Supporting Information

Takeuchi et al. 10.1073/pnas.1107719108

SI Materials and Methods

Construction of the LAGLIDADG Multiple Sequence Alignment and Phylogenetic Analysis. Putative full-length LAGLIDADG homing endonuclease sequences (pf00961, pf03161, and pf05204) were collected from the Pfam database (1). The structure-based multiple sequence alignment was built using Cn3D (http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml). The structure of I-OnuI was aligned to the structure of I-AniI complexed with DNA substrate (PDB ID code 1P8K) (2). Collected pfam sequences were aligned to the I-OnuI and I-AniI structures, at which point the homodimeric LAGLIDADG sequences were removed based on the length of the open reading frame and the occurrence of a single LAGLIDADG motif per sequence. Elimination of the homodimeric LAGLIDADG homing endonucleases (LHEs) aided in the alignment of structurally homologous regions of single chain, monomeric LHEs. The sequence alignment generated by Cn3D was subsequently validated using a modified version of Jalview (3) that calculates the MIp/Zp covariation statistic (4) in real time while the alignment is edited. Groups of misaligned sequences were realigned to minimize local covariation, as local covariation is a unique indicator of misalignment that is independent of methods used to build the multiple sequence alignment (5). Local covariation was also used as a guide to reject partial and erroneous sequences. To prepare the sequence alignment for phylogenetic analysis, the alignment was trimmed such that only regions of putative structural homology were included; gapped positions and structurally divergent regions were removed from the alignment. A region of high local covariation was identified where no alternative (and putatively correct) alignment was found; this region was removed, as it is likely misleading. PhyML was used to calculate the tree (6). The approximate likelihood-ratio test values were calculated to provide statistical support for the branching order, and values greater than 0.7 are indicated at major nodes on the tree (Fig. 1*A*). The gene tree was rendered using FigTree. An enhanced version of the tree is provided as Fig. S1*A* with branches labeled by accession numbers (http://tree/bio.ed.ac.uk/software/figtree/). Partial sequences of all the LHEs identified are provided as Dataset S1.

Protein Expression and Purification. The open reading frames of I-OnuI and its variants were inserted between BamHI and NotI sites of pGEX 6P-3 vector (GE Healthcare Life Sciences), and the GST fusion recombinant proteins were expressed in *Escherichia coli* strain BL21-CodonPlus (DE3)-RIL (Agilent Technologies). Protein expression was induced in LB medium supplemented with 0.2% glucose, 1 mM MgSO4, and 100 μg/mL ampicillin at 16 °C for approximately 20 h after the culture had achieved early log growth phase (OD_{600} = ~0.6). The harvested cells were resuspended in TDG buffer [20 mM Tris-HCl (pH 7.5), 1 mM DTT, and 5% glycerol] supplemented with 0.5 M NaCl. After adding lysozyme to 0.5 mg/mL, the cells were sonicated for 30 s 6 times and stirred on ice for 30 min. The clarified cell lysate was obtained by centrifugation at 25,000 × *g* for 30 min at 4 °C, and nucleic acids were precipitated by adding polyethylenimine (pH 7.9) to 0.25% (vol/vol). After centrifugation at 25,000 × *g* for 10 min at 4 °C again, the supernatant was filtered through a 0.45-μm PVDF membrane and mixed with glutathione sepharose 4B beads (GE Healthcare Life Sciences). The beads were extensively washed with TDG buffer supplemented with 2 M NaCl and equilibrated with Digestion buffer [50 mM Tris-HCl (pH 7.0), 0.5 M NaCl, 1 mM DTT, and 5% glycerol]. The intact I-OnuI and subsequent variant proteins were eluted by incubation with PreScission protease (GE Healthcare Life Sciences) for 16 h at 4 °C. The collected proteins were concentrated and stored at −80 °C until use.

I-LtrI protein was expressed and purified as previously described (7), with a slight modification: The N-terminal tag was eliminated by digestion with thrombin before the protein was loaded on a Superdex 75 column (GE Healthcare Life Sciences). The purified protein was concentrated and stored at −80 °C until use.

Electrophoretic Mobility Shift Assay. Each reaction mixture contained 20 mM Tris-acetate (pH 7.5), 40 mM NaCl, 1 mM CaCl2, 1 mM DTT, 0.2 mg/mL BSA, 5% glycerol, I-OnuI or the variants, unlabeled T7 terminator primer, and 10 pM radiolabeled 34 base-pair DNA substrate containing the I-OnuI or monoamine oxidase B (MAO-B) target sequence. After the reaction mixtures were incubated on ice for 10 min, the protein–DNA complexes were separated on a 5% polyacrylamide gel containing 20 mM Tris-acetate (pH 7.5) and 1 mM CaCl2 at 4 °C. Gel images were taken with a Typhoon Trio Multi-mode imager (GE Healthcare Life Sciences). DNA bands were quantified using ImageJ, and dissociation constant (K_d) values were calculated using GraphPad Prism 5 software.

In Vitro Cleavage Assay. Each reaction mixture contained 20 mM Tris-acetate (pH 7.5), 140 mM potassium glutamate (pH 7.5), 10 mM NaCl, 1 mM MgCl2, 1 mM DTT, 0.2 mg/mL BSA, I-OnuI or the variants, and 10 pM of the radiolabeled substrate used for the electrophoretic mobility shift assays. The reactions were run at 37 °C for 30 min and were terminated by adding 4× Stop solution [40 mM Tris-HCl (pH 7.5), 40 mM EDTA, 0.4% SDS, 10% glycerol, 0.1% bromophenol blue, and 0.4 mg/mL proteinase K]. After incubation at 37 °C for 15 min, each sample was loaded and run on a 20% polyacrylamide-TBE (89 mM Tris borate and 2 mM EDTA) gel. The gel images were taken and DNA bands were quantified as described above.

Identification of Target Sequences for I-OnuI Subfamily. Target sites for each LHE were predicted through comparison of the LHE-harboring host gene to related genes lacking an endonuclease. Cleavage activity against each predicted target was verified using yeast surface-displayed enzyme in both in vitro and flow-cytometry-based tethered DNA cleavage assays (8) with the following modifications: Briefly, approximately 5×10^5 cells expressing a LHE were stained with 1:250 dilution biotinylated antibody against HA-epitope tag (Covance) and 1:100 FITC-conjugated αMyc (ICL Labs) for 30 min at 4 °C in 10 mM Hepes (pH 7.5), 180 mM KCl, 10 mM NaCl, 0.2% BSA, and 0.1% galactose. The cells were then stained with preconjugated streptavidin-PE:Biotin-dsOligo-A467 in the same buffer supplemented with 400 mM KCl. The cells were washed in the buffer containing 180 mM KCl and split into two wells. Each well was then resuspended in the same buffer supplemented with 2 mM of either MgCl2 or CaCl2. After incubation at 37 °C, the cells were pelleted by centrifugation, resuspended in the buffer containing 400 mM KCl and 4 mM EDTA to enhance release of the cleaved substrates, and analyzed on a BD LSRII cytometer (BD Biosciences).

To identify the exact cleavage positions and overhangs that were generated for each target site, the verified sequences were cloned into the pCTCON2-ARL vector (pCTCON2 with modified cloning sites) between the NdeI and XhoI cloning sites. Intact plasmid containing the target sequence was first digested

with the HindIII restriction enzyme to create a 6,300-bp linear substrate. Cleavage of the target sequence in this linear substrate created two bands (1.3 and 5 kb), which were clearly distinct from the uncut substrate (6.3 kb). Each in vitro cleavage reaction contained five million yeast expressing a LHE on the surface, 10 mM DTT (to release the enzyme from the yeast surface), 5–10 μg of HindIII-linearized target plasmid, 5 mM $MgCl_2$, 10 mM Hepes (pH 7.5), 180 mM KCl, 10 mM NaCl, 0.2% BSA, and 0.1% galactose. After incubation at 37 °C for 2–4 h, the yeast cells were then spun down and the supernatant was loaded onto an agarose gel. The two product bands were purified from the gel and sequenced, using a forward primer (5′- GTT CCA GAC TAC GCT CTG CAG G -3′) for the 5-kb band, and a reverse primer (5′- GTG CTG CAA GGC GAT TAA GT -3′) for the 1.3-kb band. The sequencing reads ended abruptly at the position of each DNA strand cleaved by the enzyme. At the 3′ end of each chromatogram, a large "A" peak was observed. Because this was a characteristic of the Taq DNA polymerase in the sequencing reagent, the final A peak was ignored.

Cleavage Specificity Profiling Using Yeast Surface Display. Cleavage activity was assayed as described above. The stained cells expressing the WT I-OnuI or engineered variant protein were incubated at 37 °C for 10 or 30 min and analyzed by flow cytometry. The cleavage activities were evaluated by calculating the ratio of the noncleavage median DNA-Alexa647 fluorescence intensity to the postcleavage intensity in the matching gate. The specificity profiles of WT I-OnuI and engineered variant were obtained by measuring the relative cleavage activities of 66 target sequences containing a single base-pair substitution from their original target site.

Crystallization of I-OnuI and I-LtrI Bound to Their Cognate Target Site. To crystallize I-OnuI, the DNA oligonucleotides (5′- CTT TCC ACT TAT TCA ACC TTT TAC CC -3′ and 5′- GGT AAA AGG TTG AAT AAG TGG AAA GG-3′) were purchased from Integrated DNA Technologies (1-μmol scale, HPLC purified). The oligonucleotides were dissolved in TE buffer [10 mM Tris-HCl (pH 8.0) and 1 mM EDTA], and the complementary DNA strands were annealed by incubation at 95 °C for 10 min and slow cooling to 4 °C over a 6-h period. Two hundred micromolar I-OnuI protein in 50 mM Hepes-NaOH (pH 7.5), 150 mM NaCl, 20 mM $MgCl_2$, and 5% (vol/vol) glycerol was mixed with a 1.2-fold molar excess of the DNA substrate. The protein–DNA drops were mixed in a 1:1 volume ratio with a reservoir solution containing 100 mM sodium acetate (pH 4.6), 100 mM ammonium sulfate, and 25% (vol/vol) polyethylene glycol 300 and equilibrated at 22 °C. Crystals were soaked in 100 mM sodium acetate (pH 4.6), 100 mM ammonium sulfate, and 30% (vol/vol) polyethylene glycol 300 at 4 °C overnight and frozen by looping and submersion into liquid nitrogen. The native crystals diffracted up to approximately 2.4-Å resolution at the Advanced Light Source (ALS) beamline 8.2.1.

I-OnuI protein incorporating selenium-substituted methionine (SeMet) (120 μM) in 50 mM Hepes-NaOH (pH 7.5), 150 mM NaCl, 20 mM $MnCl_2$, and 5% (vol/vol) glycerol was mixed with a 1.2-fold molar excess of the DNA substrate. The protein–DNA drops were mixed in a 1:1 volume ratio with a reservoir solution containing 200 mM magnesium acetate, 30 mM $ZnCl_2$, and 16% (vol/vol) polyethylene glycol 3350 and equilibrated at 22 °C. Crystals were soaked in 200 mM magnesium acetate, 30 mM $ZnCl_2$, 16% (vol/vol) polyethylene glycol 3350 and 12% glycerol, and frozen by looping and submersion into liquid nitrogen. The SeMet crystals diffracted up to approximately 3.6-Å resolution at the ALS beamline 5.0.2. Dataset S1 was processed using HKL2000 package (9). The polyalanine model of I-AniI/DNA complex (PDB ID code 1P8K) was used as a search model for molecular replacement. Two copies of the search model were found using the SeMet data by PHASER (10) and refined using REFMAC5 (11). The anomalous difference Fourier map was used to determine an orientation of the protein coordinates relative to the DNA target site. After the R_{free} reached 44%, the refined model was used to search the protein/DNA complex in the native data by molecular replacement. The coordinate was further refined by REFMAC5 and Crystallography & NMR System (12). The final model was deposited in Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank with RCSB ID code 3QQY. Statistics for the crystallographic data are shown in Table S4.

To obtain I-LtrI-DNA cocrystals, the DNA oligonucleotides (5′- GGT CTA AAC GTC GTA TAG GAG CAT TTG G -3′ and 5′- CAA ATG CTC CTA TAC GAC GTT TAG ACC C -3′) were purchased from Integrated DNA Technologies (1-μmol scale, standard desalting purification). The oligonucleotides were dissolved in TE buffer, and the complementary DNA strands were annealed by incubation at 95 °C for 10 min and slow cooling to 4 °C over a 6-h period. One hundred micromolar I-LtrI protein in 50 mM Hepes-NaOH (pH 7.5), 150 mM NaCl, 5 mM $MnCl_2$ and 5% (vol/vol) glycerol was mixed with a 1.5-fold molar excess of the DNA substrate. The protein–DNA drops were mixed in a 1:1 volume ratio with a reservoir solution containing 100 mM Bis-Tris (pH 6.5), 200 mM magnesium chloride, and 20% (vol/vol) polyethylene glycol 3500 and equilibrated at 22 °C. The crystals diffracted up to approximately 2.7-Å resolution at the ALS beamline 5.0.1. Dataset S1 was processed using HKL2000 package. The polyalanine model of I-OnuI/DNA complex (PDB ID code 3QQY) was used as a search model for molecular replacement. One copy of the search model was found and refined using REFMAC5. The final model was deposited in RCSB Protein Data Bank with RCSB ID code 3R7P. Statistics for the crystallographic data are shown in Table S4.

Bacteria-Based Two-Plasmid Cleavage Assay. This cleavage assay was carried out based on a previous report (13). Open reading frames of LHEs were inserted between NcoI and NotI sites of pEndo vector. The LHE gene is tightly regulated by pBAD promoter, and addition of *L*-arabinose promotes the gene transcription. Site-directed mutagenesis or random mutagenesis on I-OnuI gene was induced by overlap extension PCR, or using GeneMorph II Random Mutagenesis Kit (Agilent Technologies), by following the manufacturer's instructions. Two copies each of the I-OnuI or MAO-B target site were inserted between AflIII and BglII and between NheI and SacII of pCcdB plasmid (four copies total, Fig. S3*A*). This plasmid encodes "control of cell death B," a toxic protein to bacteria, and the expression of this gene is induced by IPTG. Cleavage of the target sites by the expressed LHEs leads to the RecBCD-mediated degradation of the reporter plasmid and rescues the cell growth on the selective medium containing IPTG. The pEndo plasmid was transformed into NovaXGF' (Novagen) competent cells harboring pCcdB plasmid (containing four copies of the I-OnuI or MAO-B target) by electroporation. The transformants were grown in 2× YT medium (16 g/L tryptone, 10 g/L yeast extract, and 5 g/L NaCl) at 37 °C for 30 min and were 10-fold diluted with 2× YT medium supplemented with 100 μg/mL carbenicillin and 0.02% *L*-arabinose. After the culture was grown at 30 °C for 4–15 h, the cells were harvested, resuspended in sterilized water, and spread on both nonselective plates (1× M9 salt, 1% glycerol, 0.8% tryptone, 1 mM $MgSO_4$, 1 mM $CaCl_2$, 2 μg/mL thiamine, and 100 μg/mL carbenicillin) and selective plates (the nonselective plates supplemented with 0.02% *L*-arabinose and 0.4 mM IPTG). For negative selection to remove the variants active against the wild-type target sequence, the transformants were spread on the selective plates containing 33 μg/mL chloramphenicol instead of IPTG. The plates were incubated at 30 °C for 30–40 h. To proceed to the next round of selection, the pEndo plasmid was extracted

from the surviving colonies on the selective plates. The ORFs of I-OnuI variants were recovered by PCR amplification and digested with NcoI, NotI, and ScaI or PvuI. The resulting fragments were cloned into pEndo vector and subjected to further selection.

Episomal GFP Reporter Gene Conversion Assay. In the plasmid to express a LHE gene (pExodus), a LHE gene including the N-terminal HA tag followed by a nuclear localization signal was linked to mCherry gene by the 2A peptide sequence from *Thosea asigna* virus (Fig. 3*B*). The two-gene expression was driven by CMV promoter, and the cotranslated proteins were separated by ribosomal skipping (14). DR-GFP reporter codes a GFP gene sequence interrupted by a HE target site and an in-frame stop codon, followed by the truncated gene sequence (15). HEK 293T cells were grown in DMEM supplemented with 10% fetal bovine serum, 10 units/mL penicillin and 10 μg/mL streptomycin at 37 °C in 5% CO_2 atmosphere. HEK 293T cells (6×10^4) were plated 24 h prior to transfection in 12-well plates, and transfected with 0.25 μg each of DR-GFP reporter and pExodus plasmid using Fugene 6 transfection reagent (Roche Applied Science). The GFP-positive cells were detected by flow cytometry at 48 h post transfection. Western blotting was carried out using a rabbit polyclonal antibody against HA-epitope tag and a mouse monoclonal antibody against β-actin.

Targeted Mutagenesis at Chromosomal Loci. HEK 293T cells (1.3×10^5) were plated 24 h prior to transfection in six-well plates and transfected with 1 μg of pExodus plasmid. The top 25% and the following 25% of mCherry positive cells (fluorescent marker for a LHE gene expression) were separately collected using BD FACSAria cell sorter (BD Biosciences) 48 h post transfection. To extract genomic DNA, the sorted cells (approximately 1×10^5) were washed with cold PBS buffer, resuspended in TNES buffer [50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 10 mM EDTA, 1% SDS, 0.25 mg/mL proteinase K], and incubated at 50 °C for 30 min. RNase A was added to 0.25 mg/mL, and the reaction mixture was further incubated at the same temperature for 30 min. The genomic DNA was recovered by phenol/chloroform/isoamyl alchohol extraction followed by ethanol precipitation. Both on-target (i.e., MAO-B gene) and off-target loci were amplified from 50–80 ng of the extracted genomes using Phusion DNA polymerase (Finnzymes). The DNA products resulting from two rounds of PCR amplification were cleaned using PCR purification kit (Qiagen), and 150 ng of the fragments were incubated with 1.5–3.0 pmol of E2 I-OnuI recombinant protein in 20 mM Tris-acetate (pH 7.5), 100 mM potassium acetate (pH 7.5), 1 mM DTT, and 10 mM $MgCl_2$ at 37 °C for 2 h. The cleavage reactions were terminated by adding 4× Stop solution. After incubation at 37 °C for 30 min, each sample was separated on a 1.8% agarose gel containing ethidium bromide in TBE. The DNA bands were quantified using ImageJ software.

SI Results

Directed Evolution of I-OnuI to Target a Sequence in the Third Exon of the Human Monoamine Oxidase B Gene. In order to select the I-OnuI variants that cleaved the MAO-B target, we employed the two-plasmid system in bacteria, originally developed by Doyon et al. (13). A reporter plasmid containing either the I-OnuI or MAO-B target site (pCcdB/I-OnuI or pCcdB/MAO-B) encodes control of cell death B ("ccdB," a toxic protein in bacteria, which is inducible by addition of IPTG). Cleavage of the target sites in the reporter plasmid by variants of I-OnuI (encoded on pEndo plasmid) leads to RecBCD-mediated degradation of the reporter plasmid and corresponding cell survival on selective medium containing IPTG (Fig. S3*A*). In this system, I-OnuI increased a survival rate of cells harboring pCcdB/I-OnuI reporter to approximately 100%. Inactivation of catalytic activity by E22Q substitution failed to restore the cell growth on the selective plates, indicating that the hydrolysis of the target site is required to eliminate the toxic reporter plasmid (Fig. S3*B*). Although the WT I-OnuI enzyme tolerated each of four base-pair mismatches out of five in the MAO-B target to some extent, and more efficiently cleaved the other one (A at the base position −4) (Fig. 3*A*), it was unable to rescue cells harboring pCcdB/MAO-B reporter in the presence of IPTG (Fig. S3*B*).

We initially substituted all 20 possible amino acids at the positions of residues 32, 35, 40, and 229 of I-OnuI, which are adjacent to the base-pair positions −11, −10, and +2, and subjected the resulting combinatorial library to positive selection for cleavage of the MAO-B target, after preinducing an expression of the variant genes for 15 h. After two rounds of the selection, we also varied residues 48, 72, and 80 (which are adjacent to the base position −4) within the surviving variants with a collection of 13 amino acids (A, C, D, E, G, H, K, N, Q, R, S, T, and Y) and again selected active variants with the same procedures as for the former selections, with the exception that the stringency of selection was increased by preinducing enzyme expression for only 4 h before plating, instead of 15 h. The selected variant genes were further mutated by error-prone PCR and subjected to a single round of positive selection, followed by a round of negative selection using the pCcdB/I-OnuI reporter, in order to eliminate the variants with broadened, rather than altered, specificity. After another round of positive selection, we analyzed the sequences of the surviving variant genes and tested the activities of the individual variants in bacteria.

The resulting I-OnuI variant (termed "E0 I-OnuI") possessed nine amino acid substitutions (N32S, N33S, S35R, S40A, T48C, N51I, K80R, K189N, and K229R) and preferentially increased a survival rate of cells harboring pCcdB/MAO-B reporter (Fig. S3*B*). Because this variant contained three amino acid substitutions that resulted from the final random mutagenesis step, we tested an impact of each of the random substitutions on the activity of E0 I-OnuI in bacteria. An N189K variant (which reverted N189 back to a wild-type lysine residue) lost its sequence preference for the MAO-B target, whereas the other two reversions (S33N and I51N) displayed a similar specificity to the E0 variant (Fig. S3*B*).

We then measured the gene conversion activity of these I-OnuI variants in transfected HEK 293T cells, using the DR-GFP reporter containing the I-OnuI or MAO-B target site (DR-GFP/I-OnuI or DR-GFP/MAO-B). E0 I-OnuI displayed a 4-fold increase in the fraction of GFP-positive cells harboring the DR-GFP/MAO-B reporter, while promoting recombination of the reporter harboring I-OnuI target site to the same level as for catalytically inactivated I-OnuI (I-OnuI+E22Q) (Fig. S4*B*). An addition of E22Q in the E0 variant also significantly decreased the gene conversion frequency on DR-GFP/MAO-B reporter to the same level induced by the wild-type I-OnuI enzyme.

An E178D substitution was observed in several types of I-OnuI variants isolated through the directed evolution in bacterial cells. Introducing an E178D mutation in the original I-OnuI variant (E0 I-OnuI) enhanced gene conversion on both types of the DR-GFP reporters by approximately 3-fold. We also found that replacing S33 or I51 with the wild-type residue, asparagine (N), slightly influenced the gene conversion activity and therefore chose the S33N variants for further analyses. "E1 I-OnuI" was derived from E0 I-OnuI by substituting the wild-type residue (N) for S33, and the variant with additional E178D in the E1 background was termed "E2 I-OnuI."

1. Finn RD, et al. (2010) The Pfam protein families database. *Nucleic Acids Res*, 38: D211–222.
2. Bolduc JM, et al. (2003) Structural and biochemical analyses of DNA and RNA binding by a bifunctional homing endonuclease and group I intron splicing factor. *Genes Dev* 17:2875–2888.
3. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview version 2—A multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25:1189–1191.
4. Dunn SD, Wahl LM, Gloor GB (2008) Mutual information without the influence of phylogeny or entropy dramatically improves residue contact prediction. *Bioinformatics* 24:333–340.
5. Dickson RJ, Wahl LM, Fernandes AD, Gloor GB (2010) Identifying and seeing beyond multiple sequence alignment errors using intra-molecular protein covariation. *PLoS One* 5:e11082.
6. Guindon S, et al. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59:307–321.
7. Sethuraman J, Majer A, Friedrich NC, Edgell DR, Hausner G (2009) Genes within genes: Multiple LAGLIDADG homing endonucleases target the ribosomal protein S3 gene encoded within an rnl group I intron of Ophiostoma and related taxa. *Mol Biol Evol* 26:2299–2315.
8. Jarjour J, et al. (2009) High-resolution profiling of homing endonuclease binding and catalytic specificity using yeast surface display. *Nucleic Acids Res* 37:6871–6880.
9. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276:307–326.
10. McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Cryst* 40:658–674.
11. Winn MD, Murshudov GN, Papiz MZ (2003) Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol* 374:300–321.
12. Brunger AT, et al. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54:905–921.
13. Doyon JB, Pattanayak V, Meyer CB, Liu DR (2006) Directed evolution and substrate specificity profile of homing endonuclease I-SceI. *J Am Chem Soc* 128:2477–2484.
14. Szymczak AL, et al. (2004) Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral vector. *Nat Biotechnol* 22:589–594.
15. Pierce AJ, Johnson RD, Thompson LH, Jasin M (1999) XRCC3 promotes homology-directed repair of DNA damage in mammalian cells. *Genes Dev* 13:2633–2638.

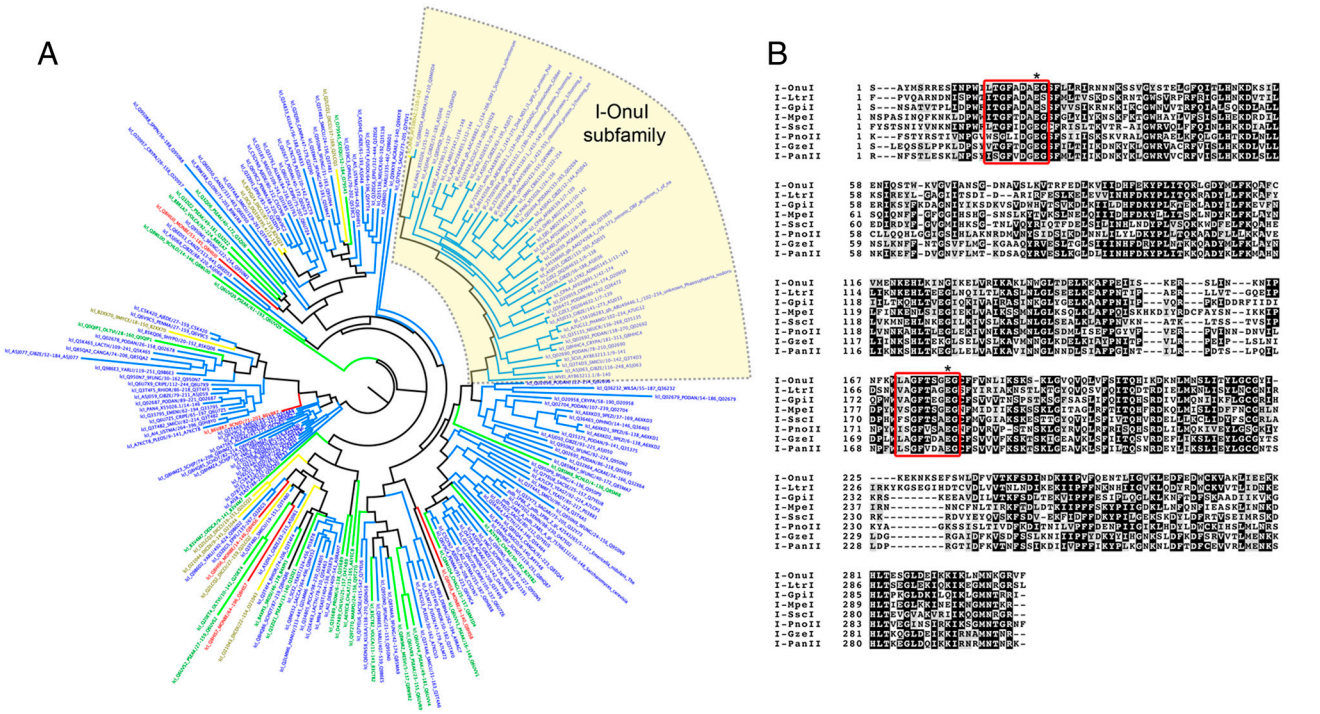

Fig. S1. (*A*) Phylogenetic tree of monomeric LAGLIDADG homing endonucleases with individual accession numbers. (*B*) The sequence alignment of I-OnuI subfamily was carried out using the MAFFT program (http://www.ebi.ac.uk/Tools/msa/). The conserved LAGLIDADG motifs are highlighted in red boxes. Asterisks (*) indicate the C-terminal acidic residues of each LAGLIDADG motif. The accession numbers for I-OnuI, I-LtrI, I-GpiI, I-MpeI, I-SscI, I-PnoII, I-GzeI, and I-PanII are AAY59060, ACV41164, ACV41152, AAQ74268, AAC48982, ABU49446, ABC86623, and CAA38767, respectively.

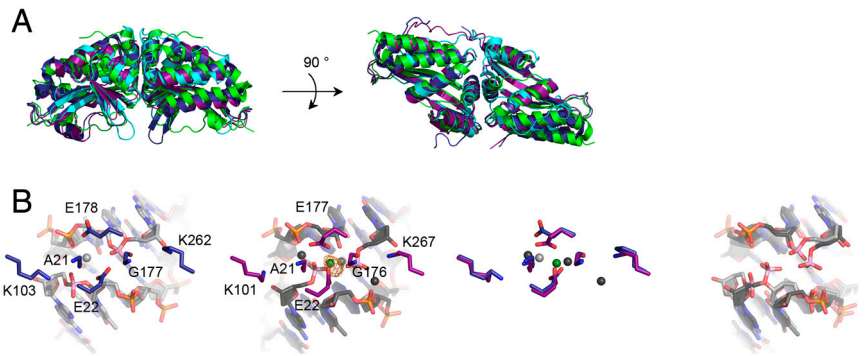

Fig. S2. (*A*) Superposition of I-OnuI, I-LtrI, I-AniI, and I-CreI. I-OnuI, I-LtrI, I-AniI (PDB ID code 1P8K), and I-CreI (PDB ID code 1G9Y) are colored in dark blue, purple, cyan, and green, respectively. Superposition of I-AniI and I-CreI on I-OnuI/I-LtrI yields rmsd values of 1.85/1.77 Å and 2.05/2.21 Å, respectively. (*B*) Active sites of I-OnuI and I-LtrI. The DNA substrates bound to I-OnuI and I-LtrI are colored in light gray and dark gray, and the scissile phosphates in light pink. Magnesium ions bound to I-OnuI and I-LtrI are depicted as light gray and dark gray spheres. A manganese ion (green sphere) situated in the center of the I-LtrI active site is verified by an anomalous Fourier map (orange mesh, contoured at 5.0 σ). Residue numbers of I-LtrI crystal shown here are shifted from the numbers assigned in the deposited PDB file, in order to fit the residue numbers of the first LAGLIDADG motif to the corresponding numbers of I-OnuI crystal. E29 in the original PDB file are labeled as E22.

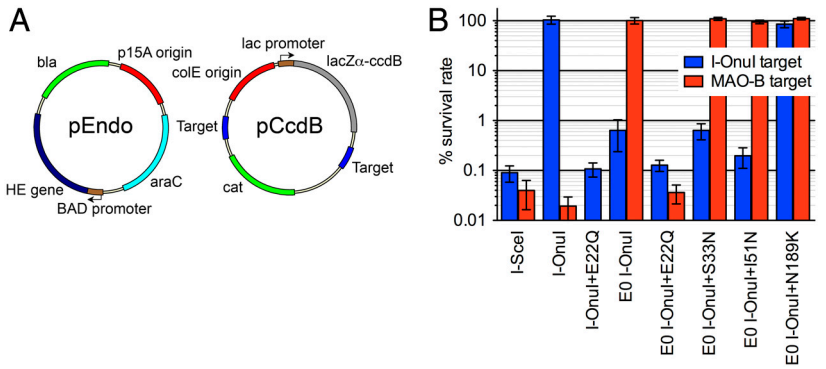

Fig. S3. (*A*) Schematic representation of two plasmids used for selection of I-OnuI variants with altered specificity. The plasmid maps were adapted from ref. 13. (*B*) The bacterial cells that carried pEndo plasmid expressing I-SceI, I-OnuI, or I-OnuI variants and pCcdB plasmid containing the I-OnuI target or MAO-B target were grown at 30 °C for 4 h and spread both on nonselective plates and on selective plates where an expression of *ccdB* gene, a poison to bacteria, was induced. Survival rates were determined by dividing the number of colonies on the selective plates by that on the nonselective plates. Error bars refer to ±SD of three to four independent experiments. E0 I-OnuI+S33N corresponds to E1 I-OnuI.

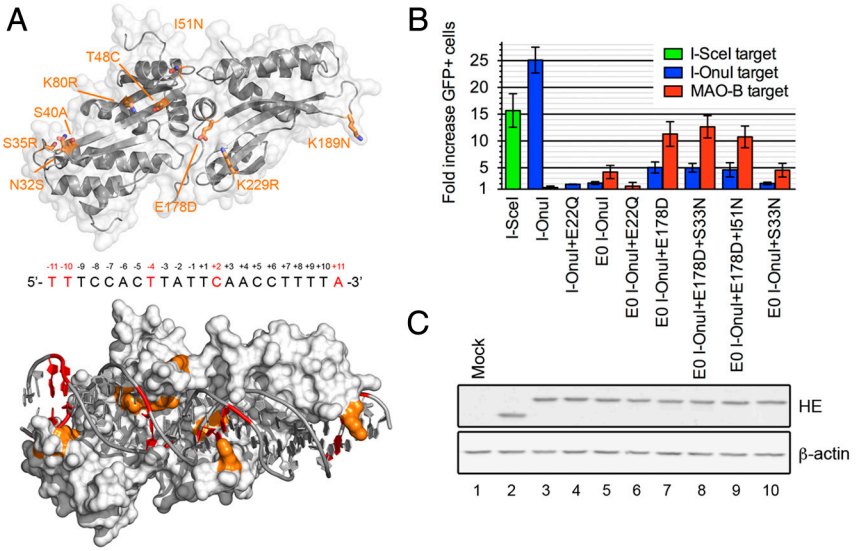

Fig. S4. (*A*) Amino acid substitutions introduced in E2 I-OnuI variant (orange) are mapped on the wild-type I-OnuI crystal structure. A target sequence for the WT I-OnuI is shown with base-pair positions mismatched to the MAO-B target colored in red. (*B*) Gene conversion activity was assayed using an episomal DR-GFP reporter. The data represents an increase in a fraction of GFP-positive cells by cotransfection with a HE gene compared to the background observed by transfection with the DR-GFP reporter alone. Error bars refer to ±SD of three independent experiments. E0 I-OnuI+S33N and E0 I-OnuI+E178D+S33N correspond to E1 I-OnuI and E2 I-OnuI, respectively. (*C*) Whole cell lysates were extracted from the cotransfected cells with a HE expression plasmid (indicated on top of the panels) and DR-GFP reporter containing the I-SceI (lane 2), I-OnuI (lanes 3 and 4), or MAO-B target (lanes 5–10). Western blotting was carried out using antibodies against HA-epitope tag (for HE proteins) and β-actin.

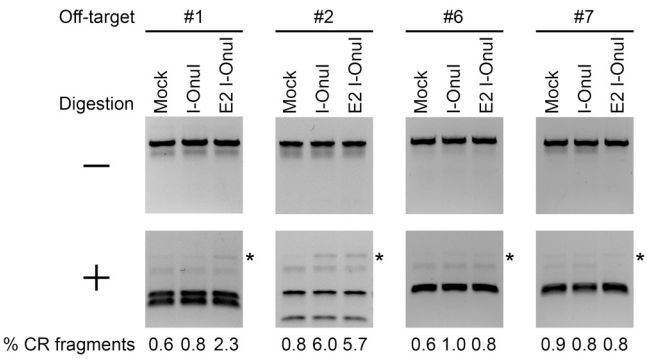

Fig. S5. Potentially cleavable off-target sites from the cells collected in the sorting gate H (Fig. 5*A*) were amplified by PCR and subjected to a single round of in vitro digestion with E2 I-OnuI recombinant protein. Enzymes expressed in the transfected cells are indicated on top of panels. Asterisks (*) indicate the cleavage-resistant (CR) fragments. These results were consistent between experiments using independently transfected and sorted cells.

Table S1. Dissociation constant (K_d) values of I-OnuI and its variants

Protein	Target	K_d, pM
I-OnuI	I-OnuI	193 ± 15
E1 I-OnuI		380 ± 220
E2 I-OnuI		380 ± 70
I-OnuI	MAO-B	1,300 ± 180
E1 I-OnuI		350 ± 120
E2 I-OnuI		350 ± 160

The K_d values were obtained from three independent experiments (means ± SD).

Table S2. Mutation frequency induced by the I-OnuI variants at the endogenous MAO-B locus

Experiment	Gate	Transfected cells	CR fragments, %		Mutation frequency, %	Fold increase
			(1st round)	(2st round)		
#1	H	Mock	1.37	0.38	0.0052	1
	H	I-OnuI	1.39	0.35	0.0048	0.9
	H	E1 I-OnuI	1.09	10.78	0.1173	22.6
	H	E2 I-OnuI	7.31	68.01	4.9704	956.2
	M	E2 I-OnuI	4.10	68.99	2.8279	544.0
#2	H	mock	0.87	0.52	0.0045	1
	H	I-OnuI	1.19	0.66	0.0078	1.7
	H	E1 I-OnuI	1.22	5.98	0.0730	16.2
	H	E2 I-OnuI	6.25	64.63	4.0390	898.2
	M	E2 I-OnuI	2.85	58.90	1.6764	372.8

The mutation frequency was estimated from the following equation: (Mutation frequency) = (fraction of the CR fragments in the first round of digestion) × (fraction of the CR fragments in the second round of digestion) × 100.

Table S3. List of the sequences closely matched to the MAO-B target in human genome

	Sequence	Chromosome
MAO-B target	GGTCCACATATTTAACCTTTTG	X
I-OnuI target	TT.....T....C.....A	
Off-target #1	TA.....	13
Off-target #2	.T.....C.....T	4
Off-target #3	AC....A.....	X
Off-target #4	TC.....T.....	1
Off-target #5	.T.....T....C	2
Off-target #6	T..A.....A	6
Off-target #7	TC.....CA	4
Off-target #8	AA.....C.....T	1
Off-target #9	TT.....C.....C	1
Off-target #10	T.....C.AC	4
Off-target #11	AT....A.....T	Y
Off-target #12	TC.....G....GC	4
Off-target #13	TA.....A....CC	4
Off-target #14	AA.....A...CT	12
Off-target #15	TA..A.....GC	1
Off-target #16	TA.....G.....GT	6
Off-target #17	TA.....C.....CC	7
Off-target #18	TA.....A...CT	14
Off-target #19	TTG.....CT	12

Table S4. Data collection and refinement statistics

	I-OnuI (native)	I-OnuI (Se-Met)	I-LtrI
Data collection			
Beamline	ALS 8.2.1	ALS 5.0.2	ALS 5.0.1
Wavelength, Å	1.000	0.979	0.977
Space group	$P2_12_12_1$	C_12_1	C_12_1
Unit cell dimension, Å	$a = 37.948$, $b = 73.933$, $c = 166.930$	$a = 93.068$, $b = 168.470$, $c = 78.422$, $\beta = 124.232°$	$a = 113.100$, $b = 42.594$, $c = 103.191$, $\beta = 110.634°$
Total reflections	124,406	88,838	86,865
Unique reflections	18,568	11,845	12,740
Resolution, Å	50.0–2.40 (2.49–2.40)	50.0–3.60 (3.73–3.60)	50–2.70 (2.80–2.70)
Completeness, %	96.8 (91.0)	99.9 (99.9)	99.2 (98.7)
Redundancy	6.7 (4.8)	7.5 (7.6)	6.8 (6.5)
R_{merge}	0.080 (0.143)	0.102 (0.434)	0.065 (0.240)
Average $I/\sigma(I)$	18.6 (9.62)	21.0 (4.95)	27.0 (6.97)
Refinement			
R_{work}, %	18.9		19.2
R_{free}, %	24.0		27.0
Number of atoms	3,553		3,507
Protein	2,419		2,356
Nucleic acid	1,060		1,103
Water	73		43
Metal ions	1		5
rmsd bond length, Å	0.018		0.008
rmsd bond angle, °	2.217		1.337
Average B factor, Å^2	30.6		52.5
Protein, Å^2	30.0		51.2
Nucleic acid, Å^2	33.1		56.2
Water, Å^2	11.3		29.3
Metal ion, Å^2	42.7		38.8
Ramachandran distribution			
Most favored, %	90.4		87.9
Additionally allowed, %	8.5		11.0
Generously allowed, %	0.0		1.1
Disallowed, %	1.1		0.0

Other Supporting Information Files

Dataset S1 (PDF)