

Supplemental Material

Targeted forward genetics: Population-scale analyses of allele replacements spanning thousands of base pairs in fission yeast

Aaron J. Storey, Hsin-Ping Wang, Reine U. Protacio, Mari K. Davidson, and Wayne P. Wahls*

Department of Biochemistry and Molecular Biology, University of Arkansas for Medical Sciences, Little Rock, AR 72205-7199, U.S.A.

This file contains

- Figure S1 Effects of mutation position and homology lengths upon precise allele replacements
- Figure S2 DNA mutations within individual clones of the gene targeting vector library
- Figure S3 Amino acid substitutions encoded by individual clones of the gene targeting vector library
- Figure S4 Nature of mutations within Ade6 protein discovered by population-scale, precise genome editing

Figure S1

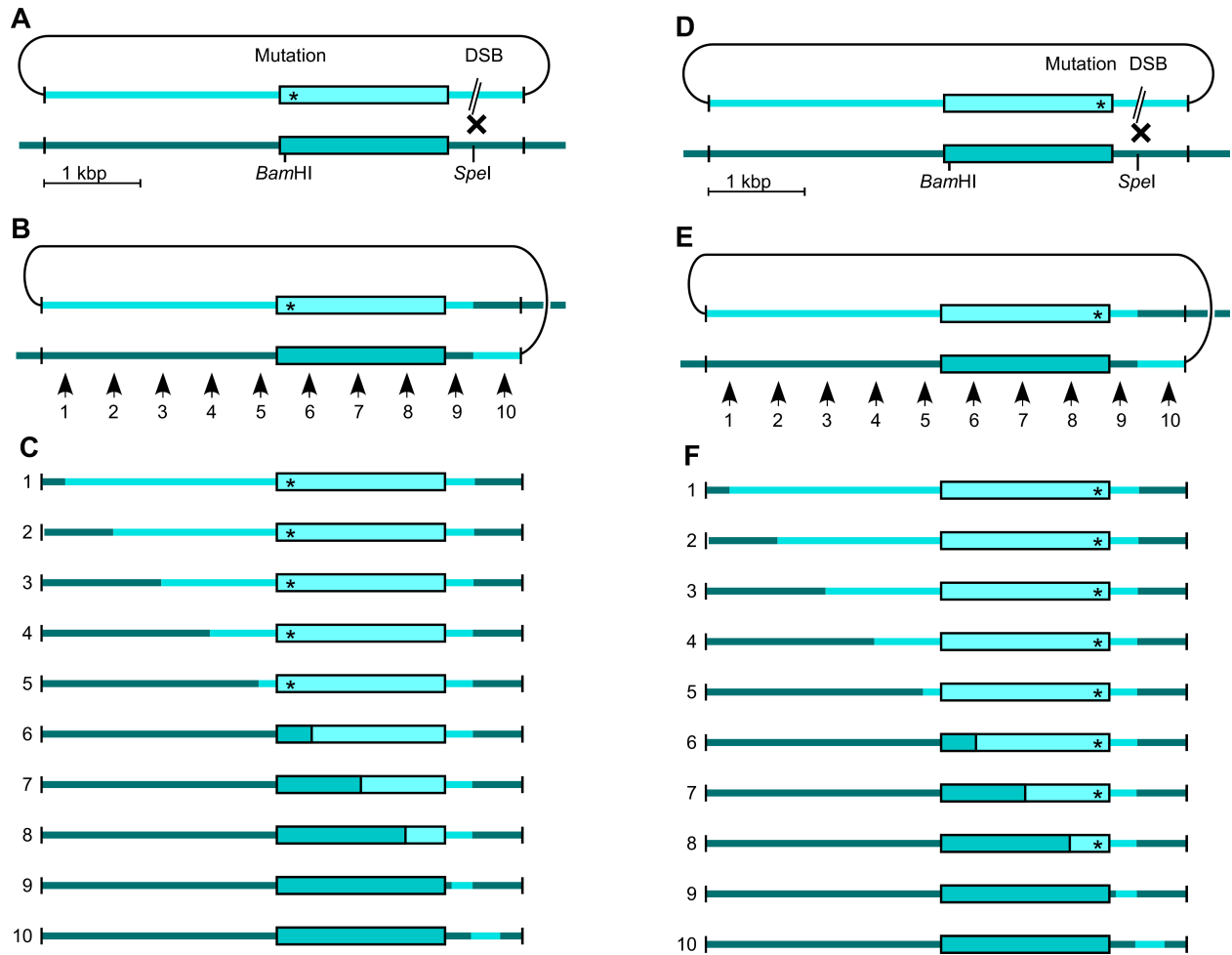


Figure S1. Effects of mutation position and homology lengths upon precise allele replacements. As in **Figure 2**, but showing the fate of individual clones in the gene targeting vector library. **(A, D)** A mutation or other modification (asterisk) can be placed anywhere in the region of homology, such as to the left or right end of the *ade6* ORF (*boxes*). In each case, pop-in recombination (×) generates a tandem duplication with one wild-type and one mutant copy **(B, E)**. Subsequent pop-out recombination events can occur anywhere within the region of homology (exemplified by *numbered arrows*). **(C, F)** The excision events will leave either the wild-type or mutant copy in the genome. Note that the likelihood of leaving a given mutation in the genome is proportional to the fraction of homology on the productive (i.e., non-DSB) side of that mutation.

Figure S2

	1									100
ade6	ATGAGCGAAA	AACAGGTGT	AGGGATCCTT	GGAGGTGGTC	AATTGGGCCG	AATGATGGTA	GAGGCAGCCC	ATCGCTTAAA	CATCAAATGC	ATCATCTTGG
	4
	1
	8
	2
	3
	5
	7
	10
	9
	101									200
ade6	ATGCAGCAAA	TTCTCCTGCC	AAACAAATG	ATGGAGGACG	TGAGCACATT	GATGCATCAT	TACTGACCC	CGATGCAATT	GTTGAACTGT	CTAAGAAGTG
	4
	1
	8
	2
	3
	5
	7
	10
	9
	201									300
ade6	CACGTTATTA	ACAACTGAAA	TTGAGCATAT	TAACACTGAT	GCCTTGGCAG	CCGTTACGAA	ATCTGTTGCT	GTTGAACCCT	CTCCTGCAAC	TCTGCGATGC
	4
	1
	8
	2
	3
	5
	7
	10
	9
	301									400
ade6	ATTCAAGACA	AATATCTTCA	AAAACAGCAT	TTACAGGTTT	TTAAGATTGC	ACTTCCTGAA	TTTTGCGATG	CACCTGACCA	GGAAAAGTGT	GAAAAAGCAG
	4
	1
	8
	2
	3
	5
	7
	10
	9
	401									500
ade6	GCCAAGAGTT	TGGTTATCCT	TTTGTACTGA	AAAGTAAAAC	ATTGGCTTAC	GACGGTCGTG	GAAATTACGT	TGTTTCATCA	CCATCTGAGA	TTCTACTGTC
	4
	1
	8
	2
	3
	5
	7
	10
	9
	501									600
ade6	CATCAAAGCA	CTTGGTGATC	GTCGGCTTTA	TGTTGAAAAG	TTCGTTCCCT	TCTCCATGGA	AATTGCAGTG	ATGGTAGTAC	GCAGTTTAGA	CGGAAAAGTT
	4
	1
	8
	2
	3
	5
	7
	10
	9

```

601
ade6 TATGCTTATC CTACAACCTGA GACCATTCAA AAGGATAATG TTTGTCATTT AGTATATGCC CCTGCTCGTC TTCCTTCTC AATCAACAG CGTGCTCAAA 700
4 .....
1 .....
8 .....
2 .....
3 .....
5 .....
7 .....
10 .....
9 .....

701
ade6 CCCTTGCCAT GGATGCGGTG CGCACTTTTG AAGGCGCTGG TATATATGGT GTAGAGATGT TTGTTTGGCA AGATGGTGAG ACCATTTTAC TCAACGAAAT 800
4 .....
1 .....
8 .....
2 .....
3 .....
5 .....
7 .....
10 .....
9 .....

801
ade6 TGCTCCTCGG CCTCACAATT CAGGTCACCTA CACCATTGAA GCTTGCCCAA CTTCTCAGTT TGAAGCTCAC TTACGGGCCA TATGTGGTCT TCCTTTGAGC 900
4 .....
1 .....
8 .....
2 .....
3 .....
5 .....
7 .....
10 .....
9 .....

901
ade6 GAAATCAACA CCCAACTCTC GACTTCCACA ACTCATGCGT TGATGGTAAA TATTTTAGGT ACTGATGATC CTGATTATGT TTCAAAGATC GCTAAACGTT 1000
4 .....
1 .....
8 .....
2 .....
3 .....
5 .....
7 .....
10 .....
9 .....

1001
ade6 CTCTGTCCAT TCCCGGTGCA ACTTTGCATC TTTATGGTAA AGCTGAATCT AGAAAGGGTC GCAAGATGGG ACACGTTACC ATCATTCTG ATTCACCTCA 1100
4 .....
1 .....
8 .....
2 .....
3 .....
5 .....
7 .....
10 .....
9 .....

1101
ade6 AGAATGTGAA CGTAGGTATC AGATGCTTCT TGACGTCAA GATCCTGTCT AATCACCTGT TGTGGTATT ATCATGGGTT CGGATTCTGA TTTAAGCAAG 1200
4 .....
1 .....
8 .....
2 .....
3 .....
5 .....
7 .....
10 .....
9 .....

1201
ade6 ATGAAAGATG CTGCCGTGAT TTTAGATGAA TTCAAGGTGC CTTACGAACT TACTATTGTT TCAGCTCACC GCACACCAGA TCGCATGGTT ACTTATGCTC 1300
4 .....
1 .....
8 .....
2 .....
3 .....
5 .....
7 .....

```

```

10 .....
9 .....
1301 ..... 1400
ade6 GTACCGCAGC TTCAAGAGGG TTGCGTGTGA TTATTGCTGG TGCTGGTGGT GCCGCTCATT TGCCTGGTAT GGTGCTGCA ATGACACCTC TTCCAGTAAT
4 .....
1 .....
8 .....
2 .....
3 .....
5 .....
7 .....
10 .....
9 .....
1401 ..... 1500
ade6 CGGCGTTCCT GTAAAAGGAA GCACTCTTGA CCGAGTTGAC TCTCTTCACT CTATTGTTC AATGCCTCGA GGTGTCCTG TCGCCACTGT TGCTATCAAT
4 .....
1 .....
8 .....
2 .....
3 .....
5 .....
7 .....
10 .....
9 .....
1501 ..... 1600
ade6 AATAGCCAAA ATGCCGGTAT TTTAGCCTGT CGTATACTTG CTACATTTC ACCCTCCCTT TTGGCTGCTA TGGAGAGCTT TATGGACAAT ATGAAAGAGA
4 .....
1 .....
8 .....
2 .....
3 .....
5 G.....
7 .....
10 .....
9 .....

1601 ..... 1659
ade6 TTGTTTTAGA AAAGGCTGAT AAATTAGAAA AAGTTGGTTG GAAAAATTAT TCTGCATAG
4 .....
1 .....
8 .....
2 .....
3 .....
5 .....
7 .....
10 .....
9 .....

```

Figure S2. DNA mutations within individual clones of the gene targeting vector library. The *ade6* (top) line indicates DNA sequence of the wild-type *ade6* ORF, numbered lines indicate position and type of mutations within individual clones from the gene targeting vector library that were sequenced.

Figure S3

ade6	1	MSEKQVVGIL	GGGQLGRMMV	EAAHRLNIKC	IILDAANSPA	KQIDGGREHI	DASFDPDAI	VELSKKCTLL	TTEIEHINTD	ALAAVTKSVA	VEPSPATLRC	100
	4	
	1	
	2	
	8	
	5	
	3	
	10G	
	7	
	9H	
ade6	101	IQDKYLQKQH	LQVFKIALPE	FCDAPDQESV	EKAGQEFQYP	FVLKSKTLAY	DGRGNVYVHQ	PSEIPTAIKA	LGDRPLYVEK	FVPFSMEIAV	MVVRSLDGKV	200
	4	
	1E	
	2	
	8G	
	5	
	3	
	10AK	
	7E	
	9	T.....A.N	
ade6	201	YAYPTTETIQ	KDNVCHLVYA	PARLPFSIQQ	RAQTLAMDAV	RTFEGAGIYG	VEMFVLQDGE	TILLNEIAPR	PHNSGHYTIE	ACPTSQFEAH	LRAICGLPFS	300
	4	
	1	
	2	
	8	
	5	
	3	
	10	
	7	
	9	
ade6	301	EINTQLSTST	THALMVNILG	TDDPDYVSKI	AKRSL SIPGA	TLHLYGKAES	RKGRKMGHVT	IISDSPQECE	RRYQMLLDVK	DPVESPVVGI	IMGSDSDL SK	400
	4	
	1	
	2	
	8A	
	5	
	3	
	10	
	7E	
	9	
ade6	401	MKDAAVILDE	FKVPYELTIV	SAHRTPDRMV	TYARTAASRG	LRVIIAGAGG	AAHLPGMVAA	MTPLPVI GVP	VKGSTLDGVD	SLHSIVQMPR	GVPVATVAIN	500
	4	
	1	
	2	
	8	
	5	
	3	
	10	
	7	
	9	
ade6	501	NSQNAGILAC	RILATFQPSL	LAAMESFMDN	MKEIVLEKAD	KLEKVGWKNY	SA	552				
	4				
	1				
	2				
	8				
	5	D.....				
	3				
	10				

7
9

Figure S3. Amino acid substitutions encoded by individual clones of the gene targeting vector library. The *ade6* (top) line indicates amino acid sequence of the Ade6 protein, numbered lines indicate positions and types of amino acid substitutions or stop codons (-) encoded by individual clones from the gene targeting vector library that were sequenced.

Figure S4

	1									100
ade6	MSEKQVVGIL	GGGQLGRMMV	EAAHRLNIKC	IILDAANSPA	KQIDGGREHI	DASFTDPDAI	VELSKKCTLL	TTEIEHINTD	ALAAVTKSVA	VEPSPATLRC
Red-6I
Pink-8Y
Pink-1
Red-10
Red-8T
Pink-11S
Red-3
Pink-4
Pink-9
Pink-12A
Pink-3
Pink-7VG
Red-2-
Red-5
	101									200
ade6	IQDKYLQKQH	LQVFKIALPE	FCDAPDQESV	EKAGQEFGYP	FVLKSKTLAY	DGRGNYVVHQ	PSEIPTAIKA	LGDRPLYVEK	FVPFSMEIAV	MVVRSLDGKV
Red-6
Pink-8
Pink-1	...E
Red-10H
Red-8L
Pink-11V
Red-3E
Pink-4L
Pink-9	.R	G
Pink-12G
Pink-3EL
Pink-7Y
Red-2
Red-5	.-S
	201									260
ade6	YAYPTTETIQ	KDNVCHLVYA	PARLPFSIQQ	RAQTLAMDAV	RTFEGAGIYG	VEMFVLQDGE				
Red-6
Pink-8
Pink-1
Red-10
Red-8
Pink-11
Red-3	...A
Pink-4
Pink-9
Pink-12
Pink-3
Pink-7
Red-2
Red-5

Figure S4. Nature of mutations within Ade6 protein discovered by population-scale, precise genome editing. Targeted forward genetics identified presumptive *ade6* mutants of fission yeast. The *ade6* locus from individual mutant candidates was sequenced and translated *in silico*. Top line indicates amino acid sequence of wild-type Ade6 protein from residues 1 to 260; subsequent lines show amino acid substitutions or stop codons (-) within individual clones. Every mutant that was sequenced had at least one coding mutation within the *ade6* locus, demonstrating that the allele replacement process is robust and precise.