FIGURES AND TABLES

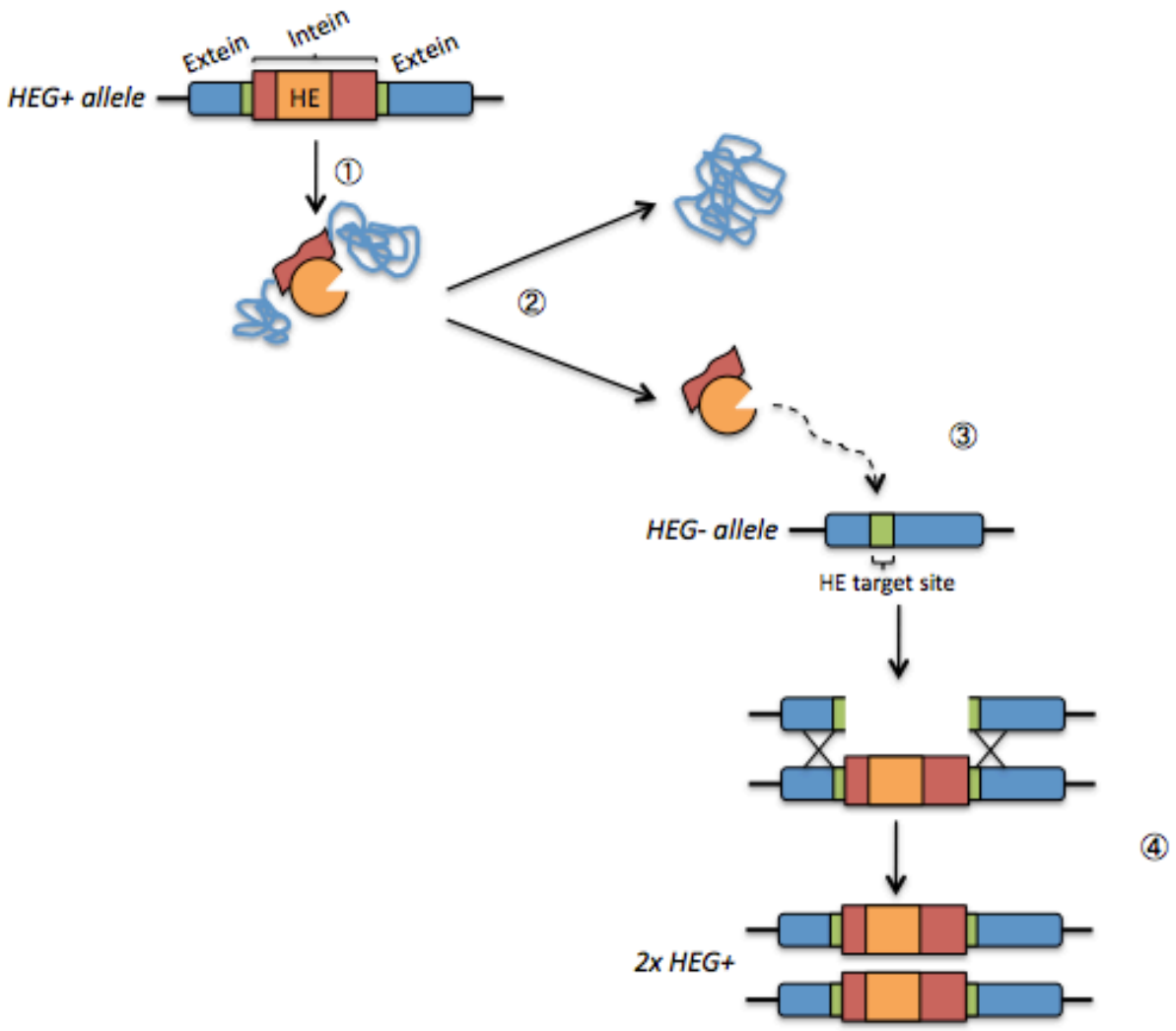


Figure 1. Diagram of the homing process. 1) Transcription and translation of a gene containing an intervening homing endonuclease gene (HEG), denoted HEG+. 2) Following proteolysis and protein ligation catalyzed by the splicing domains of the intein, this protein gives rise to the host extein and the intein, which contains a functional endonuclease domain. 3) If a second, HEG- copy of the same gene is present, the intein will home to that sequence and cleave its target. 4) Double-strand break repair using the HEG+ allele as a template results in transfer of the intein.

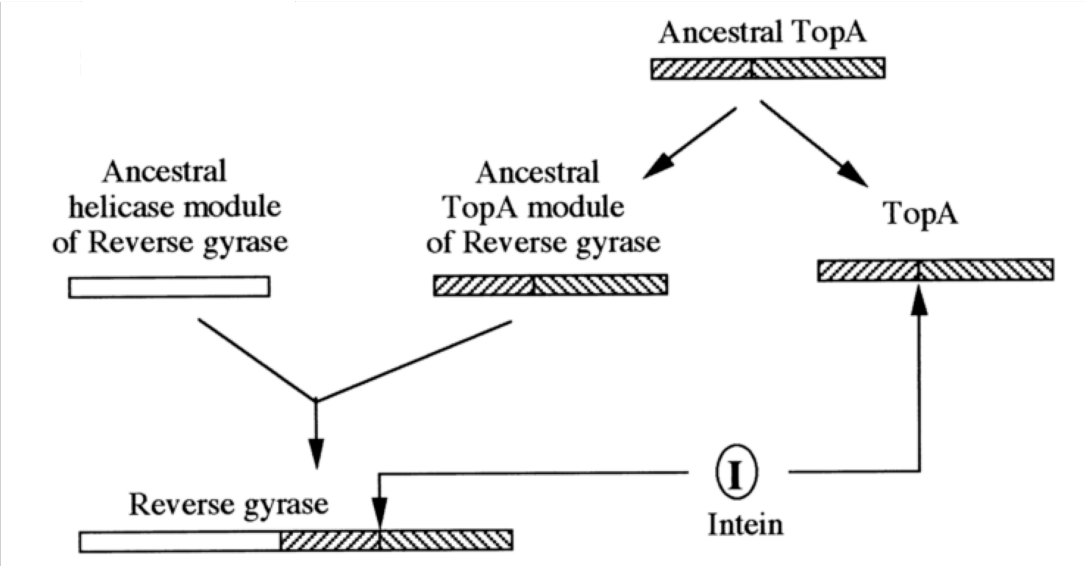


Figure 2. Proposed evolutionary history of the reverse gyrase and topoisomerase I (TopA) genes in hyperthermophiles (adapted from Chute *et al.*, 1998).

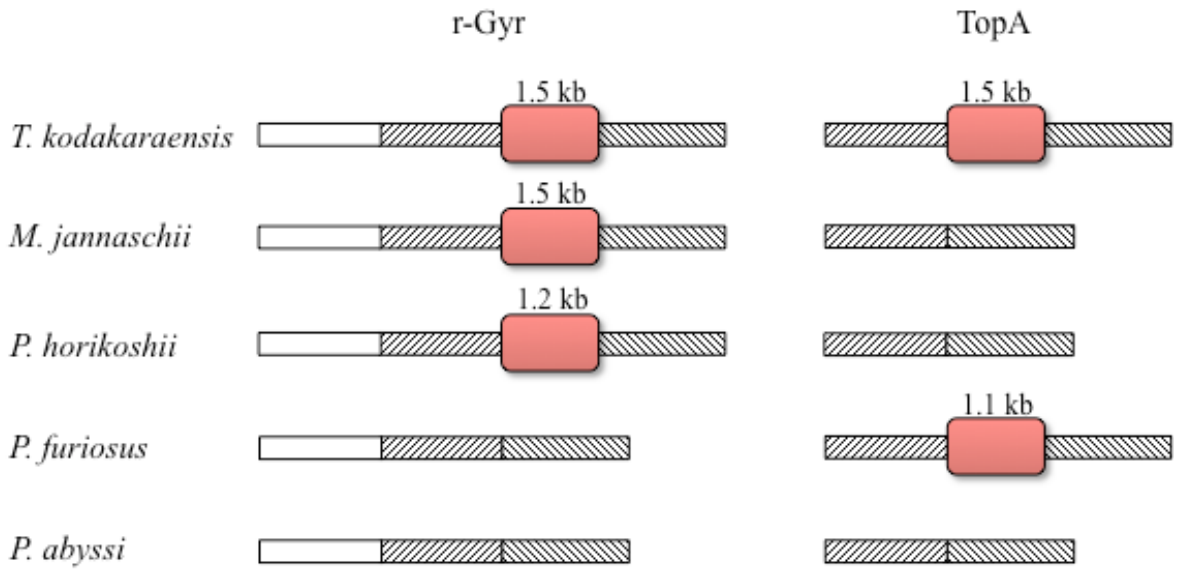


Figure 3. Uneven distribution of the r-Gyr/TopA intein among hyperthermophilic archaea. Among five species that harbor multiple common inteins, only five out of ten potential insertion sites possess the intein, shown in pink. The approximate size of the intervening sequence, in kilobases, is represented above each.

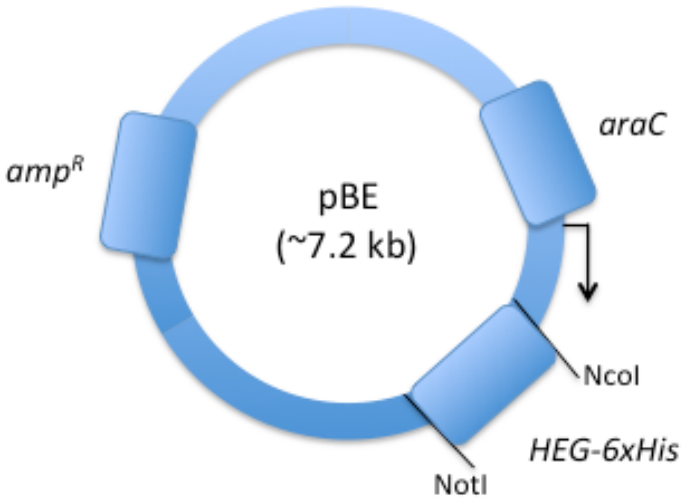


Figure 4. Map of cloned intein expression plasmid, which is derived from pBE and contains a marker for Amp^R. The plasmid contains a poly-His-tagged homing endonuclease gene (HEG) under control of an arabinose-inducible promoter (araC coupled to the pBAD promoter).

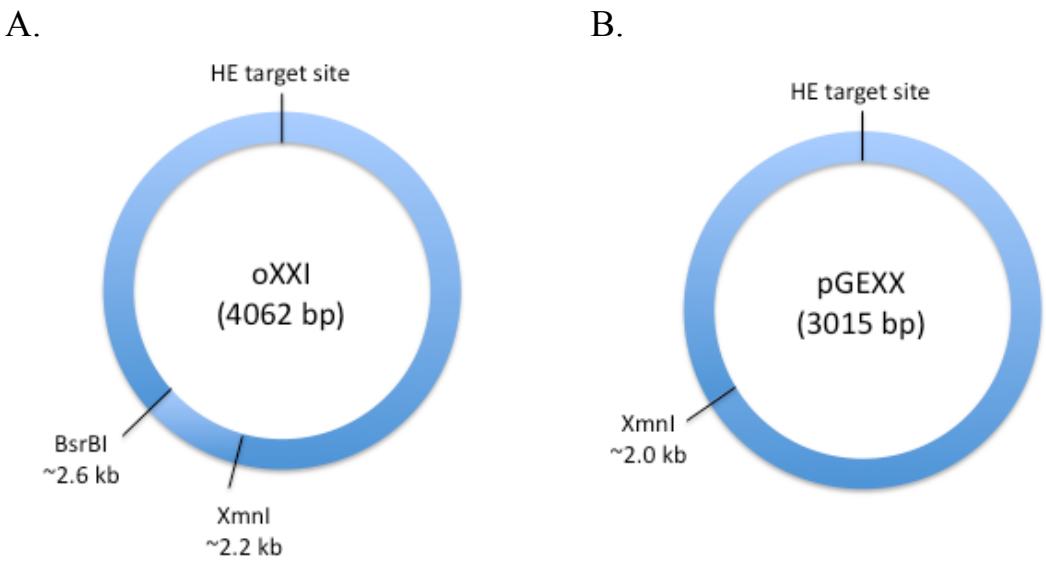


Figure 5. Cut maps of cloned target site plasmids used in this experiment: (A) the oXXI plasmid, which is derived from pBR322 and contains a selectable marker for Tet^R, and (B) the pGEXX plasmid, which is derived from pGEM and contains a marker for Amp^R.

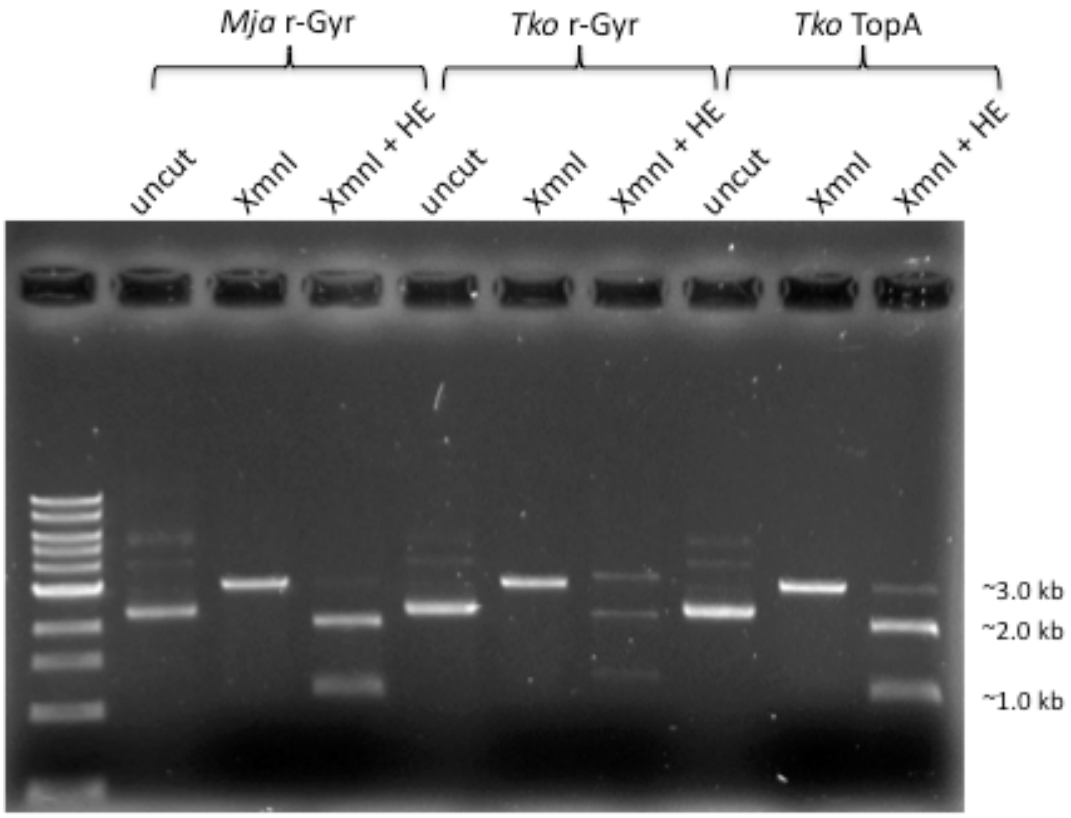


Figure 6. Results of homing endonuclease digests of native target sites cloned into the pGEXX plasmid. Sites from *M. jannaschii* r-Gyr, *T. kodakaraensis* r-Gyr, and *T. kodakaraensis* TopA were linearized at 37°C, then digested with the native HE at 80°C. Gel electrophoresis was used to visualize digest results, along with uncut and linearized (XmnI) controls. Ladder, 1 kb.

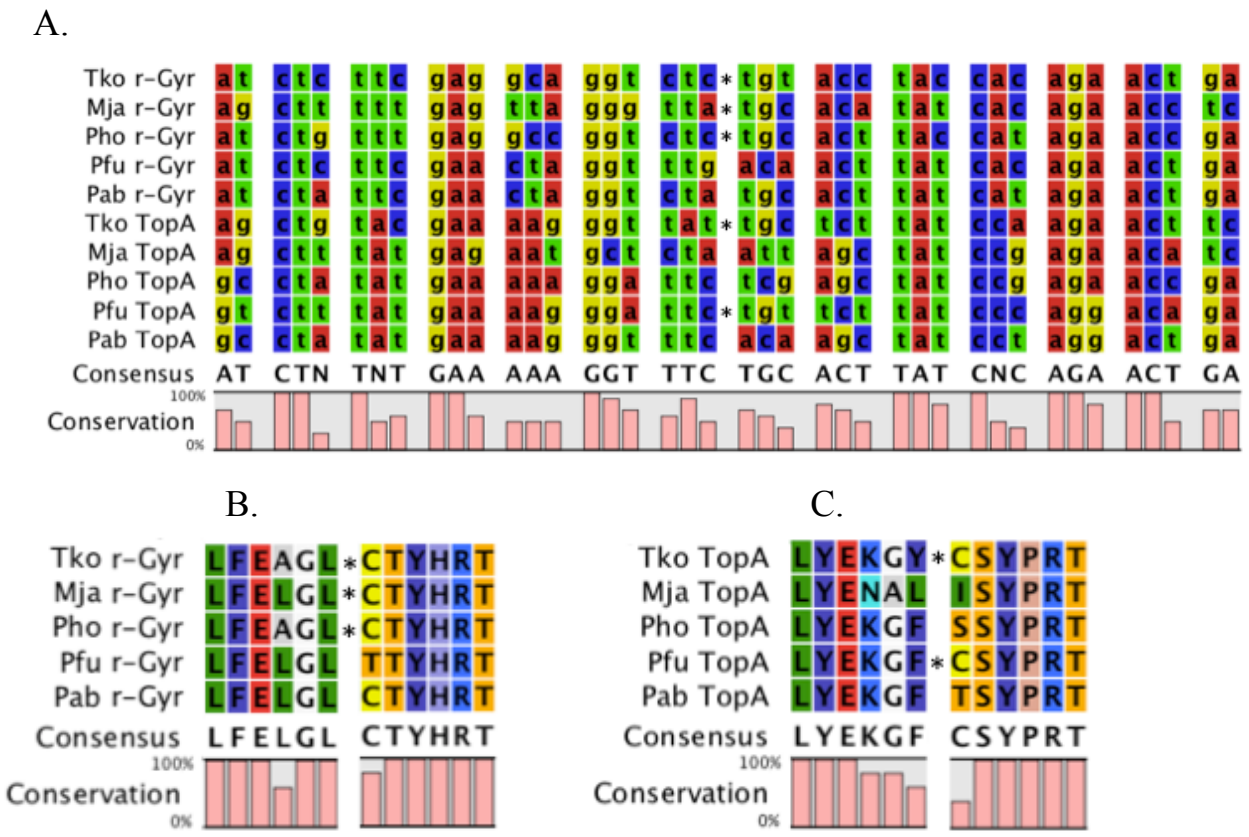


Figure 7. HE target site alignments across the r-Gyr and TopA genes in the five species studied. (A) Alignment of the 40 bases flanking the potential site of insertion demonstrates divergence at the DNA level, while alignment of the corresponding in-frame translation in (B) r-Gyr and (C) TopA demonstrates conservation at the protein level. Asterisks indicate the presence of an intein *in vivo.* Species, *T. kodakaraensis* (Tko), *M. jannaschii* (Mja), *P. horikoshii* (Pho), *P. furiosus* (Pfu), *P. abyssi* (Pab).

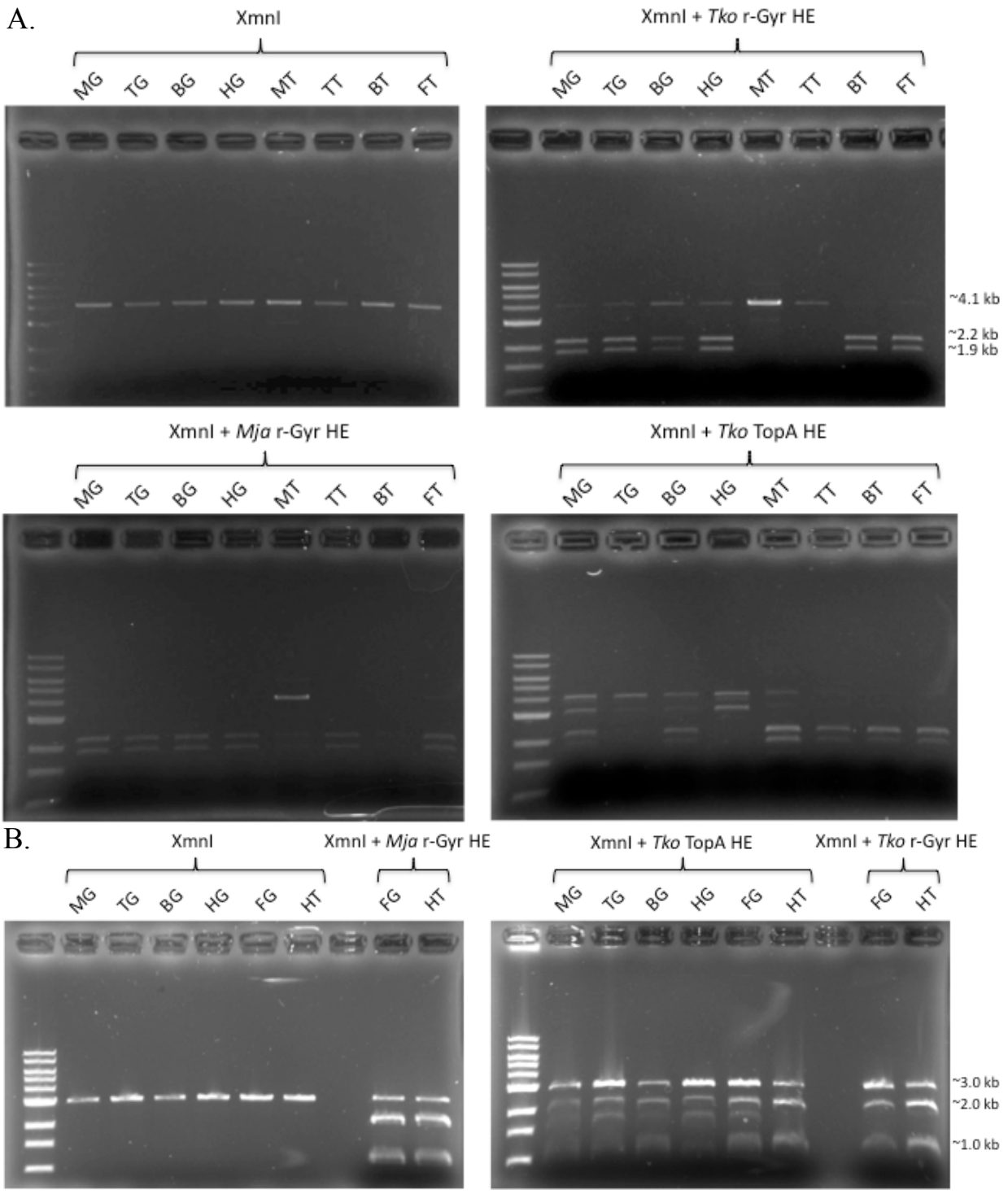


Figure 8. Results of HE digests of ten target sites from five species, cloned into (A) the oXXI site plasmid or (B) the pGEXX site plasmid. Sites were linearized at 37°C, then digested with the indicated HE at 80°C. For each lane, the target site is listed by two letter abbreviation: *Mja* r-Gyr (MG), *Tko* r-Gyr (TG), *Pab* r-Gyr (BG), *Pho* r-Gyr (HG), *Pfu* r-Gyr (FG), *Mja* TopA (MT), *Tko* TopA (TT), *Pab* TopA (BT), *Pho* TopA (HT), *Pfu* TopA (FT). Ladder, 1 kb.

A. B.

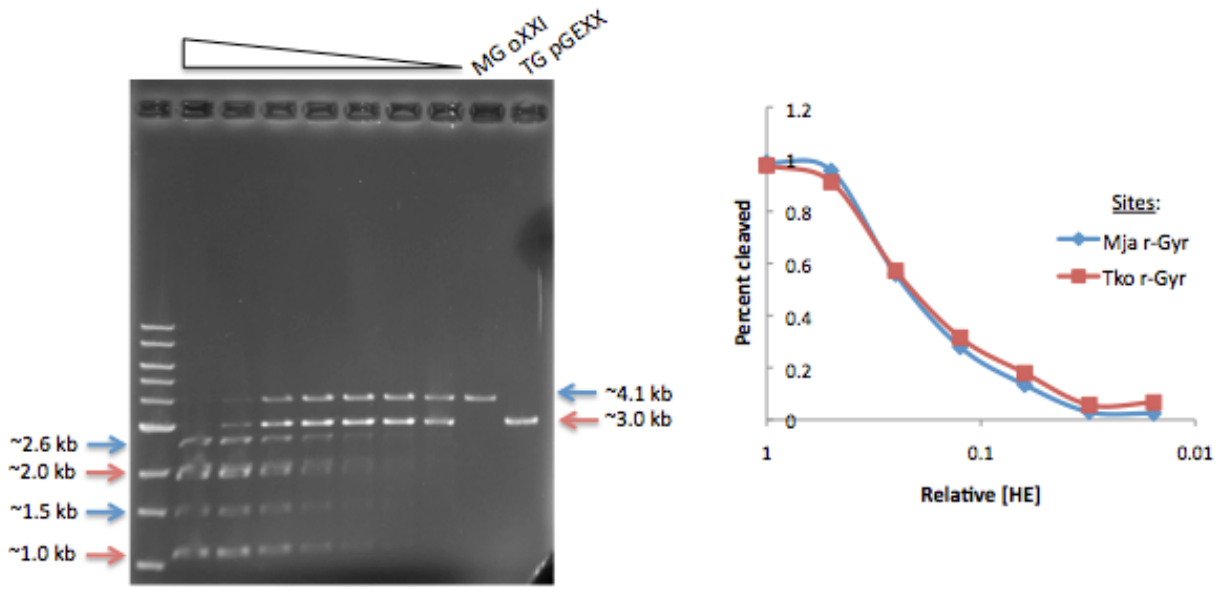
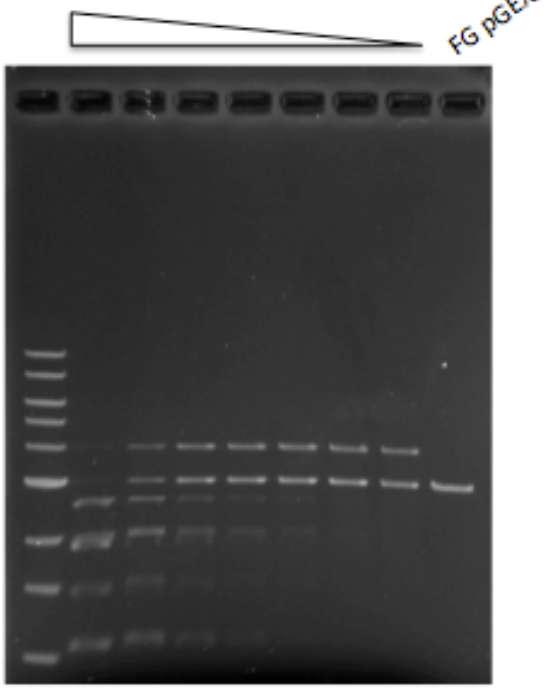


Figure 9. Competition assay results for the *M. jannaschii* r-Gyr HE with its native site and the *T. kodakaraensis* r-Gyr site. (A) Gel electrophoresis of HE digests with equimolar amounts of *Mja* r-Gyr oXXI and *Tko* r-Gyr pGEXX. Linearized site plasmids (last two lanes) were pooled and digested with serial dilutions of *Mja* r-Gyr HE (first seven lanes after 1 kb ladder). (B) Graph of percent cleaved for each site at each relative HE concentration, normalized to the highest concentration tested. Fraction cleaved in each lane was quantified based on band intensity. *C*, the relative *Mja* r-Gyr HE concentration required for 50% cleavage of the nonnative *Tko* r-Gyr site compared to its native site, is approximately 0.96.

A.



B.

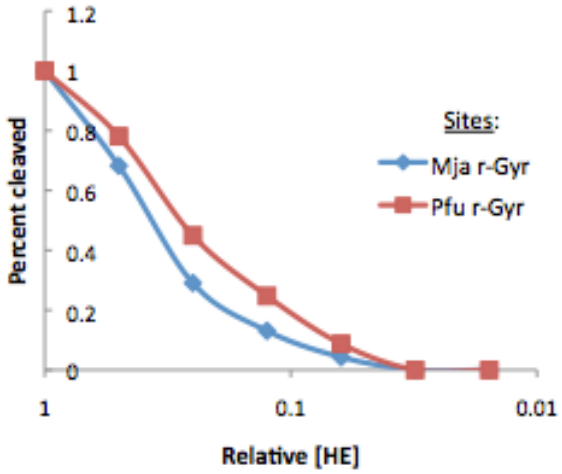
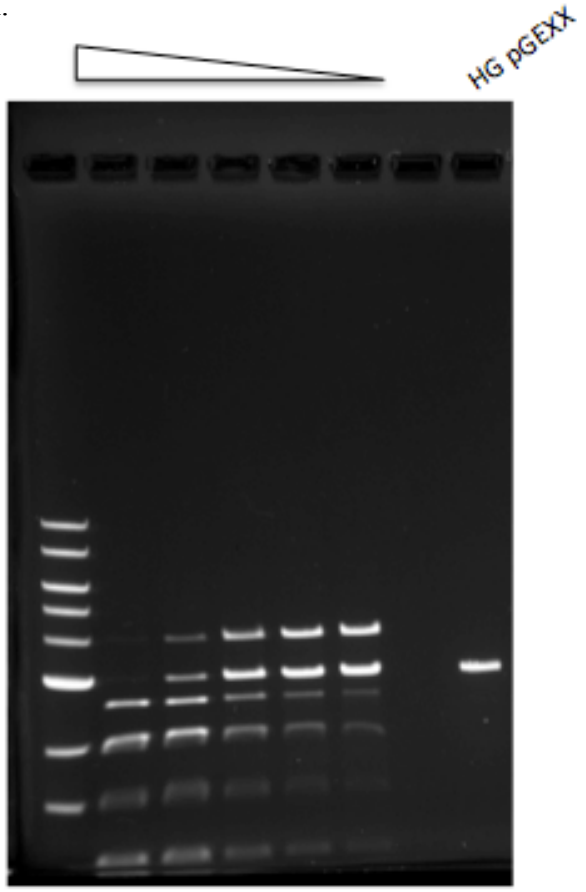


Figure 10. Competition assay results for the *M. jannaschii* r-Gyr HE with its native site and the *P. furiosus* r-Gyr site. (A) Gel electrophoresis of HE digests with equimolar amounts of *Mja* r-Gyr oXXI and *Pfu* r-Gyr pGEXX. Linearized site plasmids (last lane, nonnative) were pooled and digested with serial dilutions of *Mja* r-Gyr HE (first seven lanes after 1 kb ladder). (B) Graph of percent cleaved for each site at each relative HE concentration, normalized to the highest concentration tested. $C \approx 0.75$.

A.



B.

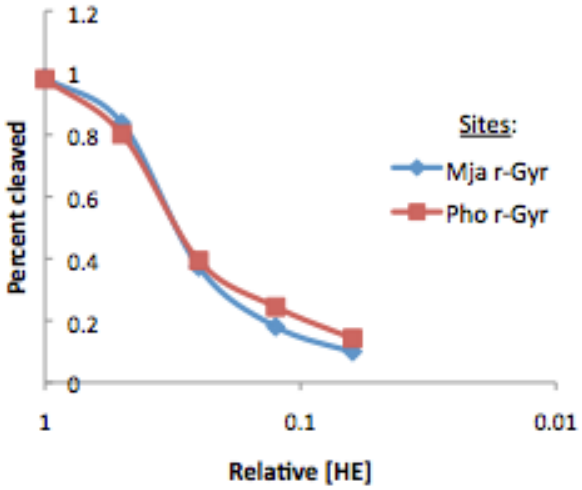
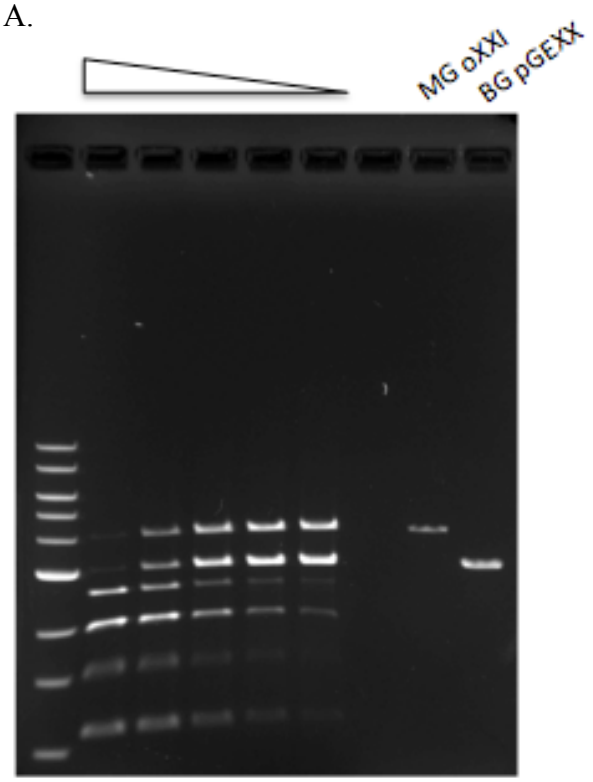


Figure 11. Competition assay results for the *M. jannaschii* r-Gyr HE with its native site and the *P. horikoshii* r-Gyr site. (A) Gel electrophoresis of HE digests with equimolar amounts of *Mja* r-Gyr oXXI and *Pho* r-Gyr pGEXX. Linearized site plasmids (last lane, nonnative) were pooled and digested with serial dilutions of *Mja* r-Gyr HE (first five lanes after 1 kb ladder). (B) Graph of percent cleaved for each site at each relative HE concentration, normalized to the highest concentration tested. $C \approx 0.99$.

A.



B.

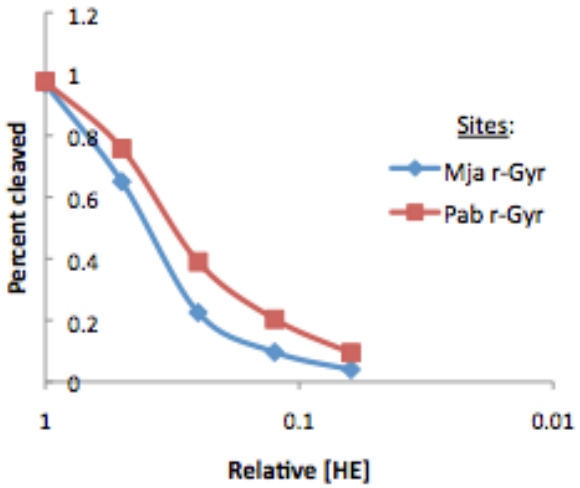
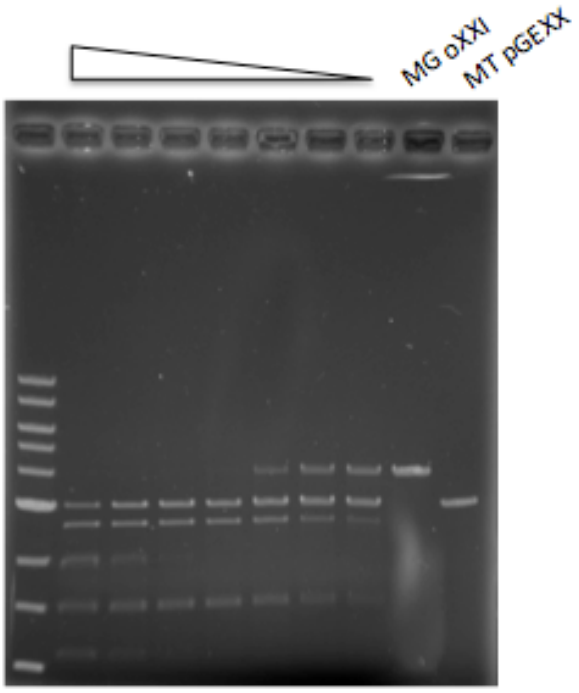


Figure 12. Competition assay results for the *M. jannaschii* r-Gyr HE with its native site and the *P. abyssi* r-Gyr site. (A) Gel electrophoresis of HE digests with equimolar amounts of *Mja* r-Gyr oXXI and *Pab* r-Gyr pGEXX. Linearized site plasmids (last two lanes) were pooled and digested with serial dilutions of *Mja* r-Gyr HE (first five lanes after 1 kb ladder). (B) Graph of percent cleaved for each site at each relative HE concentration, normalized to the highest concentration tested. $C \approx 0.79$.

A.



B.

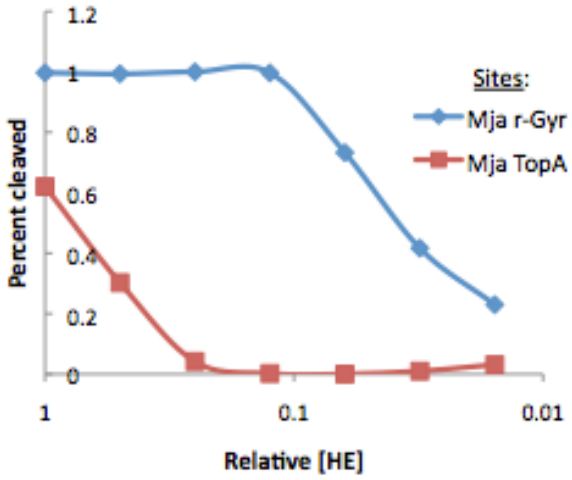
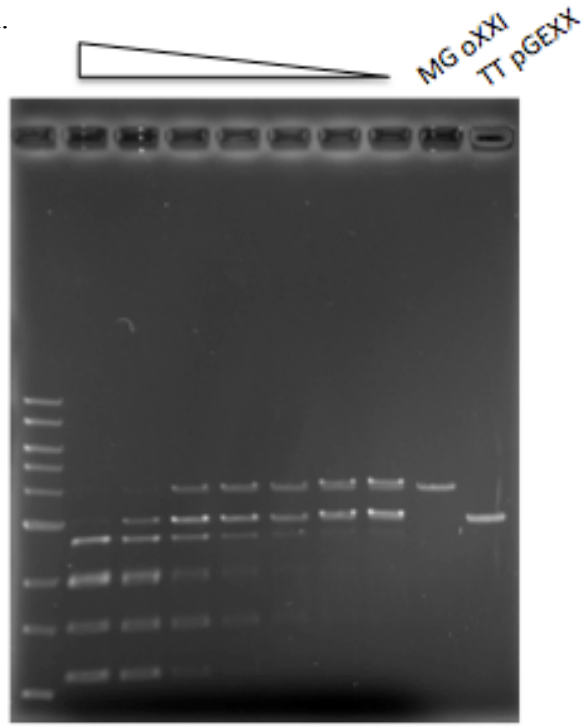


Figure 13. Competition assay results for the *M. jannaschii* r-Gyr HE with its native site and the *M. jannaschii* TopA site. (A) Gel electrophoresis of HE digests with equimolar amounts of *Mja* r-Gyr oXXI and *Mja* TopA pGEXX. Linearized site plasmids (last two lanes) were pooled and digested with serial dilutions of *Mja* r-Gyr HE (first seven lanes after 1 kb ladder). (B) Graph of percent cleaved for each site at each relative HE concentration, normalized to the highest concentration tested. $C \approx 20.53$.

A.



B.

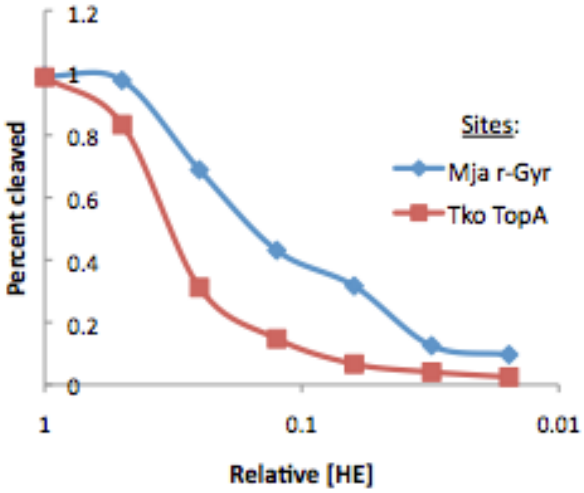
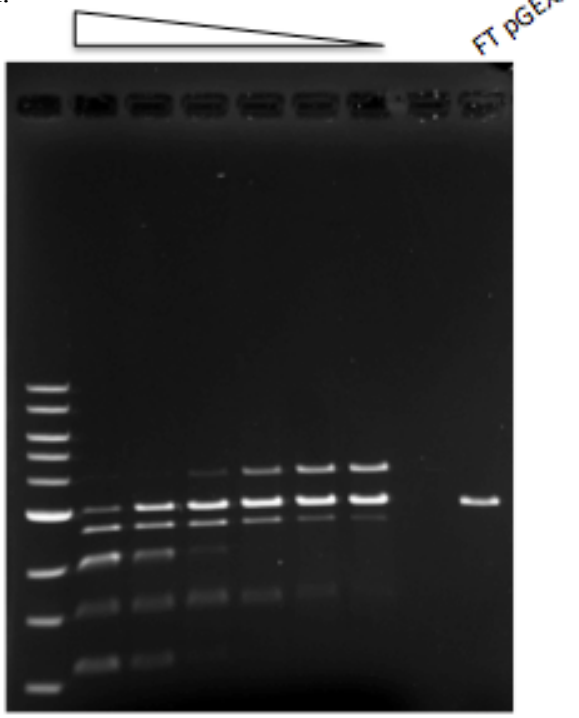


Figure 14. Competition assay results for the *M. jannaschii* r-Gyr HE with its native site and the *T. kodakaraensis* TopA site. (A) Gel electrophoresis of HE digests with equimolar amounts of *Mja* r-Gyr oXXI and *Tko* TopA pGEXX. Linearized site plasmids (last two lanes) were pooled and digested with serial dilutions of *Mja* r-Gyr HE (first seven lanes after 1 kb ladder). (B) Graph of percent cleaved for each site at each relative HE concentration, normalized to the highest concentration tested. $C \approx 2.15$.

A.



B.

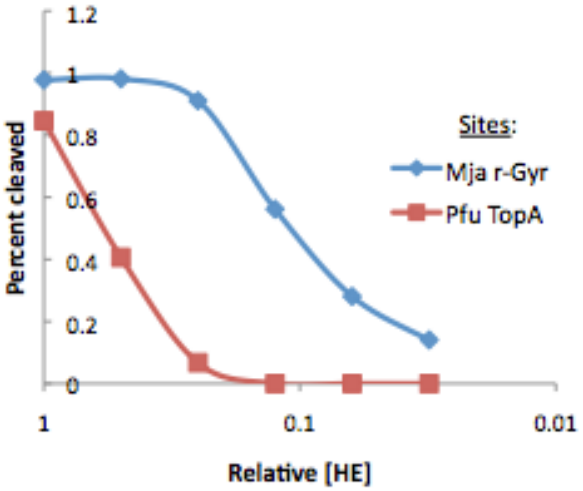
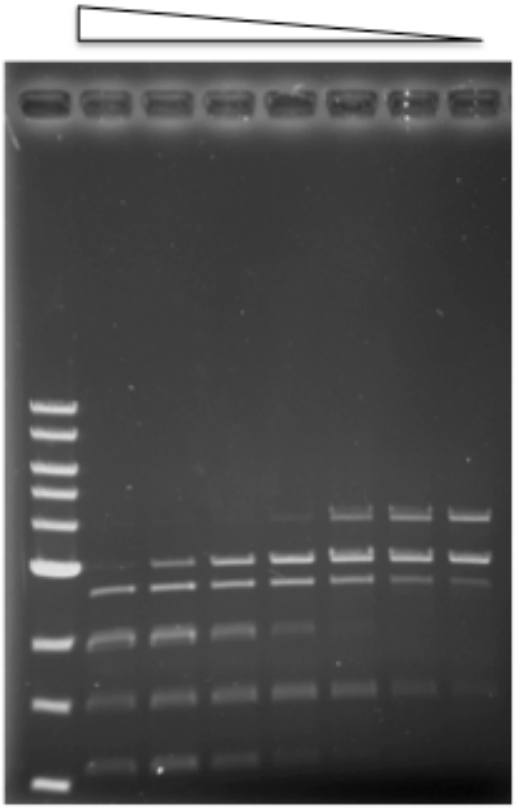


Figure 15. Competition assay results for the *M. jannaschii* r-Gyr HE with its native site and the *P. furiosus* TopA site. (A) Gel electrophoresis of HE digests with equimolar amounts of *Mja* r-Gyr oXXI and *Pfu* TopA pGEXX. Linearized site plasmids (last lane, nonnative) were pooled and digested with serial dilutions of *Mja* r-Gyr HE (first six lanes after 1 kb ladder). (B) Graph of percent cleaved for each site at each relative HE concentration, normalized to the highest concentration tested. $C \approx 5.45$.

A.



B.

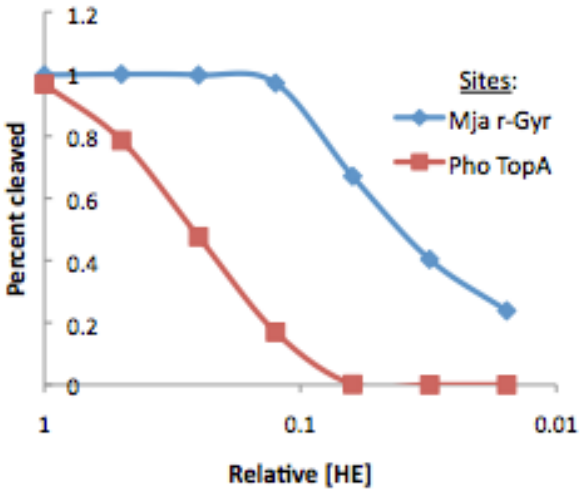
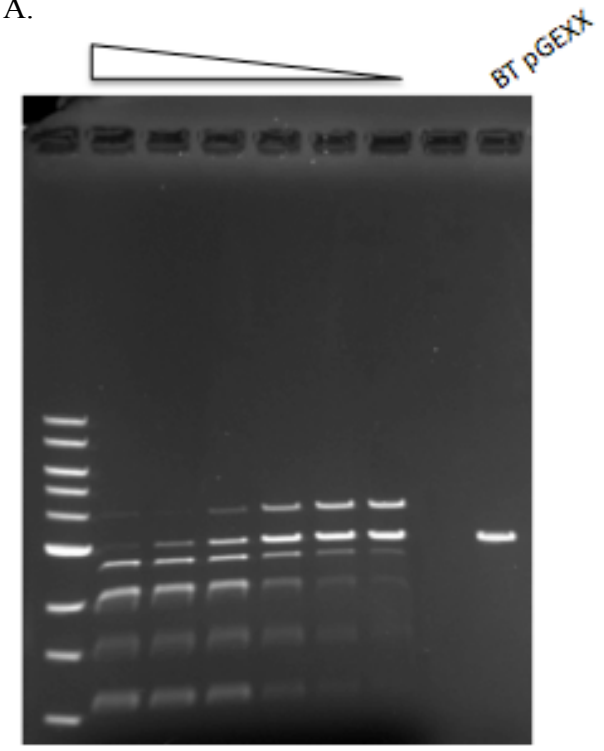


Figure 16. Competition assay results for the *M. jannaschii* r-Gyr HE with its native site and the *P. horikoshii* TopA site. (A) Gel electrophoresis of HE digests with equimolar amounts of *Mja* r-Gyr oXXI and *Pho* TopA pGEXX. Linearized site plasmids were pooled and digested with serial dilutions of *Mja* r-Gyr HE (seven lanes after 1 kb ladder). (B) Graph of percent cleaved for each site at each relative HE concentration, normalized to the highest concentration tested. $C \approx 6.33$.

A.



B.

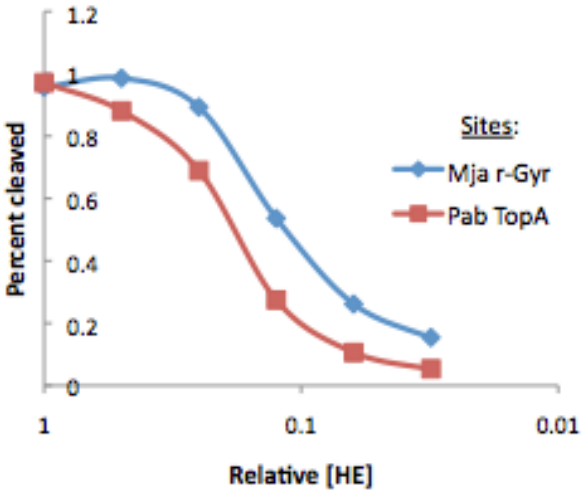
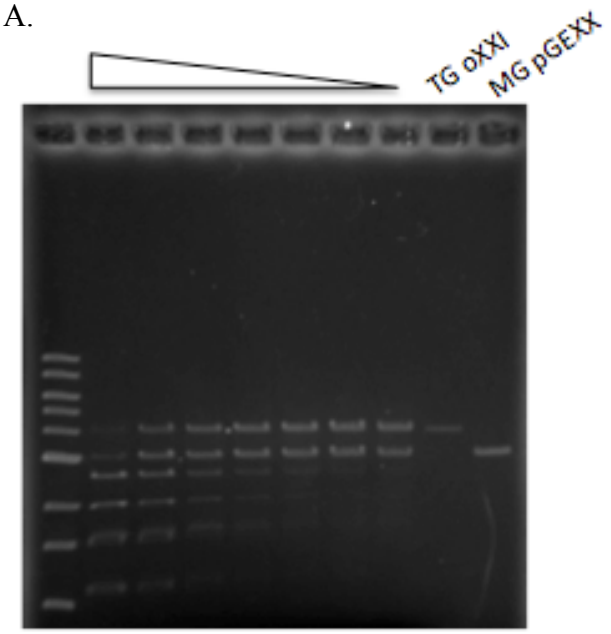


Figure 17. Competition assay results for the *M. jannaschii* r-Gyr HE with its native site and the *P. abyssi* TopA site. (A) Gel electrophoresis of HE digests with equimolar amounts of *Mja* r-Gyr oXXI and *Pab* TopA pGEXX. Linearized site plasmids (last lane, nonnative) were pooled and digested with serial dilutions of *Mja* r-Gyr HE (first six lanes after 1 kb ladder). (B) Graph of percent cleaved for each site at each relative HE concentration, normalized to the highest concentration tested. $C \approx 1.65$.

A.



B.

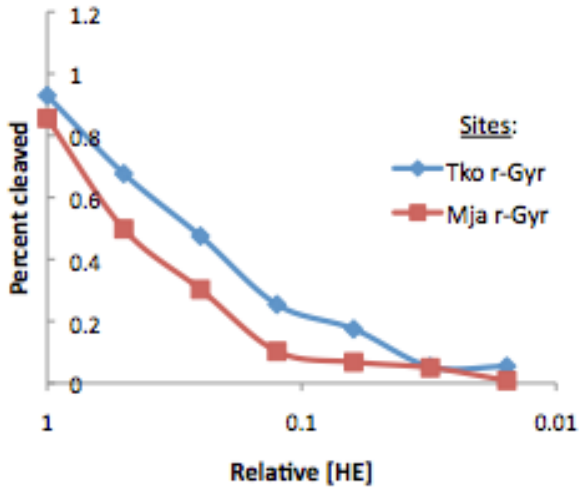
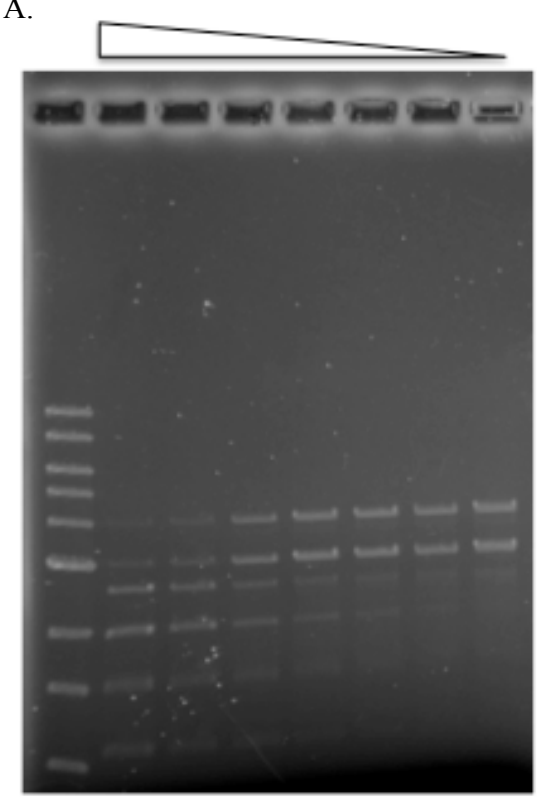


Figure 18. Competition assay results for the *T. kodakaraensis* r-Gyr HE with its native site and the *M. jannaschii* r-Gyr site. (A) Gel electrophoresis of HE digests with equimolar amounts of *Tko* r-Gyr oXXI and *Mja* r-Gyr pGEXX. Linearized site plasmids (last two lanes) were pooled and digested with serial dilutions of *Tko* r-Gyr HE (first seven lanes after 1 kb ladder). (B) Graph of percent cleaved for each site at each relative HE concentration, normalized to the highest concentration tested. $C \approx 1.79$.

A.



B.

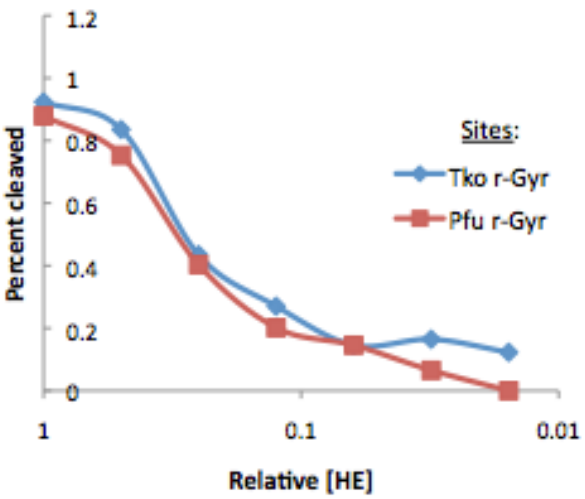
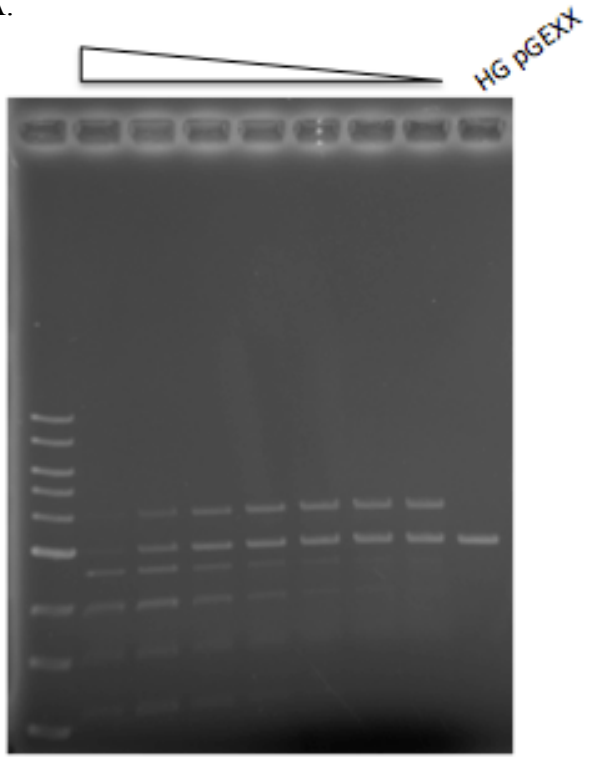


Figure 19. Competition assay results for the *T. kodakaraensis* r-Gyr HE with its native site and the *P. furiosus* r-Gyr site. (A) Gel electrophoresis of HE digests with equimolar amounts of *Tko* r-Gyr oXXI and *Pfu* r-Gyr pGEXX. Linearized site plasmids were pooled and digested with serial dilutions of *Tko* r-Gyr HE (seven lanes after 1 kb ladder). (B) Graph of percent cleaved for each site at each relative HE concentration, normalized to the highest concentration tested. $C \approx 1.10$.

A.



B.

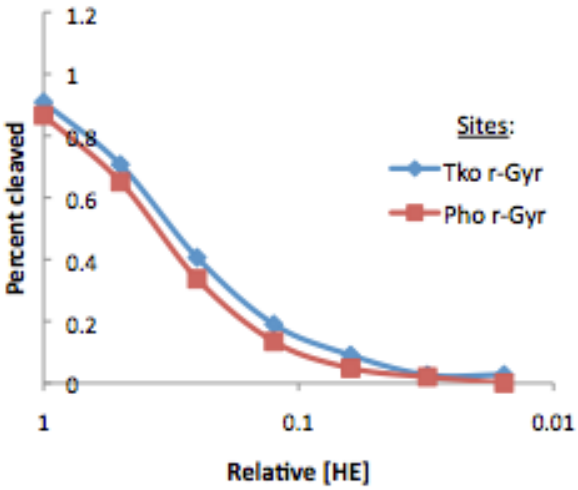
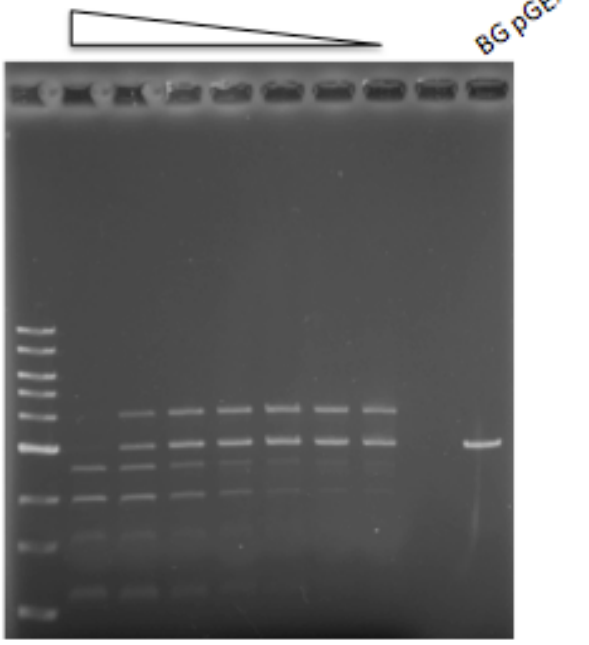


Figure 20. Competition assay results for the *T. kodakaraensis* r-Gyr HE with its native site and the *P. horikoshii* r-Gyr site. (A) Gel electrophoresis of HE digests with equimolar amounts of *Tko* r-Gyr oXXI and *Pho* r-Gyr pGEXX. Linearized site plasmids (last lane, nonnative) were pooled and digested with serial dilutions of *Tko* r-Gyr HE (first seven lanes after 1 kb ladder). (B) Graph of percent cleaved for each site at each relative HE concentration, normalized to the highest concentration tested. $C \approx 1.16$.

A.



B.

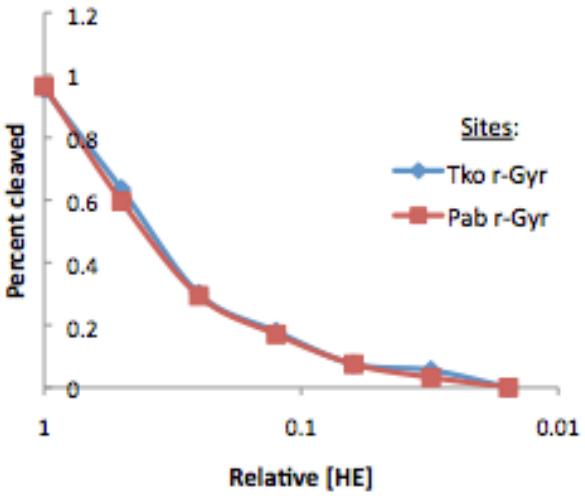
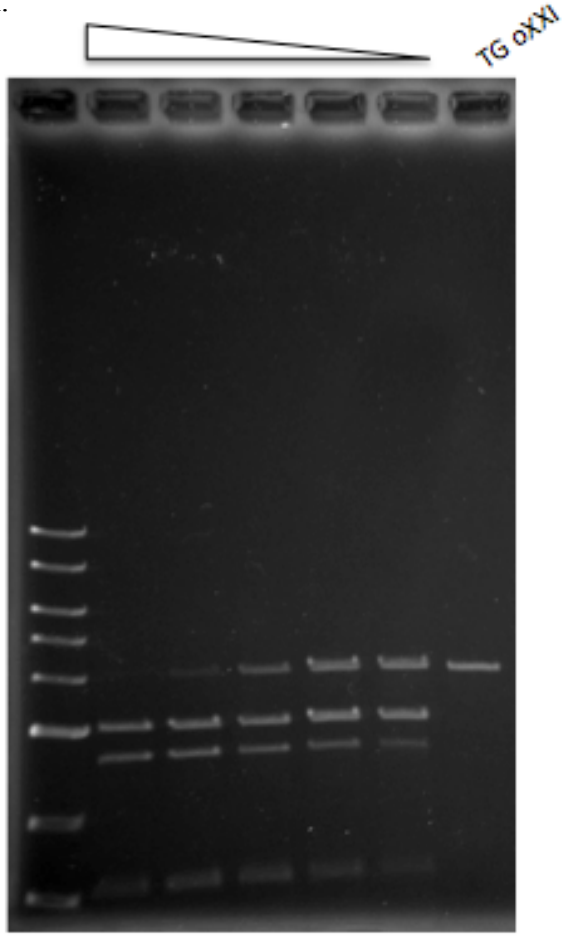


Figure 21. Competition assay results for the *T. kodakaraensis* r-Gyr HE with its native site and the *P. abyssi* r-Gyr site. (A) Gel electrophoresis of HE digests with equimolar amounts of *Tko* r-Gyr oXXI and *Pab* r-Gyr pGEXX. Linearized site plasmids (last lane, nonnative) were pooled and digested with serial dilutions of *Tko* r-Gyr HE (first seven lanes after 1 kb ladder). (B) Graph of percent cleaved for each site at each relative HE concentration, normalized to the highest concentration tested. $C \approx 1.05$.

A.



B.

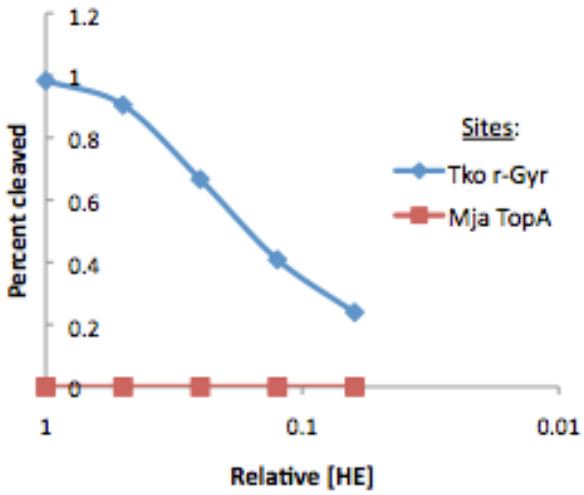
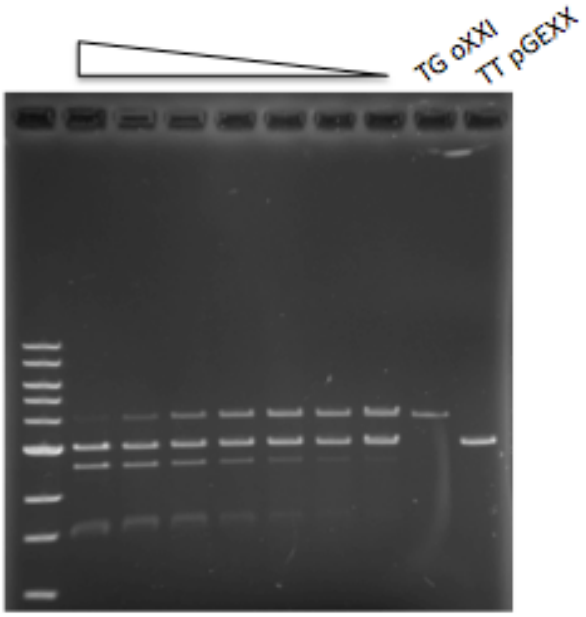


Figure 22. Competition assay results for the *T. kodakaraensis* r-Gyr HE with its native site and the *M. jannaschii* TopA site. (A) Gel electrophoresis of HE digests with equimolar amounts of *Tko* r-Gyr oXXI and *Mja* TopA pGEXX. Linearized site plasmids (last lane, native) were pooled and digested with serial dilutions of *Tko* r-Gyr HE (first five lanes after 1 kb ladder). (B) Graph of percent cleaved for each site at each relative HE concentration, normalized to the highest concentration tested. No cleavage of the nonnative site was observed.

A.



B.

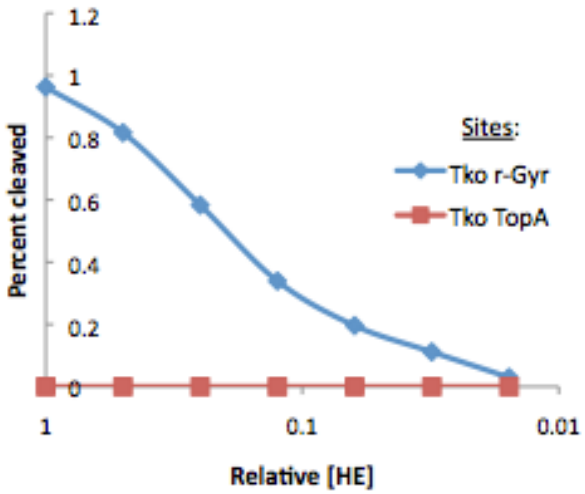
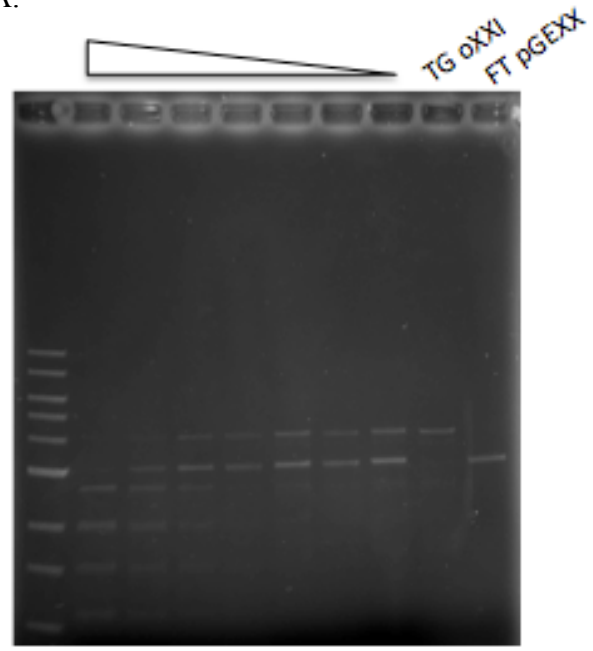


Figure 23. Competition assay results for the *T. kodakaraensis* r-Gyr HE with its native site and the *T. kodakaraensis* TopA site. (A) Gel electrophoresis of HE digests with equimolar amounts of *Tko* r-Gyr oXXI and *Tko* TopA pGEXX. Linearized site plasmids (last two lanes) were pooled and digested with serial dilutions of *Tko* r-Gyr HE (first seven lanes after 1 kb ladder). (B) Graph of percent cleaved for each site at each relative HE concentration, normalized to the highest concentration tested. No cleavage of the nonnative site was observed.

A.



B.

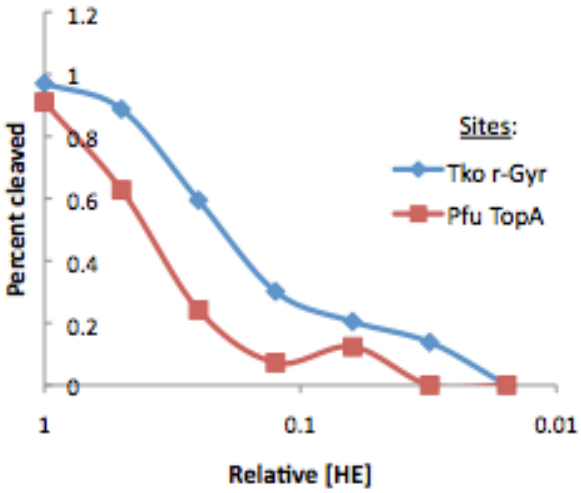
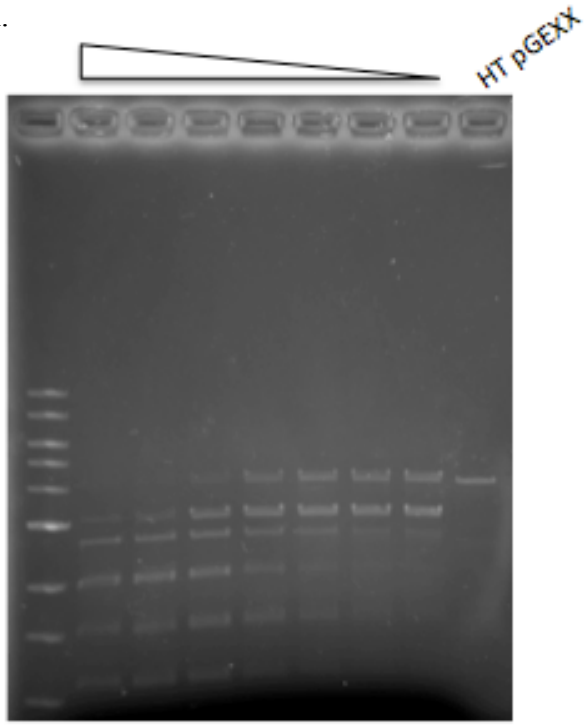


Figure 24. Competition assay results for the *T. kodakaraensis* r-Gyr HE with its native site and the *P. furiosus* TopA site. (A) Gel electrophoresis of HE digests with equimolar amounts of *Tko* r-Gyr oXXI and *Pfu* TopA pGEXX. Linearized site plasmids (last two lanes) were pooled and digested with serial dilutions of *Tko* r-Gyr HE (first seven lanes after 1 kb ladder). (B) Graph of percent cleaved for each site at each relative HE concentration, normalized to the highest concentration tested. $C \approx 1.99$.

A.



B.

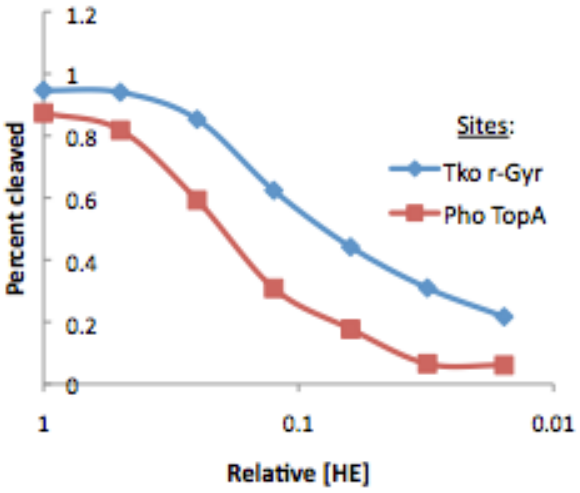
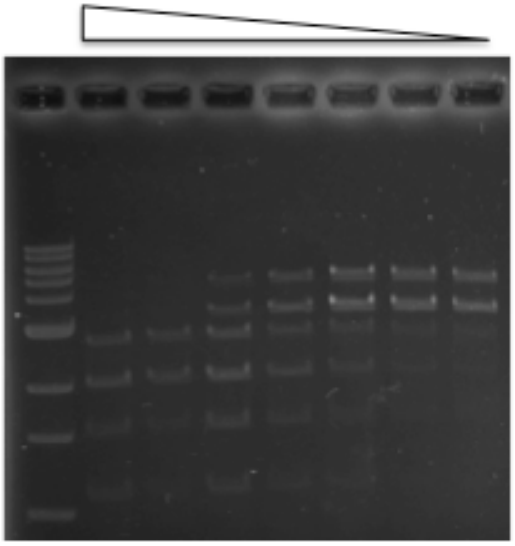


Figure 25. Competition assay results for the *T. kodakaraensis* r-Gyr HE with its native site and the *P. horikoshii* TopA site. (A) Gel electrophoresis of HE digests with equimolar amounts of *Tko* r-Gyr oXXI and *Pho* TopA pGEXX. Linearized site plasmids (last lane, nonnative) were pooled and digested with serial dilutions of *Tko* r-Gyr HE (first seven lanes after 1 kb ladder). (B) Graph of percent cleaved for each site at each relative HE concentration, normalized to the highest concentration tested. $C \approx 2.54$.

A.



B.

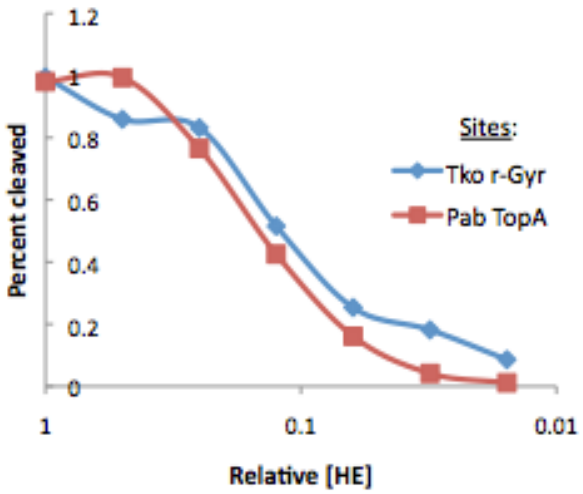


Figure 26. Competition assay results for the *T. kodakaraensis* r-Gyr HE with its native site and the *P. abyssi* TopA site. (A) Gel electrophoresis of HE digests with equimolar amounts of *Tko* r-Gyr oXXI and *Pab* TopA pGEXX. Linearized site plasmids were pooled and digested with serial dilutions of *Tko* r-Gyr HE (seven lanes after 1 kb ladder). (B) Graph of percent cleaved for each site at each relative HE concentration, normalized to the highest concentration tested. $C \approx 1.26$.

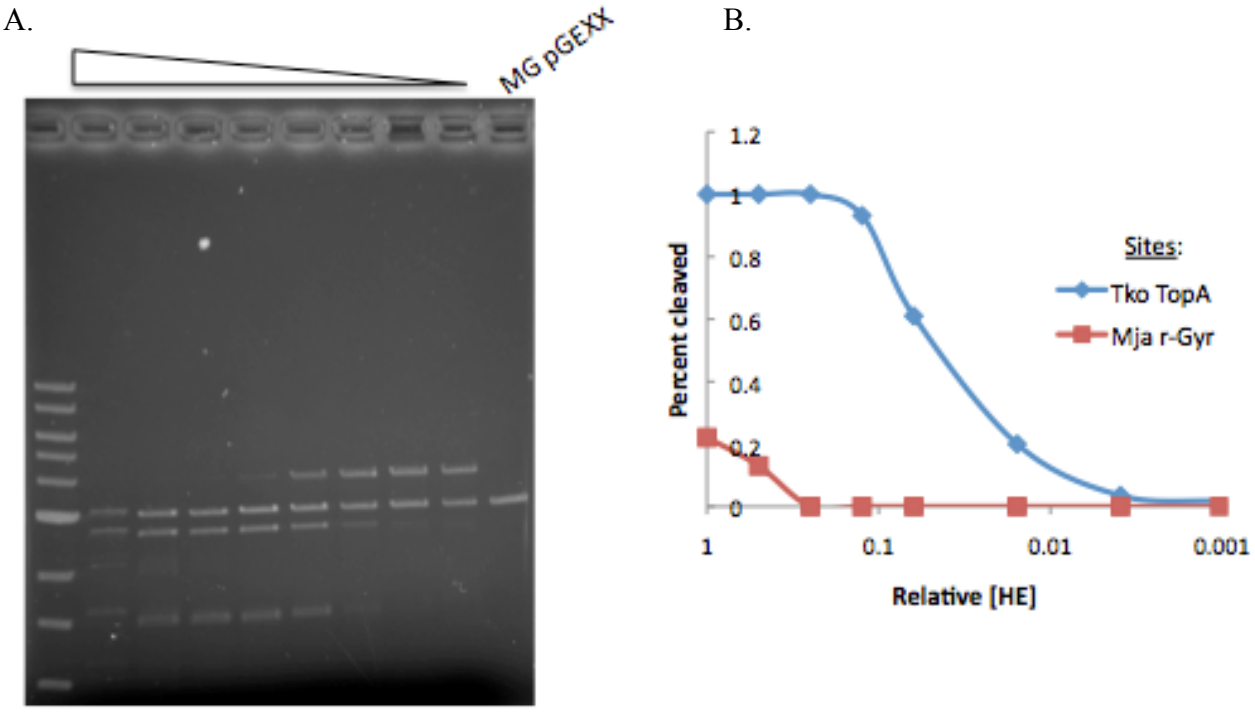
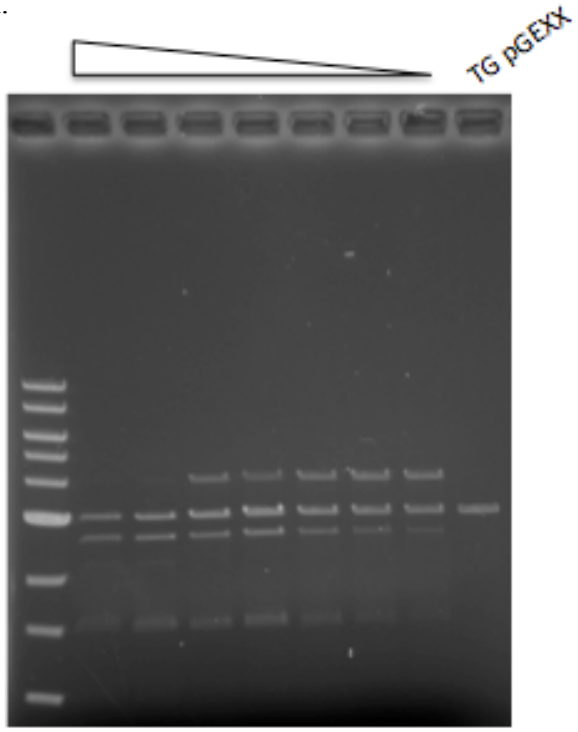


Figure 27. Competition assay results for the *T. kodakaraensis* TopA HE with its native site and the *M. janaschii* r-Gyr site. (A) Gel electrophoresis of HE digests with equimolar amounts of *Tko* TopA oXXI and *Mja* r-Gyr pGEXX. Linearized site plasmids (last lane, nonnative) were pooled and digested with serial dilutions of *Tko* TopA HE (first eight lanes after 1 kb ladder). (B) Graph of percent cleaved for each site at each relative HE concentration, normalized to the highest concentration tested. Only 20% cleavage of the nonnative site was observed, so *C* could not be calculated.

A.



B.

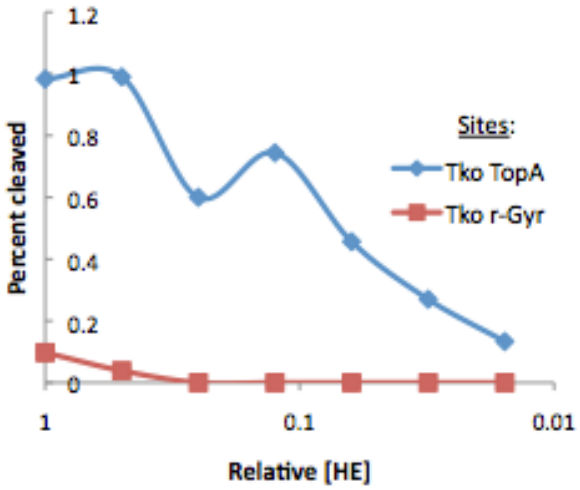
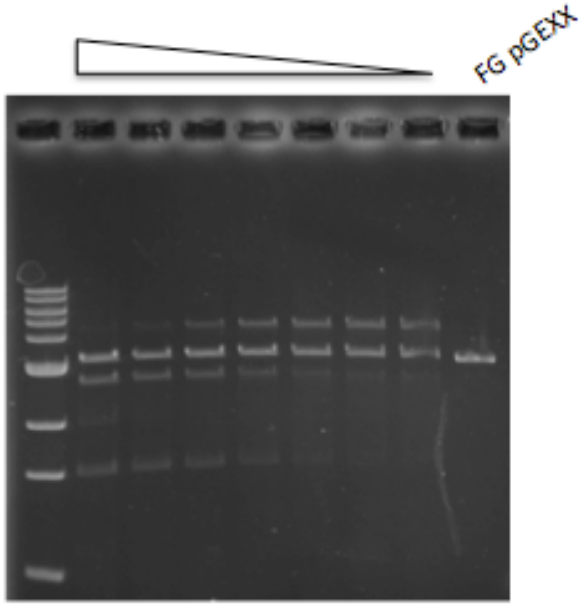


Figure 28. Competition assay results for the *T. kodakaraensis* TopA HE with its native site and the *T. kodakaraensis* r-Gyr site. (A) Gel electrophoresis of HE digests with equimolar amounts of *Tko* TopA oXXI and *Tko* r-Gyr pGEXX. Linearized site plasmids (last lane, nonnative) were pooled and digested with serial dilutions of *Tko* TopA HE (first seven lanes after 1 kb ladder). (B) Graph of percent cleaved for each site at each relative HE concentration, normalized to the highest concentration tested. Only 10% cleavage of the nonnative site was observed, so *C* could not be calculated.

A.



B.

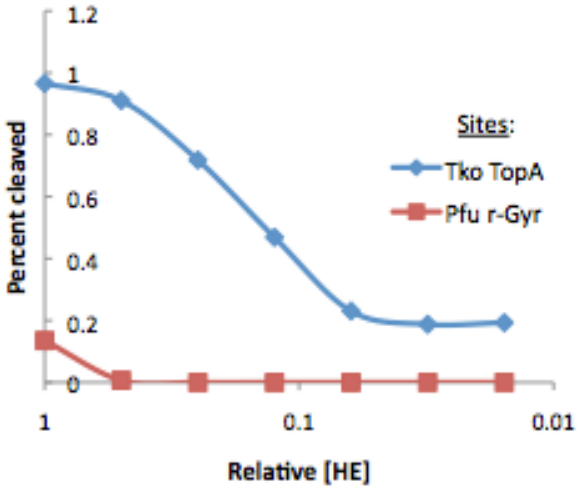
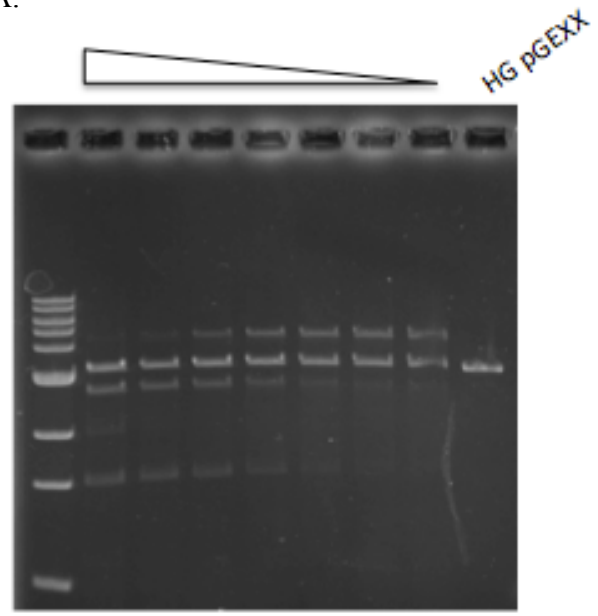


Figure 29. Competition assay results for the *T. kodakaraensis* TopA HE with its native site and the *P. furiosus* r-Gyr site. (A) Gel electrophoresis of HE digests with equimolar amounts of *Tko* TopA oXXI and *Pfu* r-Gyr pGEXX. Linearized site plasmids (last lane, nonnative) were pooled and digested with serial dilutions of *Tko* TopA HE (first seven lanes after 1 kb ladder). (B) Graph of percent cleaved for each site at each relative HE concentration, normalized to the highest concentration tested. Only 14% cleavage of the nonnative site was observed, so *C* could not be calculated.

A.



B.

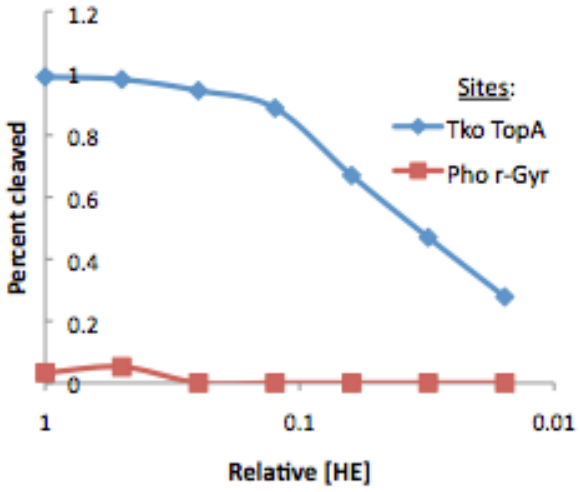
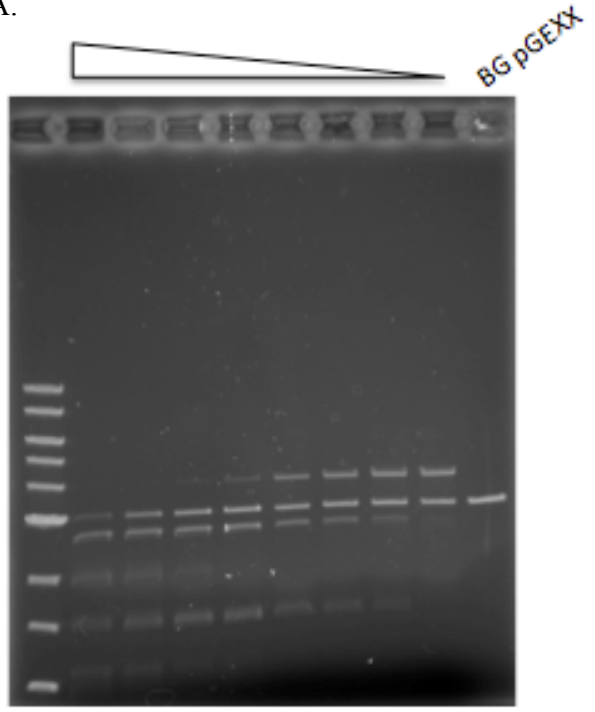


Figure 30. Competition assay results for the *T. kodakaraensis* TopA HE with its native site and the *P. horikoshii* r-Gyr site. (A) Gel electrophoresis of HE digests with equimolar amounts of *Tko* TopA oXXI and *Pho* r-Gyr pGEXX. Linearized site plasmids (last lane, nonnative) were pooled and digested with serial dilutions of *Tko* TopA HE (first seven lanes after 1 kb ladder). (B) Graph of percent cleaved for each site at each relative HE concentration, normalized to the highest concentration tested. Only 5% cleavage of the nonnative site was observed, so *C* could not be calculated.

A.



B.

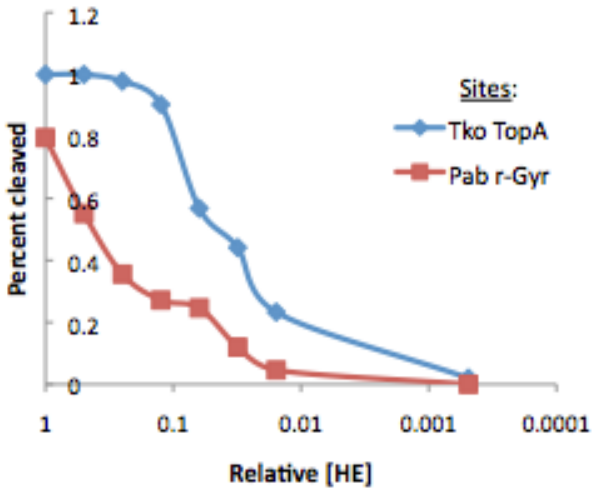
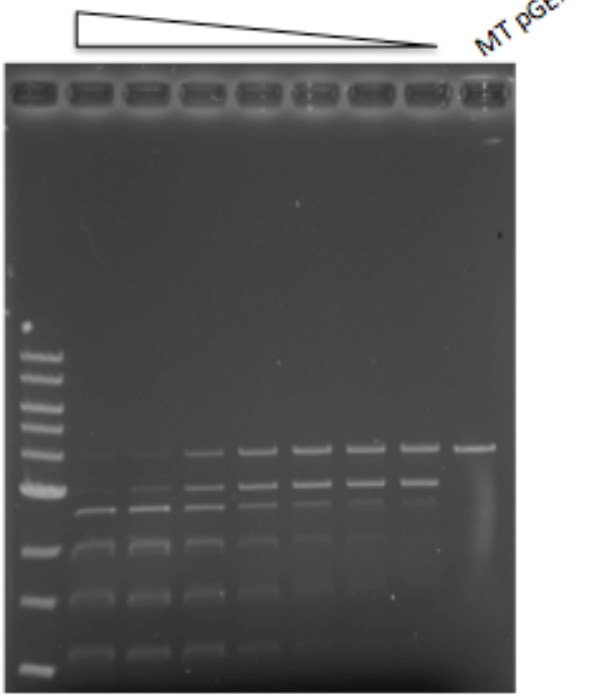


Figure 31. Competition assay results for the *T. kodakaraensis* TopA HE with its native site and the *P. abyssi* r-Gyr site. (A) Gel electrophoresis of HE digests with equimolar amounts of *Tko* TopA oXXI and *Pab* r-Gyr pGEXX. Linearized site plasmids (last lane, nonnative) were pooled and digested with serial dilutions of *Tko* TopA HE (first eight lanes after 1 kb ladder). (B) Graph of percent cleaved for each site at each relative HE concentration, normalized to the highest concentration tested. $C \approx 9.52$.

A.



B.

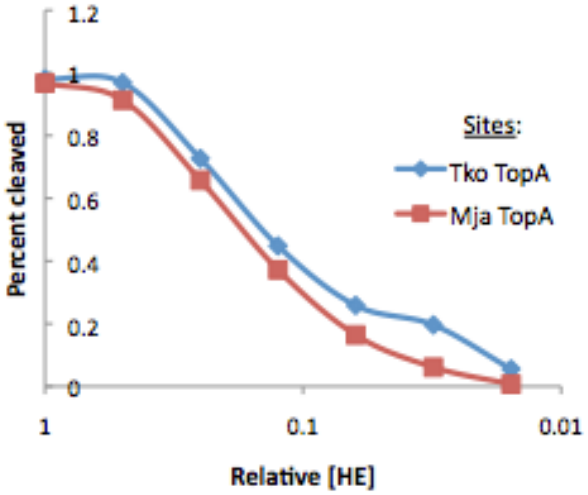
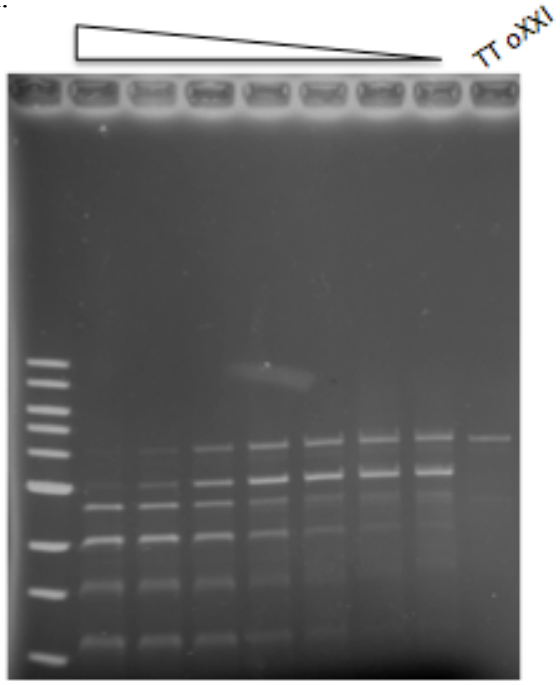


Figure 32. Competition assay results for the *T. kodakaraensis* TopA HE with its native site and the *M. jannaschii* TopA site. (A) Gel electrophoresis of HE digests with equimolar amounts of *Tko* TopA oXXI and *Mja* TopA pGEXX. Linearized site plasmids (last lane, nonnative) were pooled and digested with serial dilutions of *Tko* TopA HE (first seven lanes after 1 kb ladder). (B) Graph of percent cleaved for each site at each relative HE concentration, normalized to the highest concentration tested. $C \approx 1.22$.

A.



B.

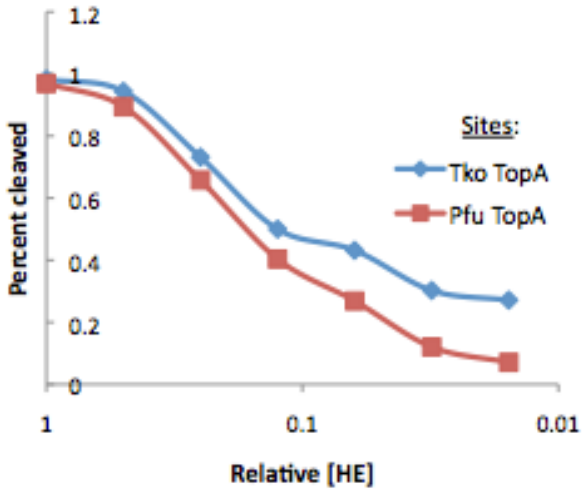
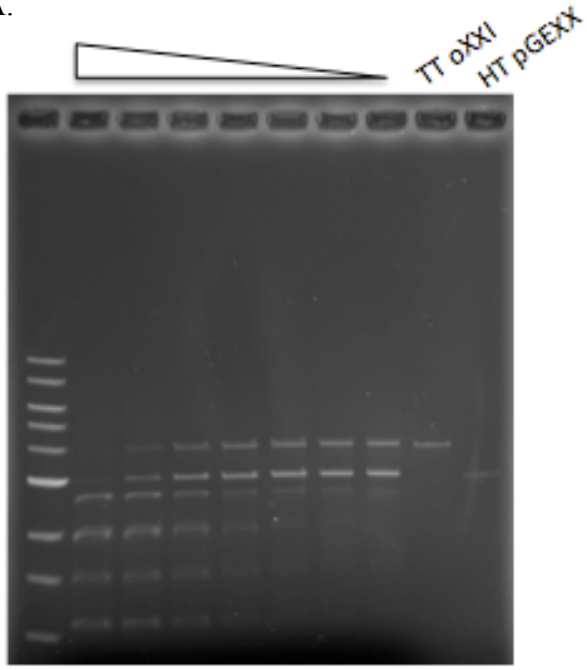


Figure 33. Competition assay results for the *T. kodakaraensis* TopA HE with its native site and the *P. furiosus* TopA site. (A) Gel electrophoresis of HE digests with equimolar amounts of *Tko* TopA oXXI and *Pfu* TopA pGEXX. Linearized site plasmids (last lane, native) were pooled and digested with serial dilutions of *Tko* TopA HE (first seven lanes after 1 kb ladder). (B) Graph of percent cleaved for each site at each relative HE concentration, normalized to the highest concentration tested. $C \approx 1.38$.

A.



B.

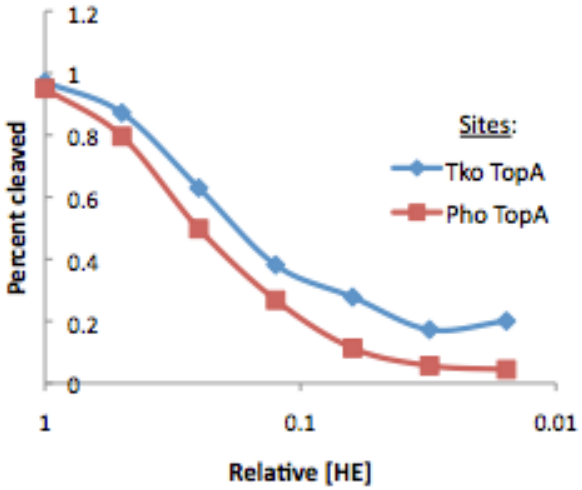
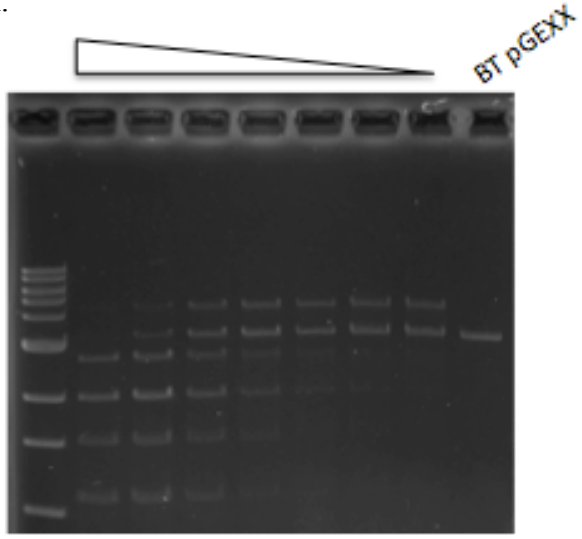


Figure 34. Competition assay results for the *T. kodakaraensis* TopA HE with its native site and the *P. horikoshii* TopA site. (A) Gel electrophoresis of HE digests with equimolar amounts of *Tko* TopA oXXI and *Pho* TopA pGEXX. Linearized site plasmids (last two lanes) were pooled and digested with serial dilutions of *Tko* TopA HE (first seven lanes after 1 kb ladder). (B) Graph of percent cleaved for each site at each relative HE concentration, normalized to the highest concentration tested. $C \approx 1.36$.

A.



B.

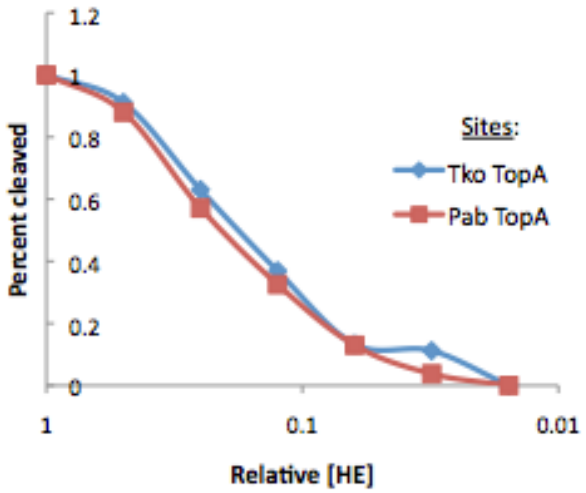


Figure 35. Competition assay results for the *T. kodakaraensis* TopA HE with its native site and the *P. abyssi* TopA site. (A) Gel electrophoresis of HE digests with equimolar amounts of *Tko* TopA oXXI and *Pab* TopA pGEXX. Linearized site plasmids (last lane, nonnative) were pooled and digested with serial dilutions of *Tko* TopA HE (first seven lanes after 1 kb ladder). (B) Graph of percent cleaved for each site at each relative HE concentration, normalized to the highest concentration tested. $C \approx 1.14$.

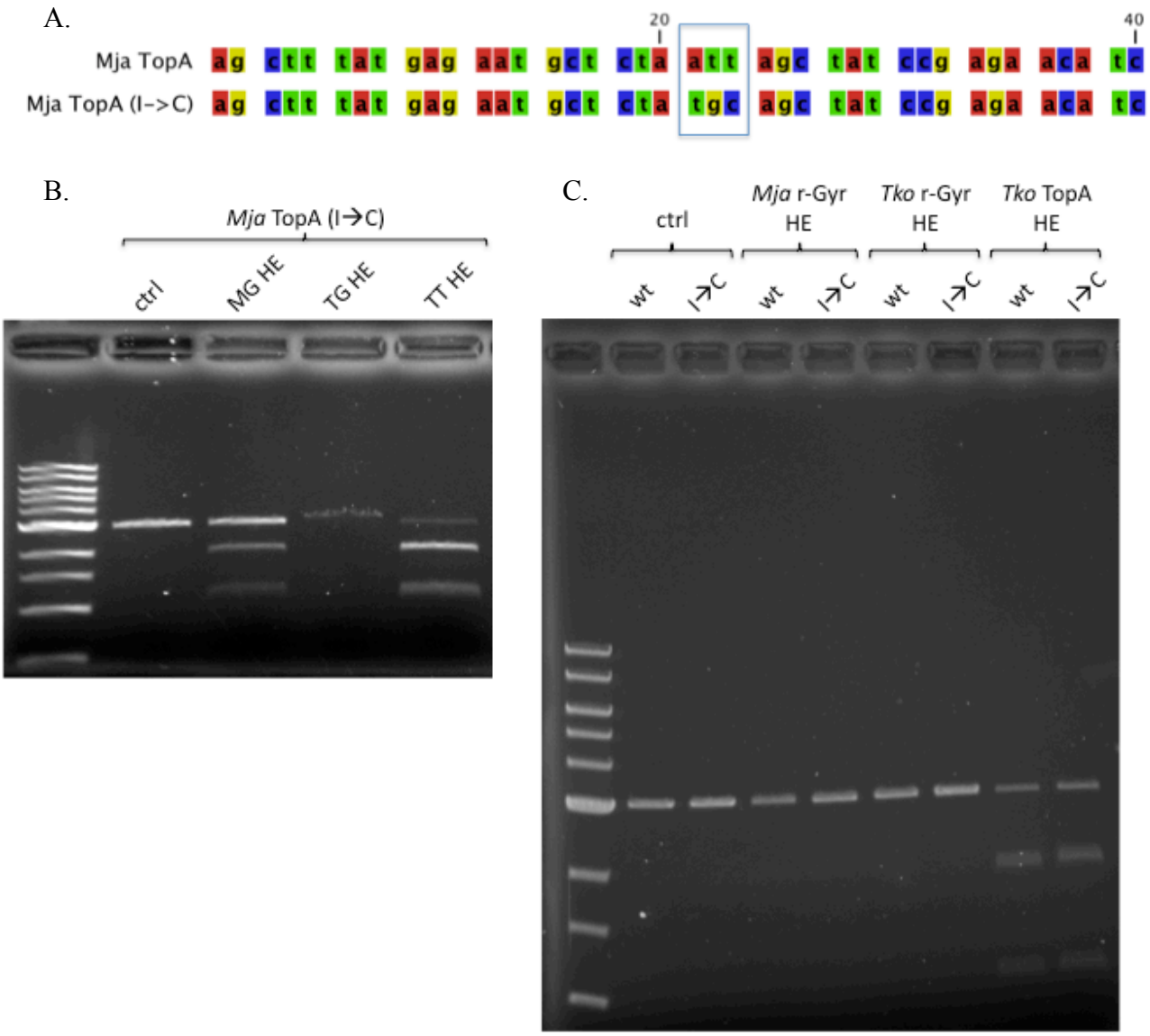


Figure 36. Effects of introducing an Ile→Cys mutation at the residue of the *M. jannaschii* TopA site immediately downstream of the hypothetical site of intein insertion. (A) The MT(I→C) site plasmid in pGEXX was cloned based on the original 40 bp *Mja* TopA site by mutating the boxed codon to the corresponding codon from the *Mja* r-Gyr site. (B) Gel electrophoresis of HE digest of the MT(I→C) site. Target site plasmid was linearized with XmnI (ctrl) and digested with either *Mja* r-Gyr (MG HE), *Tko* r-Gyr (TG HE), or *Tko* TopA (TT HE) at 80°C. (C) Gel electrophoresis of HE digest of the MT(I→C) site, along with MT wild-type (wt) controls. Equimolar target site plasmids were linearized with XmnI (ctrl) and digested with one of the three HEs under identical conditions.

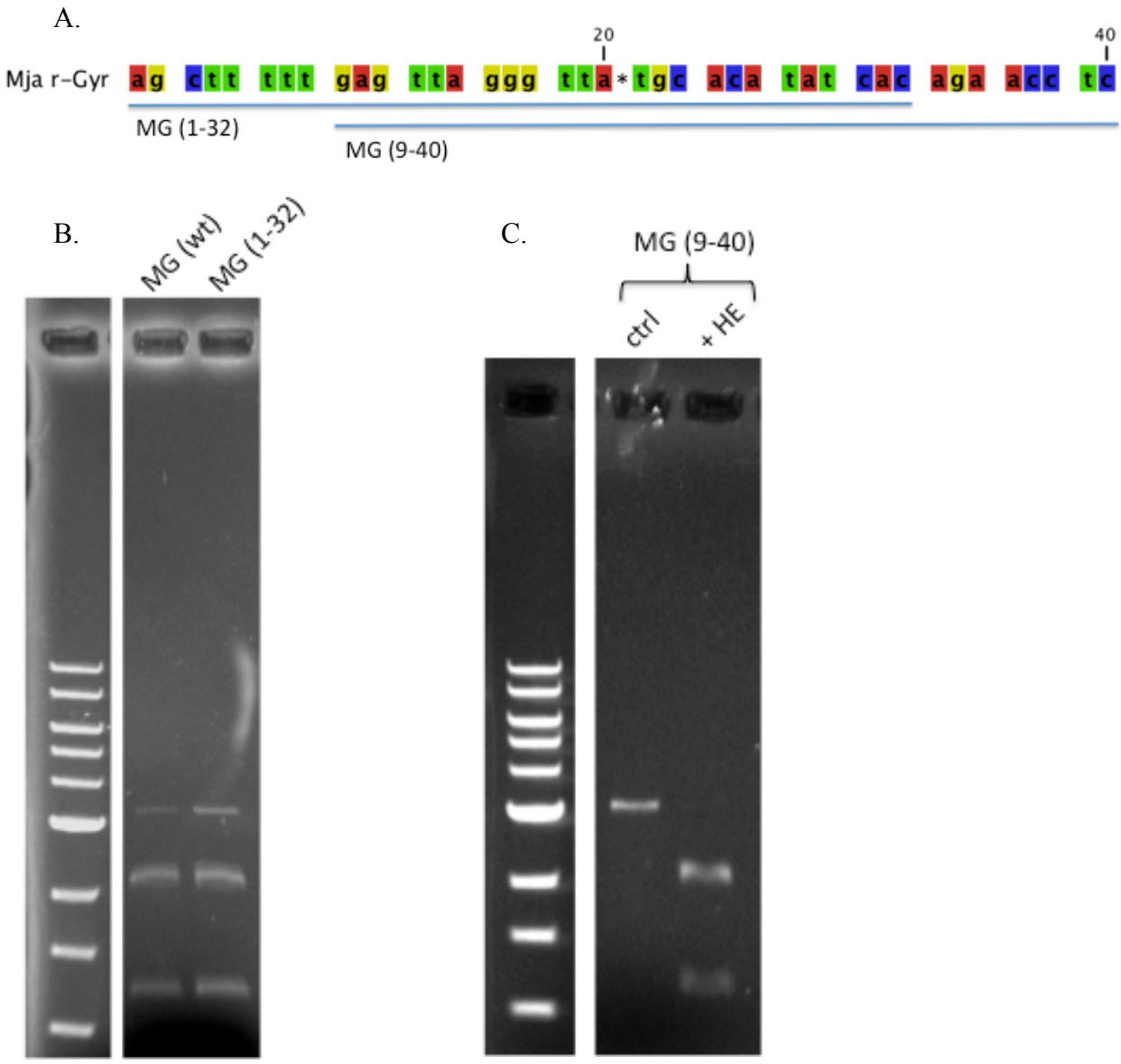


Figure 37. Initial round of site narrowing to determine the minimal target site for *M. jannaschii* r-Gyr. (A) MG(1-32) and MG(9-40) site plasmids in pGEXX were cloned based on trimming 8 bp from either end of the original 40 bp *Mja* r-Gyr site. (B) Gel electrophoresis of HE digest of MG (1-32) narrowed site, compared to wild-type (wt) site. Target site plasmids in pGEXX were linearized with XmnI and digested with *Mja* r-Gyr HE at 80°C. Here, the top band represents linearized plasmid. (C) Gel electrophoresis of HE digest of MG(9-40) narrowed site. Target site plasmid in pGEXX was linearized with XmnI (ctrl) and digested with *Mja* r-Gyr HE at 80°C (+HE).

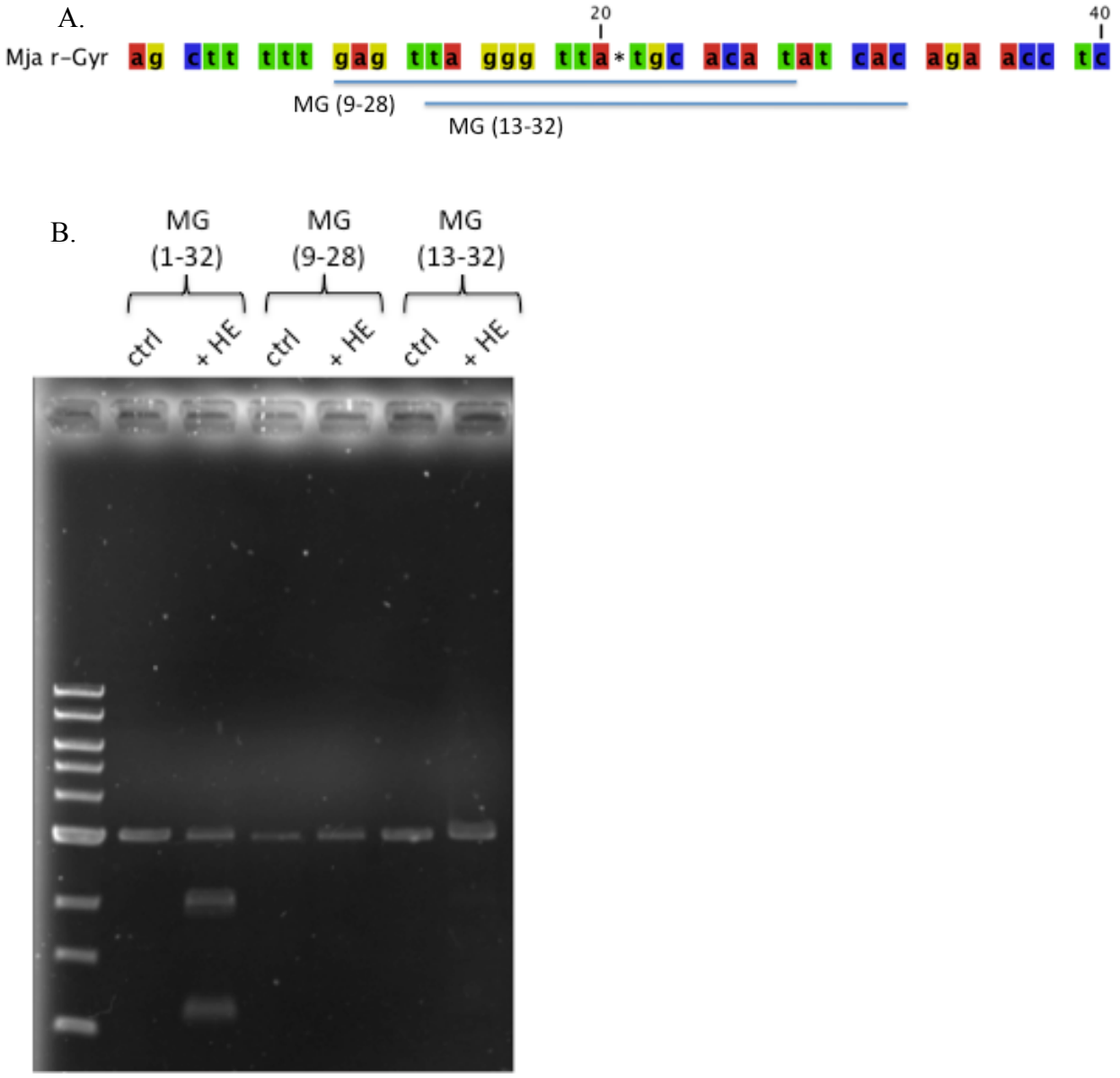


Figure 38. Second round of site narrowing to determine the minimal target site for *M. jannaschii* r-Gyr. (A) MG(9-28) and MG(13-32) site plasmids were cloned to generate 20 bp targets based on the *M. jannaschii* r-Gyr sequence. (B) Gel electrophoresis of HE digests of new narrowed sites, compared to a previously narrowed site. Target site plasmids in pGEXX were linearized with XmnI (ctrl) and digested with *Mja* r-Gyr HE at 80°C (+HE).

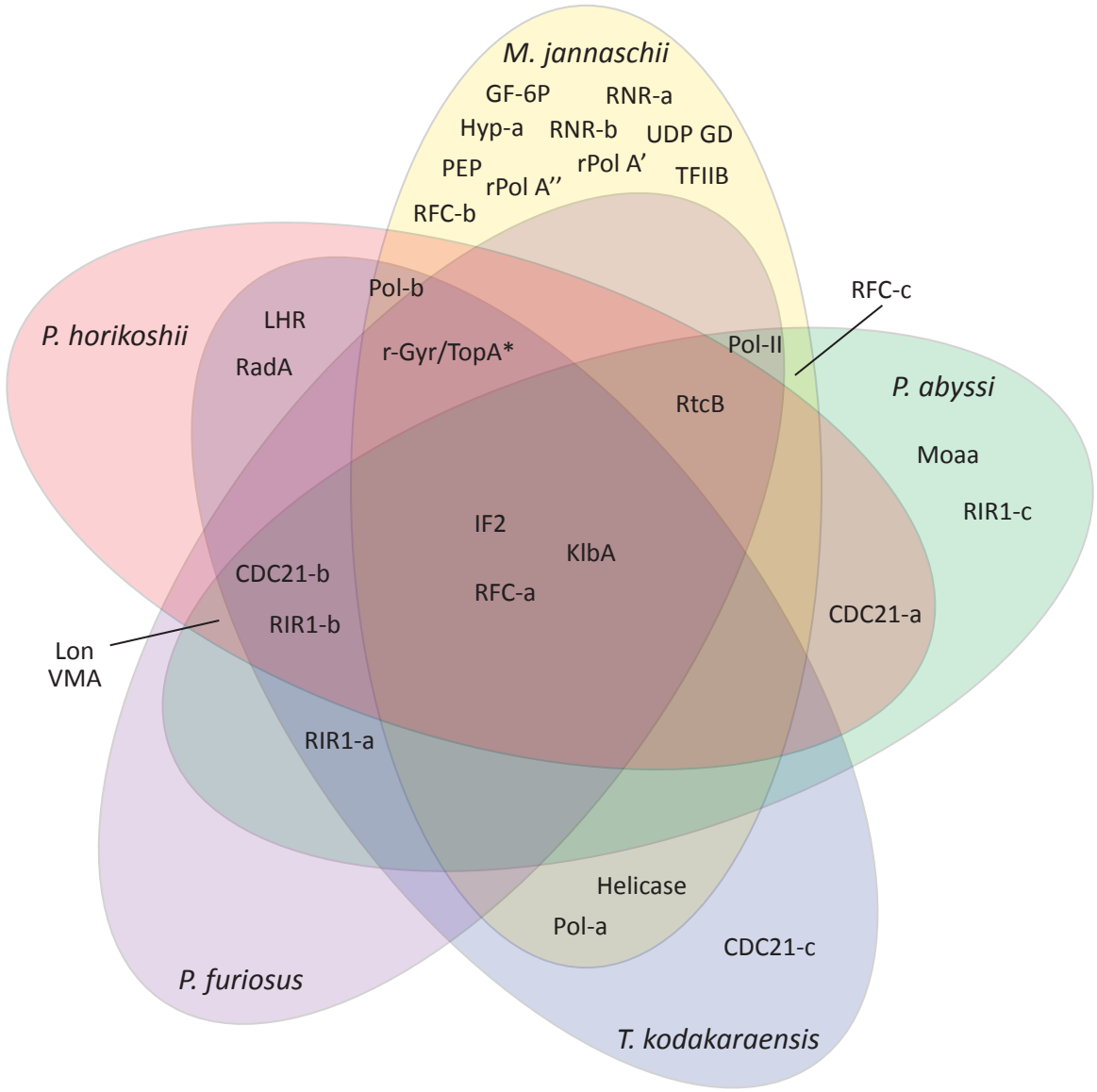


Figure 39. The distribution of common inteins among the five species in this investigation, according to InBase, the NEB Intein Database (Perler, 2002). For exteins that contain multiple inteins, inteins are listed by site of insertion (suffix) in order to group allelic inteins. The r-Gyr/TopA group of inteins that is the subject of this investigation is starred.

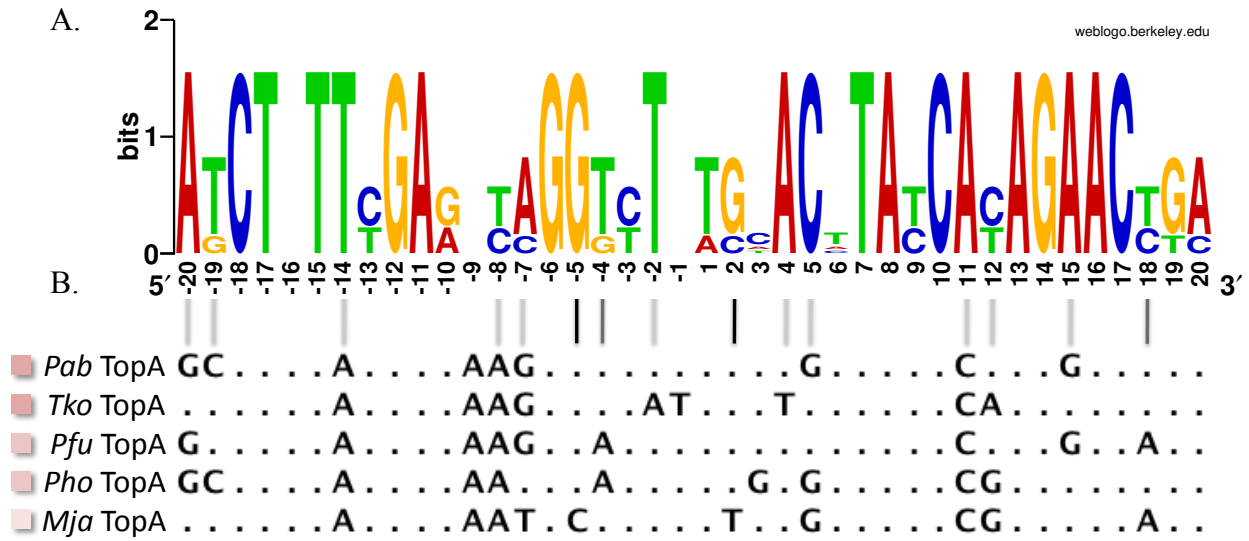


Figure 40. Mapping potential sites of protein-DNA contact for the *M. jannaschii* r-Gyr HE. (A) Sequence logo based on the five sites cleaved with roughly equal affinity, including the native site. Overall height at each position – numbered by distance from hypothetical site of intein insertion – corresponds to the degree of conservation at that base, while relative symbol heights correspond to the frequency of each nucleotide (Crooks & Hon, 2004). (B) Alignment of the five sites cleaved most poorly by the *Mja* r-Gyr HE, in order of decreasing preference. Relative preference is indicated to the left of each site according to heat map coloring (Table 1). The only nucleotides shown are those that do not occur at that position among the sites cleaved best by the HE. Lines represent positions that are both conserved in cleaved sites (≥0.5 bits) and changed in one or more poorly cleaved sites. Darker lines indicate bases that are changed only in sites with the poorest cleavage (i.e. *Mja* TopA).

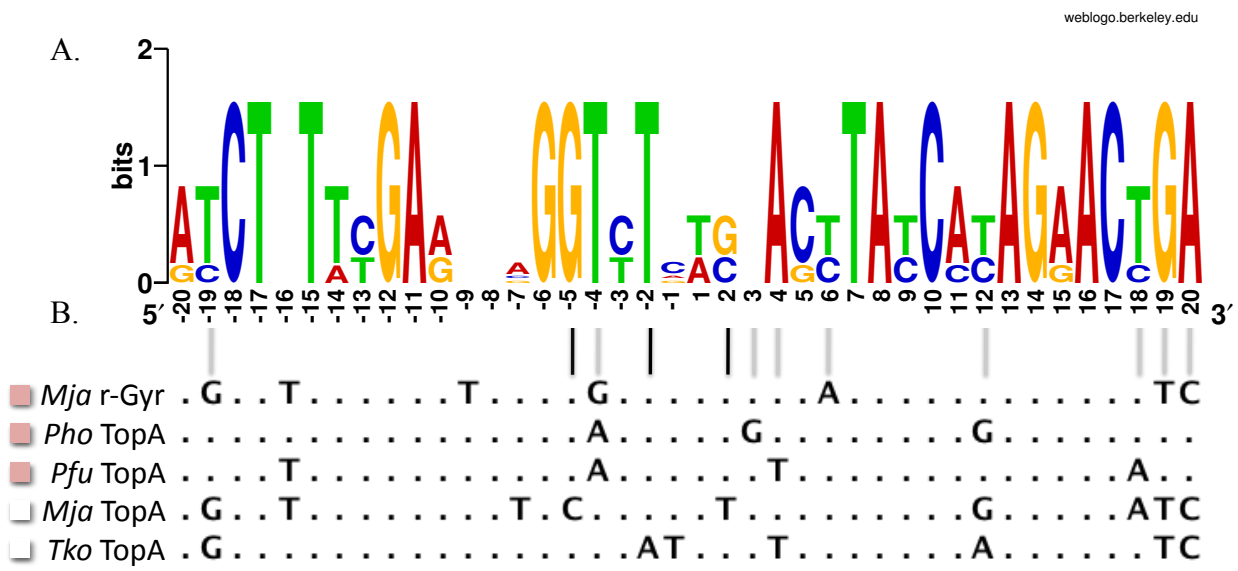


Figure 41. Mapping potential sites of protein-DNA contact for the *T. kodakaraensis* r-Gyr HE. (A) Sequence logo based on the five sites cleaved with roughly equal affinity, including the native site. Overall height at each position – numbered by distance from hypothetical site of intein insertion – corresponds to the degree of conservation at that base, while relative symbol heights correspond to the frequency of each nucleotide (Crooks & Hon, 2004). (B) Alignment of the five sites cleaved most poorly by the *Tko* r-Gyr HE, in order of decreasing preference. Relative preference is indicated to the left of each site according to heat map coloring (Table 1). The only nucleotides shown are those that do not occur at that position among the sites cleaved best by the HE. Lines represent positions that are both conserved in cleaved sites (≥0.5 bits) and changed in one or more poorly cleaved sites. Darker lines indicate bases that are changed only in sites with the poorest cleavage (i.e. *Mja* TopA and *Tko* TopA).

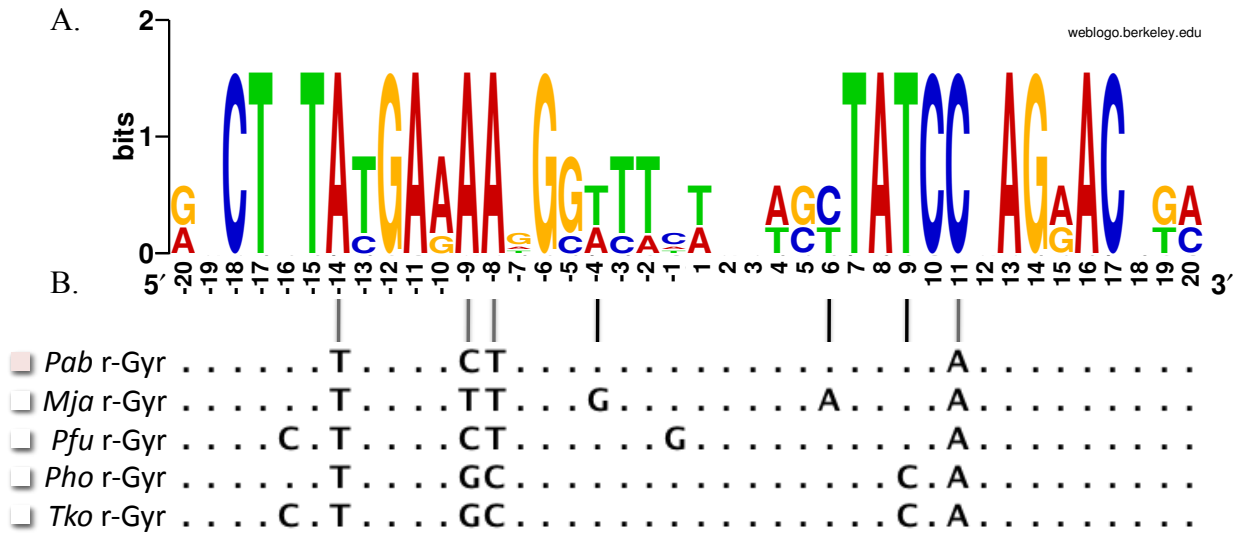


Figure 42. Mapping potential protein-DNA contacts for the *T. kodakaraensis* TopA HE. (A) Sequence logo based on the five sites cleaved with roughly equal affinity, including the native site. Overall height at each position – numbered by distance from hypothetical site of intein insertion – corresponds to the degree of conservation at that base, while relative symbol heights correspond to the frequency of each nucleotide (Crooks & Hon, 2004). (B) Alignment of the five sites cleaved most poorly by the *Tko* TopA HE, in order of decreasing preference. Relative preference is indicated to the left of each site according to heat map coloring (Table 1). The only nucleotides shown are those that do not occur at that position among the sites cleaved best by the HE. Lines represent positions that are both conserved in cleaved sites (≥0.5 bits) and changed in one or more poorly cleaved sites. Darker lines indicate bases that are changed only in sites with the poorest cleavage (i.e. *Mja* r-Gyr, *Pfu* r-Gyr, *Pho* r-Gyr, and *Tko* r-Gyr).

Table 1. Results of competitive cleavage assays with the *M. jannaschi* r-Gyr, *T. kodakaraensis* r-Gyr, and *T. kodakaraensis* TopA HEs to compare cleavage preferences of each for nine nonnative sites based on r-Gyr and TopA loci from five species of archaea. HE protein isolates were serially diluted into aliquots of linearized equimolar native and nonnative sites. Following 80°C incubation, digests were analyzed by gel electrophoresis and quantification of band intensity. Calculated values (*C*) represent the relative amount of HE required to digest the nonnative site to 50%, normalized to that required for equivalent digest of the native site. Heat map from light to dark represents relative affinity for the corresponding target site, from weak to strong. In order to compare ratios, map was colored based on log(*C*), where 0 corresponds to equal preference for native and nonnative sites.

		Mja r-Gyr HE	*Tko* r-Gyr HE	*Tko* TopA HE
r-Gyr	*Mja**	1.0[a]	1.8	N/A[c]
	*Tko**	1.0	1.0[a]	N/A[c]
	Pfu	0.8	1.1	N/A[c]
	*Pho**	1.0	1.2	N/A[c]
	Pab	0.8	1.1	9.5
TopA	*Mja*	20.5	N/A[b]	1.2
	*Tko**	2.1	N/A[b]	1.0[a]
	*Pfu**	5.4	2.0	1.4
	Pho	6.3	2.5	1.4
	Pab	1.7	1.3	1.1

* Indicates the presence of an intein at that locus *in vivo*

[a] Native site, therefore *C*=1

[b] No cleavage of nonnative site observed

[c] Insufficient cleavage of nonnative site (<50%) to calculate *C*

Table 2. Potential members of the r-Gyr/TopA group of inteins, selected from among the top 100 hits of protein BLAST of the *M. jannaschii* r-Gyr intein. Candidates were selected for inclusion if they were annotated as reverse gyrase (r-Gyr) or DNA topoisomerase I (TopA) genes. Loci are sorted by E value.

Species	Gene	In InBase?[1]	Accession
Methancaldococcus jannaschii	r-Gyr	✓	NP_248519.1
Methanotorris formicicus	r-Gyr		ZP_09706935.1
Archaeoglobus profundus	r-Gyr		YP_003400660.1
Thermococcus kodakaraensis	r-Gyr	✓	YP_182883.1
Thermococcus sp. AM4	r-Gyr		YP_002582729.1
Thermococcus zilligii	r-Gyr		ZP_11215194.1
Pyrococcus horikoshii	r-Gyr	✓	NP_142736.1
Methanocaldococcus vulcanius	r-Gyr		YP_003247600.1
Thermococcus litoralis	TopA		ZP_09729318.1
Thermococcus kodakaraensis	TopA	✓	YP_183504.1
Pyrococcus yayanosii	TopA		YP_004623434.1
Thermococcus sibiricus	TopA		YP_002994394.1
Pyrococcus furiosus	TopA	✓	NP_578223.1
Methanocaldococcus sp. FS406-22	TopA		YP_003458026.1
Pyrococcus sp. NA2	TopA		YP_004424150.1

[1]Sequences with a checkmark represent genes that are catalogued in InBase, the NEB Intein Database, and are the subject of this investigation.