Development of a method to produce functional chimeric genes via single-strand annealing pathway[§]

Kyu-Jin Ahn, Ju-Hee Choi, Ye-Seul Lim, Min-Ku Kim, and Sung-Ho Bae[*] ⓘD

Department of Biological Sciences, College of Natural Science, Inha University, Incheon 22212, Republic of Korea

Single-strand annealing 경로를 이용한 기능성 chimeric 유전자 제조 방법의 개발[§]

안규진 · 최주희 · 임예슬 · 김민규 · 배성호[*] ⓘD

인하대학교 자연과학대학 생명과학과

Duplicated genes with similar sequences are commonly observed in eukaryotic genomes, some of which show sufficient homology to allow for recombination. *ASN1* and *ASN2* are an example of duplicated genes that share highly homologous sequence throughout the open reading frame. In this study, we devised a genetic method to produce functional chimeric *ASN* genes via single-strand annealing pathway. Creation of a DNA double-strand break between *ASN1* and *ASN2* using HO endonuclease generated mutant cells with chimeric genes at significantly high frequency. All mutant cells exhibited normal growth in the medium lacking asparagine, suggesting that the chimeric genes are functional. Sequence analysis of chimeric genes revealed that they had different single junctions. Our results provide a method to produce a variety of chimeric genes simultaneously by randomly fusing homologous genes.

Keywords: *ASN1*, *ASN2*, asparagine synthetase, chimeric gene, single-strand annealing

Single-strand annealing (SSA) is a major recombination pathway for repairing DNA double-strand breaks (DSBs) occurring between direct long repeats (Bhargava *et al*., 2016; Hanscom and McVey, 2020). The initiating event of SSA is end resection that creates long 3' single-stranded tails that are complementary to each other. This is a common process with homologous recombination (HR), but subsequent processes do not involve strand invasion and Holliday junction formation observed in HR. Instead, single-stranded homologous sequences are annealed together, and the nonhomologous 3' single-stranded tails are subsequently cleaved, followed by ligation to complete the repair process (Li *et al*., 2019). This pathway inevitably accompanies a deletion mutation between repeated sequences, but allows rapid repair of breaks within tandem repeat arrays (Ranjha *et al*., 2018).

Duplicated genes are commonly observed in all sequenced eukaryotic genomes. In *Saccharomyces cerevisiae*, approximately 30% genes are present in at least two copies (Dujon, 1998; Katju *et al*., 2009). Most of them share homology in short stretches of open reading frame (ORF) or show homology only at the amino acid sequence level, but some gene pairs such as *ASN1* and *ASN2*, encoding asparagine synthetases, exhibit high homology even at the nucleotide sequence level.

Asparagine synthetases, the enzymes responsible for producing asparagine from aspartic acid and glutamine-derived nitrogen,

are well conserved from *Escherichia coli* to mammals (Lomelino *et al.*, 2017). *Saccharomyces cerevisiae* contains two highly similar asparagine synthetases, Asn1p and Asn2p, encoded by two unlinked genes, *ASN1* and *ASN2* (Dang *et al.*, 1996). Both enzymes are composed of 572 amino acids and show 88% identity (Supplementary data Fig. S1). In addition, the nucleotide sequences of *ASN1* and *ASN2* genes are 81% identical throughout the ORFs (1,394 base-pairs [bp] of 1,719 bp match). Lack of asparagine synthetase activity leads to auxotrophy for asparagine, however, due to the redundant functions of Asn1p and Asn2p, disruption of *asn1* or *asn2* alone has no effect on auxotrophic phenotype. Combination of *asn1* and *asn2* mutations are required to lead to total asparagine auxotrophy in yeast (Fig. 1; Dang *et al.*, 1996).

All *S. cerevisiae* strains used in this study are listed in Table 1. Yeast extract-peptone (YP) media containing glucose (YPD) and synthetic drop-out (SD) media were prepared as described previously (Choi *et al.*, 2013). Insertion of a genetic marker at a specific position of chromosome was conducted by one-step gene replacement methods as previously described (Baudin *et al.*, 1993). Yeast colonies that grew on selective medium were
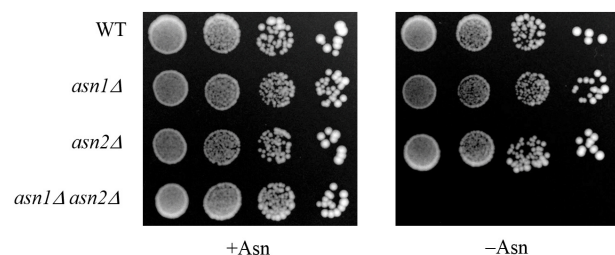

Fig. 1. Asparagine auxotrophic phenotype of the *asn1Δ asn2Δ* double mutant strain. Wild-type strain YJH40 (WT) and strains containing the indicated mutation were spotted in 10-fold serial dilutions onto SD containing (+Asn) and lacking (-Asn) asparagine plates and allowed to grow at 30°C for 3 days. The strains containing each mutation are as follows: *asn1Δ*: YKJ7; *asn2Δ*: YKJ9; *asn1Δ asn2Δ*: YKJ10.

analyzed by PCR for correct insertion. Induction of HO endonuclease was performed as described previously (Myung and Kolodner, 2003), with some modifications. Briefly, yeast cells were cultured at 30°C in SD media lacking amino acids required for elimination of spontaneous auxotrophic mutants, washed with sterile water two times, and suspended in an equal volume of YP media containing 2% (w/v) glycerol and 1% succinic acid. After additional culture for 5 h, 20% galactose was added to a final concentration of 2% to induce HO endonuclease and cells were incubated for an additional 2 h. Cells were then washed with sterile water two times and suspended in a 10 × volume of YPD and then incubated at 30°C for 16 h. The cells were then plated onto YPD plates and SD plates containing 5-fluoroorotic acid (5-FOA, 1 mg/ml). The SSA frequency was determined by dividing the number of colonies growing on SD containing 5-FOA by the number of colonies growing on YPD. For PCR and sequencing analyses of deletion structures, genomic DNAs were isolated from cells using the standard glass beads/chloroform-phenol procedure.

Fusion of two homologous sequences by SSA repair pathway produces a chimeric sequence (Anand *et al.*, 2014; Choi *et al.*, 2020). Based on these observations, we devised a genetic method to produce functional chimeric *ASN* genes. To construct HO-inducible SSA assay system, a 36-bp *MAT*a-derived HO cleavage site (HOcs) was introduced adjacent to *URA3* (HOcs-*URA3*) and then *HOcs-URA3-ASN1* integration cassette generated by overlap PCR was inserted downstream of *ASN2* in *asn1Δ* strain (Fig. 2A). Galactose induction of HO endonuclease resulted in efficient cleavage of HOcs (Choi *et al.*, 2020). A DSB between *ASN* genes can be repaired via two pathways. In non-homologous end-joining (NHEJ) pathway, the *URA3* marker remains intact and cells are sensitive to 5-FOA. On the other hand, DSB repair by SSA results in a deletion between *ASN*

Table 1. Yeast strains used in this study

Strain	Genotype	Reference
YJH40	*MATα hoΔ ade1-110 leu2-3,112 lys5 trp1::hisG ura3-52 hml::ADE1 hmr::ADE1 ade3::GAL::HO met2Δ can1Δ ura3Δ*, HOcs was mutated	Choi *et al.* (2020)
YKJ7	YJH40 *asn1Δ::ADE3*	This study
YKJ9	YJH40 *asn2Δ::URA3*	This study
YKJ10	YJH40 *asn1Δ::ADE3 asn2Δ::URA3*	This study
YKJ11	YKJ7 *HOcs-URA3-ASN1* inserted downstream of *ASN2*	This study
YLS15	YJH40 *HOcs-URA3* inserted downstream of *HXT10*	This study

genes and cells become resistant to 5-FOA (Fig. 2A). Therefore, chimeric *ASN* genes can be obtained by selecting colonies growing on 5-FOA-containing medium. The HOcs-*URA3* marker flanked by non-homologous sequence (YLS15) was so stable that no 5-FOA^r colonies were observed after plating 5×10^6 cells, and galactose induction of HO endonuclease resulted in only a moderate increase in the SSA frequency (Fig. 2B). In contrast, placing HOcs-*URA3* between *ASN* genes (YKJ11) produced a large number of 5-FOA^r colonies without galactose induction and the addition of galactose further increased the SSA frequency by 1,000-fold. This result suggests that the high SSA frequency in YKJ11 is due to *ASN* genes flanking HOcs-*URA3*. The 5-FOA^r colonies from YLS15 are presumed to result from extensive resection of DSB ends, and no further analysis was carried out.
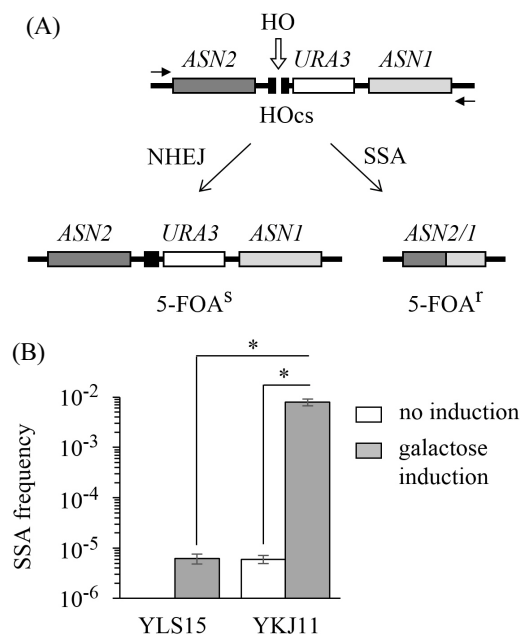

Fig. 2. HO-inducible SSA assay. (A) Overview of the HO-inducible SSA assay. HOcs denotes HO cleavage site. A downward arrow indicates the cleavage by HO endonuclease. Horizontal arrows indicate the position and direction of oligonucleotides used for PCR amplification. The sequences are as follows: Forward: 5'-GAC TCA TGG CAA GAT TTC TC-3'; Reverse: 5'-AAG TCA AGT GAG GTA TGT AG-3'. In this assay, DSB can be repaired by NHEJ or SSA. NHEJ restores HOcs leaving *URA3* intact, whereas SSA removes *URA3*, generating a chimeric *ASN* gene. (B) SSA frequencies in the presence (grey squares) and absence (open square) of galactose induction. SSA frequency was calculated by dividing the number of colonies growing on SD plates containing 5-FOA by the number of colonies growing on YPD plates. The results are the average (± SD) of 5 independent experiments. Asterisks indicate statistical significance (*P < 0.001).

The structures of recombined *ASN* genes were analyzed by PCR, using chromosomal DNA isolated from each 5-FOA^r cell derived from YKJ11. To ensure that all mutants were independently isolated and that multiple descendants of a single mutant cell were not analyzed, only one colony per experiment was randomly selected and analyzed. PCR analysis of 10 independent colonies (C1–C10) revealed that they all have only one type of deletion mutations (Fig. 3A) and the estimated size of PCR products was consistent with the structure predicted in Fig. 2A. Further PCR analysis confirmed that they all lacked both the intervening *URA3* marker and the intact *ASN* genes (data not shown). These results indicate that they were generated via SSA pathway. Additionally, the absence of intact *ASN* genes suggests that they show Asn$^+$ phenotype only if the chimeric gene is functional. When the mutant cells designated as C1–C10 were examined for asparagine auxotrophy by spotting assay, all were found to exhibit growth comparable to wild type strain on the plates lacking asparagine (Fig. 3B).

Next, we performed sequence analysis with C1–C10. The intervening sequence between *ASN* genes was completely removed, and all chimeric genes were found to have a single junction between two recombined *ASN* genes (Fig. 3C). The junctions are short stretches of fully matched sequences and ranged from 3 to 23 bp in lengths. Most junctions were located downstream of the gene, which is thought to be due to the uneven distribution of mismatches between *ASN* genes. The upstream 951 bp has 76% identity, whereas the downstream 765 bp has 88% identity (Fig. 3C). *E. coli* asparagine synthetase B (ASNB) consists of two domains: an N-terminal domain responsible for the hydrolysis of glutamine and a C-terminal domain required for the production of the aspartyl-AMP intermediate (Larsen *et al.*, 1999). Using ASNB as a reference, positions of the two domains in yeast Asn proteins were predicted (Fig. 3C). Most junctions of chimeric genes were in the C-terminal domain, but one in the N-terminal domain. This suggests that chimeric N-terminal domains may also be functional. Asn1p and Asn2p are highly homologous throughout the entire amino acid sequences (Supplementary data Fig. S1) and hence recombination at any point may produce an active chimeric enzyme without significant changes of protein structure. Consistent with this, more than 200 colonies resistant to 5-FOA were further investigated, and all showed Asn$^+$ phenotype (data
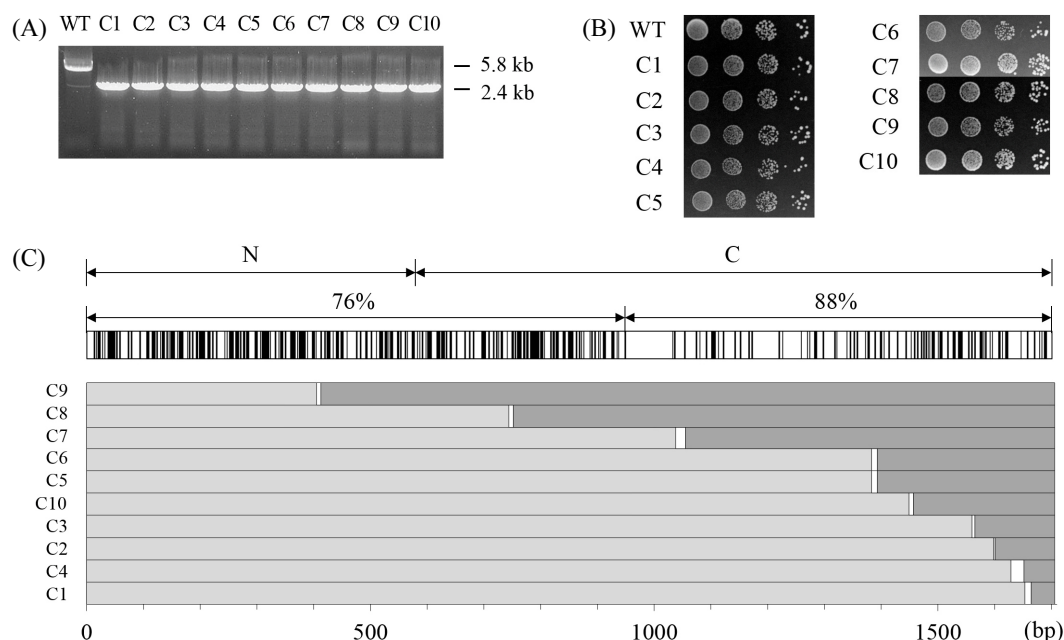
Fig. 3. Analyses of chimeric *ASN* genes. (A) PCR analysis of chimeric genes. DNA samples were extracted from YKJ11 (WT) and each mutant cell (C1–C10), followed by PCR reactions using primers indicated as horizontal arrows in Fig. 2A. (B) Growth of each mutant cell on SD plates lacking asparagine. YKJ11 (WT) and mutant cells containing chimeric genes (C1–C10) were spotted in 10-fold serial dilutions onto SD lacking asparagine and allowed to grow at 30°C for 3 days. (C) Sequence analysis of the chimeric genes. DNA sequence of *ASN1* is aligned with that of *ASN2* and base-pair differences are indicated to scale by black lines in the open box at the top of the graph. DNA sequences of the chimeric genes C1–C10 are represented by the lanes on the graph. Light and dark grey parts denote *ASN2* and *ASN1* sequences, respectively. The white areas indicate the junctions of gene fusion. Note that the junctions are the regions of fully matched sequences. N and C denote the deduced N- and C-terminal domains, respectively. Percentages denote levels of nucleotide identities.

not shown).

In this study, we developed a genetic method to produce functional chimeric genes based on a property that the homologous sequences are annealed together during DSB repair by SSA. Our results showed that all or near all chimeric *ASN* genes randomly generated by SSA were functional. Since a long heteroduplex intermediate is generated during SSA (George and Alani, 2012), production of functional chimeric genes without frameshift mutations suggests the possibility that this method could be applied to recombining other genes with lower homology than *ASN* genes. In addition, this method could provide a useful tool of screening for recombinant proteins with improved function by randomly fusing homologous genes. In this study, only a single junction was observed in each chimeric gene. If a method for generating multiple junctions is developed, it may be compatible with *in vitro* techniques such as DNA shuffling method (Acevedo-Rocha *et al.*, 2018). We recently found that some of the chimeric genes generated in mismatch repair-defective cells had multiple junctions (unpublished data) and are currently developing a method to

increase the frequency of multiple junctions.

Screening method for the proteins with desired properties is an important issue for protein engineering. Being able to select viable cells under specific conditions is a great advantage of *in vivo* mutagenesis. In this study, we have demonstrated that yeast auxotrophic phenotype is an effective selection method for functional proteins. We anticipate that improvement of this assay system will allow more screening methods to be applied to this method. Since random mutagenesis produces numerous combinations of mutations, library size is as important as library quality to improve or alter the function of proteins. Library sizes generated by *in vitro* techniques are typically limited by DNA cloning protocols and/or transformation efficiency (Packer and Liu, 2015). The genetic method described in this study does not involve these processes, making it easier to generate many variants than *in vitro* techniques. The relatively low mutagenesis efficiencies of most *in vivo* random mutagenesis protocols are the main reason for favoring *in vitro* methods such as error-prone PCR for library generation (Packer and Liu, 2015). However, in this study, due to the efficient site-specific

cleavage by HO endonuclease and direct selection of 5-FOA^r colonies, HO-inducible SSA assay provided a mutagenesis efficiency of nearly 100%. Thus, we expect that successful development of this method could break barriers to library size and library quality.

적 요

유사한 염기 서열을 가진 중복된 유전자(duplicated genes)는 진핵 생물 유전체에서 흔하게 발견되며, 이중에서 어떤 것들은 재조합을 허용할 만큼 충분한 상동성을 보여주기도 한다. *ASN1*과 *ASN2*는 open reading frame 전체에 걸쳐서 높은 상동성을 보이는 중복된 유전자의 한 예이다. 본 연구에서는 single-strand annealing 경로를 통하여 기능성 chimeric *ASN* 유전자를 제조하는 유전학적 방법을 고안하였다. HO endonuclease를 이용하여 *ASN1*과 *ASN2* 사이에 DNA double-strand break을 만들면 chimeric 유전자를 가진 돌연변이가 높은 빈도로 생성되었다. 돌연변이들은 모두 asparagine이 없는 배지에서 야생형과 비슷한 성장을 보였으며, 이는 chimeric *ASN* 유전자가 정상적인 기능을 한다는 것을 의미한다. Chimeric 유전자의 염기서열 분석 결과, 이들은 서로 다른 단일 접합부를 가지고 있었다. 이러한 결과는 상동성이 있는 유전자를 무작위로 융합하여 다양하고 많은 chimeric 유전자를 동시에 제조할 수 있는 방법을 제시하고 있다.

Acknowledgments

References

Acevedo-Rocha CG, Ferla M, and Reetz MT. 2018. Directed evolution of proteins based on mutational scanning. *Methods Mol. Biol.* **1685**, 87–128.

Anand RP, Tsaponina O, Greenwell PW, Lee CS, Du W, Petes TD, and Haber JE. 2014. Chromosome rearrangements via template switching between diverged repeated sequences. *Genes Dev.* **28**, 2394–2406.

Baudin A, Ozier-Kalogeropoulos O, Denouel A, Lacroute F, and

Cullin C. 1993. A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **21**, 3329–3330.

Bhargava R, Onyango DO, and Stark JM. 2016. Regulation of single-strand annealing and its role in genome maintenance. *Trends Genet.* **32**, 566–575.

Choi DH, Lee R, Kwon SH, and Bae SH. 2013. Hrq1 functions independently of Sgs1 to preserve genome integrity in *Saccharomyces cerevisiae*. *J. Microbiol.* **51**, 105–112.

Choi JH, Lim YS, Kim MK, and Bae SH. 2020. Analyses of DNA double-strand break repair pathways in tandem arrays of *HXT* genes of *Saccharomyces cerevisiae*. *J. Microbiol.* **58**, 957–966.

Dang VD, Valens M, Bolotin-Fukuhara M, and Daignan-Fornier B. 1996. Cloning of the *ASN1* and *ASN2* genes encoding asparagine synthetases in *Saccharomyces cerevisiae*: differential regulation by the CCAAT-box-binding factor. *Mol. Microbiol.* **22**, 681–692.

Dujon B. 1998. European Functional Analysis Network (EUROFAN) and the functional analysis of the *Saccharomyces cerevisiae* genome. *Electrophoresis* **19**, 617–624.

George CM and Alani E. 2012. Multiple cellular mechanisms prevent chromosomal rearrangements involving repetitive DNA. *Crit. Rev. Biochem. Mol. Biol.* **47**, 297–313.

Hanscom T and McVey M. 2020. Regulation of error-prone DNA double-strand break repair and its impact on genome evolution. *Cells* **9**, 1657.

Katju V, Farslow JC, and Bergthorsson U. 2009. Variation in gene duplicates with low synonymous divergence in *Saccharomyces cerevisiae* relative to *Caenorhabditis elegans*. *Genome Biol.* **10**, R75.

Larsen TM, Boehlein SK, Schuster SM, Richards NG, Thoden JB, Holden HM, and Rayment I. 1999. Three-dimensional structure of *Escherichia coli* asparagine synthetase B: a short journey from substrate to product. *Biochemistry* **38**, 16146–16157.

Li J, Sun H, Huang Y, Wang Y, Liu Y, and Chen X. 2019. Pathways and assays for DNA double-strand break repair by homologous recombination. *Acta Biochim. Biophys. Sin.* **51**, 879–889.

Lomelino CL, Andring JT, McKenna R, and Kilberg MS. 2017. Asparagine synthetase: Function, structure, and role in disease. *J. Biol. Chem.* **292**, 19952–19958.

Myung K and Kolodner RD. 2003. Induction of genome instability by DNA damage in *Saccharomyces cerevisiae*. *DNA Repair* **2**, 243–258.

Packer MS and Liu DR. 2015. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **16**, 379–394.

Ranjha L, Howard SM, and Cejka P. 2018. Main steps in DNA double-strand break repair: an introduction to homologous recombination and related processes. *Chromosoma* **127**, 187–214.