Seq 8	St b174	CTTAGTAA
Seq 8	St b210	TTTACTAA
Seq 8	St b249	ATTACTTA
Seq 8	St f583	CTTAGTAA
Seq 8	St f619	CTTACTAA
Seq 9	St f35	ATTAGTAA
Seq 9	St b52	ATTACTGA
Seq 9	St b376	CTGACTAA
Seq 9	St b758	TTTACTAA
Seq 10	St b141	CTGACTAA
Seq 10	St b357	CTTAGTAA
Seq 10	St f436	ATTACTAA
Seq 10	St f633	CTGACTAA
Seq 11	St b137	CTTAGTAA
Seq 11	St f656	ATTACTAA
Seq 12	St f34	ATTAGTAA
Seq 12	St b187	ATTAGTAA
Seq 12	St f606	CTTACTAA
Seq 12	St b759	ATTACTAA
Seq 13	St b162	ATTAGTAA
Seq 13	St f631	ATTACTAA
Seq 13	St b665	CTTAGTTA
Seq 14	St f287	CTGACTAA
Seq 15	St f251	ATTACTAA
Seq 15	St b542	CTTAGTAA
Seq 16	St b160	CTTAGTAA
Seq 16	St b193	CTTAGTGA

```
Seq 16  St f483  CTGACTAA
Seq 16  St f633  ATTACTAA
Total time 0:1:23.
```

Thanks for using MDscan! For questions, contact Xiaole Liu xliu@smi.stanford.edu

MD SCAN RESULTS

Downregulated

```
Date: Thu, 30 May 2002 19:30:43 -0700 (PDT)
From: Web Administrator <web@SMI.Stanford.EDU>
Apparently-To: stephent@Stanford.EDU

Pm 0.2638        Minimum match (6/8)

Top 3 motifs      Wid       Score1  Segment          Con       Deg
Mtf 1     8          4.319    122       TTTTTTTT        TTTTTTTT
Final Motif 1: Wid 8 Score1 4.319 Segment 122
         A     C     G     T      Con  rCon Deg  rDeg
1        1     1     1    97      T    A    T    A
2        1     1     1    97      T    A    T    A
3        1     1     1    97      T    A    T    A
4        1    13     1    85      T    A    T    A
5        1    11     1    87      T    A    T    A
6        1     1     1    97      T    A    T    A
7        1     1     1    97      T    A    T    A
8        1     1     1    97      T    A    T    A
Seq 1   St b121TTTTTTTT
Seq 1   St b122TTTTTTTT
Seq 1   St b123TTTTTTTT
Seq 1   St b124TTTTTTTT
Seq 1   St b125TTTTTTTT
Seq 1   St b126TTTTTTTT
Seq 1   St b127TTTTTTTT
Seq 1   St b128TTTTTTTT
Seq 1   St b129TTTTTTTT
Seq 1   St b257TTTTCTTT
Seq 2   St b116TTTTTTTT
Seq 3   St b160TTTTCTTT
Seq 3   St b161TTTCTTTT
Seq 3   St b165TTTTTTTT
Seq 3   St b166TTTTTTTT
Seq 3   St b167TTTTTTTT
Seq 3   St b168TTTTTTTT
Seq 3   St b169TTTTTTTT
Seq 3   St b170TTTTTTTT
Seq 3   St f593  TTTTTTTT
Seq 3   St f594  TTTTTTTT
Seq 3   St f595  TTTTTTTT
Seq 3   St f596  TTTTTTTT
Seq 3   St f597  TTTTTTTT
Seq 3   St f598  TTTTTTTT
Seq 3   St f599  TTTTTTTT
Seq 3   St f600  TTTTTTTT
Seq 3   St f601  TTTTTTTT
```

Seq 3	St f602	TTTTTTTT
Seq 3	St f663	TTTCTTTT
Seq 4	St f335	TTTTCTTT
Seq 4	St f336	TTTCTTTT
Seq 4	St f705	TTTTCTTT
Seq 4	St f706	TTTCTTTT
Seq 5	St b100	TTTTTTTT
Seq 5	St b101	TTTTTTTT
Seq 5	St b102	TTTTTTTT
Seq 5	St b106	TTTTCTTT
Seq 5	St b686	TTTTCTTT
Seq 7	St b102	TTTCTTTT
Seq 7	St b243	TTTCTTTT
Seq 7	St b247	TTTTTTTT
Seq 7	St b248	TTTTTTTT
Seq 7	St b249	TTTTTTTT
Seq 7	St b250	TTTTTTTT
Seq 7	St b251	TTTTTTTT
Seq 7	St b252	TTTTTTTT
Seq 7	St b253	TTTTTTTT
Seq 7	St b254	TTTTTTTT
Seq 7	St b255	TTTTTTTT
Seq 7	St b256	TTTTTTTT
Seq 7	St b257	TTTTTTTT
Seq 7	St b258	TTTTTTTT
Seq 7	St b259	TTTTTTTT
Seq 7	St b260	TTTTTTTT
Seq 7	St b261	TTTTTTTT
Seq 7	St b262	TTTTTTTT
Seq 7	St b263	TTTTTTTT
Seq 7	St b293	TTTCTTTT
Seq 8	St b169	TTTTTTTT
Seq 8	St b170	TTTTTTTT
Seq 8	St f235	TTTCTTTT
Seq 8	St f284	TTTTTTTT
Seq 8	St f285	TTTTTTTT
Seq 8	St f289	TTTTCTTT
Seq 8	St f290	TTTCTTTT
Seq 8	St b567	TTTTTTTT
Seq 9	St f655	TTTTTTTT
Seq 9	St f656	TTTTTTTT
Seq 9	St f657	TTTTTTTT
Seq 9	St f658	TTTTTTTT
Seq 11	St b204	TTTCTTTT
Seq 11	St f604	TTTTCTTT
Seq 11	St f605	TTTCTTTT
Seq 11	St f617	TTTTTTTT
Seq 11	St f618	TTTTTTTT
Seq 11	St f619	TTTTTTTT
Seq 11	St f620	TTTTTTTT
Seq 11	St f621	TTTTTTTT
Seq 11	St f622	TTTTTTTT
Seq 11	St f695	TTTTTTTT
Seq 11	St f761	TTTTCTTT
Seq 12	St b335	TTTTTTTT
Seq 12	St f359	TTTTTTTT

```
Seq 12  St f360  TTTTTTTT
Seq 12  St f361  TTTTTTTT
Seq 12  St f662  TTTTTTTT
Seq 12  St f663  TTTTTTTT
Seq 12  St f664  TTTTTTTT
Seq 13  St b26   TTTTCTTT
Seq 13  St b27   TTTCTTTT
Seq 13  St b291TTTCTTTT
Seq 16  St b94   TTTTTTTT
Seq 17  St b114TTTTTTTT
Seq 17  St b612TTTTCTTT
Seq 18  St b753TTTCTTTT
Seq 19  St b770TTTTTTTT
Seq 19  St b771TTTTTTTT
Seq 19  St b772TTTTTTTT
Seq 19  St b773TTTTTTTT
Seq 19  St b774TTTTTTTT
Seq 19  St b775TTTTTTTT
Seq 21  St b117TTTTTTTT
Seq 21  St b118TTTTTTTT
Seq 21  St b119TTTTTTTT
Seq 21  St b120TTTTTTTT
Seq 21  St b121TTTTTTTT
Seq 21  St b122TTTTTTTT
Seq 21  St b184TTTTCTTT
Seq 21  St b185TTTCTTTT
Seq 21  St b480TTTTCTTT
Seq 22  St b138TTTTTTTT
Seq 23  St b123TTTTTTTT
Seq 23  St b124TTTTTTTT
Seq 23  St b125TTTTTTTT
Seq 23  St b126TTTTTTTT
Seq 23  St b127TTTTTTTT
Seq 23  St b128TTTTTTTT
Seq 23  St f616  TTTTTTTT
Seq 23  St f617  TTTTTTTT
Seq 23  St f618  TTTTTTTT
Seq 23  St f619  TTTTTTTT
Mtf 2     8           4.291    138       TTTTTTTT         TTTTTTTT
Final Motif 2: Wid 8 Score1 4.291 Segment 138
         A     C     G     T     Con  rCon Deg  rDeg
1       14     9     1    76      T    A    T    A
2        1     1     1    98     T    A    T    A
3        1    10     1    88      T    A    T    A
4        1     1     1    98     T    A    T    A
5        1     1     1    98     T    A    T    A
6        1     1     1    98     T    A    T    A
7        1     1     1    98     T    A    T    A
8        1     1     1    98     T    A    T    A
Seq 1    St b118TTCTTTTT
Seq 1    St b120CTTTTTTT
Seq 1    St b121TTTTTTTT
Seq 1    St b122TTTTTTTT
Seq 1    St b123TTTTTTTT
Seq 1    St b124TTTTTTTT
Seq 1    St b125TTTTTTTT
```

Seq 1	St b126	TTTTTTTT
Seq 1	St b127	TTTTTTTT
Seq 1	St b128	TTTTTTTT
Seq 1	St b129	TTTTTTTT
Seq 2	St b115	ATTTTTTT
Seq 2	St b116	TTTTTTTT
Seq 2	St b640	TTCTTTTT
Seq 3	St b162	TTCTTTTT
Seq 3	St b164	CTTTTTTT
Seq 3	St b165	TTTTTTTT
Seq 3	St b166	TTTTTTTT
Seq 3	St b167	TTTTTTTT
Seq 3	St b168	TTTTTTTT
Seq 3	St b169	TTTTTTTT
Seq 3	St b170	TTTTTTTT
Seq 3	St f592	CTTTTTTT
Seq 3	St f593	TTTTTTTT
Seq 3	St f594	TTTTTTTT
Seq 3	St f595	TTTTTTTT
Seq 3	St f596	TTTTTTTT
Seq 3	St f597	TTTTTTTT
Seq 3	St f598	TTTTTTTT
Seq 3	St f599	TTTTTTTT
Seq 3	St f600	TTTTTTTT
Seq 3	St f601	TTTTTTTT
Seq 3	St f602	TTTTTTTT
Seq 4	St f337	TTCTTTTT
Seq 4	St f707	TTCTTTTT
Seq 5	St b99	CTTTTTTT
Seq 5	St b100	TTTTTTTT
Seq 5	St b101	TTTTTTTT
Seq 5	St b102	TTTTTTTT
Seq 5	St f143	ATTTTTTT
Seq 5	St b323	TTCTTTTT
Seq 7	St b244	TTCTTTTT
Seq 7	St b246	CTTTTTTT
Seq 7	St b247	TTTTTTTT
Seq 7	St b248	TTTTTTTT
Seq 7	St b249	TTTTTTTT
Seq 7	St b250	TTTTTTTT
Seq 7	St b251	TTTTTTTT
Seq 7	St b252	TTTTTTTT
Seq 7	St b253	TTTTTTTT
Seq 7	St b254	TTTTTTTT
Seq 7	St b255	TTTTTTTT
Seq 7	St b256	TTTTTTTT
Seq 7	St b257	TTTTTTTT
Seq 7	St b258	TTTTTTTT
Seq 7	St b259	TTTTTTTT
Seq 7	St b260	TTTTTTTT
Seq 7	St b261	TTTTTTTT
Seq 7	St b262	TTTTTTTT
Seq 7	St b263	TTTTTTTT
Seq 7	St b294	TTCTTTTT
Seq 8	St b168	CTTTTTTT
Seq 8	St b169	TTTTTTTT

Seq 8	St b170	TTTTTTTT
Seq 8	St f283	CTTTTTTT
Seq 8	St f284	TTTTTTTT
Seq 8	St f285	TTTTTTTT
Seq 8	St f291	TTCTTTTT
Seq 8	St b566	ATTTTTTT
Seq 8	St b567	TTTTTTTT
Seq 8	St f648	ATTTTTTT
Seq 9	St b107	ATTTTTTT
Seq 9	St f655	TTTTTTTT
Seq 9	St f656	TTTTTTTT
Seq 9	St f657	TTTTTTTT
Seq 9	St f658	TTTTTTTT
Seq 11	St b132	ATTTTTTT
Seq 11	St b205	TTCTTTTT
Seq 11	St f606	TTCTTTTT
Seq 11	St f616	ATTTTTTT
Seq 11	St f617	TTTTTTTT
Seq 11	St f618	TTTTTTTT
Seq 11	St f619	TTTTTTTT
Seq 11	St f620	TTTTTTTT
Seq 11	St f621	TTTTTTTT
Seq 11	St f622	TTTTTTTT
Seq 11	St f634	ATTTTTTT
Seq 11	St f694	ATTTTTTT
Seq 11	St f695	TTTTTTTT
Seq 11	St b716	ATTTTTTT
Seq 12	St b334	CTTTTTTT
Seq 12	St b335	TTTTTTTT
Seq 12	St f358	CTTTTTTT
Seq 12	St f359	TTTTTTTT
Seq 12	St f360	TTTTTTTT
Seq 12	St f361	TTTTTTTT
Seq 12	St f374	ATTTTTTT
Seq 12	St f661	CTTTTTTT
Seq 12	St f662	TTTTTTTT
Seq 12	St f663	TTTTTTTT
Seq 12	St f664	TTTTTTTT
Seq 13	St b28	TTCTTTTT
Seq 15	St f681	ATTTTTTT
Seq 16	St b93	CTTTTTTT
Seq 16	St b94	TTTTTTTT
Seq 17	St b113	ATTTTTTT
Seq 17	St b114	TTTTTTTT
Seq 19	St b769	CTTTTTTT
Seq 19	St b770	TTTTTTTT
Seq 19	St b771	TTTTTTTT
Seq 19	St b772	TTTTTTTT
Seq 19	St b773	TTTTTTTT
Seq 19	St b774	TTTTTTTT
Seq 19	St b775	TTTTTTTT
Seq 20	St f663	ATTTTTTT
Seq 21	St b116	ATTTTTTT
Seq 21	St b117	TTTTTTTT
Seq 21	St b118	TTTTTTTT
Seq 21	St b119	TTTTTTTT

```
Seq 21  St b120TTTTTTTT
Seq 21  St b121TTTTTTTT
Seq 21  St b122TTTTTTTT
Seq 21  St b186TTCTTTTT
Seq 22  St b125ATTTTTTT
Seq 22  St b137ATTTTTTT
Seq 22  St b138TTTTTTTT
Seq 23  St b122ATTTTTTT
Seq 23  St b123TTTTTTTT
Seq 23  St b124TTTTTTTT
Seq 23  St b125TTTTTTTT
Seq 23  St b126TTTTTTTT
Seq 23  St b127TTTTTTTT
Seq 23  St b128TTTTTTTT
Seq 23  St f615  ATTTTTTT
Seq 23  St f616  TTTTTTTT
Seq 23  St f617  TTTTTTTT
Seq 23  St f618  TTTTTTTT
Seq 23  St f619  TTTTTTTT
Mtf 3     8           4.289     116       AAAAAAAA       AAAAAAAA
Final Motif 3: Wid 8 Score1 4.289 Segment 116
         A     C     G     T     Con  rCon Deg  rDeg
1       97     1     1     1      A    T    A    T
2       91     7     1     1      A    T    A    T
3       97     1     1     1      A    T    A    T
4       97     1     1     1      A    T    A    T
5       85     1    13     1      A    T    A    T
6       97     1     1     1      A    T    A    T
7       97     1     1     1      A    T    A    T
8       97     1     1     1      A    T    A    T
Seq 1    St f665  AAAAAAAA
Seq 1    St f666  AAAAAAAA
Seq 1    St f667  AAAAAAAA
Seq 1    St f668  AAAAAAAA
Seq 1    St f669  AAAAAAAA
Seq 1    St f670  AAAAAAAA
Seq 1    St f671  AAAAAAAA
Seq 1    St f672  AAAAAAAA
Seq 1    St f673  AAAAAAAA
Seq 2    St f678  AAAAAAAA
Seq 3    St b131AAAAGAAA
Seq 3    St b192AAAAAAAA
Seq 3    St b193AAAAAAAA
Seq 3    St b194AAAAAAAA
Seq 3    St b195AAAAAAAA
Seq 3    St b196AAAAAAAA
Seq 3    St b197AAAAAAAA
Seq 3    St b198AAAAAAAA
Seq 3    St b199AAAAAAAA
Seq 3    St b200AAAAAAAA
Seq 3    St b201AAAAAAAA
Seq 3    St f624  AAAAAAAA
Seq 3    St f625  AAAAAAAA
Seq 3    St f626  AAAAAAAA
Seq 3    St f627  AAAAAAAA
Seq 3    St f628  AAAAAAAA
```

Seq 3	St f629	AAAAAAAA
Seq 3	St f633	AAAAGAAA
Seq 4	St b84	ACAAAAAA
Seq 4	St b88	AAAAGAAA
Seq 4	St b124	ACAAAAAA
Seq 4	St b458	AAAAGAAA
Seq 5	St f692	AAAAAAAA
Seq 5	St f693	AAAAAAAA
Seq 5	St f694	AAAAAAAA
Seq 7	St f501	AAAAGAAA
Seq 7	St f531	AAAAAAAA
Seq 7	St f532	AAAAAAAA
Seq 7	St f533	AAAAAAAA
Seq 7	St f534	AAAAAAAA
Seq 7	St f535	AAAAAAAA
Seq 7	St f536	AAAAAAAA
Seq 7	St f537	AAAAAAAA
Seq 7	St f538	AAAAAAAA
Seq 7	St f539	AAAAAAAA
Seq 7	St f540	AAAAAAAA
Seq 7	St f541	AAAAAAAA
Seq 7	St f542	AAAAAAAA
Seq 7	St f543	AAAAAAAA
Seq 7	St f544	AAAAAAAA
Seq 7	St f545	AAAAAAAA
Seq 7	St f546	AAAAAAAA
Seq 7	St f547	AAAAAAAA
Seq 7	St f551	AAAAGAAA
Seq 7	St f692	AAAAGAAA
Seq 8	St f225	ACAAAAAA
Seq 8	St f227	AAAAAAAA
Seq 8	St b504	AAAAGAAA
Seq 8	St b509	AAAAAAAA
Seq 8	St b510	AAAAAAAA
Seq 8	St b559	AAAAGAAA
Seq 8	St f624	AAAAAAAA
Seq 8	St f625	AAAAAAAA
Seq 9	St b136	AAAAAAAA
Seq 9	St b137	AAAAAAAA
Seq 9	St b138	AAAAAAAA
Seq 9	St b139	AAAAAAAA
Seq 9	St b596	ACAAAAAA
Seq 11	St b97	ACAAAAAA
Seq 11	St b99	AAAAAAAA
Seq 11	St b172	AAAAAAAA
Seq 11	St b173	AAAAAAAA
Seq 11	St b174	AAAAAAAA
Seq 11	St b175	AAAAAAAA
Seq 11	St b176	AAAAAAAA
Seq 11	St b177	AAAAAAAA
Seq 11	St b189	AAAAGAAA
Seq 11	St f590	AAAAGAAA
Seq 12	St b130	AAAAAAAA
Seq 12	St b131	AAAAAAAA
Seq 12	St b132	AAAAAAAA
Seq 12	St b433	AAAAAAAA

```
Seq 12  St b434AAAAAAAA
Seq 12  St b435AAAAAAAA
Seq 12  St f459  AAAAAAAA
Seq 13  St f503  AAAAGAAA
Seq 13  St f767  AAAAGAAA
Seq 16  St f700  AAAAAAAA
Seq 17  St f680  AAAAAAAA
Seq 18  St f41    AAAAGAAA
Seq 18  St f549  ACAAAAAA
Seq 19  St f19    AAAAAAAA
Seq 19  St f20    AAAAAAAA
Seq 19  St f21    AAAAAAAA
Seq 19  St f22    AAAAAAAA
Seq 19  St f23    AAAAAAAA
Seq 19  St f24    AAAAAAAA
Seq 21  St f609  AAAAGAAA
Seq 21  St f672  AAAAAAAA
Seq 21  St f673  AAAAAAAA
Seq 21  St f674  AAAAAAAA
Seq 21  St f675  AAAAAAAA
Seq 21  St f676  AAAAAAAA
Seq 21  St f677  AAAAAAAA
Seq 22  St f353  ACAAAAAA
Seq 22  St f656  AAAAAAAA
Seq 23  St b175AAAAAAAA
Seq 23  St b176AAAAAAAA
Seq 23  St b177AAAAAAAA
Seq 23  St b178AAAAAAAA
Seq 23  St f666  AAAAAAAA
Seq 23  St f667  AAAAAAAA
Seq 23  St f668  AAAAAAAA
Seq 23  St f669  AAAAAAAA
Seq 23  St f670  AAAAAAAA
Seq 23  St f671  AAAAAAAA
Total time 0:1:37.
```

Thanks for using MDscan! For questions, contact Xiaole Liu xliu@smi.stanford.edu

Upregulated

```
Date: Thu, 30 May 2002 19:31:03 -0700 (PDT)
From: Web Administrator <web@SMI.Stanford.EDU>
Apparently-To: stephent@Stanford.EDU

Pm 0.2638        Minimum match (6/8)

Top 3 motifs      Wid       Score1  Segment           Con       Deg
Mtf 1     8          3.706     58         TTCTAGAA      TTCTAGAA
Final Motif 1: Wid 8 Score1 3.706 Segment 58
         A     C     G     T     Con  rCon Deg  rDeg
1        2     2     2    95     T    A    T    A
2        2     2     2    95     T    A    T    A
3       10    86     2     2     C    G    C    G
4       10    11     8    71     T    A    T    A
5       71     8    11    10     A    T    A    T
6        2     2    86    10     G    C    G    C
```

7	93	2	3	2	A	T	A	T
8	93	2	2	4	A	T	A	T
Seq 1	St b226	TTCTAGAA						
Seq 1	St b240	TTCGAGAA						
Seq 1	St f554	TTCTCGAA						
Seq 1	St f568	TTCTAGAA						
Seq 2	St b41	TTCAAGAA						
Seq 2	St b178	TTCGAGAA						
Seq 2	St b210	TTCTAGAA						
Seq 2	St f584	TTCTAGAA						
Seq 2	St f616	TTCTCGAA						
Seq 2	St f753	TTCTTGAA						
Seq 3	St b61	TTCTCGAA						
Seq 3	St f148	TTCTAGGA						
Seq 3	St f301	TTATAGAA						
Seq 3	St b493	TTCTATAA						
Seq 3	St f733	TTCGAGAA						
Seq 4	St b134	TTCTCGAA						
Seq 4	St b254	TTCTGGAA						
Seq 4	St b347	TTCTAGAA						
Seq 4	St b357	TTCTGGAA						
Seq 4	St f437	TTCCAGAA						
Seq 4	St f447	TTCTAGAA						
Seq 4	St f540	TTCCAGAA						
Seq 4	St f660	TTCGAGAA						
Seq 5	St b134	TTCTAGAT						
Seq 6	St f221	TTCTTGAA						
Seq 6	St b322	TTCCAGAA						
Seq 6	St b332	TTCTGGAA						
Seq 6	St f379	TTCTATAA						
Seq 6	St b415	TTATAGAA						
Seq 6	St f462	TTCCAGAA						
Seq 6	St f472	TTCTGGAA						
Seq 6	St b573	TTCAAGAA						
Seq 7	St b347	TTCTAGAA						
Seq 7	St f447	TTCTAGAA						
Seq 8	St f277	TTCAAGAA						
Seq 8	St b285	TTCTAGAA						
Seq 8	St b295	TTCTGGAA						
Seq 8	St f499	TTCCAGAA						
Seq 8	St f509	TTCTAGAA						
Seq 8	St b517	TTCTTGAA						
Seq 9	St b177	TTCTATAA						
Seq 9	St f617	TTATAGAA						
Seq 10	St b204	TTCTGGAA						
Seq 10	St f590	TTCCAGAA						
Seq 11	St b131	TTCTAGAA						
Seq 11	St b176	TTCTATAA						
Seq 11	St f618	TTATAGAA						
Seq 11	St f663	TTCTAGAA						
Seq 12	St b60	TTCTTGAA						
Seq 12	St b228	TTCTAGAA						
Seq 12	St f566	TTCTAGAA						
Seq 12	St f734	TTCAAGAA						
Seq 13	St f258	TTCTAGAA						
Seq 13	St b536	TTCTAGAA						

```
Seq 15  St f367  TTCTTGAA
Seq 15  St b427TTCAAGAA
Seq 17  St f67    TTATAGAA
Seq 17  St b727TTCTATAA
Mtf 2      8            3.703     63          TTCTAGAA        TTCTAGAA
Final Motif 2: Wid 8 Score1 3.703 Segment 63
         A      C      G      T       Con  rCon Deg  rDeg
1        2      1      1     95       T     A     T     A
2        2      1      1     95       T     A     T     A
3       11     85      1      2       C     G     C     G
4       13     10      7     70       T     A     T     A
5       66     15     10     10       A     T     A     T
6        2      1     87     10       G     C     G     C
7       91      1      6      2       A     T     A     T
8       93      1      1      4       A     T     A     T
Seq 1    St b226TTCTAGAA
Seq 1    St b240TTCGAGAA
Seq 1    St b264TTCTCGGA
Seq 1    St f554  TTCTCGAA
Seq 1    St f568  TTCTAGAA
Seq 2    St b41    TTCAAGAA
Seq 2    St b178TTCGAGAA
Seq 2    St b210TTCTAGAA
Seq 2    St f584  TTCTAGAA
Seq 2    St f616  TTCTCGAA
Seq 2    St f753  TTCTTGAA
Seq 3    St b61    TTCTCGAA
Seq 3    St f148  TTCTAGGA
Seq 3    St f301  TTATAGAA
Seq 3    St b493TTCTATAA
Seq 3    St f733  TTCGAGAA
Seq 4    St b134TTCTCGAA
Seq 4    St b254TTCTGGAA
Seq 4    St b347TTCTAGAA
Seq 4    St b357TTCTGGAA
Seq 4    St f437  TTCCAGAA
Seq 4    St f447  TTCTAGAA
Seq 4    St f540  TTCCAGAA
Seq 4    St f660  TTCGAGAA
Seq 5    St b134TTCTAGAT
Seq 6    St f221  TTCTTGAA
Seq 6    St b322TTCCAGAA
Seq 6    St b332TTCTGGAA
Seq 6    St f379  TTCTATAA
Seq 6    St b415TTATAGAA
Seq 6    St f462  TTCCAGAA
Seq 6    St f472  TTCTGGAA
Seq 6    St b573TTCAAGAA
Seq 7    St b347TTCTAGAA
Seq 7    St f447  TTCTAGAA
Seq 8    St f277  TTCAAGAA
Seq 8    St b285TTCTAGAA
Seq 8    St b295TTCTGGAA
Seq 8    St f499  TTCCAGAA
Seq 8    St f509  TTCTAGAA
Seq 8    St b517TTCTTGAA
```

```
Seq 9    St b177TTCTATAA
Seq 9    St f617  TTATAGAA
Seq 10  St b204TTCTGGAA
Seq 10  St b235TTCACGAA
Seq 10  St f590  TTCCAGAA
Seq 11  St b131TTCTAGAA
Seq 11  St b176TTCTATAA
Seq 11  St f618  TTATAGAA
Seq 11  St f663  TTCTAGAA
Seq 12  St b60   TTCTTGAA
Seq 12  St b228TTCTAGAA
Seq 12  St f566  TTCTAGAA
Seq 12  St f734  TTCAAGAA
Seq 13  St f258  TTCTAGAA
Seq 13  St f283  TTATCGAA
Seq 13  St b536TTCTAGAA
Seq 15  St f367  TTCTTGAA
Seq 15  St b427TTCAAGAA
Seq 17  St f67    TTATAGAA
Seq 17  St b727TTCTATAA
Seq 18  St f332  TTCTCGGA
Seq 18  St b351TTCACGAA
Mtf 3     8            3.698    62          TTCTAGAA        TTCTAGAA
Final Motif 3: Wid 8 Score1 3.698 Segment 62
         A     C     G     T     Con  rCon Deg  rDeg
1        2     1     1    95      T    A    T    A
2        2     1     1    95      T    A    T    A
3       10    82     1     7      C    G    C    G
4       13    10     7    69      T    A    T    A
5       69    10    10    10      A    T    A    T
6        2     1    87    10      G    C    G    C
7       95     1     1     2      A    T    A    T
8       93     1     1     4      A    T    A    T
Seq 1    St b226TTCTAGAA
Seq 1    St b240TTCGAGAA
Seq 1    St f554  TTCTCGAA
Seq 1    St f568  TTCTAGAA
Seq 2    St b41   TTCAAGAA
Seq 2    St b178TTCGAGAA
Seq 2    St b210TTCTAGAA
Seq 2    St f584  TTCTAGAA
Seq 2    St f616  TTCTCGAA
Seq 2    St f753  TTCTTGAA
Seq 3    St b61   TTCTCGAA
Seq 3    St f301  TTATAGAA
Seq 3    St b493TTCTATAA
Seq 3    St f733  TTCGAGAA
Seq 4    St b134TTCTCGAA
Seq 4    St b254TTCTGGAA
Seq 4    St b347TTCTAGAA
Seq 4    St b357TTCTGGAA
Seq 4    St f437  TTCCAGAA
Seq 4    St f447  TTCTAGAA
Seq 4    St f540  TTCCAGAA
Seq 4    St f660  TTCGAGAA
Seq 5    St b134TTCTAGAT
```

```
Seq 5   St b190TTTTAGAA
Seq 6   St f221  TTCTTGAA
Seq 6   St b322TTCCAGAA
Seq 6   St b332TTCTGGAA
Seq 6   St f379  TTCTATAA
Seq 6   St b415TTATAGAA
Seq 6   St f462  TTCCAGAA
Seq 6   St f472  TTCTGGAA
Seq 6   St b573TTCAAGAA
Seq 7   St b347TTCTAGAA
Seq 7   St f447  TTCTAGAA
Seq 8   St f277  TTCAAGAA
Seq 8   St b285TTCTAGAA
Seq 8   St b295TTCTGGAA
Seq 8   St f350  TTTTAGAA
Seq 8   St f499  TTCCAGAA
Seq 8   St f509  TTCTAGAA
Seq 8   St b517TTCTTGAA
Seq 9   St b177TTCTATAA
Seq 9   St f617  TTATAGAA
Seq 10  St b204TTCTGGAA
Seq 10  St b235TTCACGAA
Seq 10  St f590  TTCCAGAA
Seq 11  St b131TTCTAGAA
Seq 11  St b176TTCTATAA
Seq 11  St f618  TTATAGAA
Seq 11  St f663  TTCTAGAA
Seq 12  St b60   TTCTTGAA
Seq 12  St b228TTCTAGAA
Seq 12  St f566  TTCTAGAA
Seq 12  St f734  TTCAAGAA
Seq 13  St f258  TTCTAGAA
Seq 13  St b536TTCTAGAA
Seq 15  St f367  TTCTTGAA
Seq 15  St b427TTCAAGAA
Seq 16  St b758TTTTAGAA
Seq 17  St f67    TTATAGAA
Seq 17  St b727TTCTATAA
Seq 18  St b351TTCACGAA
Total time 0:1:39.
```

Thanks for using MDscan! For questions, contact Xiaole Liu xliu@smi.stanford.edu

Yap1p

Date: Thu, 30 May 2002 11:30:58 -0700 (PDT)
From: Web Administrator <web@SMI.Stanford.EDU>
Apparently-To: stephent@Stanford.EDU

```
****************************************
*                                      *
*      BioProspector Search Result     *
*                                      *
****************************************

The highest scoring 3 motifs are:
```

```
Motif #1:
******************************
Width (8, 0); Gap [0, 0]; MotifScore 4.301; Segments 36

Blk1   A     C     G     T     Con  rCon Deg  rDeg
1    0.00  0.00  0.00  1.00     T    A    T    A
2    0.00  0.00  0.00  1.00     T    A    T    A
3    1.00  0.00  0.00  0.00     A    T    A    T
4    0.00  0.00  1.00  0.00     G    C    G    C
5    0.00  0.00  0.00  1.00     T    A    T    A
6    0.78  0.22  0.00  0.00     A    T    A    T
7    1.00  0.00  0.00  0.00     A    T    A    T
8    0.00  0.00  0.47  0.53     T    A    K    M

>Gtt2   seg 1   r92     TTAGTAAG
>Aad6   seg 1   r236    TTAGTAAG
>Aad6   seg 2   f578    TTAGTAAG
>Aad4   seg 1   f445    TTAGTAAG
>Aad4   seg 2   r447    TTAGTAAT
>Gpx2   seg 1   r411    TTAGTAAG
>Gpx2   seg 2   f541    TTAGTAAT
>Trx2   seg 1   r175    TTAGTAAG
>Lap4   seg 1   f418    TTAGTCAG
>Flr1   seg 1   r358    TTAGTAAT
>Flr1   seg 2   f653    TTAGTCAG
>Flr1   seg 3   r161    TTAGTCAG
>Ttr1   seg 1   r138    TTAGTAAT
>Ydl124w        seg 1   r188    TTAGTAAG
>Ydl124w        seg 2   f35     TTAGTAAT
>Ylr108c        seg 1   r163    TTAGTAAT
>YLL055w        seg 1   r507    TTAGTCAG
>SDL1  seg 1    r543    TTAGTAAT
>YLR460c        seg 1   r161    TTAGTAAT
>YLR460c        seg 2   r311    TTAGTCAG
>YCR102c        seg 1   r185    TTAGTAAT
>YCR102c        seg 2   r335    TTAGTCAG
>YNL260c        seg 1   r17     TTAGTAAG
>YKL071w        seg 1   f579    TTAGTAAT
>YKL071w        seg 2   f603    TTAGTAAT
>YKL071w        seg 3   f615    TTAGTAAT
>YKL071w        seg 4   f627    TTAGTAAT
>YKL071w        seg 5   f639    TTAGTAAT
>YDR132c        seg 1   r184    TTAGTAAG
>Aad15seg 1     r640    TTAGTAAT
>YKR071c        seg 1   r256    TTAGTAAT
>YKR071c        seg 2   f701    TTAGTAAT
>YML131w        seg 1   f294    TTAGTCAT
>YNL134c        seg 1   f613    TTAGTAAG
>YNL134c        seg 2   r742    TTAGTAAT
>YNL134c        seg 3   r283    TTAGTCAG
******************************

Motif #2:
******************************
Width (8, 0); Gap [0, 0]; MotifScore 4.301; Segments 36
```

```
Blk1   A     C     G     T      Con  rCon Deg  rDeg
1    0.00  0.00  0.00  1.00     T    A    T    A
2    0.00  0.00  0.00  1.00     T    A    T    A
3    1.00  0.00  0.00  0.00     A    T    A    T
4    0.00  0.00  1.00  0.00     G    C    G    C
5    0.00  0.00  0.00  1.00     T    A    T    A
6    0.78  0.22  0.00  0.00     A    T    A    T
7    1.00  0.00  0.00  0.00     A    T    A    T
8    0.00  0.00  0.47  0.53     T    A    K    M

>Gtt2   seg 1   r92     TTAGTAAG
>Aad6   seg 1   r236    TTAGTAAG
>Aad6   seg 2   f578    TTAGTAAG
>Aad4   seg 1   f445    TTAGTAAG
>Aad4   seg 2   r447    TTAGTAAT
>Gpx2   seg 1   r411    TTAGTAAG
>Gpx2   seg 2   f541    TTAGTAAT
>Trx2   seg 1   r175    TTAGTAAG
>Lap4   seg 1   f418    TTAGTCAG
>Flr1   seg 1   r358    TTAGTAAT
>Flr1   seg 2   f653    TTAGTCAG
>Flr1   seg 3   r161    TTAGTCAG
>Ttr1   seg 1   r138    TTAGTAAT
>Ydl124w        seg 1   r188    TTAGTAAG
>Ydl124w        seg 2   f35     TTAGTAAT
>Ylr108c        seg 1   r163    TTAGTAAT
>YLL055w        seg 1   r507    TTAGTCAG
>SDL1   seg 1   r543    TTAGTAAT
>YLR460c        seg 1   r311    TTAGTCAG
>YLR460c        seg 2   r161    TTAGTAAT
>YCR102c        seg 1   r335    TTAGTCAG
>YCR102c        seg 2   r185    TTAGTAAT
>YNL260c        seg 1   r17     TTAGTAAG
>YKL071w        seg 1   f579    TTAGTAAT
>YKL071w        seg 2   f603    TTAGTAAT
>YKL071w        seg 3   f615    TTAGTAAT
>YKL071w        seg 4   f627    TTAGTAAT
>YKL071w        seg 5   f639    TTAGTAAT
>YDR132c        seg 1   r184    TTAGTAAG
>Aad15seg 1     r640    TTAGTAAT
>YKR071c        seg 1   r256    TTAGTAAT
>YKR071c        seg 2   f701    TTAGTAAT
>YML131w        seg 1   f294    TTAGTCAT
>YNL134c        seg 1   f613    TTAGTAAG
>YNL134c        seg 2   r742    TTAGTAAT
>YNL134c        seg 3   r283    TTAGTCAG
******************************

Motif #3:
******************************
Width (8, 0); Gap [0, 0]; MotifScore 4.242; Segments 36

Blk1   A     C     G     T      Con  rCon Deg  rDeg
1    0.53  0.47  0.00  0.00     A    T    M    K
2    0.00  0.00  0.00  1.00     T    A    T    A
```

```
3    0.00  0.00  0.22  0.78     T    A    T    A
4    1.00  0.00  0.00  0.00     A    T    A    T
5    0.00  1.00  0.00  0.00     C    G    C    G
6    0.00  0.00  0.00  1.00     T    A    T    A
7    1.00  0.00  0.00  0.00     A    T    A    T
8    1.00  0.00  0.00  0.00     A    T    A    T

>Gtt2    seg 1    f702     CTTACTAA
>Aad6    seg 1    f558     CTTACTAA
>Aad6    seg 2    r216     CTTACTAA
>Aad4    seg 1    r349     CTTACTAA
>Aad4    seg 2    f347     ATTACTAA
>Gpx2    seg 1    f383     CTTACTAA
>Gpx2    seg 2    r253     ATTACTAA
>Trx2    seg 1    f619     CTTACTAA
>Lap4    seg 1    r376     CTGACTAA
>Flr1    seg 1    f436     ATTACTAA
>Flr1    seg 2    r141     CTGACTAA
>Flr1    seg 3    f633     CTGACTAA
>Ttr1    seg 1    f656     ATTACTAA
>Ydl124w          seg 1    f606     CTTACTAA
>Ydl124w          seg 2    r759     ATTACTAA
>Ylr108c          seg 1    f631     ATTACTAA
>YLL055w          seg 1    f287     CTGACTAA
>SDL1  seg 1    f251     ATTACTAA
>YLR460c          seg 1    f633     ATTACTAA
>YLR460c          seg 2    f483     CTGACTAA
>YCR102c          seg 1    f459     CTGACTAA
>YCR102c          seg 2    f609     ATTACTAA
>YNL260c          seg 1    f777     CTTACTAA
>YKL071w          seg 1    r167     ATTACTAA
>YKL071w          seg 2    r155     ATTACTAA
>YKL071w          seg 3    r179     ATTACTAA
>YKL071w          seg 4    r191     ATTACTAA
>YKL071w          seg 5    r215     ATTACTAA
>YDR132c          seg 1    f610     CTTACTAA
>Aad15seg 1    f154     ATTACTAA
>YKR071c          seg 1    f538     ATTACTAA
>YKR071c          seg 2    r93      ATTACTAA
>YML131w          seg 1    r500     ATGACTAA
>YNL134c          seg 1    r181     CTTACTAA
>YNL134c          seg 2    f52      ATTACTAA
>YNL134c          seg 3    f511     CTGACTAA
*****************************

Thanks for using BioProspector.
For questions, please contact Xiaole Liu at xliu@smi.stanford.edu.
```