DOI: 10.1534/genetics.109.101253

# Novel Nucleotide Sequence Motifs That Produce Hotspots of Meiotic Recombination in *Schizosaccharomyces pombe*

**Walter W. Steiner,[1] Estelle M. Steiner, Angela R. Girvin and Lauren E. Plewik**

*Department of Biology, Niagara University, Lewiston, New York 14109*

Manuscript received January 28, 2009
Accepted for publication April 9, 2009

ABSTRACT

In many organisms, including yeasts and humans, meiotic recombination is initiated preferentially at a limited number of sites in the genome referred to as recombination hotspots. Predicting precisely the location of most hotspots has remained elusive. In this study, we tested the hypothesis that hotspots can result from multiple different sequence motifs. We devised a method to rapidly screen many short random oligonucleotide sequences for hotspot activity in the fission yeast *Schizosaccharomyces pombe* and produced a library of ~500 unique 15- and 30-bp sequences containing hotspots. The frequency of hotspots found suggests that there may be a relatively large number of different sequence motifs that produce hotspots. Within our sequence library, we found many shorter 6- to 10-bp motifs that occurred multiple times, many of which produced hotspots when reconstructed *in vivo*. On the basis of sequence similarity, we were able to group those hotspots into five different sequence families. At least one of the novel hotspots we found appears to be a target for a transcription factor, as it requires that factor for its hotspot activity. We propose that many hotspots in *S. pombe*, and perhaps other organisms, result from simple sequence motifs, some of which are identified here.

MEIOSIS is a form of cell division common to all sexually reproducing organisms. It differs from mitosis in that two cell divisions follow a single round of DNA replication, resulting in four haploid cells known as gametes or spores. The first division of meiosis differs from mitosis in that paternal and maternal homologs segregate to opposite poles, reducing by half the number of chromosomes in the resulting gametes. Prior to the first division, homologous chromosomes recombine with each other at a greatly elevated frequency compared to mitosis (Esposito and Wagstaff 1981). This recombination serves at least two important functions. First, it forms crossovers (chiasmata) between chromosomes, which are required in most organisms for the proper segregation of homologous chromosomes (Baker *et al.* 1976). Second, the random shuffling of maternal and paternal alleles at each generation increases genetic diversity, which enhances the ability of a species to adapt to its environment through natural selection.

Meiotic recombination events are not distributed evenly throughout the genome of most organisms. Rather, they occur at high frequency at some sites and low frequency at others. Sites that recombine at a frequency significantly higher than the genomic average are known as hotspots. These hotspots coincide with the formation of DNA double-strand breaks (DSBs) in both the fission and budding yeasts and likely many other organisms (Sun *et al.* 1989; Cao *et al.* 1990; Fan *et al.* 1995; Cervantes *et al.* 2000; Mahadevaiah *et al.* 2001; Petes 2001; Steiner *et al.* 2002; Young *et al.* 2002; Cromie *et al.* 2007). Formation of DSBs requires a number of different proteins, including Spo11 (Rec12 in *Schizosaccharomyces pombe*), a widely conserved protein among eukaryotes that has the active site for cleaving the phosphodiester backbone (Keeney *et al.* 1997; Cervantes *et al.* 2000; Malik *et al.* 2007). After the formation of DSBs, the two broken ends of the DNA initiate recombination by invading intact homologous DNA to form joint molecules, which can be resolved to produce both crossover and noncrossover exchanges (Pâques and Haber 1999).

DNA breaks occur at preferred genomic positions during meiosis, but the factors determining the positions of most break sites are not clearly understood. A global analysis of the distribution of DSB sites in the budding yeast *Saccharomyces cerevisiae* showed that most breaks occur in 5-kb regions where the GC content exceeds the average GC content for the genome (Gerton *et al.* 2000), but the causal relation, if any, between elevated GC content and DSB formation is unknown. In addition, DSBs occur primarily in intergenic regions (IGRs) (Baudat and Nicolas 1997; Gerton *et al.* 2000), which supports the possibility that many hotspots are associated with the binding of

Supporting information is available online at http://www.genetics.org/cgi/content/full/genetics.109.101253/DC1.

[1]*Corresponding author:* Department of Biology, Box 2032, Niagara University, Lewiston, NY 14109. E-mail: wsteiner@niagara.edu

transcription factors. Indeed, several well-characterized hotspots are known to require transcription factor binding (White *et al.* 1993; Kon *et al.* 1997; Petes 2001; Mieczkowski *et al.* 2006).

Like *S. cerevisiae*, the fission yeast *S. pombe* shows a nonrandom distribution of meiotic DSBs (Cervantes *et al.* 2000; Young *et al.* 2002). A genomewide analysis of the distribution of DSBs in *S. pombe* revealed that DSBs occur primarily in intergenic regions, particularly in exceptionally large intergenic regions (Cromie *et al.* 2007). In fact, it was noted that large IGRs are strongly predictive of DSBs. The median IGR size in *S. pombe* is 0.7 kb (Wood *et al.* 2002), but approximately half of all prominent DSBs occur in IGRs >3 kb, and 44% of IGRs >3 kb contain prominent DSBs (Cromie *et al.* 2007). However, the basis for the association of DSBs and IGRs remains unclear. One possibility is that some IGRs contain specific sequence motifs that create recombination hotspots whereas others do not.

One simple sequence motif, ATGACGTCA, known as *M26* or *CRE* (**c**yclic-AMP **r**esponse **e**lement), acts as a recombination hotspot at multiple sites in the *S. pombe* genome. These sites include both IGRs and protein-coding regions, for example, the *ade6* gene where this hotspot motif was originally discovered (Gutz 1971; Fox *et al.* 1997; Steiner and Smith 2005a). The *CRE* sequence is a binding site for a heterodimeric transcription factor, Atf1-Pcr1, which is required for activity of the hotspot (Kon *et al.* 1997). Since the *CRE* hotspot can predict only a small fraction of the DSBs in the *S. pombe* genome (Cromie *et al.* 2007 and W. Steiner, unpublished observation), we hypothesized that there may be other simple sequence motifs that contribute to recombination hotspots. The results shown in this study confirm that hypothesis.

## MATERIALS AND METHODS

**Strains and genetic procedures:** All strains used in this study with assigned *ade6* allele numbers are shown in Table 1. The *ade6-4001* allele was generated by amplification of the *kanMX6-ura4*$^+$ construct contained in the plasmid *pura4-kanMX6* (Steiner and Smith 2005b). This construct was amplified by PCR using oligonucleotides with 80 base 5′ extensions homologous to the *ade6* insertion site (Table 1). The resulting PCR product was used for linear transformation of strain WS121 selecting for uracil prototrophy. Transformants were confirmed by resistance to G418, sequencing, and Southern blot hybridization.

The plasmid pWS35 was constructed by PCR amplification of a 613-bp fragment of *ade6* with primers 5′-NNNNNCT CGAGCTTGGAAATGTAACGATGAC-3′ and 5′-NNNNNCTC GAGTAAGCCAATGTTTTACTTTTCAG-3′, containing *Xho*I restriction sites near their 5′ ends. (The additional random bases at the 5′ ends were added to permit efficient cleavage by *Xho*I.) The resulting PCR product was digested with *Xho*I and ligated into the *Xho*I site of pSP1 (Cottarel *et al.* 1993), producing pWS35. This plasmid contains the *S. cerevisiae LEU2* gene as a selectable marker. The 613-bp *ade6* fragment of pWS35 is the same fragment deleted in the *ade6-4001* allele (Table 1).

Growth and sporulation medium have been previously described (Gutz *et al.* 1974; Steiner and Smith 2005a,b). Heterothallic strains were grown in rich medium supplemented with adenine (A), uracil (U), leucine, histidine, and lysine (YEL-5S). Strains containing the plasmid pWS35 were grown in minimal medium supplemented with adenine and uracil (NBL-AU). Crosses between heterothallic strains were performed on SPA-5S, and homothallic crosses were performed on SPA-AU. Crosses were incubated for 2 days at 25° before harvesting and analysis as described (Steiner and Smith 2005b). Any given cross was performed a minimum of three times for determination of recombination frequencies.

**Screen for hotspots:** WS129 was grown overnight in NBL-A to 0.5-1 × $10^7$ cells/ml and transformed using a lithium acetate-mediated transformation procedure (Bähler *et al.* 1998). Transforming DNA consisted of a 1.6-kb *ade6* fragment containing a 15- or 30-bp random sequence substitution from nucleotides 125–139 or 125–154, respectively (nucleotides from start of *ade6* open reading frame). Both substitutions also include an A→T substitution at nucleotide 121, which produces a stop codon and insures that all transformants are adenine auxotrophs. The linear DNA used for transformation was generated by overlap extension PCR (Vallejo *et al.* 1995). The primers used were oWS202: ACGAACATCATTAAGCG CGAAGCG, oWS203: ACGCATGAGTTGTGGAAGTCGAGA, oWS208: GTTAGGCAGGAGAATTTGCTGCA, oWS209: TGC AGCAAATTCTCCTGCCTAACNNNNNNNNNNNNNNNNN NNNNNNNNNNNNNCATCATTTACTGACCCCGATGCAA TTG, and oWS241: TGCAGCAAATTCTCCTGCCTAACNNNN NNNNNNNNNNGTGAGCACATTGATGCATCATTTAC. The product of oWS202 and oWS208 was combined with the product of oWS209 (or 241) and oWS203 for the overlap extension reaction. All PCR products were performed with the high-fidelity polymerases Vent (New England Biolabs) or *Pfx* (Invitrogen), using conditions recommended by the manufacturers. A plasmid clone of the *ade6* gene (pAS1; Szankasi *et al.* 1988) was used as the starting template for PCR reactions. Approximately 5 μg of linear *ade6* DNA was used per $10^8$ cells transformed.

Following transformation, cells were plated on NBA-AU (2–4 × $10^7$ cells/plate) and incubated at 32° for 2 days. Those plates were replica plated to NBA-AU containing 1 mg/ml 5-FOA (Toronto Research Chemicals) to select for cells with *ade6* replacements. After 3 days growth, individual colonies became visible. These plates were replica plated a second time to NBA-AU + 5-FOA to reduce background growth. After 4 days growth at 32°, colonies were replica plated to SPA-AU and incubated 2 days at 25° to induce meiosis. Those colonies were then exposed to acetone vapor (Egel 1977) to kill remaining vegetative cells, replica plated to YEA-4S and incubated 3 days at 32°. Red colonies showing many white papillae were picked and streaked for single colonies to NBA containing uracil and limiting adenine (10 μg/ml) to distinguish *ade6*$^-$ colonies from *ade6*$^+$ recombinants. Red colonies from each streak were patched in a grid pattern to NBA-AU along with 5 control strains: WS129, WS3, WS135, WS136, and WS137 (Table 1). After 2 days growth at 32°, those patches were replica plated to three different media:

1. SPA-AU to confirm the original meiotic hotspot phenotype as described above and qualitatively estimate the activity of each hotspot relative to *ade6-M375*, *ade6-M26*, and *ade6-3074* by comparing the density of white papillae. The heterothallic strain, WS3, which cannot sporulate, served as a control for the effectiveness of the subsequent acetone treatment for this step.
2. YEA-4S to test for a possible mitotic hotspot phenotype. Strains containing a mitotic hotspot should show many

**TABLE 1**

**Strains**

| *ade6* allele | Mating type | Description or sequence[a] | Base pairs[b] | Motif[c] | Strain[d] |
|---|---|---|---|---|---|
| + | $h^{90}$ | | | NA | WS121 |
| *52* | $h^-$ | G796A[e] | | NA | WS3 |
| *M375* | $h^{90}$ | **T**GAGGACGTGAG[f] | 133–144 | NA | WS135 |
| *M26* | $h^{90}$ | GGA**T**GACGTGAG[f] | 133–144 | *CRE* | WS136 |
| *M26* | $h^+$ | *his7-366* | | | WS322 |
| *M26* | $h^+$ | *his7-366 php2Δ::kanMX6*[g] | | | WS310 |
| *M26* | $h^+$ | *his7-366 php3Δ::kanMX6*[g] | | | WS324 |
| *M26* | $h^+$ | *php5Δ::kanMX6*[g] | | | WS295 |
| *469* | $h^-$ | C1468T[f] | | NA | WS315 |
| *469* | $h^-$ | *php2Δ::kanMX6* | | | WS313 |
| *469* | $h^-$ | *php3Δ::kanMX6* | | | WS317 |
| *469* | $h^-$ | *php5Δ::kanMX6* | | | WS311 |
| *3074* | $h^{90}$ | GGA**T**GACGT**C**AG[h] | 133–144 | *CRE* | WS137 |
| *4001* | $h^{90}$ | Δ(-162-451)::*kanMX6-ura4*$^+$ | — | NA | WS129 |
| *4002* | $h^{90}$ | CCA<u>A</u>TCA[i] | 129–134 | 7-7 | WS224 |
| *4002* | $h^+$ | *his7-366* | | | WS326 |
| *4002* | $h^+$ | *his7-366 php2Δ::kanMX6* | | | WS328 |
| *4002* | $h^+$ | *his7-366 php3Δ::kanMX6* | | | WS320 |
| *4002* | $h^+$ | *his7-366 php5Δ::kanMX6* | | | WS293 |
| *4003* | $h^{90}$ | A__**C**ATG<u>A</u>CATCAT | 148–160 | 6-6/8-1 | WS330 |
| *4003* | $h^-$ | | | | WS182 |
| *4005* | $h^{90}$ | ACGT<u>A</u>A_T | 138–145 | 7-1 | WS149 |
| *4006* | $h^{90}$ | C**G**TCA<u>T</u>A[i] | 149–155 | 7-15 | WS150 |
| *4008* | $h^{90}$ | T<u>A</u>TTAC<u>G</u>T**A**A**T** | 160–168 | 6-1 (Con) | WS331 |
| *4008* | $h^-$ | | | | WS184 |
| *4009* | $h^{90}$ | AT<u>A</u>GC**G**TCAT**A**TACT | 152–165 | 7-15 (Con) | WS332 |
| *4009* | $h^-$ | | | | WS186 |
| *4010* | $h^{90}$ | GATG<u>A</u>CAT**A**A | 151–159 | 6-21 (Con) | WS365 |
| *4010* | $h^-$ | | | | WS188 |
| *4071* | $h^{90}$ | ATGATG<u>T</u>CA**C**[i] | 152–161 | 7-2 (Con) | WS237 |
| *4072* | $h^{90}$ | ACCCCG<u>C</u>A**C**GCA | 167–177 | 7-4 (Con) | WS240 |
| *4073* | $h^{90}$ | ACG**G**CC**C**CCA_CAATT[i] | 126–141 | 7-31 (Con) | WS241 |
| *4094* | $h^{90}$ | G**GATGTAAGT**[j] | 130–139 | 10-1 | WS374 |
| *4095* | $h^{90}$ | G**GTCTG**GA**C**C[j] | 130–139 | 10-2 | WS376 |
| *4096* | $h^{90}$ | GATG**ACATCA**[j] | 130–139 | 8-1 | WS378 |
| *4099* | $h^{90}$ | TGA**ACCCCG**C**AC**TGA[j] | 129–143 | 7-4 (Con) | WS382 |
| *4100* | $h^{90}$ | **GCCCCC**ACA[j] | 132–140 | 7-31 (Con) | WS384 |
| *4101* | $h^{90}$ | ACCCCG<u>C</u>A**C**G**T**AAT | 167–179 | 7-4 | WS386 |
| *4102* | $h^{90}$ | A**T**G**G**CC**C**CCA_C**T**ATT[i] | 126–141 | 7-31 | WS394 |
| *4103* | $h^{90}$ | **TGACCCCGCACGT**[j] | 130–142 | 7-4 | WS409 |
| *4104* | $h^{90}$ | **TGGCCCCC**AC**TAT**[j] | 130–142 | 7-31 | WS410 |

[a] Bases are numbered starting at the first nucleotide of the *ade6* open reading frame. Bases in regular type indicate wild-type sequence. Boldface type indicates base substitutions. Insertions or deletions are underlined. For any given *ade6* allele, the relevant sequence is shown only once. Additional genotypic information for strains containing the same allele of *ade6* is also indicated in this column.

[b] The nucleotide positions of the *ade6* sequence shown in the third column. Numbering indicates wild-type sequence before insertions or deletions.

[c] Indicates the motif (Table 2, Table S2a, and Table S3) on which the allele is based. NA, not applicable. Con, consensus sequence (Figure S1).

[d] In addition to the *ade6* mutations shown in the third column, all strains also contain *ura4-D18* and *leu1-32*, except for WS3 (*ura4*$^+$ *leu1*$^+$). All homothallic strains ($h^{90}$) also contain the plasmid pWS35 (Figure 1).

[e] M. Fox and G. Smith, personal communication.

[f] Szankasi *et al.* (1988).

[g] Mercier *et al.* (2006).

[h] Steiner and Smith (2005b).

[i] Complement strand shown.

[j] These strains also contain a closely linked nonsense mutation, A121T.

white papillae prior to undergoing meiosis (only one strain with a potential mitotic hotspot was found).

3. YEA-5S containing 100 μg/ml G418 to confirm homologous replacement of the *ade6* gene (Storici *et al.* 2001). Homologous gene replacement results in the simultaneous loss of both the *ura4*$^+$ and *kanMX6* markers found in the *ade6-4001* allele, resulting in both resistance to 5-FOA and sensitivity to G418.

**Sequence analysis:** G418-sensitive strains that showed an obviously higher density of white papillae than an *ade6-M375* control strain were allowed to lose the plasmid, pWS35, by two successive nonselective streaks onto YEA-4S. Plasmid-free derivatives were identified by the inability to grow in the absence of leucine and used for preparation of genomic DNA, which was used as a template for PCR amplification of *ade6* using primers oWS202 and oWS203 (above). PCR products were sequenced by the High Throughput Genomics Unit (University of Washington, Seattle).

The sequence substitutions in *ade6* were analyzed for common 6- to 10-bp motifs using YMF3.0 (Sinha and Tompa 2002, 2003) (http://wingless.cs.washington.edu/YMF/YMFWeb/YMFInput.pl) and MEME (Bailey *et al.* 1996) (http://meme.nbcr.net/meme/intro.html).

Sequence randomization was accomplished by using an algorithm (Pearson and Lipman 1988) available through the San Diego Super Computer Center (http://workbench.sdsc.edu/).

**Reconstruction of specific motifs:** Specific motifs were reconstructed in the *ade6* gene by overlap-extension PCR (Vallejo *et al.* 1995) using inner primers containing the desired mutations and outside primers oWS202 and oWS203 (above). All reconstructed motifs listed in Table 1 were confirmed by sequencing and Southern blot hybridization.

## RESULTS

**The screen for recombination hotspots:** Screening for sequence-dependent recombination hotspots from a large pool of random sequences required a means of rapidly identifying potential candidates. We therefore took advantage of the fact that strains containing mutations in the *ade6* gene produce red-colored colonies on medium containing limiting quantities of adenine (Gutz 1971), whereas *ade6*$^+$ strains form white colonies on the same medium. As shown in Figure 1, strain WS129 was transformed with *ade6* DNA containing either a 15- or 30-bp random nucleotide sequence substitution. Transformed cells were selected by loss of the *ura4*$^+$ gene inserted into *ade6* (resistance to 5-FOA) and homologous gene replacement was later confirmed by the simultaneous loss of the *kanMX6* gene (G418 sensitivity; Storici *et al.* 2001). After transformed cells formed colonies on 5-FOA medium, they were replica plated to sporulation medium, producing colonies containing mostly spores. Following treatment with acetone vapor to kill remaining unsporulated cells (Egel 1977), these colonies were replica plated to growth medium containing limiting quantities of adenine. (See materials and methods for full details of the screen.) Since the cells carry a plasmid containing a fragment of the *ade6* gene, recombination between

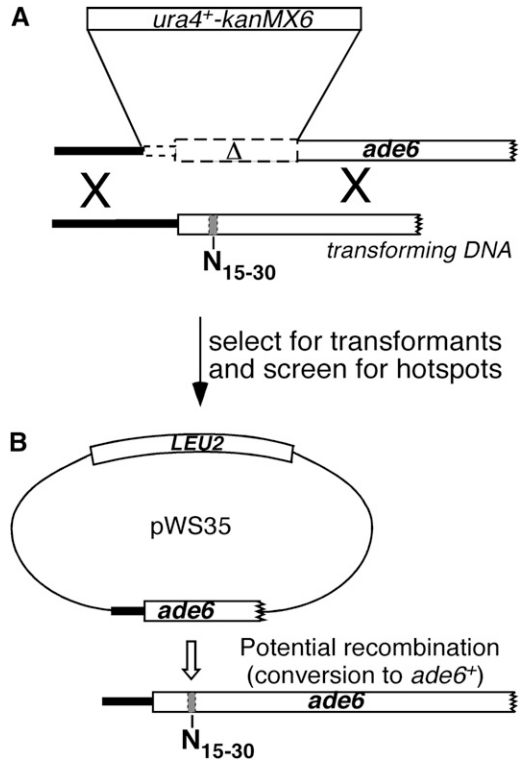


Figure 1.—Method for screening a large number of unique 15- or 30-bp sequences for hotspot activity. (A) A strain (WS129) with a partial *ade6* deletion and *ura4*$^+$*-kanMX6* insertion and plasmid pWS35 (not shown) is transformed with linear *ade6* DNA containing a 15- or 30-bp random sequence substitution (shaded bar). (B) Homologous recombinants lose the *ura4*$^+$*-kanMX6* marker and now share homology to a fragment of *ade6* carried on the plasmid pWS35. After transfer to sporulation medium, these cells will self-mate and form spores. Strains containing a hotspot in the random sequence region will recombine with the plasmid at high frequency (open arrow) to form *ade6*$^+$ spores, which can be identified as white papillae on red colonies after replica plating to the appropriate medium (Figure 2). Although this figure implies noncrossover conversion to *ade6*$^+$, crossover recombinants would also likely produce *ade6*$^+$ spores, because the plasmid carries 164 bp of the *ade6* promoter, which is adequate for *ade6* expression (Zahn-Zabal *et al.* 1995). *ade6* genes and fragments are drawn to scale; other genes and constructs are not. Open rectangles represent open reading frames.

chromosome and plasmid can produce *ade6*$^+$ spores during meiosis. Plasmid × chromosome recombination has previously been shown to accurately reflect chromosome × chromosome recombination in *S. pombe* (Ponticelli and Smith 1989). These recombination events can be seen at this stage as white papillae in otherwise red colonies. Strains with no recombination hotspot in *ade6*, for example *ade6-M375* (Gutz 1971), produce red colonies with few white papillae, whereas strains containing a recombination hotspot, such as *ade6-M26* or *ade6-3074* (Steiner and Smith 2005b), produce red colonies with a large number of papillae (Figure 2). Those colonies are easily distinguished from neighboring colonies, most of which resembled *ade6-M375* in our screen.

We screened ~27,600 colonies each containing a 30-bp random sequence and 18,300 colonies each containing a 15-bp random sequence. Among the 30-bp and 15-bp random sequences, we identified 393 and 102 strains, respectively, showing recombination frequencies obviously higher than *ade6-M375*. This number includes only strains capable of plasmid loss (see MATERIALS AND METHODS), since plasmid integration, particularly if it occurs within the *ade6* gene itself, might create a recombination hotspot unrelated to any specific sequence (VIRGIN *et al.* 1995). The higher frequency of hotspots among the 30-bp random sequences (1.4%) than the 15-bp sequences (0.6%) is consistent with the expectation that longer sequences have a greater chance of containing any given sequence motif. The complete list of hotspot-containing sequences and their relative activities is shown in supporting information, Table S1.

To confirm that the hyper-rec phenotype observed in our transformed strains was due to the sequence substitution in *ade6*, we crossed plasmid-free derivatives of the hyper-rec strains with the parent strain WS129. Spores from the cross that were 5-FOA resistant were tested for hotspot activity (papillation) by repeating the steps of the screen described above. In 36 randomly chosen strains, the hyper-rec phenotype showed complete linkage to *ade6* among the 50–100 spores tested from each cross. That is, every spore containing the sequence substitution from each of the 36 crosses showed the hyper-rec phenotype. This result is strong evidence that the observed hotspot activity in those strains, and probably all of our hotspot strains, is almost certainly due to the sequence substitution within *ade6* and not some other cause.

**Common sequence motifs appear frequently:** Since it is unlikely that the entire 15- or 30-bp sequence substitution is required for hotspot activity in any of our hotspot-containing strains, we analyzed those sequences for common 6- to 10-bp motifs as described in MATERIALS AND METHODS. The most common motifs found contained the 6-bp sequence TGACGT, referred to here as the *CRE*-core sequence (Table S2). This result demonstrates that our screen had, in fact, identified sequence-dependent recombination hotspots, as expected. However, the high frequency of the *CRE* motif was surprising since there was no prior reason to expect that the *CRE* hotspot would appear more frequently than other potential hotspot motifs. A likely interpretation of this result is that the *CRE* hotspot has the shortest sequence capable of creating a recombination hotspot, <7 bp in the region of *ade6* utilized in our screen (FOX *et al.* 2000). Hotspots requiring a longer sequence than this would occur at lower frequency.

Since the goal of our screen was to identify novel hotspot motifs, we eliminated 97 sequences from our pool that contained the 6-bp *CRE*-core sequence, TGACGT, which is required for all known Atf1-Pcr1-dependent hotspots in *S. pombe* (FOX *et al.* 2000; STEINER and SMITH 2005a). These strains accounted for almost 20% of our total sequence pool. We analyzed the remaining 398 sequences for common motifs as described above (Table 2 and Table S3). To determine the likelihood that any of the observed high-frequency motifs formed hotspots, we compared our results to an analysis of the same set of sequences following sequence randomization. Among random sequences, multiple occurrences of any motif would result from chance and not from any property of the motif, *e.g.*, hotspot activity. For any given motif size, we found a significantly higher number of motifs among the actual sequences, *i.e.*, those containing hotspots, compared to randomized sequences. For example, we found 17 different eight-base motifs that occurred four or more times among our pool of 398 sequences. Among random sequences, there were only four eight-base motifs that occurred at the same frequency and none that occurred more frequently (Table S3b). This result suggests that the majority of those motifs are hotspots, or perhaps form part of a larger hotspot.

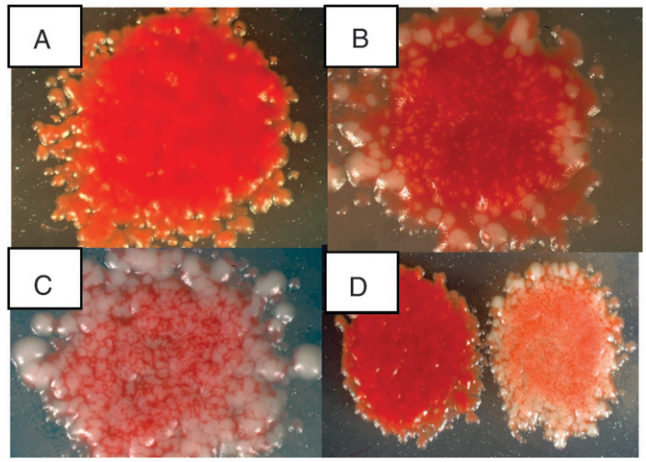


FIGURE 2.—Visual assay for hotspots. Strains containing known hotspot alleles (*ade6-M26* or *-3074*) and one control allele (*ade6-M375*) were put through the steps of the screen described in the text. (A) *ade6-M375*, (B) *ade6-M26*, (C) *ade6-3074*, (D) *ade6-M375* (left), and *ade6-3074* (right). A larger version of this figure is shown in Figure S2.

**Confirmation of hotspot activity by motif reconstruction:** To test whether any of the high-frequency sequence motifs shown in Table 2 were hotspots, we reconstructed some of them in the *ade6* gene by specific base-pair changes to the wild-type sequence, for example motifs 6-6, 8-1, 7-1, and 7-15. (The first number of each motif name refers to the length of the motif and the second number refers to the number in that series; Table 2 and Table S3). These experiments generated hotspots in some cases (*ade6-4002*, *-4003*, and *-4096*; Figure 3), but not in others (*ade6-4005*, *4006*; Table 1 and data not shown). We also tested two 10-base motifs that occurred three times each among our pool of

**TABLE 2**

**Common motifs among hotspot sequences lacking *CRE* hotspot**

| 10-base motif number and sequence | | Count[a] | 8-base motif number and sequence | | Count | 7-base motif number and sequence | | Count | 6-base motif number and sequence | | Count |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | GGATGTAAGT | 3 | 1 | TGACATCA | 6 | 1 | ACGTAAT | 14 | 1 | ACGTAA | 22 |
| 2 | GGTCTGGACC | 3 | 4 | CCAATGAG | 4 | 2 | ATGTCAC | 9 | 6 | ACATGA | 16 |
| | | | 5 | AGAGCTCT[b] | 4 | 4 | CCCCGCA | 8 | 9 | AAAGAT | 15 |
| | | | 6 | TCGGCCGA[b] | 4 | 7 | CCAATCA | 7 | 10 | GTATGA | 15 |
| | | | 7 | AGACGCAG | 4 | 8 | CCCACCC | 7 | 45 | CTATTA | 11 |
| | | | 8 | GTCTAGAC[b] | 4 | 15 | CGTCATA | 6 | 57 | CATCCC | 10 |
| | | | 9 | ATAATTGG | 4 | 31 | CCCCCAC | 5 | 21[c] | GATGAC | 18 |
| | | | 10 | AACAGGCG | 4 | | | | | | |
| | | | 11 | ATTGGCGG | 4 | | | | | | |
| | | | 12 | AAGCATGA | 4 | | | | | | |
| | | | 13 | CGCAGTAA | 4 | | | | | | |
| | | | 21 | AATGGATA | 3 | | | | | | |
| | | | 22 | CCATTACG | 3 | | | | | | |
| | | | 41[c] | AGGGATGA | 4 | | | | | | |

Some of the most common 6- to 10-base motifs among 398 sequences lacking the *CRE* core sequence, TGACGT. Only two motifs longer than 8-bp that occurred more than twice were found, 10-1 and 10-2. Only sequences tested for hotspot activity (Figure 3 and Table S4) are shown. Motif numbers correspond to those shown in Table S3a.

[a] Number of times that a given motif is found among the 398 sequences lacking the *CRE*-core sequence, TGACGT.

[b] A palindrome. Palindromes are counted twice, because they occur on both strands.

[c] These motifs are found only among our total pool of sequences and are shown in Table S2a.

sequences (*ade6-4094* and *-4095*; Figure 3), one of which, *ade6-4095*, produced a hotspot. A hotspot in this case is considered to be any allele that produces a significantly greater frequency of recombinants than *ade6-M375* ($P < 0.01$; Student's *t*-test), a common control allele for the *ade6-M26* hotspot (Gutz 1971).

Motifs that occurred at high frequency but failed to produce hotspots in our reconstruction experiments (*ade6-4005*, *-4006*, and *-4094*) could be explained by either of two possibilities: (1) those motifs are simply not hotspots, that is, they occurred by chance (for example, see Table S3b), or (2) those motifs form only part of a larger sequence necessary for hotspot activity. Reasoning that the second possibility may be true for many motifs, we streamlined our tests for hotspot activity by first testing most of the motifs shown in Table 2 in a "subscreen." Each subscreen involved repeating the original screen with a given motif flanked by several random nucleotides on one or both sides of the motif (Table S4). Transformants producing hotspots were sequenced and aligned to identify potential consensus sequences (Figure S1). Thus, many of the reconstructed motifs shown in Figure 3 represent consensus sequences identified in those subscreen experiments. Five of the six consensus sequences tested produced hotspots with activity significantly greater than *ade6-M375* ($P < 0.01$, Student's *t*-test; consensus sequences for motifs 6-1, 6-21, 7-2, 7-4, 7-31; Figure 3), suggesting that the consensus sequence is sufficient for hotspot activity, at least within this narrow region of the genome. However, we cannot infer that the entire consensus sequence is necessary for activity, which can be determined only by systematic mutagenesis.

For some hotspots, we observed an additional level of complexity. For example, the consensus sequence for motif 7-4 found by subscreen was DACCCCGCACD (Figure S1; D = A, G, or T). When the consensus sequence was reconstructed at its original location (bp 131–141), it produced a strong hotspot (*ade6-4099*, Figure 3). However, when moved only 36 bp away, the same motif produced less than one-tenth as many $Ade^+$ recombinants (*ade6-4072*). This reduced activity suggests either (1) that additional nucleotides outside of the consensus sequence are required for full activity or (2) that there is a position-dependent effect on hotspot activity. Since each hotspot sequence found by subscreen of motif 7-4 was 13 bp long (Figure S1), we tested whether one of these slightly longer sequences would create a more active hotspot at the new location. One of those motifs (7-4-11, Figure S1) is identical to the relatively weak *ade6-4072* hotspot over its 13-bp length except that the last base is a T rather than a C. Making this single base substitution more than doubled activity of the hotspot at that position (*ade6-4072 vs. -4101*, Figure 3). However, hotspot activity of the *ade6-4101* allele still remains significantly lower than the identical 13-bp sequence at its original position (*ade6-4103*). Thus, neither a position-effect nor an effect of more distant nucleotides can be excluded.

Similar complexity was observed for the 7-31 motif. In that case, the 9-bp consensus sequence, RCCCCCACA, was reconstructed at approximately the same position as

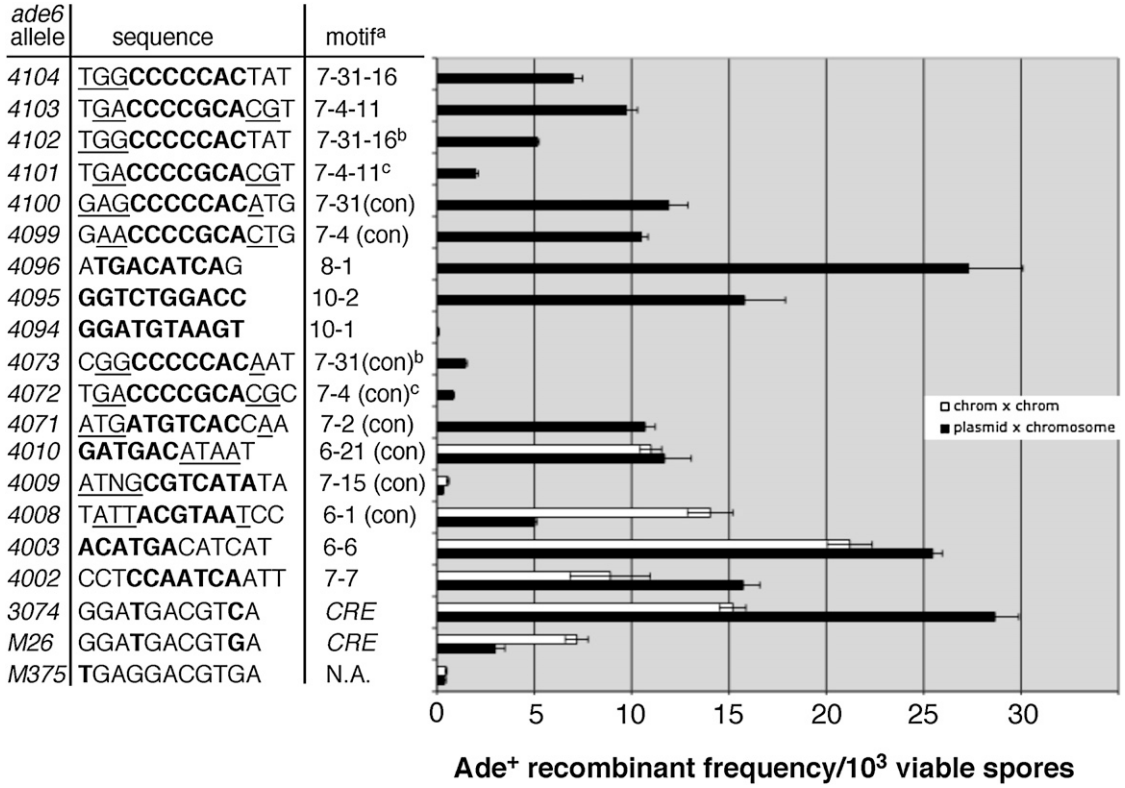


Figure 3.—Reconstructed sequence motifs produce hotspots. High-frequency sequence motifs (Table 2 and Figure S1) were reconstructed in the *ade6* gene and tested for hotspot activity in homothallic crosses (plasmid pWS35 × chromosome; solid bars) or heterothallic crosses (chromosome × chromosome; open bars). Heterothallic strains were crossed with WS315 (*ade6-469*; Table 1). Each bar represents the average of at least three crosses ±1 SEM. Sequences show the original motif (Table 2; boldface type) with some flanking nucleotides (regular type). Consensus sequence nucleotides (Figure S1) are underlined when applicable. In this table, only *ade6-4009* and *ade6-4094* are not significantly more active than *ade6-M375* ($P > 0.01$, Student's *t*-test). Notes: [a]The motif (Table 2) on which the allele is based. Con, consensus sequence. NA, not applicable. Motifs with three numbers indicate reconstruction of a particular allele from a subscreen (Figure S1), *e.g.*, 7-31-16 = sequence number 16 from subscreen of motif 7-31. [b]These motifs are inverted relative to *ade6-4100* and *-4104*. [c]These motifs are located 36 bp downstream relative to *ade6-4099* and *-4103*.

in the subscreen, but in the inverted orientation. Activity of that hotspot (*ade6-4073*) was significantly lower than the same sequence in the forward orientation (*ade6-4100*). However, two additional base changes (*ade6-4102*) more than tripled the number of Ade$^+$ recombinants. The *ade6-4102* and *-4104* alleles have the same 13-bp sequence as motif 7-31-16 found in a subscreen (Figure S1), but are inverted relative to each other. Both alleles also include the additional K (G or T) consensus nucleotide at the first position (Figure S1), which may increase their activities relative to the weaker *ade6-4073* hotspot. The similar activities of the *ade6-4102* and *-4104* alleles suggests that the nucleotides flanking the hotspot motif have a greater influence on hotspot activity than the motif orientation *per se*.

**Potential hotspot families:** Upon close inspection of some of the sequences in Figure 3, it became apparent that several of them looked quite similar to the previously characterized *CRE* hotspot. The most active form of this hotspot is a 10-bp palindrome, ATGACGTCAT (Steiner and Smith 2005b). Figure 3 shows that seven other hotspot motifs differ from the 10-bp *CRE* palindrome at only one or two positions. Given the previously demonstrated flexibility of the *CRE* hotspot sequence (Fox *et al.* 2000), it is conceivable that these related sequences could also be Atf1-Pcr1-dependent hotspots.

Since several of the hotspots we identified could be grouped into one family (*CRE*) on the basis of sequence, we also compared other hotspots (Figure 3 or consensus sequence hotspots from Figure S1) to see whether similar groupings were possible. Five motifs were grouped on the basis of their common CCAAT sequence. We speculated that this group of hotspots could be recognized by the CCAAT-binding factor. The CCAAT-binding factor is an evolutionarily conserved heteromeric transcription factor encoded by the *php2*, *php3*, and *php5* genes and is required for expression of many eukaryotic genes (McNabb *et al.* 1997). Hotspot activity of one member of this group, *ade6-4002*, was reduced significantly by deletion of each gene encoding a subunit of the CCAAT-binding factor, while *ade6-M26* activity was largely unaffected (Figure 4). In fact, the very low level of recombination observed in these experiments suggests that the *ade6-4002* allele is completely dependent on the CCAAT-binding factor for hotspot activity, much as *ade6-M26* hotspot activity is

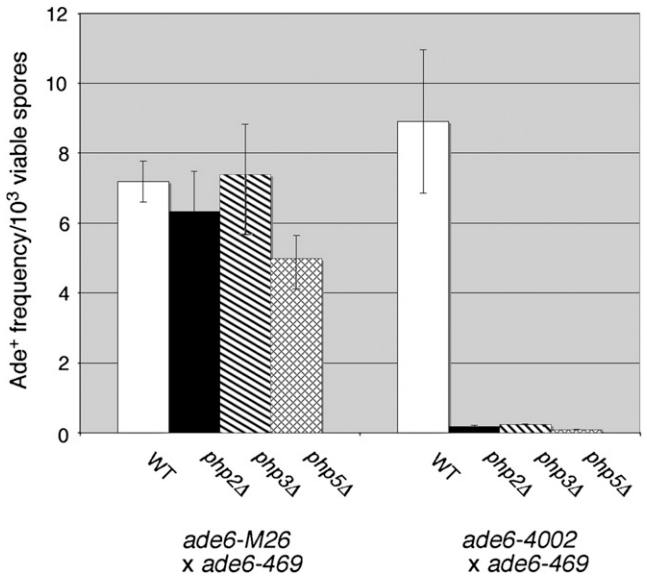


FIGURE 4.—The CCAAT-binding factor is required specifically for activity of the *ade6-4002* hotspot, but not *ade6-M26*. Crosses were performed between heterothallic strains containing the indicated *ade6* alleles and mutations in the *php2*, *php3*, or *php5* genes, which encode subunits of the CCAAT-binding factor. Bars represent the average Ade$^+$ recombinant frequencies ± SEM from a minimum of three experiments for each cross.

completely dependent on the Atf1-Pcr1 transcription factor (KON *et al.* 1997).

A third potential family of hotspots is based on the set of alleles derived from the 7-4 and 7-31 motifs, referred to here as the oligo-C family of hotspots. A search of the TransFac database (http://www.gene-regulation.com/pub/databases.html) suggests that these sequences may be targets for the MIG1 DNA-binding factor in *S. cerevisiae*, which has several orthologs in *S. pombe* including Scr1, Rsv1, Hsr1, and Rst2. Yet a fourth family contains two closely related sequences, *ade6-4095* and the consensus sequence of motif 8-8. The last potential family has only a single representative, motif 8-6, which shows no obvious similarity to any other identified motif (Figure 5). The very high frequency of hotspots observed in a subscreen of this motif (61%; Table S4), suggests that virtually all of the essential bases are contained within the eight-base motif tested.

**CRE-group**

| | |
|---|---|
| ade6-3083 | ATGACGTCAT |
| ade6-4003 | ATGACATCAT |
| ade6-4008 | ATTACGTAAT |
| ade6-4010 | ATGACATAAT |
| ade6-4071 | ATGATGTCAC |
| ade6-4096 | ATGACATCAG |
| Motif 7-15 | ATGACGCNAT (Comp) |
| Motif 8-12 | AAGCATGACGT |
| Motif 8-13 | ATNNCGCAGT |
| Motif 8-22 | ATTACGTAAG |
| Motif 8-41 | AGGGATGACGT |

**CCAAT-group**

| | |
|---|---|
| ade6-4002 | CCTCCAATCA |
| Motif 8-11 | CCGCCAATCA |
| Motif 8-4 | RRCCAATGAGRGG |
| Motif 8-9 | RACCAATTATAT (Comp) |
| Motif 8-21 | RCCAATGGATA G |

**oligo-C-group**

| | |
|---|---|
| Motif 7-4 | WACCCCGCACD |
| Motif 7-31 | KNRCCCCCACA |
| ade6-4099 | GAACCCCGCACTG |
| ade6-4100 | GAGCCCCCACATG |
| ade6-4101 | TGACCCCGCACGT |
| ade6-4102 | TGGCCCCCACTAT |
| ade6-4103 | TGACCCCGCACGT |
| ade6-4104 | TGGCCCCCACTAT |

**ade6-4095 group**

| | |
|---|---|
| ade6-4095 | GGTCTGGACC |
| motif 8-8 | GGTCTAGAC |

**Motif 8-6** W TCGGCCGA

FIGURE 5.—Potential hotspot families. Hotspot motifs (Figure 3) and hotspot consensus sequences (Figure S1) are aligned to show similarities. In the *CRE* family of hotspots, bases differing from the 10-bp *CRE* palindrome (*ade6-3083*, STEINER and SMITH 2005b) are underlined. Comp, complement of consensus sequence (Figure S1).

## DISCUSSION

In this study, we devised a method to rapidly screen short nucleotide sequences for hotspot activity and produced a sizable library of 15 and 30 bp sequences containing recombination hotspots. Within those sequences, we identified many shorter motifs ≥6 bp in length that occurred multiple times and hence may form all or part of recombination hotspot sequences. On the basis of our results, we can conclude that the previously characterized *CRE* hotspot is clearly not the only sequence motif capable of creating a recombination hotspot. However, it was surprising to us how frequently that previously identified hotspot occurred in our library of sequences. The 6-base sequence TGACGT (the *CRE* core sequence) common to all known Atf1-Pcr1-dependent hotspots (SCHUCHERT *et al.* 1991; FOX *et al.* 2000; STEINER and SMITH 2005b) was found 75 times, or in 15% of our total pool of hotspot sequences (Table S1 and Table S2a). This frequency increases to almost 20% if one includes sequences from our pool in which the *CRE* core is formed at the junction between random and nonrandom nucleotides. For example, if the first 4 bases of our

15- or 30-base random region reads GTCA, the 6-base sequence ACGTCA is formed at the junction, which is the complement of the *CRE* core sequence shown above (see sequence of oligonucleotides oWS209 and oWS241 in MATERIALS AND METHODS and Table S1). However, the frequency of *CRE*-like hotspots may be even greater than the observed frequency of the traditional *CRE*-core sequence, as several hotspots lacking that core still showed obvious sequence similarity to *CRE* (Figure 5).

Since our data indicate that other sequence motifs unrelated to *CRE* also create hotspots, it is possible that the overrepresentation of *CRE* is due to it having the shortest sequence necessary for observable hotspot activity, which may be fewer than 7 bp in some locations (FOX *et al.* 2000). Consistent with this, we have observed that the *ade6-4002* hotspot requires not only the CCAATCA sequence shown in Figure 3, but also three partially degenerate bases to the left (C. KALINOWSKI and W. STEINER, unpublished observation). The results from subscreens of other motifs (Table S4) also suggests that hotspots other than *CRE* may require more than 7 bp for activity. In those experiments, none of the seven or eight base motifs tested produced hotspots in 100% of transformed cells, suggesting that one or more specific nucleotides are required in the random regions flanking each motif.

**The number of different hotspot motifs:** What does the frequency of hotspots we observed say about the potential number of different hotspot motifs? We observed that ~0.6% of random 15mers produced an observable hotspot. In a random 15-bp sequence, there are as many as 9 unique 7-bp sequences when viewed in 7-bp windows moving in steps of 1 bp. Therefore, the probability of finding any unique 7-bp sequence in a random 15mer is $(0.25)^7 \times 9 = 0.055\%$. The observed frequency of hotspots was ~10-fold higher than this, suggesting that a minimum of 10 unique 7-bp motifs are required to account for the observed frequency of hotspots. By the same reasoning, ~50 8-bp motifs, or 225 9-bp motifs would be required. On the basis of the frequency of hotspots among random 30mers, the same calculations produce slightly lower estimates of the number of different hotspots. In either case, however, our data suggest that the number of different hotspot motifs is potentially large.

Without knowing the precise nucleotide sequence required for any given hotspot, it is not possible to determine precisely which of the sequences in our library can have their hotspot activity attributed to a particular motif, with the exception of many of the well-characterized *CRE*-like hotspots. However, even under generous assumptions about which motifs may create hotspots (for example, that all *CRE*-core sequences and all CCAAT sequences are hot; see highlighted motifs in Table S1), 197 sequences remain with no identified hotspot motif. Thus, additional hotspot motifs remain to be identified.

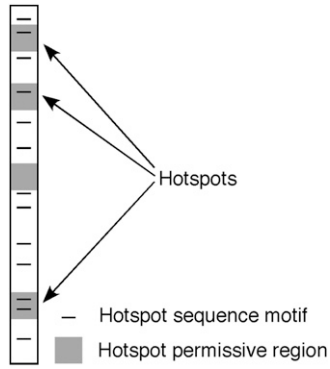


FIGURE 6.—Model to explain the location of recombination hotspots. The large rectangle indicates a portion of a chromosome. Shaded regions indicate regions of the genome that are permissive for recombination hotspots. In *S. pombe*, these regions coincide primarily with large intergenic regions (CROMIE *et al.* 2007). Solid lines indicate hotspot sequence motifs. Hotspots occur where these two chromosomal features coincide.

**The basis of recombination hotspots:** Meiotic recombination hotspots have been most thoroughly analyzed in the budding and fission yeasts *S. cerevisiae* and *S. pombe*, respectively. In both organisms, hotspots of recombination are sites of DSBs, which occur preferentially at a limited number of positions throughout their respective genomes. What determines the location of hotspots is incompletely understood, though some correlations have been made. For example, in *S. cerevisiae* DSBs occur predominantly in intergenic regions where the GC content modestly exceeds the genomic average GC content (BAUDAT and NICOLAS 1997; GERTON *et al.* 2000; MIECZKOWSKI *et al.* 2006). In *S. pombe* DSBs also occur predominantly in intergenic regions, particularly in large intergenic regions, but unlike *S. cerevisiae* the correlation between GC content and DSBs is very weak (CROMIE *et al.* 2007). In both yeasts, tested DSB hotspots also colocalize to sites of "open" chromatin, *i.e.*, sites that are sensitive to cleavage with S1 or micrococcal nucleases (WU and LICHTEN 1994; MIZUNO *et al.* 1997; HIROTA *et al.* 2007).

Beyond the aforementioned genomic features, predicting the precise location of recombination hotspots in most cases remains elusive. For example, the well-characterized *mbs1* hotspot of *S. pombe* occurs in an unusually large 7-kb intergenic region (YOUNG *et al.* 2002; CROMIE *et al.* 2005). Nevertheless, the breaks within that large region are distributed over a much narrower ~2-kb region. And within that smaller region, the breaks are further focused to four discrete clusters (CROMIE *et al.* 2005). What directs DSBs to these particular sites is unknown. However, a 9-bp *CRE* motif does predict the location of multiple DSB sites scattered across the *S. pombe* genome (STEINER and SMITH 2005a). This sequence is a binding site for the Atf1-Pcr1 transcription factor, which is required for its hotspot

activity (Kon *et al.* 1997). We propose that there may be many other unrelated short sequence motifs that are responsible for many, and potentially most or all, of the DSB hotspots in *S. pombe* and perhaps other organisms. However, such hotspot motifs alone are unlikely to be sufficient for hotspot activity at all sites in the genome, since DSBs are not observed at all *CRE* sites in the genome (Steiner and Smith 2005a). Thus, other factors, such as chromatin structure, also play a role in promoting or permitting hotspots at particular sites. In Figure 6, we propose a model that recombination hotspots are found at positions where these two factors, a hotspot sequence motif and permissive chromatin structure, intersect.

How do sequence motifs create recombination hotspots? One possibility is that the nucleotide sequence itself possesses some property that makes it unusually susceptible to cleavage during meiosis. For example, tandem repeats of a pentanucleotide sequence reported to inhibit nucleosome formation can create a recombination hotspot in *S. cerevisiae* (Kirkpatrick *et al.* 1999). It has also been reported that polypurine/polypyrimidine tracts (PPTs) of ≥12 bp are associated with hotspots in *S. cerevisiae* (Bagshaw *et al.* 2006). However, we found no extensive tandem repeats and only a handful of PPTs ≥12 bp among our hotspot sequences (Table S1). Further, a direct test of a 30-bp random PPT in our experimental system did not produce hotspots in any of eight independent transformants (W. Steiner, unpublished observation). Thus, we favor instead the model that most of the sequences in our hotspot library contain target sequences for DNA binding proteins, for example transcription factors, that promote DSBs when bound to DNA. This model is also consistent with current data. First, we are aware of only two previously reported examples of defined sequence motifs creating recombination hotspots, the *CRE* hotspot of *S. pombe* (Schuchert *et al.* 1991) and the Bas1 target sequence, TGACTC, of *S. cerevisiae* (Mieczkowski *et al.* 2006). Both of these motifs require the binding of a transcription factor for their hotspot activity (Kon *et al.* 1997; Mieczkowski *et al.* 2006). Second, it has been recently observed in *S. pombe* that hotspots of recombination show significant colocalization to sites expressing noncoding RNAs (Wahls *et al.* 2008), suggesting that these sites are bound by transcription factors.

How might transcription factors promote recombination when bound to their target sequence? Yamada *et al.* (2004) showed that the *ade6-M26* hotspot was dependent on both a histone acetyl transferase (Gcn5) and an ATP-dependent chromatin remodeling factor (Snf22). These researchers suggested that binding of the Atf1-Pcr1 transcription factor to the *M26* motif recruits these chromatin modifying enzymes, resulting in localized chromatin remodeling and making the site accessible to the DSB machinery. It would be interesting to determine whether other hotspot-associated transcription factors operate by a similar mechanism, and what such factors might have in common that results in the recruitment of chromatin modifying enzymes.

It is possible that simple sequence motifs produce hotspots in many different organisms. For example, several sequence motifs have been reported as potential hotspots in humans (Zhang *et al.* 2004; Myers *et al.* 2005, 2008). Since recombination hotspots disrupt gene linkages, they complicate efforts to find human disease genes by linkage analysis (Hey 2004; Nishant and Rao 2005). Thus, the ability to identify potential hotspots solely on the basis of sequence is of practical significance. It is possible that some hotspots we have identified in our analysis are also active in other organisms. Two of the motifs reported as potential human hotspots, CGCCCCCGC and CCCCACCCC, show strong similarity to motifs found in our screen, motifs 7-4 and 7-8, respectively (Table 2), at least one of which (7-4) was confirmed to be a hotspot (Figure 3 and Table S4). This result suggests the intriguing possibility that some of the motifs identified here may also be active in other organisms.

We thank Simon Labbé for the donation of strains. This work was supported by a National Institutes of Health grant, R15GM078065 (to W.W.S.). W.W.S. was supported by a Leukemia and Lymphoma Society special fellowship, 3230-05, during a portion of this work.

## LITERATURE CITED

Bagshaw, A. T., J. P. Pitt and N. J. Gemmel, 2006 Association of poly-purine/poly-pyrimidine sequences with meiotic recombination hotspots. BMC Genomics **7:** 179–188.

Bailey, T. L., N. Williams, C. Misleh and W. W. Li, 1996 MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res. **34:** W369–W373.

Baker, B. S., A. T. C. Carpenter, M. S. Esposito, R. E. Esposito and L. Sandler, 1976 The genetic control of meiosis. Annu. Rev. Genet. **10:** 53–134.

Baudat, F., and A. Nicolas, 1997 Clustering of meiotic double-strand breaks on yeast chromosome III. Proc. Natl. Acad. Sci. USA **94:** 5213–5218.

Bähler, J., J.-Q. Wu, M. S. Longtine, N. G. Shah, A. McKenzie, III *et al.*, 1998 Heterologous modules for efficient and versatile PCR-based gene targeting in *Schizosaccharomyces pombe*. Yeast **14:** 943–951.

Cao, L., E. Alani and N. Kleckner, 1990 A pathway for generation and processing of double-strand breaks during meiotic recombination in *S. cerevisiae*. Cell **61:** 1089–1101.

Cervantes, M. D., J. A. Farah and G. R. Smith, 2000 Meiotic DNA breaks associated with recombination in *S. pombe*. Mol. Cell **5:** 883–888.

Cottarel, G., D. Beach and U. Deuschle, 1993 Two new multipurpose multicopy *Schizosaccharomyces pombe* shuttle vectors, pSP1 and pSP2. Curr. Genet. **23:** 547–548.

Cromie, G. A., R. W. Hyppa, H. P. Cam, J. A. Farah, S. I. S. Grewal *et al.*, 2007 A discrete class of intergenic DNA dictates meiotic DNA break hotspots in fission yeast. PLoS Genet. **3:** 1496–1507.

Cromie, G. A., C. A. Rubio, R. W. Hyppa and G. R. Smith, 2005 A natural meiotic DNA break site in *Schizosaccharomyces pombe* is a hotspot of gene conversion, highly associated with crossing over. Genetics **169:** 595–605.

Egel, R., 1977 Selective spore survival during replica-plating of fission yeast. Arch. Microbiol. **112:** 109–110.

Esposito, M. S., and J. E. Wagstaff, 1981 Mechanisms of mitotic recombination, pp. 341–370 in *The Molecular Biology of the Yeast Saccharomyces*, edited by J. N. Strathern, E. W. Jones and J. R.

Broach. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Fan, Q., F. Xu and T. D. Petes, 1995 Meiosis-specific double-strand DNA breaks at the *HIS4* recombination hot spot in the yeast *Saccharomyces cerevisiae*: control in *cis* and *trans*. Mol. Cell. Biol. **15:** 1679–1688.

Fox, M. E., J. B. Virgin, J. Metzger and G. R. Smith, 1997 Position- and orientation-independent activity of the *Schizosaccharomyces pombe* meiotic recombination hot spot *M26*. Proc. Natl. Acad. Sci. USA **94:** 7446–7451.

Fox, M. F., T. Yamada, K. Ohta and G. R. Smith, 2000 A family of CRE-related DNA sequences with meiotic recombination hotspot activity in *Schizosaccharomyces pombe*. Genetics **156:** 59–68.

Gerton, J. L., J. DeRisi, R. Shroff, M. Lichten, P. O. Brown *et al.*, 2000 Global mapping of meiotic recombination hotspots and coldspots in the yeast *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **97:** 11383–11390.

Gutz, H., 1971 Site specific induction of gene conversion in *Schizosaccharomyces pombe*. Genetics **69:** 317–337.

Gutz, H., H. Heslot, U. Leupold and N. Loprieno, 1974 *Schizosaccharomyces pombe*, pp. 395–446 in *Handbook of Genetics*, edited by R. C. King. Plenum, New York.

Hey, J., 2004 What's so hot about recombination hotspots? PLoS Biol. **2:** 0730–0733.

Hirota, K., W.W. Steiner, T. Shibata and K. Ohta, 2007 Chromatin configuration at natural meiotic recombination hot spots in fission yeast. Eukaryot. Cell **6:** 2072–2080.

Keeney, S., C. N. Giroux and N. Kleckner, 1997 Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. Cell **88:** 375–384.

Kirkpatrick, D. T., Y. H. Wang, M. Dominska, J. D. Griffith and T. D. Petes, 1999 Control of meiotic recombination and gene expression in yeast by a simple repetitive DNA sequence that excludes nucleosomes. Mol. Cell. Biol. **19:** 7661–7671.

Kon, N., M. D. Krawchuk, B. G. Warren, G. R. Smith and W. P. Wahls, 1997 Transcription factor Mts1/Mts2 (Atf1/Pcr1, Gad7/Pcr1) activates the *M26* meiotic recombination hotspot in *Schizosaccharomyces pombe*. Proc. Natl. Acad. Sci. USA **94:** 13756–13770.

Mahadevaiah, S. K., J. M. A. Turner, F. Baudat, E. P. Rogakou, P. deBoer *et al.*, 2001 Recombinational DNA double strand breaks in mice precede synapsis. Nat. Genet. **27:** 271–276.

Malik, S.-B., M. A. Ramesh, A. M. Hulstrand and J. M. Logsdon Jr., 2007 Protist homologs of the meiotic *Spo11* gene and topoisomerase VI reveal an evolutionary history of gene duplication and lineage-specific loss. Mol. Biol. Evol. **24:** 2827–2841.

McNabb, D. S., K. A.-S. Tseng and L. Guarente, 1997 The *Saccharomyces cerevisiae* Hap5p homolog from fission yeast reveals two conserved domains that are essential for assembly of heterotetrameric CCAAT-binding factor. Mol. Cell. Biol. **17:** 7008–7018.

Mercier, A., B. Pelletier and S. Labbé, 2006 A transcription factor cascade involving Fep1 and the CCAAT-binding factor Php4 regulates gene expression in response to iron deficiency in the fission yeast *Schizosaccharomyces pombe*. Eukaryotic Cell **5:** 1866–1881.

Mieczkowski, P. A., M. Dominska, M. J. Buck, J. L. Gerton, J. D. Lieb *et al.*, 2006 A global analysis of the relationship between the binding of the Bas1p transcription factor and meiosis-specific double-strand DNA breaks in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **26:** 1014–1027.

Mizuno, K.-i., Y. Emura, M. Baur, J. Kohli, K. Ohta *et al.*, 1997 Remodeling of chromatin structure around a single nucleotide mutation in *ade6–M26* that creates meiotic recombination hotspot in fission yeast. Genes Dev. **11:** 876–886.

Myers, S., L. Bottolo, C. Freeman, G. McVean and P. Donnelly, 2005 A fine-scale map of recombination rates and hotspots across the human genome. Science **310:** 321–324.

Myers, S., C. Freeman, A. Auton, P. Donnelly and G. McVean, 2008 A common sequence motif associated with recombination hot spots and genome instability in humans. Nat. Genet. **40:** 1124–1129.

Nishant, K. T., and M. R. S. Rao, 2005 Molecular features of meiotic recombination hotspots. Bioessays **28:** 45–56.

Pearson, W., and D. Lipman, 1988 Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85:** 2444–2448.

Petes, T. D., 2001 Meiotic recombination hot spots and cold spots. Nat. Rev. Genet. **2:** 360–370.

Pâques, F., and J. E. Haber, 1999 Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. Microbiol. Mol. Biol. Rev. **63:** 349–404.

Ponticelli, A. S., and G. R. Smith, 1989 Meiotic recombination deficient mutants of *Schizosaccharomyces pombe*. Genetics **123:** 45–54.

Schuchert, P., M. Langsford, E. Käslin and J. Kohli, 1991 A specific DNA sequence is required for high frequency of recombination in the *ade6* gene of fission yeast. EMBO J. **10:** 2157–2163.

Sinha, S., and N. Tompa, 2002 Discovery of novel transcription factor binding sites by statistical overrepresentation. Nucleic Acids Res. **30:** 5549–5560.

Sinha, S., and N. Tompa, 2003 YMF: a program for discovery of novel transcription factor binding sites by statistical overrepresentation. Nucleic Acids Res. **31:** 3586–3588.

Steiner, W. W., R. W. Schreckhise and G. R. Smith, 2002 Meiotic DNA breaks at the *S. pombe* recombination hotspot *M26*. Mol. Cell **9:** 847–855.

Steiner, W. W., and G. R. Smith, 2005a Natural meiotic recombination hotspots in the *S. pombe* genome successfully predicted from the simple sequence motif *M26*. Mol. Cell. Biol. **25:** 9054–9062.

Steiner, W. W., and G. R. Smith, 2005b Optimizing the nucleotide sequence of a meiotic recombination hotspot in *Schizosaccharomyces pombe*. Genetics **169:** 1973–1983.

Storici, F., L. K. Lewis and M. A. Resnick, 2001 In vivo site-directed mutagenesis using oligonucleotides. Nat. Biotechnol. **19:** 773–776.

Sun, H., D. Treco, N. P. Schultes and J. W. Szostak, 1989 Double-strand breaks at an initiation site for meiotic gene conversion. Nature **338:** 87–90.

Szankasi, P., W. D. Heyer, P. Schuchert and J. Kohli, 1988 DNA sequence analysis of the *ade6* gene of *Schizosaccharomyces pombe*: wild-type and mutant alleles including the recombination hotspot allele *ade6–M26*. J. Mol. Biol. **204:** 917–925.

Vallejo, A. N., R. J. Pogulis and L. R. Pease, 1995 Mutagenesis by PCR, pp. 603–612 in *PCR Primer: A Laboratory Manual*, edited by C. W. Dieffenbach and G. S. Dveksler. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Virgin, J. B., J. Metzger and G. R. Smith, 1995 Active and inactive transplacement of the *M26* recombination hotspot in *Schizosaccharomyces pombe*. Genetics **141:** 33–48.

Wahls, W. P., E. R. Siegel and M. K. Davisdson, 2008 Meiotic recombination hotspots of fission yeast are directed to loci that express non-coding RNA. PLoS ONE **3:** e2887.

White, M. A., M. Dominska and T. D. Petes, 1993 Transcription factors are required for the meiotic recombination hotspot at the *HIS4* locus in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **90:** 6621–6625.

Wood, V., R. G. William, M.-A. Rajandream, M. Lyne, R. Lyne *et al.*, 2002 The genome sequence of the eukaryote fission yeast *Schizosaccharomyces pombe*. Nature **415:** 871–880.

Wu, T.-C., and M. Lichten, 1994 Meiosis-induced double-strand break sites determined by yeast chromatin structure. Science **263:** 515–518.

Yamada, T., K. Mizuno, K. Hirota, N. Kon, W. P. Wahls, *et al.*, 2004 Roles of histone acetylation and chromatin remodeling factor in a meiotic recombination hotspot. EMBO J. **23:** 1792–1803.

Young, J. A., R. W. Schreckhise, W. W. Steiner and G. R. Smith, 2002 Meiotic recombination remote from prominent DNA break sites in *S. pombe*. Mol. Cell **9:** 253–263.

Zahn-Zabal, M., E. Lehmann and J. Kohli, 1995 Hot spots of recombination in fission yeast: inactivation of the *M26* hot spot by deletion of the *ade6* promoter and the novel hotspot *ura4-aim*. Genetics **140:** 469–478.

Zhang, J., F. Li, J. Li, M. Q. Zhang and X. Zhang, 2004 Evidence and characteristics of putative human alpha recombination hotspots. Hum. Mol. Genet. **13:** 2823–2828.

Communicating editor: G. R. Smith

GENETICS

**Supporting Information**
http://www.genetics.org/cgi/content/full/genetics.109.101253/DC1

# Novel Nucleotide Sequence Motifs that Produce Hotspots of Meiotic Recombination in *Schizosaccharomyces pombe*

**Walter W. Steiner, Estelle M. Steiner, Angela R. Girvin and Lauren E. Plewik**


DOI: 10.1534/genetics.109.101253

| | **Motif 6-1** | Activity | |
|---|---|---|---|
| | HNNNHACGTAANNND | Test 1 | Test 2 |
| 1 | ACATTACGTAAACGT | +++ | +++ |
| 2 | ATGTTACGTAATAAT | +++ | ++ |
| 3 | TTATTACGTAATACT | +++ | +++ |
| 4 | AAACAACGTAAAATA | ++ | + |
| 5 | AGGCTACGTAATAAG | ++ | ++ |
| 6 | CCAATACGTAATACG | +++ | ++ |
| 7 | TTAATACGTAATACG | ++ | ++ |
| 8 | CGTATACGTAATAAA | ++ | ++ |
| 9 | TGATTACGTAAGAGA | +++ | +++ |
| 10 | AGTTTACGTAATGCA | +++ | ++ |
| 11 | AATTTACGTAATATG | + | + |
| 12 | AGGTTACGTAAGGGT | ++ | ++ |
| 13 | AATTAACGTAAAACA | ++ | + |
| 14 | TATATACGTAATAGG | +++ | ++ |
| 15 | AGTATACGTAATTAG | ++ | ++ |
| 16 | AGTATACGTAATAAG | ++ | +++ |
| 17 | AAATTACGTAAACCG | ++ | +++ |
| 18 | AAGATACGTAATGGG | ++ | + |
| 19 | CATTAACGTAACCCA | + | + |
| 20 | ATACCACGTAATCCG | + | ++ |
| 21 | TTAATACGTAATAGG | ++ | ++ |
| 22 | TAATAACGTAATATG | + | + |
| 23 | CGATAACGTAATAAT | + | + |
| 24 | CAATTACGTAAAAAT | ++ | ++ |
| 25 | CAAATACGTAATACT | ++ | ++ |
| 26 | CAAATACGTAATCGG | ++ | ++ |
| 27 | AGAGCACGTAATAGG | ++ | ++ |
| 28 | CAATTACGTAAATTT | ++ | ++ |
| 29 | ATCCCACGTAATATG | + | + |
| 30 | TTCTTACGTAATTAG | ++ | ++ |
| 31 | TGTGTACGTAATAAA | + | + |
| 32 | CTATTACGTAAAAAA | ++ | +++ |
| 33 | CGAATACGTAATGGT | ++ | ++ |
| 34 | AATTTACGTAATACT | +++ | +++ |
| 35 | ATATTACGTAATGAA | +++ | +++ |
| 36 | TCGTTACGTAATTGT | ++ | ++ |
| 37 | ACATAACGTAATACA | ++ | ++ |
| 38 | CCATTACGTAACAAA | +++ | +++ |
| 39 | ATACCACGTAATAAT | + | + |
| 40 | TAATTACGTAATACT | +++ | ++++ |
| 41 | ACAACACGTAATTGG | ++ | + |
| 42 | ATGTTACGTAAAATA | ++ | + |
| 43 | TCGTTACGTAACTTA | ++ | ++ |
| 44 | AGACAACGTAACACA | ++ | + |
| 45 | ACGTTACGTAACGGG | ++ | + |
| 46 | AGATTACGTAAAGAA | ++ | ++ |
| 47 | AGAACACGTAATGGT | + | + |
| 48 | CTACTACGTAAGAAT | + | + |
| 49 | CGGATACGTAATAGA | ++ | + |
| 50 | AAATTACGTAAAGGG | ++ | ++ |
| 51 | CCGTAACGTAATAGA | + | + |
| 52 | AGTACACGTAATGCA | + | + |
| CONSENSUS | ATTACGTAATA | CRE-like | |

**motif 6-21**

| | HNNNNGATGACHNND | Activity Test 1 | Test 2 |
|---|---|---|---|
| 1 | AAGTCGATGACATAA | ++ | ++ |
| 2 | CATTGGATGACAACT | +++ | +++ |
| 3 | AAGGCGATGACATAA | + | ++ |
| 4 | AGGTGGATGACATAA | ++ | ++ |
| 5 | ACACAGATGACATAA | ++ | + |
| 6 | ACCCAGATGACATAA | + | + |
| 7 | ACCGAGATGACATAA | ++ | + |
| 8 | CAATGGATGACAGCG | | +++ |
| 9 | ATAACGATGACTAAG | ++ | + |
| 10 | CTGGGGATGACATAA | ++ | + |
| 11 | CGAGAGATGACATAA | +++ | ++ |
| 12 | AATAAGATGACATAA | ++ | + |
| 13 | CAATAGATGACCGCG | ++++ | +++ |
| 14 | AGATTGATGACATAA | ++ | ++ |
| 15 | TAAAGATGACATAA | ++ | + |
| 16 | TCCGAGATGACATAA | ++ | ++ |
| 17 | ATTATGATGACGAGG | ++ | ++ |
| 18 | TATTGGATGACTGAA | +++ | +++ |
| 19 | CACATGATGACATAA | ++ | ++ |
| 20 | AAACGGATGACATAA | ++ | ++ |
| 21 | AGGGTGATGACATAT | ++ | ++ |
| 22 | CAATAGATGACAGCAT | + | + |
| 23 | TTACAGATGACATAA | + | + |
| 24 | CATTGGATGACACTT | +++ | ++ |
| 25 | CGCGGGATGACATAA | +++ | ++ |
| 26 | CAATGGATGACCCCG | +++ | ++ |
| 27 | TGGAGGATGACATAA | ++ | ++ |
| 28 | ATTATGATGACATGG | ++ | ++ |
| 29 | TCAGGGATGACCTCA | + | + |
| 30 | CAATCGATGACCGCT | ++ | ++ |
| 31 | CATTGGATGACCCAT | ++ | ++ |
| 32 | AGGGTGATGACATAA | +++ | ++ |
| CONSENSUS | GATGACATAA | CRE-like | |

**Motif 6-45**

| | HNNNNCTATTANNND | Activity Test 1 | Test 2 |
|---|---|---|---|
| 1 | CCCGCCTATTAGCCA | ++ | +++ |
| 2 | AAGGACTATTACGTA | ++ | + |
| 3 | TCGGACTATTACGTA | ++ | ++ |
| 4 | CCCGCCTATTATTGG | ++ | ++ |
| 5 | CCCCGCTATTAATGG | ++ | ++ |
| 6 | TCACACTATTACGTA | + | ++ |
| 7 | TAGACCTATTACGTA | ++ | ++ |
| 8 | AGTTGCTATTACGTA | ++ | ++ |
| 9 | CCCGCCTATTAGTAG | ++ | ++ |

**Motif 7-2**

| | NNNATGTCACNNN | Activity Test 1 | Test 2 |
|---|---|---|---|
| 1 | GTGATGTCACTAG | ++ | ++ |
| 2 | ATGATGTCACAAG | +++ | ++++ |
| 3 | ATGATGTCACCCG | ++++ | ++++ |
| 4 | ATGATGTCACACA | ++++ | ++++ |
| 5 | ATGATGTCACAAC | ++++ | ++++ |
| 6 | GTGATGTCACAGG | +++ | +++ |
| 7 | GTGATGTCACTTC | +++ | ++ |
| 8 | ATGATGTCACGAG | ++++ | ++++ |
| 9 | ATGATGTCACTAT | ++++ | ++++ |
| 10 | ATGATGTCACGCG | ++ | ++ |
| 11 | ATGATGTCACTAT | +++ | ++++ |
| 12 | CTGATGTCACGAC | + | ++ |
| 13 | ATGATGTCACGAA | +++ | ++++ |
| 14 | ATGATGTCACTGT | +++ | ++++ |
| 15 | GTGATGTCACTAG | ++ | + |
| 16 | ATGATGTCACTTA | ++++ | ++++ |
| 17 | ATGATGTCACTAG | ++++ | ++++ |
| 18 | ATGATGTCACAAG | +++ | ++++ |
| 19 | ATGATGTCACGTT | +++ | ++++ |
| CONSENSUS | ATGATGTCACNA | CRE-like | |

**Motif 7-4**

| | NNNCCCCGCANNN | Activity Test 1 | Test 2 |
|---|---|---|---|
| 1 | CTTCCCCGCACAG | + | +++ |
| 2 | TTACCCCGCAATG | ++ | ++++ |
| 3 | GTACCCCGCACAG | ++ | ++++ |
| 4 | TAACCCCGCACGG | ++ | ++++ |
| 5 | ATACCCCGCACTT | ++ | ++++ |
| 6 | GAACCCCGCATGT | + | ++ |
| 7 | TTACCCCGCAGTG | ++ | ++++ |
| 8 | TGACCCCGCACTT | ++++ | ++++ |
| 9 | AGACCCCGCACGG | ++++ | ++++ |
| 10 | GAGCCCCGCACGA | ++++ | ++++ |
| 11 | TGACCCCGCACGT | ++++ | ++++ |
| 12 | ATTCCCCGCACGG | +++ | +++ |
| 13 | GAGCCCCGCACAC | +++ | +++ |
| 14 | AAGCCCCGCACGG | +++ | +++ |
| 15 | CAACCCCGCACTG | +++ | +++ |
| 16 | ATTCCCCGCACTA | +++ | ++ |
| 17 | TTACCCCGCAGGA | +++ | ++ |
| 18 | TTACCCC-CACGT | +++ | +++ |
| 19 | GTACCCCGCACGA | ++++ | ++++ |
| 20 | GAACCCCGCACAA | ++++ | +++ |
| 21 | TAACCCCGCAATT | +++ | ++ |
| 22 | TAACCCCGCACGG | ++++ | ++++ |
| 23 | GTACCCCGCACGG | ++++ | ++++ |
| 24 | GTACCCCGCATAT | ++++ | +++ |
| CONSENSUS | DACCCCGCACD | oligo-C like | |

**motif 7-15**

| | NNNBCGTCATANNND | Activity Test 1 | Test 2 |
|---|---|---|---|
| 1 | GTTGCGTCATACGGG | +++ | +++ |
| 2 | ATAGCGTCATAGAAA | ++ | ++ |
| 3 | AGCGCGTCATACTTT | + | + |
| 4 | ATATCGTCATAGGTG | ++ | +++ |
| 5 | ATCGCGTCATAAAAG | ++ | +++ |
| 6 | TTGGCGTCATAACGA | ++ | + |
| 7 | AACTCGTCATAGCAG | ++ | ++ |
| 8 | ACTGCGTCATAGAGA | +++ | +++ |
| 9 | AGTGCGTCATAGAGA | ++ | ++ |
| 10 | TTTGCGTCATAGCAG | ++ | +++ |
| 11 | CTGTCGTCATATGAG | ++ | + |
| 12 | ATCGCGTCATAAATT | ++ | ++ |
| 13 | AATTCGTCATAGGGA | +++ | +++ |
| 14 | ATGTCGTCATAGAGG | +++ | +++ |
| CONSENSUS | ATNGCGTCATA | CRE-like | |

**Motif 7-31**

| | NNNCCCCCACNNN | Activity Test 1 | Test 2 |
|---|---|---|---|
| 1 | GGGCCCCCACTCT | ++ | +++ |
| 2 | GGGCCCCCACATC | ++ | +++ |
| 3 | TCGCCCCCACATG | +++ | +++ |
| 4 | TTACCCCCACAGG | ++ | ++ |
| 5 | ACGCCCCCACACG | ++ | +++ |
| 6 | GTACCCCCACGTA | ++ | +++ |
| 7 | TAGCCCCCACTAA | ++ | ++ |
| 8 | GAACCCCCACTCA | ++ | +++ |
| 9 | GAACCCCCACATA | ++ | ++ |
| 10 | GATCCCCCACATG | ++ | ++ |
| 11 | TGACCCCCACAAT | ++ | ++ |
| 12 | ACGCCCCCACATA | +++ | +++ |
| 13 | TAACCCCCACAGA | ++ | ++ |
| 14 | TTTCCCCCACATA | ++ | + |
| 15 | TGGCCCCCACACC | +++ | +++ |
| 16 | TGGCCCCCACTAT | +++ | +++ |
| 17 | TTACCCCCACGTG | +++ | +++ |
| 18 | CTACCCCCACAGG | ++ | ++ |
| 19 | AAGCCCCCACATA | +++ | +++ |
| 20 | TCGCCCCCACTAG | ++ | ++ |
| 21 | GTGCCCCCACAAG | +++ | +++ |
| 22 | GAGCCCCCACATA | +++ | +++ |
| 23 | GTGCCCCCACGTA | +++ | +++ |
| 24 | GAGCCCCCACACA | +++ | +++ |
| 25 | GAGCCCCCACTAA | ++ | ++ |
| 26 | TGGCCCCCACACT | ++ | +++ |
| 27 | TTGCCCCCACTCC | ++ | +++ |
| 28 | GTACCCCCACAAG | ++ | +++ |
| 29 | TGGCCCCCACGCG | ++ | +++ |
| 30 | GTGCCCCCACAAC | ++ | ++ |
| 31 | ACACCCCCACTTA | ++ | ++ |
| 32 | AAACCCCCACAAA | ++ | + |
| 33 | GGGCCCCCACGTC | +++ | ++ |
| 34 | TAACCCCCACAGG | ++ | ++ |
| Consnesus | KNRCCCCCACA | oligo-C like | |

**Motif 8-4**

| | NNNCCAATGAGNNN | Activity Test 1 | Test 2 |
|---|---|---|---|
| 1 | TAACCAATGAGGAG | + | + |
| 2 | CAGCCAATGAGCAC | ++ | ++ |
| 3 | GGGCCAATGAGGGG | + | + |
| 4 | CGACCAATGAGACA | + | + |
| 6 | CGGCCAATGAGAGA | + | + |
| 8 | TGACCAATGAGGGG | + | + |
| 9 | CGGCCAATGAGGGG | + | + |
| 10 | CAACCAATGAGAGA | ++ | ++ |
| 11 | TAACCAATGAGGGG | ++ | + |
| 12 | CGACCAATGAGCAC | ++ | + |
| 13 | AGCCAATGAGTGG | ++ | ++ |
| 14 | TAGCCAATGAGTGG | + | + |
| 15 | GAGCCAATGAGGGC | + | + |
| 16 | GGACCAATGAGAGC | ++ | ++ |
| 17 | GGGCCAATGAGGGG | + | + |
| 18 | TGCCCAATGAGGGG | + | + |
| 19 | CAGCCAATGAGAAA | ++ | ++ |
| 20 | TAGCCAATGAGACG | + | + |
| 21 | TGTCCAATGAGAGG | + | + |
| 22 | CGGCCAATGAGGGG | + | + |
| 23 | CAGCCAATGAGTGC | + | + |
| 27 | CGGCCAATGAGGAT | + | + |
| 28 | GGGCCAATGAGAGG | ++ | + |
| 29 | TGGCCAATGAGAAG | + | + |
| 30 | TAACCAATGAGAAC | + | + |
| 31 | GGGCCAATGAGGGG | ++ | ++ |
| 32 | CAGCCAATGAGACC | + | + |
| 33 | TGTCCAATGAGGGG | + | + |
| 35 | GCACCAATGAGAGA | + | + |
| 36 | CGGCCAATGAGGGC | + | + |
| 37 | TGGCCAATGAGAGA | + | + |
| 38 | ACACCAATGAGGGG | ++ | ++ |
| 39 | GGGCCAATGAGGAG | + | + |
| CONSENSUS | RRCCAATGAGRGG | CCAAT-like | |

**Motif 8-6**

| | NNNTCGGCCGANNN | Activity Test 1 | Test 2 |
|---|---|---|---|
| 1 | AGTTCGGCCGAGGG | +++ | ++ |
| 2 | CGGTCGGCCGATCG | ++ | + |
| 3 | TAGTCGGCCGAAAA | ++ | + |
| 4 | TTCTCGGCCGACAC | ++ | ++ |
| 5 | AAGTCGGCCGAAAA | ++ | ++ |
| 6 | ATCTCGGCCGAGTA | ++ | ++ |
| 7 | TATTCGGCCGACGA | ++ | ++ |
| 8 | GTTTCGGCCGATAA | ++ | + |
| 10 | TAATCGGCCGAGCG | ++ | ++ |
| 11 | TTCTCGGCCGAGTG | ++ | ++ |
| 12 | TATTCGGCCGACTA | ++ | ++ |
| 13 | GGTTCGGCCGAATT | ++ | ++ |
| 14 | AATTCGGCCGATAG | ++ | ++ |
| 15 | ATTTCGGCCGATAG | ++ | ++ |
| 16 | TTATCGGCCGAATC | ++ | + |

| | | | |
|---|---|---|---|
| 17 | CTGTCGGCCGACCA | ++ | + |
| 18 | GTGTCGGCCGAGCG | ++ | + |
| 19 | TAATCGGCCGAGCG | ++ | ++ |
| 20 | AAATCGGCCGAGAG | ++ | ++ |
| 21 | TTATCGGCCGACAG | ++ | ++ |
| 22 | AATTCGGCCGAGACC | ++ | ++ |
| 23 | TCATCGGCCGATCA | ++ | + |
| 24 | ACTTCGGCCGATGC | ++ | ++ |
| 25 | TTTTCGGCCGAACT | ++ | ++ |
| 26 | TATTCGGCCGAGCT | ++ | + |
| 27 | GTTTCGGCCGAGGG | ++ | ++ |
| 28 | ATTTCGGCCGACAT | ++ | + |
| 30 | CTTTCGGCCGAAAA | ++ | +++ |
| 31 | CCGTCGGCCGAACA | ++ | ++ |
| 32 | GACTCGGCCGAAAG | ++ | ++ |
| CONSENSUS | W TCGGCCGA | unique | |

**Motif 8-8**

| | NNNGTCTAGACNNN | Activity Test 1 | Test 2 |
|---|---|---|---|
| 1 | GGGGTCTAGACCGC | +++ | +++ |
| 2 | GGGGTCTAGACTCA | +++ | +++ |
| 3 | AGGGTCTAGACCCG | +++ | +++ |
| 4 | GCGGTCTAGACCTT | +++ | +++ |
| 5 | ACGGTCTAGACCCG | +++ | +++ |
| 6 | CGGGTCTAGACTGG | +++ | +++ |
| 7 | TCAGTCTAGACCGG | +++ | +++ |
| 9 | CTAGTCTAGACCTC | +++ | +++ |
| 10 | ACAGTCTAGACCTG | +++ | +++ |
| 11 | GCGGTCTAGACGGT | ++ | ++ |
| 12 | CAAGTCTAGACCAA | +++ | +++ |
| 13 | CTGGTCTAGACCCC | +++ | +++ |
| 14 | AGGGTCTAGACGGT | ++ | ++ |
| 15 | GAGGTCTAGACGCC | +++ | +++ |
| 16 | GCGGTCTAGACGCG | +++ | +++ |
| 17 | GGGGTCTAGACCGT | +++ | +++ |
| 18 | ACGGTCTAGACTTG | +++ | +++ |
| 19 | TGCGTCTAGACCCT | +++ | +++ |
| 20 | TAGGTCTAGACACG | ++ | ++ |
| 21 | AGTGTCTAGACCCC | +++ | +++ |
| 22 | AGGGTCTAGACATT | ++ | ++ |
| 23 | ATGGTCTAGACAAA | +++ | +++ |
| 24 | CCGGTCTAGACGGG | ++ | ++ |
| 25 | TGAGTCTAGACTCT | +++ | +++ |
| 26 | GTAGTCTAGACCCG | +++ | +++ |
| 27 | GCGGTCTAGACACG | +++ | +++ |
| 28 | ATAGTCTAGACCGC | +++ | +++ |
| 29 | CTGGTCTAGACCAC | +++ | +++ |
| 30 | TGGGTCTAGACTAC | +++ | +++ |
| 31 | CAGGTCTAGACCGC | +++ | +++ |
| 32 | CGAGTCTAGACGCA | ++ | ++ |
| Consensus | GGTCTAGAC | Similar to ade6-4095 | |

**Motif 8-9**

| | NNNATAATTGGNNN | Activity Test 1 | Test 2 |
|---|---|---|---|
| 1 | AGTATAATTGGCCG | + | + |
| 2 | CATATAATTGGCTA | +++ | +++ |
| 3 | AGTATAATTGGTCA | + | + |
| 4 | GATATAATTGGACA | + | ++ |
| 5 | AAGATAATTGGTCA | ++ | +++ |
| 7 | CATATAATTGGTCC | +++ | +++ |
| 9 | AATATAATTGGTTA | + | ++ |
| 11 | CACATAATTGGCTA | + | ++ |
| 13 | ATTATAATTGGTTA | ++ | + |
| 14 | GATATAATTGGCTC | + | ++ |
| 15 | CATATAATTGGACG | +++ | +++ |
| 16 | GATATAATTGGCTC | + | + |
| 17 | GATATAATTGGACG | ++ | +++ |
| 18 | CATATAATTGGTTG | ++ | ++ |
| 19 | CCGATAATTGGTCG | +++ | +++ |
| 20 | AATATAATTGGACG | +++ | +++ |
| 21 | AATATAATTGGTCC | +++ | ++ |
| 22 | TCGATAATTGGTCA | ++ | ++ |
| 23 | AATATAATTGGTTG | ++ | ++ |
| 24 | CATATAATTGGTCA | ++ | +++ |
| 25 | CATATAATTGGTTG | ++ | ++ |
| CONSENSUS | ATATAATTGGTY | CCAAT-like (complement) | |

**Motif 8-11**

| | NNNATTGGCGGNNN | Activity Test 1 | Test 2 |
|---|---|---|---|
| 1 | CTGATTGGCGGGGG | ++++ | ++++ |
| 2 | GTCATTGGCGGAAA | ++++ | ++++ |
| 3 | TCTATTGGCGGCGC | ++ | ++ |
| 4 | CGGATTGGCGGGGA | + | + |
| 5 | GGGATTGGCGGTCA | + | + |
| 6 | ATGATTGGCGGGGC | ++++ | ++++ |
| 7 | CCCATTGGCGGTAA | +++ | +++ |
| 8 | TTTATTGGCGGAGG | ++ | ++ |
| 9 | TTGATTGGCGGAAT | ++++ | ++++ |
| 10 | CTGATTGGCGGCTG | ++++ | ++++ |
| 11 | GTCATTGGCGGATA | ++++ | ++++ |
| 12 | TTCATTGGCGGCCG | +++ | +++ |
| 14 | GGGATTGGCGGACG | +++ | +++ |
| 15 | ATGATTGGCGGGAG | ++++ | ++++ |
| 16 | GCCATTGGCGGGAC | +++ | +++ |
| 17 | TTGATTGGCGGAGG | ++++ | ++++ |
| 18 | GTCATTGGCGGCGA | ++ | ++ |
| 19 | CTTATTGGCGGATG | ++ | +++ |
| 20 | CCGATTGGCGGCGA | ++ | +++ |
| 21 | GCCATTGGCGGCGC | ++ | ++ |
| 22 | TTAATTGGCGGGGG | + | + |
| 23 | ATGATTGGCGGGTC | ++++ | ++++ |
| 24 | GCGATTGGCGGGTA | ++++ | +++ |
| 25 | CTGATTGGCGGACA | ++++ | ++++ |
| 26 | GCCATTGGCGGCAC | +++ | +++ |
| 27 | TCCATTGGCGGGGC | +++ | +++ |
| 28 | CCCATTGGCGGGAC | +++ | +++ |
| 29 | ATGATTGGCGGAAG | ++++ | ++++ |
| 30 | ATGATTGGCGGGGC | ++++ | ++++ |
| CONSENSUS | TGATTGGCGG | CCAAT-like (complement) | |

**Motif 8-12**
NNNAAGCATGANNN

| | |
|---|---|
| 1 | CGTAAGCATGACGA |
| 2 | AAGAAGCATGACGT |
| 3 | TACAAGCATGACGT |
| 4 | AAGAAGCATGACGT |
| 5 | CGCAAGCATGACGA |
| 7 | ATTAAGCATGACGT |
| 8 | GGTAAGCATGACGT |
| CONSENSUS | AAGCATGACGT |

CRE-like

**Motif 8-13**
NNNCGCAGTAANNN

| | |
|---|---|
| 1 | CCCCGCAGTAAAAC |
| 2 | AGGCGCAGTAAGGG |
| 3 | TAGCGCAGTAATCC |
| 6 | TGACGCAGTAATG |
| 8 | TTACGCAGTAACAC |
| 9 | TGACGCAGTAACGG |
| 10 | TGACGCAGTAATG |
| 11 | TGACGCAGTAAGAA |
| 14 | TGACGCAGTAACCC |
| 15 | TAGCGCAGTAAACC |
| Consensus | TNNCGCAGTAA |

Potential CRE-like

**Motif 8-21 (Table S2)**
NNNAATGGATANNN

| | | Activity Test 1 | Test 2 |
|---|---|---|---|
| 1 | ACCAATGGATAGGG | +++ | +++ |
| 2 | ACCAATGGATAGAG | +++ | +++ |
| 3 | ACCAATGGATAAGG | +++ | +++ |
| 4 | GCCAATGGATATAC | ++ | +++ |
| 5 | GCCAATGGATAAGG | +++ | +++ |
| 7 | GCCAATGGATAGAA | ++ | +++ |
| 9 | GCCAATGGATAGGG | +++ | +++ |
| 10 | ACCAATGGATAAGA | ++ | ++ |
| 11 | CGTAATGGATAGCG | + | ++ |
| 13 | GCCAATGGATAGAT | ++ | ++ |
| 14 | GCCAATGGATATGG | +++ | +++ |
| 15 | GCCAATGGATACGT | +++ | ++ |
| 17 | GCCAATGGATAGGG | ++++ | +++ |
| 18 | ACCAATGGATACGC | ++ | ++ |
| 19 | GCCAATGGATACTG | +++ | +++ |
| 20 | ACCAATGGATAAGC | +++ | +++ |
| 21 | ACCAATGGATATGG | ++ | ++ |
| 23 | ACCAATGGATAGGC | +++ | +++ |
| 25 | ACCAATGGATAGTG | ++ | +++ |
| 26 | ACCAATGGATAGTG | + | +++ |
| 27 | ACCAATGGATAGCC | +++ | ++++ |
| 28 | GCCAATGGATAAGT | + | ++ |
| 30 | GCCAATGGATAAGT | ++ | ++ |
| CONSENSUS | RCCAATGGATA G | | |

CCAAT-like

| | **Motif 8-22**<br>NNNCCATTACGNNN | Activity<br>Test 1 | Test 2 |
|---|---|---|---|
| 1 | CAGCCATTACGTAA | +++ | +++ |
| 2 | ACGCCATTACGTAA | +++ | +++ |
| 3 | TCACCATTACGTAA | +++ | +++ |
| 4 | CCCCCATTACGTAA | ++ | ++ |
| 5 | AGCCCATTACGTAG | ++ | ++ |
| 6 | TCACCATTACGTGA | ++ | ++ |
| 7 | ACTCCATTACGTAT | ++ | ++ |
| 8 | TACCCATTACGTAA | +++ | +++ |
| 9 | TCGCCATTACGTAA | +++ | +++ |
| 10 | CGACCATTACGTCA | +++ | +++ |
| 11 | GACCCATTACGTAT | +++ | ++ |
| 13 | TTACCATTACGTCA | +++ | +++ |
| 14 | TATCCATTACGTCA | +++ | +++ |
| Consensus | CCATTACGTAA | CRE LIKE | |

| | **Motif 8-41 (Table S2)**<br>NNNAGGGATGANNN | |
|---|---|---|
| 1 | ACAAGGGATGACGT | |
| 2 | ACAAGGGATGACGT | |
| 3 | GGGAGGGATGACGT | |
| 4 | GTGAGGGATGACGC | |
| 5 | TGCAGGGATGACGT | |
| 6 | TTTAGGGATGACTA | |
| 7 | ATTAGGGATGACTA | |
| 8 | TTTAGGGATGAGTA | |
| 9 | ACAAGGGATGACGT | |
| 10 | TTTAGGGATGACGC | |
| 11 | CACAGGGATGACGA | |
| 12 | CCGAGGGATGACGT | |
| 13 | TGGAGGGATGACGT | |
| 15 | GTAAGGGATGACGC | |
| 16 | ATTAGGGATGAATA | |
| 17 | ACAAGGGATGACGT | |
| CONSENSUS | TNAGGGATGACGT | CRE-LIKE |

FIGURE S1.—Alignment of sequences derived from subscreens. High-frequency motifs found in our primary screen (Tables S2 and S3) were re-screened for hotspot activity with 3-5 random bases flanking on either side of the motif (Table S4). Sequences were aligned to lo2ok for potential consensus sequences. Each position containing a random base was tested to see if the actual distribution of bases at that position was random (chi-squared test). If the distribution of bases at a given position is unlikely to have occurred by chance ($P<0.01$), the most frequent base(s) at that position was considered to be part of the consensus sequence. Random bases are shown in black font; fixed bases and Consensus sequences are shown in red font. D = A, G, or T; H = A, C, or T; K = G or T; W = A or T; R = A or G; Y = C or T; N = any nucleotide. The 7 base *M26* sequence, 5'-ATGACGT-3' is highlighted in yellow. Potential hotspot families (Figure 5) are listed next to the sequence. Motifs from the subscreens (Table S4) yielding very few readable sequences are not included here, e.g. motifs 7-8, 8-5, 8-7. Hotspot activity was measured qualitatively as described in Materials and Methods and is indicated relative to control *ade6* alleles as follows:
+, activity greater than *ade6-M375*
++, activity approximately equal to *ade6-M26*
+++, activity greater than *ade6-M26*, but less than *ade6-3074*
++++, activity approximately equal to *ade6-3074*.

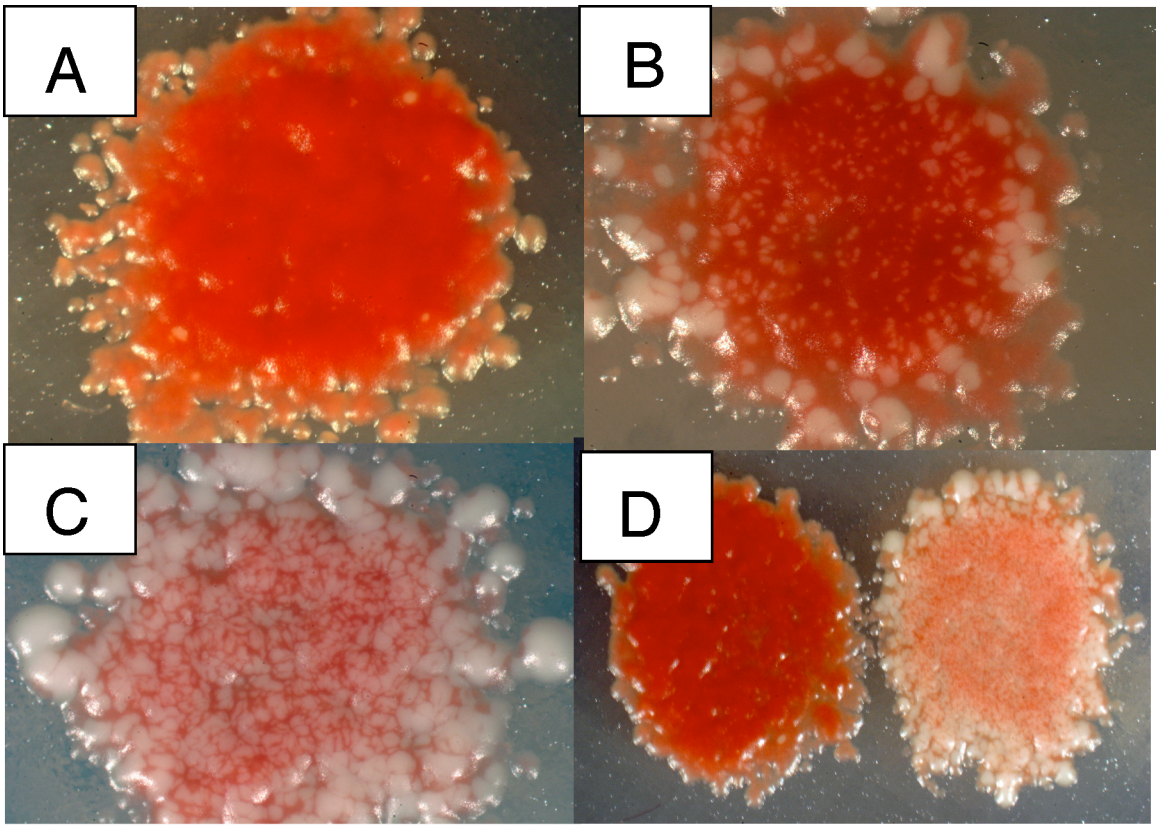


FIGURE S2.—Visual assay for hotspots. Strains containing known hotspot alleles (*ade6-M26* or *-3074*) and one control allele (*ade6-M375*) were put through the steps of the screen described in the text. A) *ade6-M375*, B) *ade6-M26*, C) *ade6-3074*, D) *ade6-M375* (left) and *ade6-3074* (right).

**TABLE S1**

**Hotspot sequences**

Random 30mer screened including wild-type flanking sequence[a]:

CCTGCC**T**AACNNNNNNNNNNNNNNNNNNNNNNNNNNNNCATCATTTAC

Random 15mer screened including wild-type flanking sequence[a]:

CCTGCC**T**AACNNNNNNNNNNNNNNGTGAGCACAT

| Strain | Sequence[b] | activity 1[c] | activity 2 |
|---|---|---|---|
| hWS7 | AATCATTACTCGGAAGTGACGTATACCATG | +++ | +++ |
| hWS8 | GTGCACGTCATACCGAGAGGAGCGTACCTA | +++ | +++ |
| hWS9 | GATTTATCTAGTGGAGTTATAAGGCAATGA | ++ | ++ |
| hWS11 | TGTAAACCTGTTACGTCATGAAGAGAAAAC | ++ | ++ |
| hWS23 | CCCCTCCCGAGGTGATAGCCGCACCAATAA | ++ | ++ |
| hWS30 | GTTCCTCTACTAACAATACGTCATGAGCAC | +++ | +++ |
| hWS32 | ATTGGTGGGATGGCAGAACACACGTTATGT | ++ | ++ |
| hWS40 | CATACAGAAACCGCCACCGAGTATAGTTGA | ++ | ++ |
| hWS41 | CCAACCGATCATAGAATGACGTTTATTTCA | ++ | ++ |
| hWS44 | GCTAATTTTCCTGCTGGCCACGTCTTATGA | ++ | ++ |
| hWS45 | GCTGACGAGTCGATAGGAAAATTGAGATTA | + | ++ |
| hWS46 | TCATGTGGGGGTAGGCAAGAAGAATACACGT | ++++ | ++++ |
| hWS47 | AAATAAGATTACGGTGACGTAATGCATCC | ++++ | ++++ |
| hWS48 | AGGAACTCGGAAATGTGTTCTTCAAGATGA | ++++ | ++++ |
| hWS50 | TTGGGCCCCAATGATTTATACTTGTCGTGA | ++ | ++ |
| hWS51 | GATAGTATCGCGTTAGTGACCCCCACGTCA | ++ | +++ |
| hWS53 | CATGGGGATTGGCGGGATGGAAGCTCATGA | +++ | +++ |
| hWS55 | GATCTTAGTTCGTTATAGGGCATATTATGA | ++++ | +++ |
| hWS57 | GTATGACTGTGTAATGACGTCATGCAAGCC | ++++ | ++++ |
| hWS58 | CCAAAGATGGCGGATCTAATCTACATATGA | ++++ | ++++ |
| hWS60 | CAGAGTCACGGTTACTACATTAATACATGA | +++ | +++ |
| hWS61 | AAAAATTCTCGAAAAGGAACGAAGTGATGA | +++ | +++ |
| hWS62 | TCATTATAATAACGGCGACTCGTTACATGA | +++ | +++ |
| hWS63 | TTTCTAGTAGCAGCGAAGCGAGCGCGGTGA | +++ | +++ |
| hWS64 | GGATTTCGGATCTAATATGGCCCGCTATTA | ++++ | ++++ |
| hWS65 | GGTTCTAGAACGGACAGGGTCTTAGCGAAT | ++ | +++ |
| hWS67 | CGAAAGGTTTAAGGTGACGTATCACGTCAA | ++ | +++ |
| hWS68 | GAAGGAGCCCCCTAAGAAGACGTATGAGGT | ++ | +++ |
| hWS69 | ATATGTTGAAGCCGTGGGCGCAGGACATGA | +++ | ++++ |
| hWS70 | AAATGTGGACAGACTGTTGGTTGTCCTTAA | ++ | ++ |

| | | | |
|---|---|---|---|
| hWS71 | GTAACTGGACGTGTCTTGTGACAGGGATGA | ++++ | ++++ |
| hWS72 | CTTTTTATCAACAAAAAAATTAAGTGACGT | ++++ | +++ |
| hWS73 | ACGGGATGCGTCTAGGTAACTCACAAGTGA | ++ | +++ |
| hWS74 | AAGCAAATCAGGTACCTAACAGAAAAATGA | +++ | +++ |
| hWS76 | TATCCCCGACACAGGCGGACTTATGTCTGA | ++ | ++ |
| hWS79 | CACGAGAAGTAGTTAAGCTTACGTCATGCG | +++ | +++ |
| hWS80 | TTTGCTAGCCGAAATGCTGCGTTCACATGA | +++ | +++ |
| hWS081 | CTAGACGTGGTTAAACTGTATCTTTGACGT | +++ | ++++ |
| hWS82 | GAAGTGACGTGCCGCTTAAGTTAATCAACG | +++ | +++ |
| hWS85 | CCCCGCACTAAAACTACATGTCCTATGCAG | ++ | ++ |
| hWS88 | GTACCCACTATGGCGTCACCGAGGCCAGCA | ++++ | ++ |
| hWS89 | CAAGTTCCGATTGTGAAGAGCAGTGGGTGA | ++ | ++ |
| hWS90 | CCAACAGCGAATATGATAGCCCAAACATGA | ++++ | +++ |
| hWS92 | CTTATTGGACCCGGCACCTTATACTTGTTA | ++++ | +++ |
| hWS93 | CAATACGTAAGTATGACGTGATCGAAGAGG | ++++ | +++ |
| hWS95 | CCGAGAATATGTAGGGATGACGCCATTACG | +++ | +++ |
| hWS96 | ACTCGACAAAGTGGCCATAGAAAAAGATGA | ++++ | +++ |
| hWS97 | GAGATTCAATTCGCTGTATTGGTGAAATGA | ++++ | +++ |
| hWS98 | CCGGGTGGTATACAAGGAGTGAACAAGTG | ++ | ++ |
| hWS100 | GCATGCCGCCGAGCCAAATAATTGACACGT | ++ | ++ |
| hWS101 | GATTTGGTCTTAACAAACGGTGATGTACGT | +++ | ++ |
| hWS102 | TGGCTATGACGAAATTGTTAGACGGCTTTA | +++ | +++ |
| hWS103 | CAGGCACACATAAAACGAGGGATGATACGT | +++ | ++++ |
| hWS107 | AAAACATTATGTGTGGCATTCACACTATG | ++++ | +++ |
| hWS109 | ACGGCGTGTGATAGGCTAACGACATGAAGT | ++ | ++ |
| hWS111 | TGCAAAAAGATGCACAAACGACATATCTGA | +++ | ++++ |
| hWS113 | GTAATCAATGGATAAATTTGCCAATTATTA | ++++ | ++++ |
| hWS115 | CCATGTAATTTCTCGAACGTCTTACCACGT | ++ | +++ |
| hWS116 | GTGCATTCCTCCACGCCTGTTTAAAGCGGC | +++ | +++ |
| hWS119 | ATCCTTCATTGCCGAAGTTAACAATGATGA | +++ | ++++ |
| hWS121 | GGATTAAAGAAATGTAACTTTGACATTCGA | +++ | ++++ |
| hWS122 | AAACGAGTGTCCCGACTTGCACCAACAAAT | ++ | ++ |
| hWS124 | TGAAATTACGTCATCCCGCTGAAAGTTTCA | +++ | +++ |
| hWS125 | CCGTTGGGAAGTTCAGAACATCGTATACGT | ++ | ++ |
| hWS126 | TATTGGAGTAGAAGGAACTTAAAATGACAT | ++++ | +++ |
| hWS127 | TTTAGAGCTTGAACCTAACTGGAAGCATGA | +++ | +++ |
| hWS128 | ATCCCCGCAGTAAGAGTGAGCACATTGATG | ++++ | ++++ |
| hWS130 | GAAAGCCAAAGGATGGCACCAGCTCTACGT | ++ | +++ |
| hWS133 | AAAAAGTTCTAGGGTACGTTAGGGGTGTGA | +++ | +++ |
| hWS134 | GTCATACATTACGAGAAATAACAAGCGTAA | ++ | +++ |
| hWS135 | GTTAGCTAAAATAACCAGTTGGTTCGATGA | +++ | +++ |

| | | | |
|---|---|---|---|
| hWS136 | CTACAATGCGGAGATTTCACATCATAATTA | +++ | +++ |
| hWS137 | TATTTATACCAATGAGGGGGGCTCTTGTTA | ++ | ++ |
| hWS138 | CAATCACAGGCAAGAGTACTTAAGCATAAA | +++ | +++ |
| hWS140 | TAAGCTGCAAGCAACCTTATAGCAACACGT | ++ | +++ |
| hWS142 | CTGAAAGGACGTCATAAGCTTACGGATGAA | ++ | +++ |
| hWS144 | GCCGAACTGCGTTGCAAGACGTTAAAATAT | ++ | ++ |
| hWS147 | TGATTGGTCGAGCACATATGGAAAGCCGAC | ++ | ++ |
| hWS149 | CGTCCAATCAGCGGTTGGAGTGAGCAGTAT | +++ | +++ |
| hWS151 | TTTCCGGGATACGTCTCCTCTTGTCGACGT | +++ | +++ |
| hWS152 | GCAACGACAATAATGGCTTAATCTTTATGA | +++ | +++ |
| hWS153 | GTGATAAAATTAGGGATCTACATGCGCTGA | +++ | +++ |
| hWS157 | CGCTTGGAACAGACGTAGGACGTCCAATGA | +++ | +++ |
| hWS158 | ACAACGGTAATTAACCGCCAATGAGGAGAG | ++ | + |
| hWS159 | CGCTCTGGTGTTTTCAGAACGAATTCATGA | +++ | ++ |
| hWS162 | GTTACGTAAGCTTTAAGAATAGCGTCACGT | ++++ | +++ |
| hWS163 | TTAATTCGTTTTTCGTACAGCTTTGCATGA | +++ | ++ |
| hWS164 | TCGGCATTTACATGGTAATCTAGTAGATGA | +++ | ++ |
| hWS166 | ATTGGACGATTAGACGTCAATTGTTTATTA | +++ | +++ |
| hWS171 | AAAAGGCCCAACCATCCCCCACGCTAGTGA | +++ | +++ |
| hWS175 | CCCTTAATGAGCCCCGAATAAGCGGGATGT | ++ | ++ |
| hWS180 | TAGAAAACTTACGGCTGGTGCGTTGTATGA | ++ | ++++ |
| hWS181 | GTAACGATGCGTCAGGCGACACAATGGACA | +++ | +++ |
| hWS184 | CGTATCGTTAAGCGGCGCAATTTGAAACAG | + | ++ |
| hWS186 | GTAAGGAGATTACGTCACGAAAAGCTACGT | ++++ | ++++ |
| hWS191 | AGTTGGCGGCCTGGGAGCCTATTAGACATA | ++++ | ++++ |
| hWS193 | TTACTGAGTCCACGTATATCCTTACAGTGA | ++ | +++ |
| hWS194 | CTCAATTAATTGGATCGAAATCTTTTGTGA | +++ | +++ |
| hWS195 | TGCTCGAGGAATACGTGTGCTGAGATAGAG | +++ | +++ |
| hWS199 | AAAGTACAAGAGATTCTCGTCATAGTATGA | +++ | ++++ |
| hWS206 | TATAAGAAGGTGTGACGTAATACGCTTATG | ++++ | +++ |
| hWS213 | CTGTGCAATTATAGACATAAATATTACAAC | +++ | +++ |
| hWS215 | TAATTTCGTCTAATAAGTCTTTTTGTGGGT | +++ | +++ |
| hWS217 | TGCCTTATGGAGAATGGGGCCACATTGAAG | +++ | +++ |
| hWS218 | TGATAGGCGTTTTCCAGTTTTGAAGCATGA | ++++ | ++++ |
| hWS220 | TATGACGCAGTAAACATCCTCGACTTACAG | +++ | +++ |
| hWS222 | CACCAAAGATTTTACGTCTTGTATGGATGA | +++ | +++ |
| hWS226 | ATTCGACAAGGGTATATAACGTCATACAAG | +++ | +++ |
| hWS229 | ATGGTCCGGCCGAAACTACGGCATAATTTA | +++ | +++ |
| hWS230 | AAAGCCTAGACACACCATAGACGAACACTA | ++ | +++ |
| hWS233 | ACATCGATCAACGTCAACCCAACGGTACGT | +++ | +++ |
| hWS236 | GGACTACTTCCCAGCGTGAGTACCTTGTGA | + | ++ |

| | | | |
|---|---|---|---|
| hWS237 | ACGTAATATTAAGAAGCAAACGAAAGCTGA | + | +++ |
| hWS239 | CGTAAAAAACCAGCTATAACGTAATAGTGA | ++++ | +++ |
| hWS246 | CGGGTGGACATGATGACAAGGTCCTCATTA | ++ | +++ |
| hWS247 | ATGCGTAACGAAAGGTAACACGTCAGTGTG | ++ | ++ |
| hWS260 | ACCGACTCACCCGTAGATAAAATGGGGTGA | ++ | ++ |
| hWS261 | GGTAAAGGGCACATAATCGCAGTATGATGA | ++ | ++ |
| hWS262 | AGTAACTTTTGTTACTATCACATAATATGA | ++ | ++ |
| hWS263 | GAAATGTCAGGAGCGGAAGAAAAGTAATGA | ++ | ++ |
| hWS264 | GTACTTGCGCGACACTCGTAAAGAGAGTGA | ++ | ++ |
| hWS265 | AGACGTGTTGGAAAGCCGAATCGGCCATGA | ++ | ++ |
| hWS267 | CGGATGAAATGCTTTAGTTTTGAAACACGG | ++ | ++ |
| hWS270 | TTTTTGGATAGCTGCCAGGTCCGGCGGTGA | ++ | ++ |
| hWS271 | ATGACGCCATTACGACAGAGCTCTGCACTC | ++ | ++ |
| hWS272 | AAGCGCATGACGCTTGCTAATAAGGATTGA | ++ | ++ |
| hWS273 | CTAGTTGACATCACCAATGCTGGACTTTTA | + | ++ |
| hWS274 | CGAAAGATGGGTAAGCGACCGCCAATGATT | ++ | ++ |
| hWS275 | TGCCTTGCGGTTCGCTATCCATTGTTGCAA | + | ++ |
| hWS277 | CATGATCCGTCAATGCCGGAATTTGTATTA | + | ++ |
| hWS278 | TAACTAGAAACTAGAGTTACACAACTGTGA | ++ | ++ |
| hWS279 | GTCATTCCACTCCCTTAGTTTGTTACACCC | + | ++ |
| hWS281 | ACCCACGCGTCCAAAACTCGGACCACGTTA | ++ | ++ |
| hWS282 | CTCATGGTTCCAGGCCCTGCACGCAGAACA | + | ++ |
| hWS285 | AAGATCTAGACAAGTGAAGTCACAGTGATA | ++ | ++ |
| hWS286 | AAGAGAACACTAATAGCAAGACGCTGACGT | +++ | ++ |
| hWS287 | TTAAGCTACGTGGCTCGTCCTATTTTGTGA | ++ | ++ |
| hWS288 | CCTTTACGTCATCGCCGACGACGTTTTGGA | ++++ | ++++ |
| hWS289 | ATAGAGGTAAGTCCCCGAGCTTTGCCACGT | ++ | ++ |
| hWS290 | GAATAGCGCAGCGATACGACAGCAAACTTA | +++ | ++ |
| hWS291 | GTTATGAAGCAACCGAGAAGGTGACATCAC | +++ | ++ |
| hWS292 | AGTATATAAATGTTGTCCGTTGGGTGATGA | +++ | ++ |
| hWS294 | GATGACGCTATGTACGGTTGAGGCGGACAA | ++ | ++ |
| hWS295 | CAACCACATTGCGGCGTGCTATAGGCCGCG | ++ | ++ |
| hWS297 | AAGAGTTTAGTCAACAATGACGTCACGTTT | +++ | +++ |
| hWS298 | CGCAGTTGTTCATTCGTCAAGACGCAGTGA | +++ | +++ |
| hWS299 | CTAGAGAATGACGACACCATTTAGAGTCCC | + | ++ |
| hWS300 | AAATGGGGTGATTACTCCGGGCATTAATGA | ++++ | ++++ |
| hWS301 | GGAGGGCCCGTATAAGGGAAAAAAAAGTGA | +++ | ++ |
| hWS304 | GTCATCAATGGTAGCCTGTGAGGGCTATCG | ++ | ++ |
| hWS305 | AATCAGTGCGAACGGGGCAATTCATAATGA | +++ | +++ |
| hWS307 | AAGTCTAGACTTTTCAATCGACGCGGACTC | + | ++ |
| hWS308 | GTCCAATCGATTAGAGGTATAGACGTAGTA | + | ++ |

| | | | |
|---|---|---|---|
| hWS309 | AACGACAGGACATTGCCGAAATGCGGACGT | + | ++ |
| hWS310 | TAGACGTCACCATCATCAAGCGCGACTCGT | ++ | ++ |
| hWS311 | ATGTCCATCTCTCGCAGTAAGATAACATGA | + | ++ |
| hWS312 | CGAGTGACGTGTGAAAACCGCATTTACT | ++ | ++ |
| hWS313 | ACCGGGAAAGAAAGATTGGAGTTAGAATGA | ++ | ++ |
| hWS314 | ATGTCCACCGACTACGGTAAAAGACGATAA | + | + |
| hWS315 | TTGAGCGGCCACTGTGATACGTAATAACTG | ++ | ++ |
| hWS316 | CCGATGAAGTGACATACACACGTAGCATTA | +++ | +++ |
| hWS317 | AAGCATTTACATCAAACCCGCGGGAGATGA | ++ | ++ |
| hWS318 | TGTCTGTTCCTATGTATGGTACAGCAGTGA | + | + |
| hWS319 | CAGACGCATTGGATCGCGAGCGCTGGGTGA | ++ | ++ |
| hWS320 | AACCAATCAGCACGAAAAAAGGCAAAAGAG | ++ | + |
| hWS321 | ACGCCATGCGACGCTCGTCAGATCACATTA | ++ | ++ |
| hWS322 | AGCATCTGCACGCGGGAAAGGTGTGAGTGA | + | ++ |
| hWS323 | ACAAGAATGGCGGTTAAGTCTCCGGTATTA | + | + |
| hWS324 | AAAGGCTAGGCGCATGTGATTGGGTGAGAG | ++ | ++ |
| hWS325 | ATAACGAAGTGAGGCATAGAAGTGCTAAAA | + | + |
| hWS326 | ATGATGAACGGTTTAGGGTATCGCCGATGA | +++ | +++ |
| hWS327 | ACCGGCACAGGGGGTATGGCTCGGAAGTGA | +++ | +++ |
| hWS328 | CCACTCCAAACTACATTGTCCGCACTTGCG | ++ | ++ |
| hWS329 | ACTGTCTGCTAATAGAGCGGCCAGCGATGA | +++ | +++ |
| hWS330 | CAAGGTATCTCACTAAGGCATAATTTATGA | +++ | +++ |
| hWS331 | TGAGGTGCAAGGTAGCACTAACGATTATGA | +++ | +++ |
| hWS332 | CGTGCGTACTTATAACAGGCGTACAAATGA | ++ | ++ |
| hWS333 | ACCGAACGATGACGCATACGGCAGCCTAGA | ++ | ++ |
| hWS334 | AGACGTCATTACTAAAGGACGCGGGCTAGC | ++ | ++ |
| hWS335 | TTGCGGAATCTTTCAGGAGGTAAAGCGTGA | + | + |
| hWS336 | TTACCTGTGACATCATGATAAGCAGAAAAA | + | + |
| hWS337 | ACAATGCGGCGCCTCGAGTTAGGAAGGTGA | ++ | + |
| hWS338 | CAGCGACTGATTTAAAAAATAAAAGCATGA | ++ | ++ |
| hWS340 | CAAGGGGCTAGACGCAGGGGGGGATGCTGA | ++++ | ++++ |
| hWS341 | ATGGAAGAAACGTCAGATAGTCCGATTTGG | + | + |
| hWS342 | ACCACGATCGAGCCTGGCTATACGTCACAG | ++ | + |
| hWS343 | TGCTGTGACCTAACGGTGGCCAATGAGGCC | + | + |
| hWS344 | TGAGGAGCAACTGCGTAAACAGGCGATATGA | +++ | +++ |
| hWS345 | ATGACGAAATTTAGA | + | |
| hWS346 | GTATCACGTAATAAA | ++ | |
| hWS347 | TATTATTACCCCCAC | + | |
| hWS348 | CAATAACATAAGGCC | ++ | |
| hWS351 | GACTACGTCACTAAC | +++ | |
| hWS352 | AATTTCAACCAATAG | + | |

| | | | |
|---|---|---|---|
| hWS353 | GAGTTAGTGCGCAGGAAAAAGATTAGGTTA | + | |
| hWS354 | GCAACACCATTCGCGCGATATAATTGGACG | + | |
| hWS356 | CGGAGGGCTCCCACACGTCATGATGTCGCA | +++ | +++ |
| hWS357 | AACGGGGTTTATAATTGGTGCATTATCGGA | ++ | + |
| hWS358 | AGGTATACGTCACTTACTGTAAAATGCATG | +++ | ++ |
| hWS359 | TGATGACATCAGTGGGATGCGTCGTACTCA | +++ | +++ |
| hWS360 | CGTGTGGGTGGGCGGTCTGGACCTAGTAC | ++ | ++ |
| hWS361 | TAAGCGCGTGGGGGAGCTATAGCAACGTGA | +++ | ++ |
| hWS362 | CAACGATTGGTCATGCGTAAATACACTTCA | ++ | + |
| hWS363 | AAGGTCGTCTACTGAAAGTCAGGTTTATGA | ++ | ++ |
| hWS364 | ACCCGGTCGGGCCTTCTGGGTCAAATATGA | +++ | +++ |
| hAG1 | ATCGGTGACGTAAGGCCGATATATATTAGT | ++++ | |
| hAG2 | ATTGAGTGGTTAAATTCTGGCAGGATATTA | +++ | |
| hAG3 | CTGTGTTTATAATTCATAAAGGAATTACGT | +++ | |
| hAG5 | TCGTAATAATGCACTATCACCAAAGATCTA | +++ | |
| hAG6 | GATCATATATACTGAAATGGGGGATTACAT | +++ | |
| hAG8 | TAAGTTCTAAGTGCAAAAGAGGTCACATCA | +++ | |
| hAG9 | GATACCTCTGGAGTAAATAACGCAACATGA | ++ | |
| hAG17 | ATGACACTATCTGTATTACTGAAATTGAGG | ++ | |
| hAG21 | GGGAGATCTCCAAAGTAACTGTTACCCCAA | N.D. | |
| hAG22 | ATGATGTTATCAAAGGACGTACTAATGTGA | +++ | |
| hAG23 | GTGACAAAAGCCCGAACTACGTATGCTTGT | +++ | |
| hAG25 | ATTATGTCATGCGCATTCGAGTGCAAACGA | +++ | |
| hAG26 | ATGCGCGGGGGAACCAGGGCGGTTTGTCAGAG | +++ | |
| hAG30 | CCCGAAACCTCCGAGTTTCGGATTCCGCGA | +++ | |
| hOES1 | ATCGGGAGCTCTACCAGCCACGACGTAATG | N.D. | |
| hOES2 | CACAAAGTGCTAGTTACGTGTCGGGGTATG | N.D. | |
| hOES3 | ACGTGATGATTATCCACCTAATAGGCACAG | N.D. | |
| hES004 | AGCTTACTTTTGTCTTAGTCAAAAAGGTGA | +++ | |
| hOES5 | AAGTGACGTTGTAAGACGAAACCAATACAC | N.D. | |
| hES005 | TAAAAGCGTGTTACAACGAAACACCGTAAA | ++++ | |
| hOES006 | TGCCTTATTCTGAATAGGGCAATAATGAAG | N.D. | |
| hOES7 | ACGTTAACAGATACGTGGCGGTCCTCTTAA | N.D. | |
| hES009 | CCCGCATCGAACGGTGTATCGATAAAACGT | N.D. | |
| hOES11 | AATTCGCTGTATCATCATGAAAACGACCCG | N.D. | |
| hOES13 | TCACCGTGGGACAAGAATCGCATAGCCGGT | N.D. | |
| hOES14 | TTTGGTATATCCACCGCCTATTACCTCACT | N.D. | |
| hOES15 | TCACTCTAAGTTCAGCCCACTACGGTCGAT | N.D. | |
| hES018 | CTAAGCATAAGGTCACTGACGTCGGCGCAA | ++ | |
| HES57 | GTACCGATATGACGC | N.D. | |
| HES58 | AAATGGATAGGTGAC | +++ | |

| | | |
|---|---|---|
| HES59 | TAAGAGCGTAGCGCA | N.D. |
| HES64 | GAGATGACGTCAACG | ++++ |
| HES65 | TGATGGATCTGACGTCAGGTCATGTTTGTT | ++ |
| HES69 | GTCATAATAGAGTGA | N.D. |
| HES82 | AGTGACGCACGCGCA | ++ |
| HES83 | GACTCGGATATTCGG | ++ |
| HES88 | GTAATAACAGGAATGAAAGTTAAAACATGA | + |
| HES90 | GTAATCCGATTTAGC | ++ |
| HES92 | GTTACGTTAATTTTG | ++++ |
| HES93 | GAATGTTGAGGGGTG | ++ |
| HES94 | CAAAGGACGTCATAA | + |
| HES95 | GCAAAAGATAGATCG | +++ |
| HES96 | TTTGCGGATAAAGCA | +++ |
| HES97 | GTGACGTGATTACCG | + |
| HES98 | GACGGAAAAACTCTA | +++ |
| HES99 | CACTCGTTCTAGCCT | +++ |
| HES100 | GATGATTGGATGACT | +++ |
| HES101 | GAAAAGATGACGTAT | +++ |
| HES102 | TCGCTTCGTCATCGC | +++ |
| HES103 | CAATCAATTACGTTC | ++ |
| HES104 | CAGAAAGCATATGAC | ++ |
| HES105 | GTCATGGCGAGCCTC | ++ |
| HES106 | TGAAGAACACCGCAT | + |
| HES107 | TATGTTAATTTTGTA | ++ |
| HES110 | CCCACCCAATAATAG | +++ |
| HES111 | TTACGTCATAAATAA | ++++ |
| HES112 | ATATGATGTCAAGGA | ++ |
| HES113 | GGGTACTATTACCCG | +++ |
| HES114 | TCAGATAAAGATGAC | ++ |
| HES115 | CCGGTTAGCGAGACC | ++ |
| HES116 | TGAGTGACGGCATAG | + |
| HES117 | GACCACTTTTGCCTT | + |
| HES120 | CAATAAAAGGGCGGG | +++ |
| HES121 | CATGACGTAACAAGA | +++ |
| HES122 | AGCCCAGATATTAGG | +++ |
| HES123 | TATTAACGAGATCCA | ++ |
| HES124 | TAATAGTGTAGGGGA | + |
| HES125 | TGTTATTGGTATGAC | ++ |
| HES127 | GGTCACACGTCAGCC | + |
| HES128 | ATCCCCGCAGTAAGA | + |
| HES131 | CGGTAGCGATCTACA | + |

| | | |
|---|---|---|
| HES132 | TCAGGGGGTGACATT | + |
| HES133 | ACCTCTCCGAAAAAA | + |
| HES134 | CGATTAGGCTAACCC | + |
| HES135 | CCTTTACATGAGTCG | + |
| HES139 | TGCATCCGATAGCTA | + |
| HES146 | TGAGGGGCAGGTGGT | + |
| hES161 | ACATATAAAGATGAC | ++ |
| hES162 | GCGCTGCGCTATTGA | + |
| hES163 | GTTACGTTAGAAAGA | ++ |
| hES165 | ATTCGCGCACAGTGA | ++ |
| hES168 | AATGCATGACGTTAG | + |
| hES172 | AGATAATTGGTTGTC | ++ |
| hES175 | CATTGGATAAGGGTA | +++ |
| hES177 | GCATCTAGACTTATC | ++ |
| hES177 | ACTTCAGGTGGACCG | N.D. |
| hES180 | CTAAGGGATTGTGAC | + |
| HES181 | GAGGTGTTACCTTAA | + |
| HES182 | ACCTGATAGGTTCTC | + |
| HES186 | TGGTTATCGGATGAC | +++ |
| HES189 | GTACCACGTTGTAGG | N.D. |
| HES193 | CGAGAACGAAGCGGA | ++ |
| HES197 | CCCACAATCCCACGA | ++ |
| HES198 | GACCATAAGCGAGGG | + |
| HES199 | CAATCAGAAATAGTC | +++ |
| HES200 | ACTGGAATAGATGAC | ++ |
| HES203 | TGGACAATCGCAATT | + |
| HES207 | TGTGACGTATGAGAA | +++ |
| HES208 | TCTAAAGTACATGAC | + |
| HES209 | AGACGATTACGTAGT | ++ |
| HES211 | AACCGGTTACCCGAA | ++ |
| HES212 | CTTATGATTGGCGGA | +++ |
| HES214 | TGCGTCATACCAGTA | ++ |
| HES221 | TTATAACGAGCTCTT | ++ |
| HES224 | TGCGGGGAAAAAAAA | ++ |
| hES225 | CGTCTCCATTACGTA | ++ |
| hES226 | GTTTCAAGCCCTCTC | +++ |
| HES229 | GATCGACCATGCGGC | + |
| hES231 | CAAAGCGACGTAATA | +++ |
| HES232 | GAGGCGCAAGCCCGA | + |
| hES236 | TCATATGACGTGACT | +++ |
| hES245 | CTCGGATGGACCGAG | ++ |

| | | |
|---|---|---|
| HES247 | CCCTCGGTAAAGTAT | + |
| HES250 | TATGACGTCAGCACC | ++ |
| HES252 | AAATAGTGAGGCTAA | + |
| hES253 | GTTAGATCAGAAAGC | +++ |
| hES254 | CTTAACCCCCGCAGAA | ++ |
| HES255 | GCGTAACGATGAATA | + |
| HES260 | CAATCAGCAAAACAT | ++ |
| HES262 | CAGATCTCAGATGAC | ++ |
| HES263 | CAATAAGAATATAAA | ++ |
| HES269 | TGTTGGGTCTGGACC | +++ |
| HES270 | CCGCCCCCACTTTGA | ++ |
| HES271 | AACCAATTAAAGGGC | +++ |
| HES272 | TCTGATTGGGTATGA | ++++ |
| hES273 | CAGATCTGGACATACCCTGAAGGATCGTGA | +++ |
| hES274 | CAAGGGGTACTGAATTCTTCAGGCTGATGA | +++ |
| hES275 | CGACCAACGTCATCCGGGGAGAATGTTCAC | ++ |
| hES276 | ATCTTACAACGCCGACGCTAATTCGCATTA | ++ |
| hES277 | ATCATTAAAGTGACATGAGGATGTAAGTAA | ++ |
| hES278 | ATCATACGAGTCTTCACGGTCTAGAGTCGC | ++ |
| hES279 | CTCGAACTAGAGTAACGAAAGTGCTTGTGA | + |
| hES280 | AAAAAGGGTATCTAGGTTCAAGGAAAATGA | +++ |
| hES281 | CCGTAGTGCACAAGAGGCCAATTGGGTAGGA | ++++ |
| hES282 | AGGGGCTGAGTAATTACGCATAAGGAGTTA | ++ |
| hES283 | CAATTACATTCGATCTGAATAGGAGCAACG | + |
| hES284 | ACGGCTAGGTAACAGACGCAGACAACAGCA | + |
| hES285 | ACCCGACAAGGCAGGGTGGGCGAAAGGAGC | + |
| hES286 | AACATTTCATGAAAGCGCTGTAGAGCATGA | ++ |
| hES287 | AGTTGTGGGATGCAACAGCTGAAGGTATGA | +++ |
| hES288 | AAGGCACCAGGCACTTATCAATGGTTGTGA | + |
| hES289 | AAGGTTAGTCGCGGATCTAAGTGGTCTAAA | + |
| hES290 | ACTTGTGACGTCGCCTCATGGAC | + |
| hES291 | TTACTCGTGACGTAAGAGTCGCCACTCACG | +++ |
| hES292 | ATGCCAACGAGCGCCGACGGTCCGGCGCTA | + |
| hES293 | AACCATGAGGTAGGAAGCGAGGTATGATAA | + |
| hES294 | AGAGCTTACAAGATAGTTGGACGAATCTGA | + |
| hES295 | AGCATTTCTCATCGGCAATCAACGCATGA | ++ |
| hES296 | TCGCTAAATGATGGGAGTCAGCCCTTTATA | + |
| hES297 | CGAAGTTGATAAAGGAGACGCAGAAGGTGA | ++ |
| hES298 | TTGTGTGTCGTTGTGACTGATACGATAGCG | + |
| hES299 | TTAGCATAACGGATTTTCATCATATGGTGA | ++ |
| hES300 | AAACGTCAAACATGCAAAGCATTGCTATGA | ++ |

| | | |
|---|---|---|
| hES301 | TGAGATAATAATACTGCAACCGGGCTATGA | ++ |
| hES303 | AGGGGGTGGGAGTGGGAAATTCAGGGGTTC | ++ |
| hES304 | CGGCCTCCACATGGAACACATAGATTGGGG | + |
| hES305 | TAAACAGATGGGTGGACCTCCTGCACGTGA | + |
| hES306 | TGAGCAAATCTCACGCCGTAACATTATGGG | + |
| hES307 | TTTCCTAGGGACGCGGATGTAAGTAAGCGG | + |
| hES308 | CAAAGTGACAAAAGTGGTAACACAAGAGGC | + |
| hES309 | AACAGGTAGCTACTCACTCTAGGAATCTGA | + |
| hES310 | TGGGAATGGCGATAATGGAGGCAGCGGTTA | + |
| hES311 | ATGAGCGGGGGAGGGGGTAGGGCATACCTTG | + |
| hES312 | AACGGTCACAAGTGTGCCTCGTACGAAAAG | + |
| hES313 | CAATCATGACCTGCTATAAACCCCTGCAAA | + |
| hES314 | TAACCGAGAATCCGGGTACTAGTCAAAATA | + |
| hES315 | TACGGTAATAGGGGGTTCACCCTCCGGCCG | + |
| hES316 | CAGAGGGAATAATACGGTAGAGTCTGGAAA | + |
| hES317 | CGCCAATGGATAAGGAGCGTTTCAGGTGAA | + |
| hES318 | ACGGTGATGTATTCCGGCTGTCAGCACCTA | + |
| hES319 | CATGCATGTCACAATGACACACATGCGGAAA | ++ |
| hES320 | TCAAAGGTACGAATTGTTATGTTTGTGTGG | + |
| hES321 | AATCTCAGAAGCTCGGAAAAGGCAGAATGA | ++ |
| hES322 | TATGAATAAGATTTAAAGATGCTGAAGTGA | + |
| hES323 | AGAGAACACTTTACAAGGCTGTCTACATGA | ++ |
| hES324 | GGCGGGTGCTTGTTTATGGAACTTGGCGTGA | ++ |
| hES325 | ATGATAGGATGCTGTGAACCAATGCCTAAC | ++ |
| hES326 | CCGCTTTGTAGAGTTGCCGAACAACCGGTC | ++ |
| hES327 | TAGCTGTGACGTACTGAGCGATCATATTTA | ++++ |
| hES328 | GGGGTGTAAGATCTTAATTTGGTGCGGTGA | + |
| hES329 | ATGAGGGGCTCAATAAGTCCCCCATAGCAA | + |
| hES330 | AACCTGAACAATGCACAACGACCGCGGTGA | + |
| hES331 | TAGGCAAGAACGGCTCACTGCGGGTGATTA | + |
| hES333 | AACTAGTCTGGGCAGGCCCATAAGTGAACA | + |
| hES334 | AGAGATTGGGCCTCCTATAGAGAAAGATGA | ++++ |
| hES335 | ATGAGGCCGGAATCTGTGTATGACGTACTG | + |
| hES336 | TACAGAGAGTCGTGAGGATAGACCAAGTGA | + |
| hES337 | AACATGTTCAATAAGCATCTATAATAATGA | ++ |
| hES338 | ATGACCGGAAAGTACTGCGTCACTCGATAA | + |
| hES339 | AACTGGCGTATCAGACAACACGGATCATGA | + |
| hES340 | AACGCTGTTGTCCGGCACGGACAGGAATGA | ++ |
| hES341 | AACGCAGGCACGCGCGATTGGATGGTATGA | +++ |
| hES342 | ATTGGGCCCAATGAGTGAACGCTGGACCTA | + |
| hES343 | AACGGAGCGATGCATTCTGTAATCGAGTGA | ++ |

| | | |
|---|---|---|
| hES344 | AGGTTTCATTGGTTGGCTAACCTACTATCC | + |
| hES345 | CCACGCCTTAAAAATCACTGCGAAGTCTTA | ++ |
| hES346 | AATAAAGGAGGTATGACGTGATAGCAAACG | ++ |
| hES347 | GAGCTCAAGTCCCGTGAAGACAGTCAATGA | ++ |
| hES348 | CCCTTGGTGAGCAAAATGCCCGCAGTATGA | ++++ |
| hES349 | AAGTGCATTAATAGTCTCCTTGACGAATGA | ++ |
| hES350 | CAATACTAGCTACCCATGTAAGTTTAATGA | ++++ |
| hES351 | AATGTAGAGCCGAGACTCGAAGGGGCATTA | ++ |
| hES353 | ACTCTACCCCGCAGAAAATTTAGGCGGCGC | + |
| hES354 | TCGGATGTAAGTGAATTGGTCCGGTGATGA | ++++ |
| hES355 | AGGTCAAGGTCAGCCAATACGCAAGCATGA | ++++ |
| hES356 | TTAATTGTGGCCGGTCTGGACCGACCCCAC | + |
| hES357 | CCCGCTTTAAAGTCCCAGGTTTAATGTTGA | + |
| hES358 | ACAACTAATTCATTACCGAGTAGGATATGA | ++ |
| hES359 | ATGCAACTACCGGCGTGCAAGGCTGGATGA | +++ |
| hES360 | TACCTCGATATAACTTCAATGTTACCATGA | ++ |
| hES361 | ATTAGGTGTAGGAACCTGTTAAATGGATGA | + |
| hES362 | TAAAGTGAAGGCGCGGTGGGAGAAATGTGA | ++ |
| hES363 | GGTGACGTCAGACTCAAAGAAGGAAACACG | ++ |
| hES365 | ATATGTCGGCCGAAGACTCTGATAACAAACA | ++ |
| hES366 | TCGTACAGATACGTTAAAAATTCTTAAGGA | + |
| hES367 | ACGGCAGAGGATAGCCGGGAGGCAAATATA | ++ |
| hES368 | GCACGTCATAGACAAAACGTCGGGTCATTA | +++ |
| hES369 | AAGACCAAAGCGGACAAATGTTTGATGAGG | ++ |
| hES370 | ACGTAACGAGCCGAATCGGAGAAACGTTAA | ++ |
| hES371 | AGTACACAGGAAAAAGACGCGAGTTCATGA | + |
| hES372 | AGTGTACAGTGCAGTACATGACGTATGTTA | N.D. |
| hES373 | AACTTAAGTAAACTGATAATATCCCTGTGA | ++ |
| hES374 | AGCCTATTGGAGGATGAGACTTCTTTAAGC | + |
| hES375 | ACGTCAATGTTGCCCAATAGGACAGTAACG | + |
| hES376 | AAGAGCTACATGATGCTTAGTCACGCATGA | ++ |
| hES377 | TGGCTGGCCAAACAAGTCAGACGGACGTGA | N.D. |
| hES378 | AAAACGTCACAATCGACCAATTGAGCACGC | N.D. |
| hES379 | ACCAGTAAGCGGTAATCCAAAAGAGGGTGA | + |
| hES380 | ACGAACAGAAACCCTGTTAGGCCTCGTTAG | + |
| hES381 | AAGGAGGGTGGGACCTAGTGAACGGTATGA | +++ |
| hES383 | TGGCGAAGAAAGAGGGAACTAGTGTGATGA | ++ |
| hES385 | ACCGAGTTAAAAAACAGGGATTAGGTTCG | N.D. |
| hES387 | CACTTAGAAGGGTGCCGCCGTTGCCGACAA | + |
| hES388 | TGTGGGTGTGCTCCACTCCAGCAATAGTGA | ++ |
| hES389 | AAACAAGCGAATGACGTCGAATTATGGAA | + |

| | | |
|---|---|---|
| hES391 | GAAAGAATGACCCAAACTGACTAGTCTTAA | + |
| hES392 | AACGAGTGGGTGGGGGTAGCATGCGTACTTA | ++ |
| hES393 | TACGTCATGGAGACAAACCTGCAATTGTA | ++++ |
| hES394 | TAAAAGGAAAGGTCTAGACAATGAGATTTA | ++ |
| hES395 | AGCAGGGGAGGTGATGAAGGAGTCTCATGA | ++ |
| hES396 | AGTTGGCGGAAGCCAGGCCGCACCGTGTGA | +++ |
| hES397 | AGGGTGGGCGTGTGA | +++ |
| hES398 | CACTGACGTATGTAAGAGGAGATGAGATTA | ++++ |
| hES400 | CCAAAATAGCACGGGACTGCAAGTAGGTGA | +++ |
| hES401 | TCGATAGTCGTGTACTGGACAGCAGTATGA | +++ |
| hES402 | ATCTTCGGCCGATAAGCTGGTTGAGTATCA | ++ |
| hES404 | TAGGAATCGGTAAATAAGACGTCATGGATA | ++++ |
| hES405 | CAATCAAAGGCTTAACTAGGAACAGGTCCA | ++ |
| hES406 | ATACGAATGCCGAGATTTAAAGGCTTATGA | ++ |
| hES407 | TCATGTGTATAAAGTGCGAAGGTGGAGGA | + |
| hES408 | AAAGCGCAAGACTAGATCAGTGAGTAATGA | ++ |
| hES409 | AAAGCTTAAAGTGGGGGTCCCCCGCACTAAA | ++ |
| hES410 | TCGCGTCTTAAGACATAAAAAGTCTGATGA | ++++ |
| hES411 | ACATGAAAGCATCTCATAACCCGGGGGGTATGA | ++ |
| hES413 | AACTACGTAGTTCTACCATGATAGGGATTA | ++ |
| hES415 | ATTACGTTCCATAGGGTGACTGAGGGATGA | ++ |
| hES416 | TATTAGTTAGCGACAGTGACGTAGGGGGAA | ++++ |
| hES417 | ATGTCACGGGTAGAGCTCTAATCCAGGGAC | + |
| hES418 | ATACCGGGATCTACCTGGAGCTGCGTGTGA | + |
| hES420 | TGAGGAGATCTATAGTATTATGAGCTAGAG | + |
| hES421 | TTTAACAGGCGGGAGAACCTAAGGGGATTA | + |
| hES422 | ACAGCCGAGGTGGATCCATAGAAGAGGTTA | + |
| hES423 | AGTGGATGAGGATAGGGAGGTAGAAGGTGA | ++ |
| hES424 | AGCCTCGAAGTCGGTAGGTAAAGAGTGTGA | + |
| hES425 | AAAAATGCAATGGACCCCCGATCGTTATGA | +++ |
| hES426 | CCCCAGGTGGGGGGACGAACCCGTCATACGA | +++ |
| hES427 | TACTCGCTTATATACGTCAGGATGTCTAGG | + |
| hES428 | ACGGTATCAACGGAAACCGAACGCACATTA | ++ |
| hES430 | CGGCCGATGTATACGACACGTAATGGATTA | ++++ |
| hES431 | AACCGTGGCACGGATTTGTGACTTGTCTTA | + |
| hES432 | AGATCGACGGGTCTCGGCAGTGGTACGTAA | + |
| hES433 | TGTGACATCACTAGCAGTTGAGCAGGGCTA | ++ |
| hES434 | CCTGAGGTGTCGGCGCTGGGAATTACGTAG | + |
| hES435 | ATGCCGAGGGGGTACGCGTGACGTGGGTAC | ++ |
| hES436 | ATCGATTGGCCAAGGACAGGCGCACGATTA | ++++ |
| hES437 | ACTAACACCGGCACAAGCTTATTATGGTGA | ++ |

| | | |
|---|---|---|
| hES438 | CGATACGTAAGCTAGTCTAGAACCAATATC | + |
| hES439 | TATATTCACTGAGGACTTAGAGGGATATGA | + |
| hES440 | CTCAACCGAGTGAGACCGCATGGTGA | +++ |
| hES441 | ACGTGTGCCGAATAGAGGTGACATAGCAGA | + |
| hES442 | TGAACATTCACTCAGTCACAATCTAGGTGA | + |
| hES443 | AATCATTGGTCGCCCGCAGAAAGAGAGGTA | ++++ |
| hES444 | TATAAGAGGGGCGTGATATTA | + |
| hES445 | AAATAAAGTTTGGCCCTCCATTCAAGATTA | ++++ |
| hES446 | AACAAAGCTTTAGAGATGAGTCTATGATTA | + |
| hES447 | ATTCGGAACATTACGTAACACGTCCGCGTA | ++ |
| hES448 | AGTACGGGGAGTCTCATCAGTTGTGACGTA | + |
| hES449 | AGCGGGGGTAGATATGAGATCACACACAAC | ++ |
| hES450 | AGTGAACGCCACCTTCAACTCGAATGA | +++ |
| hES451 | AGCTACCAAAGGATAAGCATCAAAGATGA | + |

[a]An A→T substitution, which creates a stop codon, is shown in boldface font.

[b]Potential hotspot motifs discovered in our analysis are highlighted as shown below, but may not be responsible for the observed hotspot activity. See text for further details. Note that some motifs, e.g. the *CRE* core or *ade6-4003* motifs, are often at the ends of the sequence and are completed by the invariant adjacent nucleotides. Numbers in parentheses below indicate the number of times the indicated motif or its complement is found.

*CRE* core TGACGT (109)
*ade6-4008* TTACGTA (11)
*ade6-4003* RTGACATCAT (139)
*ade6-4010* ATGACATAAT (1)
*ade6-4071* ATGATGTCAC (1)

CCAAT motif CCAAT (53)

Oligo-C motif CCCCRC (22)

*ade6-4095* motif GGTCTRGAC (4)

8-6 motif TCGGCCGA (2)

[c]The activity of most strains was determined once or twice as indicated. Plasmid x chromosome recombination was determined qualitatively by the density of papillae present on patches of each strain in comparison to strains containing alleles *ade6-M375*, *ade6-M26*, or *ade6-3074*.

\+ = activity greater than *ade6-M375* but less than *ade6-M26*
++ = activity comparable to *ade6-M26*.
+++ = activity intermediate between *ade6-M26* and *ade6-3074*.
++++ = activity comparable to *ade6-3074*.

**TABLE S2a**

**Common motifs found among total hotspot sequences[a]**

| | 8 base Motif | Count[b] | | 7 base Motif | Count | | 6 base Motif | Count |
|---|---|---|---|---|---|---|---|---|
| 1 | GTGACGTA | 14 | 1 | ACGTCAT | 31 | 1 | ACGTCA | 75 |
| 2 | TGACGTCA | 12 | 2 | TACGTCA | 27 | 2 | ATGACG | 45 |
| 3 | TGACGTAA | 11 | 3 | ACGTCAC | 25 | 3 | ACGTAA | 36 |
| 4 | ATGACGTA | 11 | 4 | GACGTCA | 21 | 4 | GACGTA | 35 |
| 5 | ACGTCATA | 11 | 5 | ACGTAAT | 18 | 5 | GACGTC | 32 |
| 6 | ATACGTCA | 10 | 6 | CGTCATA | 17 | 6 | CGTCAC | 30 |
| 7 | ATGACGTC | 10 | 7 | CACGTCA | 14 | 7 | ATTACG | 24 |
| 8 | ACGTCATG | 10 | 8 | GACGTAA | 14 | 8 | GTATGA | 24 |
| 9 | ACGTCACT | 8 | 9 | ATACGTC | 12 | 9 | TCATGA | 24 |
| 10 | ACGTCACA | 7 | 10 | ACGTCAG | 11 | 10 | ACGTAT | 22 |
| 11 | CGTCATAC | 7 | 11 | CATGACG | 11 | 11 | GTCATA | 22 |
| 12 | AAAGATGA | 6 | 12 | AGTGACG | 11 | 12 | TACGTA | 22 |
| 13 | TGACATCA | 6 | 13 | TACGTAA | 11 | 13 | ACCAAT | 21 |
| 14 | ACGTAATA | 6 | 14 | GTATGAC | 10 | 14 | CACGTC | 20 |
| 15 | ATTACGTC | 6 | 15 | GGTATGA | 10 | 15 | GTCACA | 20 |
| 16 | ATGACGTG | 6 | 16 | GTCACAA | 10 | 16 | ACATGA | 20 |
| 17 | AATGACGT | 5 | 17 | AAAGATG | 10 | 17 | AAGTGA | 19 |
| 18 | ATTACGTA | 5 | 18 | CGTCATC | 9 | 18 | TCATCA | 19 |
| 19 | ACGTAATG | 5 | 19 | ATGTCAC | 9 | 19 | TCATAA | 19 |
| 20 | ACGTCATC | 5 | 20 | AACGTCA | 9 | 20 | AAAGAT | 19 |
| 21 | AAGTGACG | 5 | 21 | CCAATGA | 9 | 21 | GATGAC | 18 |
| 22 | TCACGTCA | 5 | 22 | CCCCGCA | 8 | 22 | CATGAC | 17 |
| 23 | ACGTCACC | 5 | 23 | GTCTAGA | 8 | 23 | TCACAA | 17 |
| 24 | CTGACGTC | 5 | 24 | ACGTAAC | 8 | 24 | ACACGT | 16 |
| 25 | ATAATTGG | 4 | 25 | ATGACGC | 8 | 25 | AGTGAC | 16 |
| 26 | TTACGTAA | 4 | 26 | CATTACG | 8 | 26 | ATGTCA | 16 |
| 27 | AGTAATGA | 4 | 27 | AAGTGAC | 8 | 27 | CAATGA | 16 |
| 28 | AATGGATA | 4 | 28 | AACCAAT | 8 | 28 | TAATGA | 16 |
| 29 | AGAGCTCT | 4 | 29 | CCCGCAG | 7 | 29 | ATATGA | 16 |
| 30 | AGAAGGTG | 4 | 30 | CCCACCC | 7 | 30 | GAGTGA | 15 |
| 31 | TAAGCTTA | 4 | 31 | ACGTGTG | 7 | 31 | CCAATC | 15 |
| 32 | AAACGTCA | 4 | 32 | ACACGTC | 7 | 32 | TAACGA | 15 |
| 33 | GAAGGTGA | 4 | 33 | CGTCACA | 7 | 33 | GGATGA | 15 |
| 34 | AAGCATGA | 4 | 34 | ACGTATC | 7 | 34 | AAGCAT | 15 |
| 35 | AATTACGT | 4 | 35 | CGCAGTA | 7 | 35 | ATTATG | 15 |
| 36 | AAAGGACG | 4 | 36 | ACGTCAA | 7 | 36 | CGTCAG | 14 |
| 37 | AACGTCAT | 4 | 37 | CCAATCA | 7 | 37 | CATACC | 14 |
| 38 | GATTGTGA | 4 | 38 | GACATCA | 7 | 38 | ACGTCT | 14 |

| | | |
|---|---|---|
| 39 | ATTGTGAC | 4 |
| 40 | ACTTACAT | 4 |
| 41 | AGGGATGA | 4 |
| 42 | GACGTATA | 4 |
| 43 | CCAATGAG | 4 |
| 44 | CGCAGTAA | 4 |
| 45 | ACTACGTA | 4 |
| 46 | ATTGGCGG | 4 |
| 47 | AGACGTCA | 4 |
| 48 | ATTGGACG | 4 |
| 49 | CCATTACG | 4 |
| 50 | ACACGTCA | 4 |
| 51 | AACAGGCG | 4 |
| 52 | ATCACGTC | 4 |
| 53 | GGTATGAC | 4 |
| 54 | AGTCTAGA | 4 |
| 55 | GTCTAGAC | 4 |
| 56 | AGACGCAG | 4 |
| 57 | CATACGTC | 4 |
| 58 | AGCTTACG | 4 |
| 59 | GACGTCAC | 4 |
| 60 | TCGGCCGA | 4 |
| 61 | CGTCATGC | 4 |
| 62 | CACGTCAC | 4 |
| 63 | ACTGCGGG | 4 |
| 64 | ACGTCACG | 4 |
| 65 | CGCATGAC | 4 |
| 66 | CCCCGCAG | 4 |
| 67 | CGCCGACG | 4 |
| 68 | ATAATTGG | 4 |

| | | |
|---|---|---|
| 39 | CACCTTC | 7 |
| 40 | TACATGA | 7 |
| 41 | CGTAATA | 7 |
| 42 | AAGCTTA | 7 |
| 43 | GAAGTGA | 7 |
| 44 | AAGATGA | 7 |
| 45 | CGGCCGA | 6 |
| 46 | CCCCCAC | 6 |
| 47 | ACCCCCT | 6 |
| 48 | ACAGGCG | 6 |
| 49 | CTGCGTC | 6 |
| 50 | CGTCACC | 6 |
| 51 | CGTGTGA | 6 |
| 52 | CGCATGA | 6 |
| 53 | GCATGAC | 6 |
| 54 | GACGTGA | 6 |
| 55 | AGACGCA | 6 |
| 56 | TGACGCA | 6 |
| 57 | ACGTAAG | 6 |
| 58 | CGTCCAA | 6 |
| 59 | ACGTAGT | 6 |
| 60 | CCGCCAA | 6 |
| 61 | AGCTTAC | 6 |
| 62 | ACGTGAT | 6 |
| 63 | GTAACGA | 6 |
| 64 | AGAGCTC | 6 |
| 65 | ATTGGAC | 6 |
| 66 | CCTGTTA | 6 |
| 67 | CATCACC | 6 |
| 68 | GGTAACA | 6 |
| 69 | AATGACG | 6 |
| 70 | GATGTCA | 6 |
| 71 | AAACGTC | 6 |
| 72 | AGATGAC | 6 |
| 73 | AAGACGT | 6 |
| 74 | ACGTATA | 6 |
| 75 | CTTGTGA | 6 |
| 76 | GTGATGA | 6 |
| 77 | AAGGTGA | 6 |
| 78 | CCAATTA | 6 |
| 79 | GTAATGA | 6 |

| | | |
|---|---|---|
| 39 | GCATGA | 14 |
| 40 | CATTGG | 14 |
| 41 | CATTAC | 14 |
| 42 | TCTAGA | 14 |
| 43 | CAATCA | 14 |
| 44 | AGATCT | 14 |
| 45 | ATTGGA | 14 |
| 46 | AGATGA | 14 |
| 47 | ATGATG | 14 |
| 48 | GTAATA | 14 |
| 49 | CATGCG | 13 |
| 50 | CTAGAC | 13 |
| 51 | GACGCA | 13 |
| 52 | ACTGCG | 13 |
| 53 | GCGTCA | 13 |
| 54 | ACGTAG | 13 |
| 55 | CATCCC | 13 |
| 56 | GTGTGA | 13 |
| 57 | AACGTC | 13 |
| 58 | AGCTTA | 13 |
| 59 | AGGTGA | 13 |
| 60 | GTAACA | 13 |
| 61 | CTATTA | 13 |
| 62 | ACATCA | 13 |
| 63 | ATGTAA | 13 |
| 64 | AAAGTG | 13 |
| 65 | CCCGCA | 12 |
| 66 | CACACG | 12 |
| 67 | GGGGTA | 12 |
| 68 | ACGCAG | 12 |
| 69 | GGGTGA | 12 |
| 70 | ACGTGA | 12 |
| 71 | GATCTA | 12 |
| 72 | GTGACA | 12 |
| 73 | GTCTTA | 12 |
| 74 | ACGTTA | 12 |
| 75 | CATGTA | 12 |
| 76 | ATGAGG | 12 |
| 77 | CATAAG | 12 |
| 78 | AAGACG | 12 |
| 79 | ATAAGG | 12 |

| | | | | | |
|---|---|---|---|---|---|
| 80 | AATGGAT | 6 | 80 | GAATGA | 12 |
| 81 | ATCATGA | 6 | 81 | CCAATA | 12 |
| 82 | AATTGGT | 6 | 82 | ACGAAA | 12 |
| 83 | ATAGCAA | 6 | 83 | ATCATA | 12 |
| 84 | ACCAATA | 6 | 84 | AAGCTT | 12 |
| 85 | TCATGAA | 6 | 85 | CAATTA | 12 |
| 86 | ATTATGA | 6 | 86 | AATAAG | 12 |
| 87 | ATGATGA | 6 | 87 | TAACAA | 12 |

[a]The most common six, seven, and eight base motifs are shown above and summarized in the table below. All sequences occurring $\geq$4, $\geq$6, or $\geq$12 times are shown for eight, seven, and six base motifs, respectively. Sequences tested for hotspot activity by reconstruction (Figure 3 and Table S4) are shown in red font.

[b]Indicates the number of times that a given motif is found among the complete set of hotspot sequences shown in Table S1.

**TABLE S2b**

**Summary of motif frequency and comparison to randomized sequences**

| | 8 base motifs | | 7 base motifs | | 6 base motifs | |
|---|---|---|---|---|---|---|
| number of occurrences[a] | real[b] | random[c] | real | random | random | random |
| ≥4 | 62 | 12 | ---[d] | --- | --- | --- |
| ≥5 | 24 | 1 | --- | --- | --- | --- |
| ≥6 | 16 | 0 | 87 | 30 | --- | --- |
| ≥7 | 10 | 0 | 44 | 8 | --- | --- |
| ≥8 | 8 | 0 | 28 | 0 | --- | --- |
| ≥12 | 1 | 0 | 9 | 0 | 81 | 27 |
| ≥13 | 1 | 0 | 8 | 0 | 59 | 14 |
| ≥14 | 1 | 0 | 8 | 0 | 43 | 7 |
| ≥20 | 0 | 0 | 4 | 0 | 13 | 0 |

[a]Indicates the number of different motifs occurring at a given frequency. For example, among eight base motifs, 62 different motifs were found that occurred four or more times.

[b]Actual motifs found in the sequences from Table S1. This is for comparison to the same sequences following randomization. NOTE: Palindromes were counted only once, though their occurrence on both strands results in being counted twice in Table 2a. Hence, the values shown above are somewhat lower than what seems to be apparent from Table 2a.

[c]Indicates the number of different motifs occurring at a given frequency following randomization of the sequences in Figure S1. Palindromes were counted only once

[d]Lower frequency motifs were not counted.

**TABLE S3a**

**Common motifs among hotspot sequences lacking *CRE*[a]**

| | 8 base | | 7 base | | 6 base | |
|---|---|---|---|---|---|---|
| | Motif | Count[b] | Motif | Count | Motif | Count |
| 1 | TGACATCA | 6 | 1 | ACGTAAT | 14 | 1 | ACGTAA | 22 |
| 2 | ATTACGTA | 5 | 2 | ATGTCAC | 9 | 2 | TACGTA[c] | 18 |
| 3 | ACGTAATA | 5 | 3 | CCAATGA | 9 | 3 | ACCAAT | 18 |
| 4 | CCAATGAG | 4 | 4 | CCCCGCA | 8 | 4 | TCATGA[c] | 18 |
| 5 | AGAGCTCT[c] | 4 | 5 | GTCTAGA | 8 | 5 | ATTACG | 17 |
| 6 | TCGGCCGA[c] | 4 | 6 | TACGTAA | 8 | 6 | ACATGA | 16 |
| 7 | AGACGCAG | 4 | 7 | CCAATCA | 7 | 7 | ATGTCA | 16 |
| 8 | GTCTAGAC[c] | 4 | 8 | CCCACCC | 7 | 8 | TCATCA | 16 |
| 9 | ATAATTGG | 4 | 9 | CCCGCAG | 7 | 9 | AAAGAT | 15 |
| 10 | AACAGGCG | 4 | 10 | CGCAGTA | 7 | 10 | GTATGA | 15 |
| 11 | ATTGGCGG | 4 | 11 | ATGACGC | 7 | 11 | CAATGA | 15 |
| 12 | AAGCATGA | 4 | 12 | GGTATGA | 7 | 12 | TCATAA | 15 |
| 13 | CGCAGTAA | 4 | 13 | AACCAAT | 7 | 13 | CCAATC | 15 |
| 14 | CCCCGCAG | 4 | 14 | AAAGATG | 7 | 14 | TAACGA | 14 |
| 15 | ACTGCGGG | 4 | 15 | CGTCATA | 6 | 15 | CATTGG | 14 |
| 16 | ACTACGTA | 4 | 16 | CGGCCGA | 6 | 16 | CAATCA | 14 |
| 17 | ACGTAATG | 4 | 17 | ACAGGCG | 6 | 17 | TCTAGA[c] | 14 |
| 18 | ACTTACAT | 4 | 18 | CTGCGTC | 6 | 18 | AAGTGA | 14 |
| 19 | GAAGGTGA | 4 | 19 | AGACGCA | 6 | 19 | GACGCA | 13 |
| 20 | AGTCTAGA | 4 | 20 | TGACGCA | 6 | 20 | ACTGCG | 13 |
| 21 | AATGGATA | 3 | 21 | CCGCCAA | 6 | 21 | GAGTGA | 13 |
| 22 | CCATTACG | 3 | 22 | AGAGCTC | 6 | 22 | ATGACG | 13 |
| 23 | CGCCCACC | 3 | 23 | GTCACAA | 6 | 23 | ATTGGA | 13 |
| 24 | CGCATGAC | 3 | 24 | GATGTCA | 6 | 24 | TCACAA | 13 |
| 25 | GTCCAGAC | 3 | 25 | GACATCA | 6 | 25 | ATTATG | 13 |
| 26 | CGGATCTA | 3 | 26 | CACCTTC | 6 | 26 | AAAGTG | 13 |
| 27 | TCCCCGCA | 3 | 27 | CGTAATA | 6 | 27 | GTAATA | 13 |
| 28 | CGGCCGAA | 3 | 28 | GTGATGA | 6 | 28 | CATGCG | 12 |
| 29 | GTGGGGGA | 3 | 29 | CCAATTA | 6 | 29 | ACGCAG | 12 |
| 30 | GCGTGTGA | 3 | 30 | ATGATGA | 6 | 30 | GGGTGA | 12 |
| 31 | CCCGCAGA | 3 | 31 | CCCCCAC | 5 | 31 | GATCTA | 12 |
| 32 | CATACCCC | 3 | 32 | ACCCCCT | 5 | 32 | GTGACA | 12 |
| 33 | GGTCTGGA | 3 | 33 | TCCCCCA | 5 | 33 | AGATCT[c] | 12 |
| 34 | GGTCCAGA | 3 | 34 | CGCATGA | 5 | 34 | AAGCAT | 12 |
| 35 | AACTCGGA | 3 | 35 | ATGCGTC | 5 | 35 | ATGATG | 12 |
| 36 | GCGTTCAC | 3 | 36 | CATACCC | 5 | 36 | TAATGA | 12 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 37 | AGGGTGGG | 3 | 37 | GGATCTA | 5 | 37 | ATATGA | 12 |
| 38 | GACGCATC | 3 | 38 | CGGACAA | 5 | 38 | CCCGCA | 11 |
| 39 | CCCACCCA | 3 | 39 | ACGCAGT | 5 | 39 | CTAGAC | 11 |
| 40 | ACTGCGTC | 3 | 40 | GGCCCAA | 5 | 40 | GCGTCA | 11 |
| 41 | GGGCCCAA | 3 | 41 | GAACCTA | 5 | 41 | CATACC | 11 |
| 42 | ACGCATGA | 3 | 42 | ACCGCCA | 5 | 42 | GTCACA | 11 |
| 43 | CGCAGTGA | 3 | 43 | ACGTAAC | 5 | 43 | GCATGA | 11 |
| 44 | GTGACTGA | 3 | 44 | CTACGTA | 5 | 44 | AGGTGA | 11 |
| 45 | CGCACTAA | 3 | 45 | CGTCCAA | 5 | 45 | CTATTA | 11 |
| 46 | GCGGACAA | 3 | 46 | ACGTAGT | 5 | 46 | ACATCA | 11 |
| 47 | CCGCAGTA | 3 | 47 | CTCGTTA | 5 | 47 | ATGTAA | 11 |
| 48 | CATCCCAC | 3 | 48 | CCTATCA | 5 | 48 | CAATTA | 11 |
| 49 | AAGTCTAG | 3 | 49 | GTAACGA | 5 | 49 | CCCCGC | 10 |
| 50 | GAGCTCTA | 3 | 50 | CTAGAAC | 5 | 50 | CGGCCG | 10 |
| 51 | ATTGGGCC | 3 | 51 | CATTACG | 5 | 51 | CCCACC | 10 |
| 52 | AACGAGTG | 3 | 52 | ATTGGCG | 5 | 52 | GGTCCA | 10 |
| 53 | AGTGAACG | 3 | 53 | ACTTAGA | 5 | 53 | GGGGTA | 10 |
| 54 | GGATCTAA | 3 | 54 | ATTGGAC | 5 | 54 | CCACCC | 10 |
| 55 | GCCCACCC | 3 | 55 | CCTGTTA | 5 | 55 | CCGCCA | 10 |
| 56 | GCGTCATA | 3 | 56 | AGTGAAC | 5 | 56 | GCCGAA | 10 |
| 57 | CATTGGCG | 3 | 57 | TAACGTA | 5 | 57 | CATCCC | 10 |
| 58 | GAACCTAA | 3 | 58 | CATCACC | 5 | 58 | ACGCAT | 10 |
| 59 | ACAGGGAT | 3 | 59 | CAATCAG | 5 | 59 | AGTGAG | 10 |
| 60 | CACTTAGA | 3 | 60 | CGCAGAA | 5 | 60 | GTGTGA | 10 |
| 61 | CTATCCTC | 3 | 61 | ACGTTAA | 5 | 61 | GAGCTC | 10 |
| 62 | ATTGGACG | 3 | 62 | GGTAACA | 5 | 62 | ACTCTA | 10 |
| 63 | AACCGCCA | 3 | 63 | GATTGGA | 5 | 63 | ACGTTA | 10 |
| 64 | ATTCGCTG | 3 | 64 | AGCATGA | 5 | 64 | GACATA | 10 |
| 65 | AGATCTCC | 3 | 65 | AATGCAC | 5 | 65 | ATGAGG | 10 |
| 66 | CCAATCAG | 3 | 66 | ACTTACA | 5 | 66 | CATCAC | 10 |
| 67 | ATGACGCA | 3 | 67 | ATGACGA | 5 | 67 | AACAGG | 10 |
| 68 | GATGTCAC | 3 | 68 | ATCGTTA | 5 | 68 | GCAGTA | 10 |
| 69 | ATGTCACC | 3 | 69 | CAATGAG | 5 | 69 | ATTGGC | 10 |
| 70 | GACATCAC | 3 | 70 | TACATGA | 5 | 70 | TGGCCA | 10 |
| 71 | GCCTATTA | 3 | 71 | TGACGAA | 5 | 71 | ATAAGG | 10 |
| 72 | CATACTGC | 3 | 72 | ACCAATG | 5 | 72 | AAGGCA | 10 |
| 73 | CAATCAGC | 3 | 73 | AACGTAA | 5 | 73 | CACCAA | 10 |
| 74 | CCGCAGAA | 3 | 74 | CTTGTGA | 5 | 74 | AGTGAA | 10 |
| 75 | CTTACTGC | 3 | 75 | ATACTGC | 5 | 75 | ACATAA | 10 |
| 76 | AACTACGT | 3 | 76 | CAATGGA | 5 | 76 | AATGGA | 10 |
| 77 | ACAGCGAA | 3 | 77 | GAAGTGA | 5 | 77 | ATCATA | 10 |

| | | |
|---|---|---|
| 78 | GTGCATTA | 3 |
| 79 | GCCCTTTA | 3 |
| 80 | GTTCTAGA | 3 |
| 81 | CTATTACC | 3 |
| 82 | ATGTCACA | 3 |
| 83 | GATGCTTA | 3 |
| 84 | AGTCACAA | 3 |
| 85 | AAATGGGG | 3 |
| 86 | CTTACATC | 3 |
| 87 | CAATGAGG | 3 |
| 88 | TAACGTAA | 3 |
| 89 | CAGTATGA | 3 |
| 90 | GCCAATGA | 3 |
| 91 | GGATGTAA | 3 |
| 92 | GGTGATGA | 3 |
| 93 | AAGTGACA | 3 |
| 94 | ATAAGGCA | 3 |
| 95 | ATCCATTG | 3 |
| 96 | AATTACGT | 3 |
| 97 | ATTTCGTC | 3 |
| 98 | CAGTAAGA | 3 |
| 99 | AGACATAA | 3 |
| 100 | ATGACGAA | 3 |
| 101 | ATTATGTG | 3 |
| 102 | AACGTAAT | 3 |
| 103 | ATCATTGG | 3 |
| 104 | CGTAATAA | 3 |
| 105 | ATTGGTTG | 3 |
| 106 | ACCAATGA | 3 |
| 107 | AGAAGGTG | 3 |
| 108 | ACCAATTA | 3 |
| 109 | ATCTTTGG | 3 |
| 110 | GCCGGACC | 3 |
| 111 | AATAGTGA | 3 |
| 112 | AGCATTTC | 3 |
| 113 | AATTTCGT | 3 |

| | | |
|---|---|---|
| 78 | TCATCCA | 5 |
| 79 | ATTGGTG | 5 |
| 80 | CACTTTA | 5 |
| 81 | AAGGTGA | 5 |
| 82 | ATCCAAT | 5 |
| 83 | GATATGA | 5 |
| 84 | AATGGAT | 5 |
| 85 | AAAGTGG | 5 |
| 86 | ATCATGA | 5 |
| 87 | AATTGGT | 5 |
| 88 | ACATTTC | 5 |
| 89 | TCATGAA | 5 |
| 90 | ATTATGA | 5 |

| | | |
|---|---|---|
| 78 | CATTTC | 10 |
| 79 | CTTTAA | 10 |
| 80 | AATAAG | 10 |
| 81 | ATAATG | 10 |
| 82 | CACGCC | 9 |
| 83 | ACCGAG | 9 |
| 84 | CACACG | 9 |
| 85 | ACCGCC | 9 |
| 86 | ACGTAG | 9 |
| 87 | ACACGT | 9 |
| 88 | GTGGGA | 9 |
| 89 | ACTACG | 9 |
| 90 | AAGCGG | 9 |
| 91 | CACCCA | 9 |
| 92 | ACGACA | 9 |
| 93 | AACGAG | 9 |
| 94 | GAGGTA | 9 |
| 95 | TCGTCA | 9 |
| 96 | CAGTGA | 9 |
| 97 | GACGAA | 9 |
| 98 | CATGTA | 9 |
| 99 | GGATTA | 9 |
| 100 | ATCACC | 9 |
| 101 | AGAGCT | 9 |
| 102 | CATAAG | 9 |
| 103 | GGATGA | 9 |
| 104 | CTTACA | 9 |
| 105 | CTATCA | 9 |
| 106 | ATAACG | 9 |
| 107 | AAAGCG | 9 |
| 108 | GTGATA | 9 |
| 109 | GTAACA | 9 |
| 110 | ATCTCA | 9 |
| 111 | ATCATG | 9 |
| 112 | AATCTC | 9 |
| 113 | AAGGTG | 9 |
| 114 | CTGTTA | 9 |
| 115 | AATGTG | 9 |
| 116 | AAGGAG | 9 |
| 117 | ACAATG | 9 |
| 118 | GAATGA | 9 |

| | | |
|---|---|---|
| 119 | AATTGG | 9 |
| 120 | AACGAA | 9 |
| 121 | CCAATA | 9 |
| 122 | AACCAA | 9 |
| 123 | ACGAAA | 9 |

[a]The most common six, seven, and eight base motifs among 398 sequences lacking the *CRE* core sequence, TGACGT. Sequences tested for hotspot activity by reconstruction (Figure 3 and Table S4) are shown in red font.

[b]Indicates the number of times that a given motif is found among the 398 sequences lacking the *CRE* core sequence.

[c]Palindrome.

**TABLE S3b**

**Summary of motif frequency and comparison to randomized sequences**

| | 8 base motifs | | 7 base motifs | | 6 base motifs | |
|---|---|---|---|---|---|---|
| number of occurrences[a] | real[b] | random[c] | real | random | real | random |
| ≥4 | 17 | 4 | ---[d] | --- | --- | --- |
| ≥5 | 2 | 0 | --- | --- | --- | --- |
| ≥6 | 1 | 0 | 30 | 10 | --- | --- |
| ≥7 | 0 | 0 | 14 | 3 | --- | --- |
| ≥8 | 0 | 0 | 6 | 0 | --- | --- |
| ≥12 | 0 | 0 | 1 | 0 | 33 | 9 |
| ≥14 | 0 | 0 | 1 | 0 | 15 | 1 |
| ≥18 | 0 | 0 | 0 | 0 | 2 | 0 |

[a]Indicates the number of different motifs occurring at a given frequency. For example, among eight base motifs 17 different motifs were found that occurred four or more times.

[b]Indicates the number of different motifs occurring at a given frequency among actual hotspot sequences. This is for comparison to random sequences. NOTE: Palindromes were counted only once, though their occurrence on both strands results in being counted twice in Table 3a. Hence, the values shown above are somewhat lower than what seems to be apparent from Table 3a.

[c]Indicates the number of different motifs occurring at a given frequency following randomization of the 398 sequences lacking the *CRE* core sequence. Palindromes were counted only once.

[d]Lower frequency motifs are not listed.

**TABLE S4**

**Subscreen of high frequency motifs[a]**

| % hot[b] | sequence | Motif | |
|---|---|---|---|
| 125/204 (61%)* | NNN**TCGGCCGA**NNN | 8-6 | |
| 139/318 (44%)* | NNN**GTCTAGAC**NNN | 8-8 | >25% hot |
| 50/158 (32%)* | NNN**ATTGGCGG**NNN | 8-11 | |
| 111/395 (28%)* | NNN**CCCCCAC**NNN | 7-31 | |
| 23/118 (19%)* | NNNB**CGTCATA**NNND | 7-15 | |
| 65/612 (11%)* | NNN**CCCCGCA**NNN | 7-4 | |
| 79/852 (9.3%)* | HNNNH**ACGTAA**NNND[c] | 6-1 | >6% hot |
| 29/404(7.2%)* | NNN**CCAATGAG**NNN | 8-4 | |
| 18/292 (6.1%)* | NNN**AATGGATA**NNN | 8-21 | |
| 19/441 (4.3%)* | NNN**ATGTCAC**NNN | 7-2 | |
| 15/409 (3.7%)* | NNN**ATAATTGG**NNN | 8-9 | |
| 10/296 (3.4%)* | NNN**CCATTACG**NNN | 8-22 | |
| 15/560 (2.7%)* | NNN**AGGGATGA**NNN[d] | 8-41 | |
| 38/1621 (2.3%)* | HNNNN**GATGAC**HNND[c,d] | 6-21 | |
| 10/479 (2.1%)* | NNN**CGCAGTAA**NNN | 8-13 | |
| 25/1691 (1.2%)* | HNNNN**CTATTA**NNND[c] | 6-45 | |
| 7/449 (1.6%) | NNN**AAGCATGA**NNN | 8-12 | |
| 7/691 (1.0%) | NNN**AGAGCTCT**NNN | 8-5 | |
| 5/654 (0.8%) | NNN**CCCACCC**NNN | 7-8 | |
| 6/729 (0.8%) | HNNNNN**AAAGAT**HNN[c] | 6-9 | |
| 15/2015 (0.7%) | HNNNN**CATCCC**[c] | 6-57 | |
| 3/454 (0.5%) | NNN**AGACGCAG**NNN | 8-7 | |
| 1/334 (0.3%) | NNN**GTATGA**NNN[c] | 6-10 | |
| 0/160 (0%) | NNN**AACAGGCG**NNN | 8-10 | |

[a]Sequence motifs identified in the primary screen (bold font; Table 2 were re-screened for hotspot activity in the context of some random flanking nucleotides. The fraction of hotspots observed in these experiments is most-likely indicative of how many additional nucleotides, besides those in the indicated motif, may be required for hotspot activity. For example, if a given seven base motif requires one, two, or three specific flanking nucleotides for activity, then approximately 25%, 6.3%, or 1.5% of transformed colonies, respectively, should be observed to contain a hotspot.

[b]Numerator shows the number of hyper-rec colonies and denominator shows the total number of colonies screened. The percentage of colonies containing a hotspot are shown in parentheses. Only hyper-rec colonies capable of plasmid loss are counted.

[c]These motifs were found in an early stage of the screen prior to elimination of six base TGACGT *CRE* sequence from our pool. In some cases, random bases are not completely random in order to prevent formation of a *CRE* hotspot. B, D, H, and V indicate any nucleotide except A, C, G, and T, respectively.

[d]These motifs are found in Table S2a.

*$P<5\times10^{-4}$, the probability that the indicated frequency of hotspots is not greater than 0.6%, the frequency of hotspots observed with a completely random 15 bp sequence. (Chi-squared test with Yates correction.)