Supporting Information

Tracking the Emergence of High Affinity Aptamers for rhVEGF$_{165}$ during CE-SELEX Using High Throughput Sequencing

Meng Jing and Michael T. Bowser*

Department of Chemistry, University of Minnesota, 207 Pleasant Street SE, Minneapolis, Minnesota 55455, United States

Email: bowser@umn.edu

[1] Motifs of the Short Byproducts

[2] Gel Image of the PCR Products after Gel Extraction Purification

[3] Top 100 Abundant Oligonucleotide Fragments from 5 nt through 15 nt in Selected Pools

[4] Motifs Found in Contigs

[1] Motifs of the Short Byproducts.

Motifs of the short byproducts were discovered using MEME4.7.0 online version. The specific sequence lengths at which the short byproducts were the most abundant were first identified. All the sequences at the peak lengths were extracted to form new FASTA files. The primer regions of these sequences were removed before the sequences were uploaded to discover motifs, register position of motifs, and the abundance of the motifs. Since MEME does not allow analysis on sequences shorter than 48 nt, the first 3 nt in reverse primer region were kept for sequences at lengths below 48. The motifs were summarized in Figure S1.

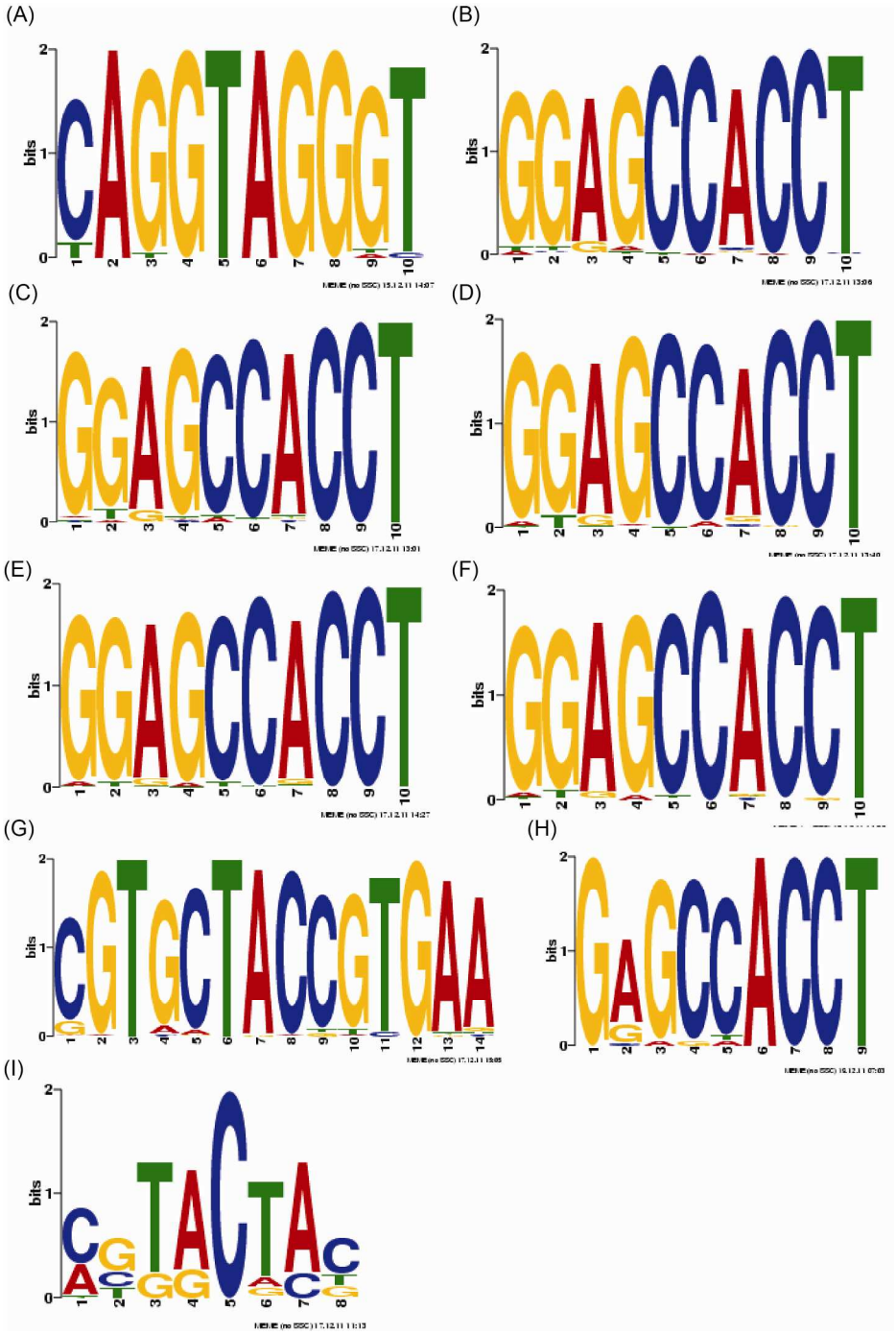

Figure S1. Motifs of the short byproducts discovered at the peak lengths. (A), motif found in pool 1, (B) motif found in pool 2, (C) motif found in pool 3, (D) motif found in pool 4, (E) motif found in pool 5, (F) motif found in pool 6, (G) motif found in pool 7 at the length of 54 nt, (H) motif found in pool 7 at the length of 46 nt, (I) motif found in control.

[2] Gel Image of the PCR Products after Gel Extraction Purification

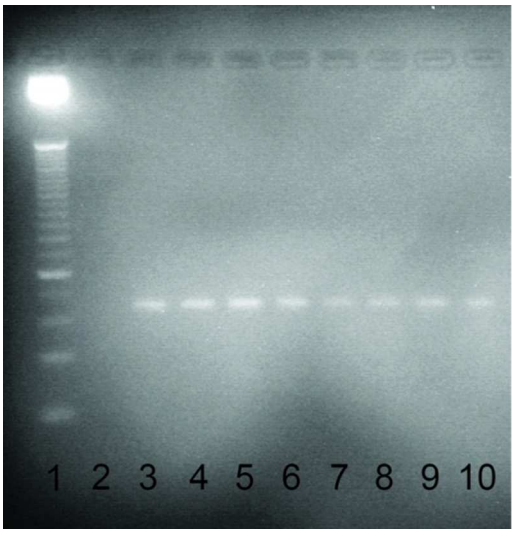

Figure S2. Gel image of 8 samples (lane 3 through lane 10) after gel extraction to remove undesired short products. Lane 1 was the 25 bp DNA ladder which indicated the size of DNA products. Lane 3 through lane 10 corresponded to the products of initial library and round 1 through round 7 after gel electrophoresis purification. This gel image showed that there were no short products detectable below the desired product bands.

[3] Top 100 Abundant Oligonucleotides Fragments from 5 nt through 15 nt in the Selected Pools

Table S1. Top 100 Abundant Oligos

5 nt Oligo	6 nt Oligo	7 nt Oligo	8 nt Oligo	9 nt Oligo	10 nt Oligo	11 nt Oligo	12 nt Oligo	13 nt Oligo	14 nt ligo	15 nt Oligo
TTTTG	GGAGCC	GGAGCCA	ATCTGGAG	GAAGCCCGC	GCCCGCTGTG	GAAGCCTGC	GAAGCCCGCTGT	GAAGCCCGCTGTG	GAAGCCCGCTGTGA	GCCCGCTGTGACATC
TTTGT	TTTGTT	TGGAGCC	TCTGGAGC	AGCCCGCTG	AGCCCGCTGT	AGCCCGCTGTG	AAGCCCGCTGTG	CTGTGACATCTGG	AGCCCGCTGTGACA	GAAGCCCGCTGTGAC
GTTTT	TTTTGT	TCTGGAG	TGTGACAT	GGAGCCGGT	AAGCCCGCTG	AAGCCCGCTGT	AGCCCGCTGTG	AGCCCGCTGTGAC	GCCCGCTGTGACAT	CGCTGTGACATCTGG
GGTTT	TTTTTG	ATCTGGA	GGTTCCG	TGGAGCCGG	CTGGAGCCG	CTGGAGCCGGT	GCCCGCTGTGAC	GCCCGCTGTGACA	GCTGTGACATCTGG	AGCCCGCTGTGACAT
TTGTT	GTTTTT	GTGACAT	ACATCTGG	GCCCGCTGT	GAAGCCCGT	GCCCGCTGTGA	TCTGGAGCCGGT	AAGCCCGCTGTGAC	AAGCCCGCCATCTG	CCGCTGTGACATCTG
TGTTT	TTTTGG	GCCGGTT	GGAGCCGG	CCCGCTGTG	TGGAGCCGGT	TCTGGACTGCCGG	TGTGACATCTGG	CCGCTGTGACATC	CTGGAGCCGGTTCC	TCTGGAGCCGGTTCC
TTTGG	TTGTTT	CATCTGG	CATCTGGA	ATCTGGAGC	TCTGGAGCCG	GTGACATCTGG	GACATCTGGAGC	CATCTGGAGCCGG	CTGTGACATCTGGA	CTGGAGCCGGTTCCC
ATTTT	TTTGGT	TGTGACA	CGCTGTGA	GGTTCCGGG	CATCTGGAGC	CCCGCTGTGAC	CTGTGACATCTGG	ATCTGGAGCCGGT	CCCGCTGTGACATC	AAGCCCGCTGTGACA
TTTTA	GGTTTT	GCTGTGA	CTGGAGCC	TCTGGAGCC	CCCGCTGTGA	GTTCCCGAGC	GGTTCCCGGAGC	GCTGTGACATCTG	CGCTGTGACATCTG	CTGTGACATCTGGAG
TTGGT	TTATTT	CGCTGTG	GAGCCGGT	CTGGAGCCG	GGAAGCCCGC	CCGCTGTGACA	ATCTGGAGCCGG	GTGACATCTGGAG	GTGACATCTGGAGC	GGAGCCGGTTCCCGG
TATTT	TGTTTT	GGAGCCG	TGGAGCCG	GACATCTGG	CTGTGACATC	CCGGTTCCCG	CCCGCTGTGAC	CCCGCTGTGACAT	CCGCTGTGACATCT	TGTGACATCTGGAGC
TTATT	ATTTTT	GGTTCCC	CTGTGACA	CCGCTGTGA	CCGCTGTGAC	GACATCTGGAG	CTGGAGCCGGT	TGGAGCCGGTTCC	CCGGTTCCCGGAGC	CCCGCTGTGACATCT

TTTA	GGGT	CTGT	GAAG	AAGCC	TGACAT	ACATCTG	CGCTGTG	TGACATCT	CATCTGGA	GCTGTGACA
T	TT	GAC	CCCG	CGCT	CTGG	GAGC	ACATC	GGAGC	GCCGGT	TCTGGA
TGGT	TTTT	CTGG	GTGAC	CGGTT	CGCTGT	CATCTGG	CCGCTGT	TCTGGAGC	TGGAGCCG	GCCGGTTCC
T	TA	AGC	ATC	CCCG	GACA	AGCC	GACAT	CGGTT	GTTCCC	CGGAGC
GGGT	TATT	AGCC	CCGCT	CATCTG	ATCTGG	CTGTGAC	GGAAGCC	GGAGCCG	GGAAGCCC	TGGAGCCGG
T	TT	GGT	GTG	GAG	AGCC	ATCT	CGCTG	GTTCCC	GCTGTG	TTCCCG
GTAT	TTGG	GGAA	CGGA	CGCTG	GTGACA	TGTGACA	GGAGCCG	CGGTTCCC	GGAGCCGG	GACATCTGG
T	TT	GCC	GCCA	TGAC	TCTG	TCTG	GTTCC	GGAGC	TTCCCG	AGCCGG
TTGG	TTTA	CCGG	GCCCG	CTGTG	CGGTTC	ATCTGGA	GTGACAT	GGAAGCCC	GAGCCGGT	GTGACATCT
G	TT	TTC	CTG	ACAT	CCGG	GCCG	CTGGA	GCTGT	TCCCGG	GGAGCC
GTGT	GTTT	ACAT	CGGTT	GGAAG	GACATC	GCTGTG	GCCGGTT	CTGGAGCC	TGTGACATC	ACATCTGGA
T	TG	CTG	CCC	CCCG	TGGA	ACATC	CCCGG	GGTTC	TGGAG	GCCGGT
GTTT	GTTT	CGGT	CCCGG	TGTGA	GGTTCC	CGCTGT	CATCTGG	CGCTGTGA	TCTGGAGC	GAGCCGGTT
G	GT	TCC	AGC	CATC	CGGA	GACAT	AGCCG	CATCT	CGGTTC	CCCGGA
TTGT	TTTG	GTTC	GTTCC	TCCCG	GTTCCC	GGTTCCC	TGACATC	GACATCTG	GACATCTG	TGACATCTG
G	TG	CCG	CGG	GAGC	GGAG	GGAG	TGGAG	GAGCC	GAGCCG	GAGCCG
GTTT	TGGG	GAGC	GCCG	CCGGA	CCGGTT	TGGAGC	GCTGTGA	TGTGACAT	ACATCTGG	GGAAGCCCG
A	TT	CGG	GTTC	GCCA	CCCG	CGGTT	CATCT	CTGGA	AGCCGG	CTGTGA
GTTA	TTTT	GACA	GCTGT	ACATCT	TTCCCG	GCCGGT	ACATCTG	CCGGTTCC	GCCGGTTCC	CCGGTTCCC
T	AT	TCT	GAC	GGA	GAGC	TCCCG	GAGCC	CGGAG	CGGAG	GGAGCC
TATG	TGGT	CCGG	GGAA	GCTGT	GGAGCC	GGAAGC	CCGGTTC	GAGCCGGT	TGACATCTG	CATCTGGAG
T	TT	AGC	GCCC	GACA	GGTT	CCGCT	CCGGA	TCCCG	GAGCC	CCGGTT
TGTG	TTTG	TGAC	CCGGT	GCCGG	CCCGGA	TGACATC	TGGAGCC	ACATCTGG	ATCTGGAG	AGCCGGTTC
T	GG	ATC	TCC	TTCC	GCCA	TGGA	GGTTC	AGCCG	CCGGTT	CCGGAG
TTTA	TTAT	TTCCC	GACAT	CCCGG	ACATCT	GAGCCG	CGGTTCC	GCCGGTTC	CGGTTCCCG	ATCTGGAGC
G	GT	GG	CTG	AGCC	GGAG	GTTCC	CGGAG	CCGGA	GAGCC	CGGTTC
ATTT	TTGG	CCGC	CCGGA	CCGGT	GCTGTG	GGAGCC	GAGCCGG	AGCCGGTT	GGTTCCCG	CGGTTCCCG
G	GT	TGT	GCC	TCCC	ACAT	GGTTC	TTCCC	CCCGG	GAGCCA	GAGCCA
TTAT	TATT	CCCG	AGCCG	GTGAC	TGTGAC	CGGTTCC	GTTCCCG	GGTTCCCG	AGCCGGTT	AGCACAGAG
G	TG	GAG	GTT	ATCT	ATCT	CGGA	GAGCC	GAGCC	CCCGGA	GTCAGA
TATT	GTTA	CGGA	CCCGC	GTTCCC	GCCGGT	AGCCGG	AGCCGGT	GTTCCCGG	CGTGCTACC	GCACAGAGG
G	TT	GCC	TGT	GGA	TCCC	TTCCC	TCCCG	AGCCA	GTGAA	TCAGAT
TGGG	GTAT	TACC	AGCCC	TTCCCG	TCCCGG	TTCCCGG	TTCCCGG	GTGCTACC	GCACAGAG	CAGCACAGA
T	TT	GTG	GCT	GAG	AGCC	AGCC	AGCCA	GTGAA	GTCAGA	GGTCAG
TGTA	GTTT	TCCC	TTCCC	TGACA	AGCCGG	TCCCGGA	GTGCTAC	CGTGCTAC	AGCACAGA	CACAGAGGT
T	AT	GGA	GGA	TCTG	TTCC	GCCA	CGTGA	CGTGA	GGTCAG	CAGATG
GGTA	TTGT	GAAG	TCCCG	GAGCC	GAGCCG	GTGCTAC	TGCTACC	GCACAGA	CACAGAGG	ACAGAGGTC
T	GT	CCC	GAG	GGTT	GTTC	CGTG	GTGAA	GGTCAG	TCAGAT	AGATGG
TTAG	ATTT	AAGC	TGACA	AGCCG	CTACCG	TGCTACC	CGTGCTA	CACAGAG	CAGCACAG	TATGCGTGC
T	TG	CCG	TCT	GTTC	TGAA	GTGA	CCGTG	GTCAGA	AGGTCA	TACCGT
GGTT	ATTT	TATG	AAGCC	CTACC	GTGCTA	GCTACCG	CACAGAG	CAGCACAG	ACAGAGGT	CTATGCGTG
A	GT	CGT	CGC	GTGA	CCGT	TGAA	GTCAG	AGGTC	CAGATG	CTACCG
GTTG	TGTT	GCCC	TACCG	TACCGT	TGCTAC	CGTGCTA	AGCACAG	CAGAGGTC	CAGAGGTC	ATGCGTGCT
T	AT	GCT	TGA	GAA	CGTG	CCGT	AGGTC	AGATG	AGATGG	ACCGTG
TTAG	TTGT	ATGC	CTACC	TGCTAC	GCTACC	GCACAG	CAGCACA	AGCACAGA	TATGCGTGC	CCTATGCGT
G	GG	GTG	GTG	CGT	GTGA	AGGTC	GAGGT	GGTCA	TACCG	GCTACC
ATTG	TTTT	CCCG	ACCGT	GTGCT	CGTGCT	CAGCAC	ACAGAGG	ACAGAGGT	CTATGCGTG	TGCGTGCTA
T	AG	CTG	GAA	ACCG	ACCG	AGAGG	TCAGA	CAGAT	CTACC	CCGTGA
TGTT	TATT	AGGT	GCTAC	GCTAC	CAGCAC	CAGAGG	GCACAGA	AGAGGTCA	TGCGTGCTA	GCGTGCTAC
A	GG	CAG	CGT	CGTG	AGAG	TCAGA	GGTCA	GATGG	CCGTG	CGTGAA
GATT	GTGT	CGTG	GTGCT	CGTGC	AGCACA	ACAGAG	CAGAGGT	GCGTGCTA	ATGCGTGCT	ACCTATGCG
T	TT	CTA	ACC	TACC	GAGG	GTCAG	CAGAT	CCGTG	ACCGT	TGCTAC
TAGG	TAGG	GTGC	TGCTA	AGAGG	GCACAG	GAGGTC	AGAGGTC	CTATGCGT	CCTATGCGT	CACCTATGC
T	GT	TAC	CCG	TCAG	AGGT	AGATG	AGATG	GCTAC	GCTAC	GTGCTA
TTGT	GTTT	ACCG	CGTGC	GCACA	CAGAGG	AGCACA	GAGGTCA	TATGCGTG	GCGTGCTA	GCAGCACAG
A	TA	TGA	TAC	GAGG	TCAG	GAGGT	GATGG	CTACC	CCGTGA	AGGTCA
ATTG	TATG	TCAG	AGGTC	AGCAC	CACAGA	AGAGGT	TATGCGT	ATGCGTGC	ACCTATGCG	AGCAGCACA
G	TT	ATG	AGA	AGAG	GGTC	CAGAT	GCTAC	TACCG	TGCTA	GAGGTC
ATGT	TGTT	GTCA	GAGG	ACAGA	GAGGTC	CACAGA	GCGTGCT	TGCGTGCT	CACCTATGC	GCCACCTAT
T	TG	GAT	TCAG	GGTC	AGAT	GGTCA	ACCGT	ACCGT	GTGCT	GCGTGC
TAGT	TTAT	AGCC	GTCAG	GGTCA	AGAGGT	AGGTCA	CCTATGC	CCTATGCG	GCAGCACA	CCACCTATG
T	TG	CGC	ATG	GATG	CAGA	GATGG	GTGCT	TGCTA	GAGGTC	CGTGCT
AGTT	TGTG	CAGA	ACAGA	GAGGT	AGGTCA	CCTATGC	CTATGCG	ACCTATGC	AGCAGCAC	AGCCACCTA
T	TT	TGG	GGT	CAGA	GATG	GTGC	TGCTA	GTGCT	AGAGGT	TGCGTG

TGTT	TGTA	CTAC	TATGC	AGGTC	ACAGAG	CTATGCG	ATGCGTG	CACCTATG	CCACCTATG	GAGCCACCT
G	TT	CGT	GTG	AGAT	GTCA	TGCT	CTACC	CGTGC	CGTGC	ATGCGT
GTAT	GTTT	GGTC	AGAG	CACAG	GGTCAG	ATGCGT	TGCGTGC	GCAGCACA	GCCACCTAT	CAGAGGTCA
G	GG	AGA	GTCA	AGGT	ATGG	GCTAC	TACCG	GAGGT	GCGTG	GATGGG
TATG	TTTA	CCGT	GCACA	CAGAG	CTATGC	GCGTGC	ACCTATG	AGCAGCAC	AGCCACCTA	CGTGCTACC
G	GG	GAA	GAG	GTCA	GTGC	TACCG	CGTGC	AGAGG	TGCGT	GTGAAG
GGTG	GATT	AGAG	GGTCA	CAGCA	CCTATG	TATGCGT	CACCTAT	CCACCTAT	GAGCCACC	GTGCTACCG
T	TT	GTC	GAT	CAGA	CGTG	GCTA	GCGTG	GCGTG	TATGCG	TGAAGC
GGGT	GAGC	CAGA	CAGA	GTCAG	TATGCG	TGCGTG	GCAGCAC	GCCACCTA	AGAGGTCA	AGAGGTCAG
A	CA	GGT	GGTC	ATGG	TGCT	CTACC	AGAGG	TGCGT	GATGGG	ATGGGA
ATTA	TTAG	GAGG	TCAGA	CTATGC	ATGCGT	ACCTATG	AGCAGCA	AGCCACCT	GTGCTACC	GAGGTCAGA
T	GT	TCA	TGG	GTG	GCTA	CGTG	CAGAG	ATGCG	GTGAAG	TGGGAG
GTTA	TTTA	ACAG	AGCAC	TATGC	GCGTGC	CACCTAT	GCCACCT	GAGCCACC	GAGGTCAG	CTACCGTGA
G	TG	AGG	AGA	GTGC	TACC	GCGT	ATGCG	TATGC	ATGGGA	AGCAGC
TGTG	TTTA	TTTTT	CAGCA	ATGCG	TGCGTG	GCAGCA	CCACCTAT	GAGGTCA	TGCTACCGT	TACCGTGAA
G	GT	GT	CAG	TGCT	CTAC	CAGAG	GCGT	GATGGG	GAAGC	GCAGCA
AGG	TTTG	TGCT	CACAG	CCTATG	ACCTAT	AGCAGC	AGCCACC	TGCTACCG	TACCGTGA	GCTACCGTG
GT	TA	ACC	AGG	CGT	GCGT	ACAGA	TATGC	TGAAG	AGCAGC	AAGCAG
GTTG	GTTG	GCTA	CTATG	GCGTG	CACCTA	CCACCTA	GAGCCAC	AGGTCAGA	CTACCGTGA	TGCTACCGT
G	TT	CCG	CGT	CTAC	TGCG	TGCG	CTATG	TGGGA	AGCAG	GAAGCA
AGGT	GGTT	GTTA	ATGCG	TGCGT	GCAGCA	GCCACCT	AGGTCAG	GCTACCGT	GCTACCGT	AGGTCAGAT
T	TG	TTT	TGC	GCTA	CAGA	ATGC	ATGGG	GAAGC	GAAGCA	GGGAGC
TATA	TGTT	AGAT	CCTAT	ACCTAT	AGCAGC	AGCCACC	GCTACCG	CCGTGAAG	ACCGTGAA	ACCGTGAAG
T	TA	GGG	GCG	GCG	ACAG	TATG	TGAAG	CAGCA	GCAGCA	CAGCAC
TGGT	GGTT	GGTT	GCGTG	CACCTA	GCCACC	GAGCCA	CTACCGT	CTACCGTG	AGGTCAGA	GGTCAGATG
A	TA	TTT	CTA	TGC	TATG	CCTAT	GAAGC	AAGCA	TGGGAG	GGAGCC
TAGG	ATGT	TTTT	TGCGT	GCAGC	CCACCT	GGTCAG	GGTCAGA	ACCGTGAA	GGTCAGAT	AAGCAGCAC
G	TT	GTT	GCT	ACAG	ATGC	ATGGG	TGGGA	GCAGC	GGGAGC	AGAGGT
GGG	GGTT	GTAG	ACCTA	AGCAG	AGCCAC	CTACCGT	CCGTGAA	TACCGTGA	AAGCAGCA	CCGTGAAGC
GT	AT	GGT	TGC	CACA	CTAT	GAAG	GCAGC	AGCAG	CAGAGG	AGCACA
GGAT	ATTT	TGGG	CACCT	GCCAC	GAGCCA	TACCGTG	TACCGTG	GGTCAGAT	CCGTGAAG	GAAGCAGCA
T	TA	TTT	ATG	CTAT	CCTA	AAGC	AAGCA	GGGAG	CAGCAC	CAGAGG
TATT	AGTT	TTTG	GCAGC	CCACCT	GTCAGA	GTCAGAT	ACCGTGA	GTCAGATG	GTGAAGCA	CGTGAAGCA
A	TT	TTT	ACA	ATG	TGGG	GGGA	AGCAG	GGAGC	GCACAG	GCACAG
TTAT	TAGG	GGTA	AGCA	AGCCA	TACCGT	GTGAAG	CGTGAAG	AAGCAGCA	GTCAGATG	GTGAAGCAG
A	TT	GGG	GCAC	CCTA	GAAG	CAGCA	CAGCA	CAGAG	GGAGCC	CACAGA
TGAT	ATTG	TGCG	GCCAC	TCAGA	TCAGAT	CCGTGA	GTCAGAT	CGTGAAGC	GAAGCAGC	TGAAGCAGC
T	GT	TGC	CTA	TGGG	GGGA	AGCAG	GGGAG	AGCAC	ACAGAG	ACAGAG
TTGG	TTGT	TTGG	CAGAT	GAGCC	ACCGTG	ACCGTG	TCAGATG	TGAAGCAG	CGTGAAGC	GGAGCCACC
A	AT	GTT	GGG	ACCT	AAGC	AAGCA	GGAGC	CACAG	AGCACA	TATGCG
ATAT	TTGT	CTAT	CCACC	ACCGT	CAGATG	CGTGAA	CAGATGG	GTGAAGCA	TGAAGCAG	GTCAGATGG
T	TG	GCG	TAT	GAAG	GGAG	GCAGC	GAGCC	GCACA	CACAGA	GAGCCA
TTTA	GTAT	GCAC	AGCCA	CAGAT	GTGAAG	TCAGATG	GAAGCAG	TCAGATGG	GGAGCCAC	AGAGGTCAG
A	TG	AGA	CCT	GGGA	CAGC	GGAG	CACAG	GAGCC	CTATGC	ATGGAG
GGTT	GGGT	TTTG	AGATG	CCGTG	CCGTGA	CAGATG	GTGAAGC	GAAGCAG	TCAGATGG	CAGAGGTCA
G	AT	GTT	GGA	AAGC	AGCA	GGAGC	AGCAC	CACAGA	GAGCCA	GATGGA
TTTG	ATTT	TTTTT	GGTA	AGATG	TGAAGC	AAGCAG	AAGCAGC	CAGATGG	GAGGTCAG	GAGGTCAGA
A	GG	TG	GGGT	GGAG	AGCA	CACAG	ACAGA	GAGCCA	ATGGAG	TGGAGC
TGGG	GGTA	TTTT	GAGCC	TGAAG	CGTGAA	AGATGG	TGAAGCA	GGAGCCAC	AGAGGTCA	AGGTCAGAT
G	TT	GTG	ACC	CAGC	GCAG	GAGCC	GCACA	CTATG	GATGGA	GGAGCC
ATGG	TTTT	GCGT	AGGTA	GAAGC	AGATGG	TGAAGC	GGAGCCA	AGGTCAGA	AGGTCAGA	TATTGGACA
T	GA	GCT	GGG	AGCA	GAGC	AGCAC	CCTAT	TGGAG	TGGAGC	CCTTCC
GTGG	GGAT	TTTT	GATG	GTGAA	GATGGG	GAAGCA	AGATGGG	GAGGTCA	GGTCAGAT	CGATTGTGG
T	TT	GGT	GGAG	GCAG	AGCC	GCACA	AGCCA	GATGGA	GGAGCC	GGACGT
AATT	GTAT	TGTT	CCGTG	CGTGA	GAAGCA	GATGGG	GGTCAGA	GGTCAGAT	GATTGTGG	GATTGTGGG
T	GT	ATT	AAG	AGCA	GCAC	AGCCA	TGGAG	GGAGC	GGACGT	GACGTA
TTAA	TATT	TTTT	ATGG	GATGG	AAGCAG	GGAGCC	GTCAGAT	GTCAGATG	ATTGGACA	GGTCAGATG
T	GT	GGG	GAGC	GAGC	CACA	ACCTA	GGAGC	GAGCC	CCTTCC	GAGCCA
ATTT	TAGT	GTTT	TGAAG	AGGTA	ATGGGA	GTCAGAT	AGGTCAG	CGATTGTG	CCGATTGTG	GACGTATTG
A	TT	TTT	CAG	GGGT	GCCA	GGAG	ATGGA	GGGAC	GGGAC	GACACC
GTGT	TTAG	ATTTT	GAAG	TGGGA	GGAGCC	GGTCAG	TCAGATG	TTGGACAC	GTCAGATG	CCGATTGTG
A	TT	TT	CAGC	GCCA	ACCT	ATGGA	GAGCC	CTTCC	GAGCCA	GGGACG
GTAG	TATT	GTTT	GTGAA	ATGGG	CAGGTA	TCAGATG	TGGACAC	TATTGGAC	TATTGGACA	CGTATTGGA
T	AT	TTG	GCA	AGCC	GGGT	GAGC	CTTCC	ACCTT	CCTTC	CACCTT

TAAT	TTGG	TATTT	AAGCA	CAGGT	TCAGAT	CAGATG	GGACACC	CGTATTGG	ATTGTGGG	ACGTATTGG
T	TG	TT	GCA	AGGG	GGAG	GAGCC	TTCCC	ACACC	GACGTA	ACACCT
ATGG	TTAT	AGCA	CGTGA	AAGCA	GTCAGA	GGACAC	GATTGTG	GATTGTGG	GGGACGTA	GGACGTATT
G	GG	CAG	AGC	GCAC	TGGA	CTTCC	GGGAC	GGACG	TTGGAC	GGACAC
TGTG	TTGT	TTTTT	CAGGT	GGAGC	CAGATG	GTGGGG	TTGGACA	ATTGTGGG	CGATTGTG	GGGACGTAT
A	TA	GG	AGG	CACC	GAGC	ACGTA	CCTTC	GACGT	GGGACG	TGGACA
GTGT	GTTG	TGTT	GGAG	TTGTG	AGATGG	TCCGATT	TTGTGGG	TTGTGGGG	ACGTATTG	GTGGGGACG
G	GT	TTT	CCAC	GGGA	AGCC	GTGG	GACGT	ACGTA	GACACC	TATTGG
GTAA	GTTT	TGGT	GGGA	TCAGA	GTGGG	TGGACA	GGGACGT	ATCCGATT	GACGTATT	GTATTGGAC
T	AG	TTT	GCCA	TGGA	GACGT	CCTTC	ATTGG	GTGGG	GGACAC	ACCTTC
TTGA	AATT	GTTT	TGGG	GATTG	GACACC	GACACCT	CGATTGT	CCGATTGT	CGTATTGG	GGGGACGTA
T	TT	GTT	AGCC	TGGG	TTCC	TCCC	GGGGA	GGGGA	ACACCT	TTGGAC
TGGT	TTAG	CACA	TTGTG	ACGTA	GGACAC	CCGATTG	GGGGACG	GGGACGT	GTGGGGAC	GATATCCGA
G	GG	GAG	GGG	TTGG	CTTC	TGGG	TATTG	ATTGGA	GTATTG	TTGTGG
GTAT	TATG	AGGT	TGTTA	AGATG	CGATTG	GGGACG	GTGGGGA	ATTGGACA	GTATTGGA	TCCGATTGT
A	GT	AGG	TTT	GAGC	TGGG	TATTG	CGTAT	CCTTC	CACCTT	GGGGAC
GTAG	ATTG	TTTG	GATTG	CAGAT	GATTGT	ATTGTG	TCCGATT	TGGACACC	GGACGTAT	TGGGGACGT
G	TT	TGT	TGG	GGAG	GGGG	GGGAC	GTGGG	TTCCC	TGGACA	ATTGGA
TGGA	TTTT	TTGT	ATTGT	ATTGT	CCGATT	GATTGT	CAGATGG	GAGGATAT	TTGTGGGG	ATTGTGGGG
T	AA	TTT	GGG	GGGG	GTGG	GGGGA	AGCCA	CCGAT	ACGTAT	ACGTAT
GTGG	TTGG	TTGT	GTATT	CGATT	TGGACA	ATCCGAT	ATCCGAT	GGACGTAT	GGGGACGT	GGATATCCG
G	TA	GGG	GGA	GTGG	CCTT	TGTG	TGTGG	TGGAC	ATTGGA	ATTGTG
TGGG	AGGT	CCTA	TGTGG	CGTATT	ACGTAT	CGATTGT	TATTGGA	TCAGATGG	TGGGGACG	GAGGATATC
A	TT	TGC	GGA	GGA	TGGA	GGGG	CACCT	AGCCA	TATTGG	CGATTG
TAGT	TGGT	TTGG	TTGGG	GGGGA	ATTGTG	TTGGACA	TGTGGGG	GTGGGGA	GATATCCG	ATATCCGATT
A	TA	TTT	TTT	CGTA	GGGA	CCTT	ACGTA	CGTATT	ATTGTG	GTGGG
AGTG	GTTA	GTGT	ACGTA	ATCCG	TCCGAT	GACGTAT	ATTGGAC	GGGGACG	ATATCCGAT	ATCCGATTG
T	GT	TTT	TTG	ATTG	TGTG	TGGA	ACCTT	TATTGG	TGTGG	TGGGGA
TGGA	TCTG	GGGT	CGTAT	TGGGG	GGGGA	GAGGAT	CCGATTG	ACGTATTG	TATCCGATT	TGTGGGGAC
G	GA	TTT	TGG	ACGT	CGTAT	ATCCG	TGGGG	GACAC	GTGGG	GTATTG
ATGT	ATTT	ATTT	GGGG	GGACA	GGGAC	GGGGAC	GACGTAT	GACGTATT	TTGGACACC	AGGATATCC
G	AT	GTT	ACGT	CCTT	GTATT	GTATT	TGGAC	GGACA	TTCCC	GATTGT
GATT	ATAT	TTTG	AGATG	GACAC	TGGGGA	TGGGGA	ATTGTGG	GTATTGGA	TCCGATTGT	TTGTGGGGA
G	TT	GGT	GAG	CTTC	CGTA	CGTAT	GGACG	CACCT	GGGGA	CGTATT
ATAT	TGTT	TTTG	GGGA	GGGAC	GACGTA	TATTGGA	CGTATTG	TGGGGAC	GAGGATAT	ATTGGACAC
G	GT	TTG	CGTA	GTAT	TTGG	CACC	GACAC	GTATTG	CCGATT	CTTCCC
ATGT	ATCT	ATTTT	TATTG	GACGT	TTGTGG	TGTGGG	GTATTGG	TATCCGAT	GGATATCC	TATCCGATT
A	GG	GT	GAC	ATTG	GGAC	GACGT	ACACC	TGTGG	GATTGT	GTGGGG
TATA	AGG	GTAT	GGAC	GTGGG	ACACCT	ACGTATT	GAGGATA	GGACACCT	ATCCGATTG	CGTGCTACC
G	GTT	TGG	GTAT	GACG	TCCC	GGAC	TCCGA	TCCCG	TGGGG	GTGAAA
GGTA	GTTA	TTTTA	CGATT	ACACCT	ATCCGA	GGACGT	GGACGTA	TGTGGGG	AGGATATC	TTGGACACC
A	TG	TT	GTG	TCC	TTGT	ATTGG	TTGGA	ACGTAT	CGATTG	TTCCCG
GGG	GATT	TTATT	GAGG	CCGATT	GGACGT	AGATGG	AGGATAT	TCCGATTG	TGTGGGGA	CGAGGATAT
AT	GT	TT	ATAT	GTG	ATTG	AGCCA	CCGAT	TGGGG	CGTATT	CCGATT
TGTA	GGGT	TTTTT	TTTTT	GATGG	GAGGAT	ATTGGAC	GACACCT	GATATCCG	TGGACACCT	GTGCTACCG
G	TA	AT	GTG	AGCC	ATCC	ACCT	TCCCG	ATTGT	TCCCG	TGAAAC
GGTT	ATGG	GTAT	TGTTT	GTATT	CGTATT	TTGTGG	ACGTATT	ATATCCGA	GTGCTACC	TGCTACCGT
C	TT	TTT	GTT	GGAC	GGAC	GGACG	GGACA	TTGTG	GTGAAA	GAAACA

[4] Motifs Found in Contigs

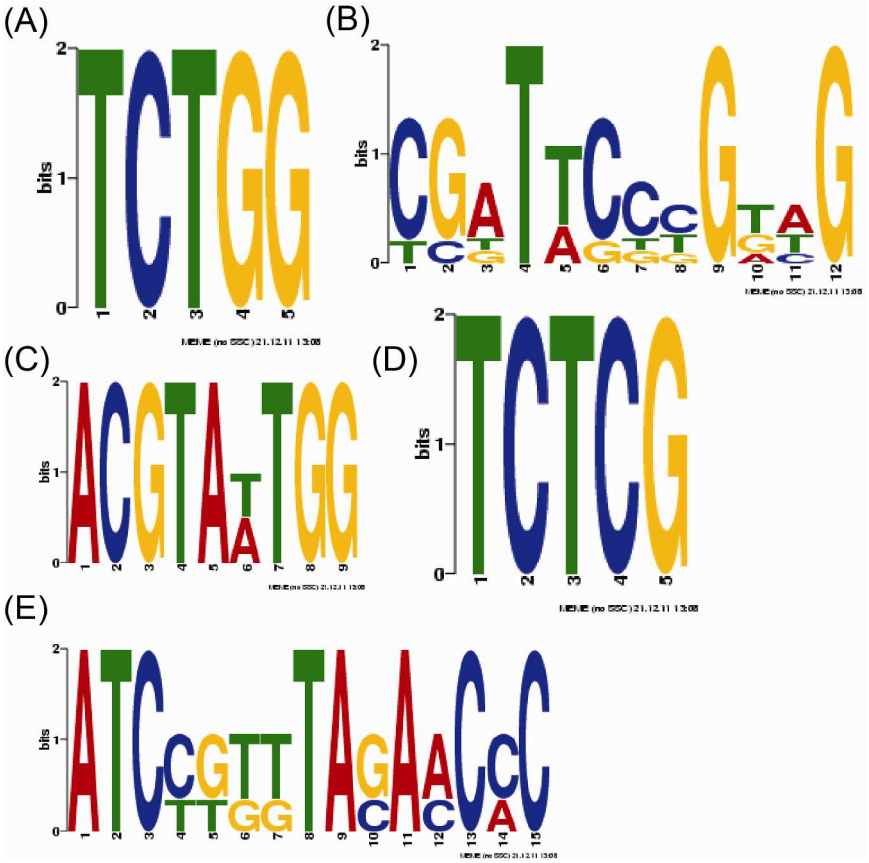

Figure S4. Motifs found in contigs. (A) Motif 1 was highlighted in pink in Table 3. (B) Motif 2 was highlighted in yellow in Table 3. (C) Motif 3 was highlighted in light blue in Table 3. (D) Motif 4 was highlighted in dark blue in Table 3. (E) Motif 5 was highlighted in green in Table 3.