

The Mitochondrial LSU rRNA Group II Intron of *Ustilago maydis* Encodes an Active Homing Endonuclease Likely Involved in Intron Mobility

Anja Pfeifer^{‡a,‡b}, Bettina Martin, Jörg Kämper, Christoph W. Basse*

Department of Genetics, Institute for Applied Biosciences of the Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

Abstract

Background: The *a2* mating type locus gene *Iga2* is critical for uniparental mitochondrial DNA inheritance during sexual development of *Ustilago maydis*. Specifically, the absence of *Iga2* results in biparental inheritance, along with efficient transfer of intronic regions in the large subunit rRNA gene between parental molecules. However, the underlying role of the predicted LAGLIDADG homing endonuclease gene *I-Umal* located within the group II intron *LRII1* has remained unresolved.

Methodology/Principal Findings: We have investigated the enzymatic activity of *I-Umal* *in vitro* based on expression of a tagged full-length and a naturally occurring mutant derivative, which harbors only the N-terminal LAGLIDADG domain. This confirmed Mg^{2+} -dependent endonuclease activity and cleavage at the *LRII1* insertion site to generate four base pair extensions with 3' overhangs. Specifically, *I-Umal* recognizes an asymmetric DNA sequence with a minimum length of 14 base pairs (5'-GACGGGAAGACCCT-3') and tolerates subtle base pair substitutions within the homing site. Enzymatic analysis of the mutant variant indicated a correlation between the activity *in vitro* and intron homing. Bioinformatic analyses revealed that putatively functional or former functional *I-Umal* homologs are confined to a few members within the Ustilaginales and Agaricales, including the phylogenetically distant species *Lentinula edodes*, and are linked to group II introns inserted into homologous positions in the LSU rDNA.

Conclusions/Significance: The present data provide strong evidence that intron homing efficiently operates under conditions of biparental inheritance in *U. maydis*. Conversely, uniparental inheritance may be critical to restrict the transmission of mobile introns. Bioinformatic analyses suggest that *I-Umal*-associated introns have been acquired independently in distant taxa and are more widespread than anticipated from available genomic data.

Citation: Pfeifer A, Martin B, Kämper J, Basse CW (2012) The Mitochondrial LSU rRNA Group II Intron of *Ustilago maydis* Encodes an Active Homing Endonuclease Likely Involved in Intron Mobility. PLoS ONE 7(11): e49551. doi:10.1371/journal.pone.0049551

Editor: Anthony R. Poteete, University of Massachusetts Medical School, United States of America

Received: September 20, 2012; **Accepted:** October 10, 2012; **Published:** November 14, 2012

Copyright: © 2012 Pfeifer et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This research was supported by the Karlsruhe Institute of Technology (KIT). In addition, the authors are grateful to the KIT publication fund for financial support. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: christoph.basse@kit.edu

‡a Current address: Université de Nice-Sophia Antipolis, Nice, France

‡b Current address: Institut National de la Santé et de la Recherche Médicale (INSERM) U1091, Institut de Biologie Valrose (IBV), Diabetes Genetics Team, Nice, France

Introduction

Homing endonuclease genes (HEGs) are widespread in microbial genomes. They frequently exist in self-splicing group I and group II introns, but also in archaeal introns, intein coding sequences and phage genomes [1–3]. On the basis of conserved amino acid motifs, at least five families of homing endonuclease (HE) proteins are distinguished [3–7]. The largest known family, termed the LAGLIDADG homing endonucleases (LHEs), is primarily encoded within archaea and in the mitochondrial or chloroplast genomes of algae and fungi [5,6]. LAGLIDADG enzymes contain one or two copies of the consensus motif. Specifically, single-motif enzymes function as homodimers, whereas double-motif enzymes are monomers with two separate domains, each resembling a subunit of a single LAGLIDADG protein [5]. LAGLIDADG motifs are not only restricted to homing endonucleases, but also exists in other proteins, such as the

HO endonuclease, which in yeast mediates the mating type switch [8,9].

HEG-containing introns are generally considered opportunistic selfish elements, with the ability to spread within or between genomes if the corresponding homing sites are present on recipient DNA [10]. Cleavage by homing endonucleases differs from restriction enzymes in that longer target sites, with lengths between 14 to 40 base pairs, are recognized [1,3,5,11]. These sites match exactly the intron insertion site in donor DNA, meaning that only DNA is cut that does not contain a copy of the intron interrupting the target site. HEG-containing introns are mobilized by gene conversion, which is initiated by double strand cleavage within the intronless allelic sites on recipient DNA. Subsequent recombination using the intron-containing copy as template confers a homozygous state with two intron-containing alleles, while the intronless allele is lost [2,3,5,11,12]. Intron homing occurs with efficiencies close to 100% as exemplified from

transfer of the I-SceI-containing mitochondrial group I intron omega (ω) in combinations of *Saccharomyces cerevisiae* ω^+ and ω^- cells [5,13,14,15]. However, in the majority of sexual eukaryotes, uniparental mitochondrial DNA (mtDNA) inheritance efficiently prevents recombination between parental mtDNAs making it difficult to address an underlying role of HEGs [16,17].

Previously, we identified a restriction length polymorphisms in the mitochondrial large subunit (LSU) ribosomal RNA (rRNA) gene of the maize smut fungus *Ustilago maydis* [18]. Within this polymorphic region, the individual mitochondrial genotypes (mitotypes) differ in the number and position of HEG-containing group I and group II introns. In particular, the W type differs from the F type by the presence of the group II intron LRIII1, but lacks flanking F type-associated group I introns (Figure 1A). Investigation of uniparental mtDNA inheritance revealed that mtDNA was preferentially transmitted from the *a2* partner by virtue of the *a2* mating type locus genes *lga2* and *rga2*. Hereby, *lga2* mediates loss of *a1*-associated mtDNA, while *rga2* protects *a2*-associated mtDNA from *lga2*-mediated elimination. This provided for conditions of biparental inheritance established either in the absence of *lga2* or in the presence of an *a1* partner ectopically expressing *rga2* [18]. Interestingly, under these conditions, recombinant mtDNA molecules were efficiently produced, and apparently, this proceeded in an unidirectional manner. Specifically, in combinations of F and W types, the F type was almost completely lost in favor of the recombinant X1 type, which matched the parental F type, but additionally carried the W type-derived LRIII1 intron (Figure 1A; [18]). This intron contains a predicted HEG (here termed I-UmaI according to the corresponding nomenclature convention; [19]), raising the question of underlying intron homing. In addition, we previously identified a W type strain (MF18) in which due to a naturally occurring frameshift mutation the second LAGLIDADG domain of I-UmaI is not expressed. Only in combinations with this strain, the parental F type was maintained and significantly less recombinant X1 type was produced [18]. This suggested a role of I-UmaI in the mobilization of the LRIII1 intron.

To provide firm evidence for intron homing among mitochondrial LSU rRNA genes, we have analyzed enzymatic activities of I-UmaI and a mutated variant lacking the second LAGLIDADG motif. In addition, we have determined the target site specificity. Here, we show that I-UmaI represents a Mg^{2+} -dependent endonuclease requiring both LAGLIDADG domains for activity and recognizing a minimum target site of 14 base pairs, which defines the LRIII1 insertion site. The gained insight has further been exploited to make predictions on the existence of putatively functional I-UmaI homologs sharing the same cleavage specificity.

Results

Expression of I-UmaI

To express the I-UmaI gene in *Escherichia coli*, the mitochondrial codon usage of *U. maydis* was determined based on its annotated mitochondrial genome sequence (NCBI accession no. DQ157700) comprising 15 genes for proteins of the respiratory chain complex and 11 putative HEGs within introns of the LSU rRNA, *cox1* and *cob* genes (Table S1). Comparative sequence analysis revealed major differences between the mitochondrial codon usage of *U. maydis* and *S. cerevisiae* (Table S2). In conclusion, except for the very rarely occurring triplets TTA/G and AGG, which were absent from the I-UmaI sequence, and the complete absence of the AGA and TGA codons, the mitochondrial code of *U. maydis* basically did not deviate from the standard code. In addition, the rarely occurring ATA codon, which occurs at nucleotide position 787 in the I-UmaI sequence, likely encodes Ile (Table S2). Therefore, the

I-UmaI open reading frame (ORF) required no adaptation for expression in *E. coli*. The I-UmaI gene was conditionally expressed with a C-terminal His tag extension. The corresponding construct (pAP1) additionally expressed an N-terminal thioredoxin (THX) domain to assure solubility of the product. In addition, a plasmid (pAP2) was used lacking the THX extension to exclude a possible influence on enzymatic activity. Immunoblot analysis confirmed proper expression of both fusion proteins (here termed AP1 and AP2, respectively) as well as their affinity-based purification by Ni^{2+} -NTA column chromatography (Figure 2A).

To assess enzymatic activities of AP1 and AP2, *in vitro* assays were conducted based on cleavage of supercoiled double-stranded plasmid DNA (pSL521), which contains the putative target site region (Figure 1B and Materials and Methods). This revealed that AP1 and AP2 from both crude (CF) or affinity-purified (PF) fractions were able to linearize substrate DNA. As expected, cleavage activity was neither detected in protein extracts from non-induced *E. coli* cells (lanes 1,4 in Figure 2A) nor from non-transformed cells incubated under inducing conditions (lane 6 in Figure 2A).

Determination of the I-UmaI Target Site and Cleavage Conditions

To determine the I-UmaI target site, cleaved pSL521 reaction products were blunt ended in the presence of T4 DNA polymerase and religated either in the presence or absence of an intervening DNA fragment (see Materials and Methods). Sequence analysis of three independent clones received from cleavage with either AP1 or AP2 revealed a deletion of four nucleotides (5'-GGAA-3') at the joining site in all cases. This is in accordance with the feature of LHEs to leave four nucleotide 3'overhangs upon cleavage [3]. Furthermore, this showed that cleavage exactly occurred at the LRIII1 insertion site (Figure 1B).

With respect to cleavage conditions, the enzyme worked equally well at 30°C and 37°C, and was active throughout a physiological pH range from 5.5 to 8.5, although cleavage was most efficient at pH values of 8 or 8.5 (Figure S1A). In addition, I-UmaI cleaved irrespective of whether the substrate plasmid was linearized or supercoiled (Figure S1B). Homing endonucleases efficiently work in the presence of divalent metal ions like Mg^{2+} , Mn^{2+} and Co^{2+} [5], [20] and references therein. The absence of Mg^{2+} in the reaction buffer strongly reduced activity of I-UmaI and additional application of 5 mM EGTA abolished substrate cleavage, whereas 0.5 mM Mg^{2+} was sufficient for efficient substrate cleavage. Among additional metal ions (Mn^{2+} , Zn^{2+} , Co^{2+} , Ca^{2+} , Cu^{2+} , Ni^{2+}), only Mn^{2+} provided for cleavage albeit less efficiently than Mg^{2+} (Figure S1C and data not shown).

Both LAGLIDADG Domains are Required for Enzymatic Activity of I-UmaI

The I-UmaI sequence contains two predicted LAGLIDADG domains from amino acid positions 48 to 149 and 199 to 308, respectively. While the corresponding ORF sequences of four wild-type isolates were identical and predicted a protein of 336 residues, the ORF sequence of strain MF18 predicted a protein with only the first LAGLIDADG domain due to a frameshift mutation (deletion of nucleotide positions 531–537; [18]). To assess enzymatic activity of this mutant form (here termed I-UmaI_{mut}), plasmid pAP10 was constructed providing for expression of I-UmaI_{mut} up to the naturally occurring stop codon at position 610 (U₆₁₀AG). Due to initial uncertainty about the codon usage for ATA, the single A₅₄₇TA codon was substituted to ATG (see Materials and Methods). However, since this codon lies within the

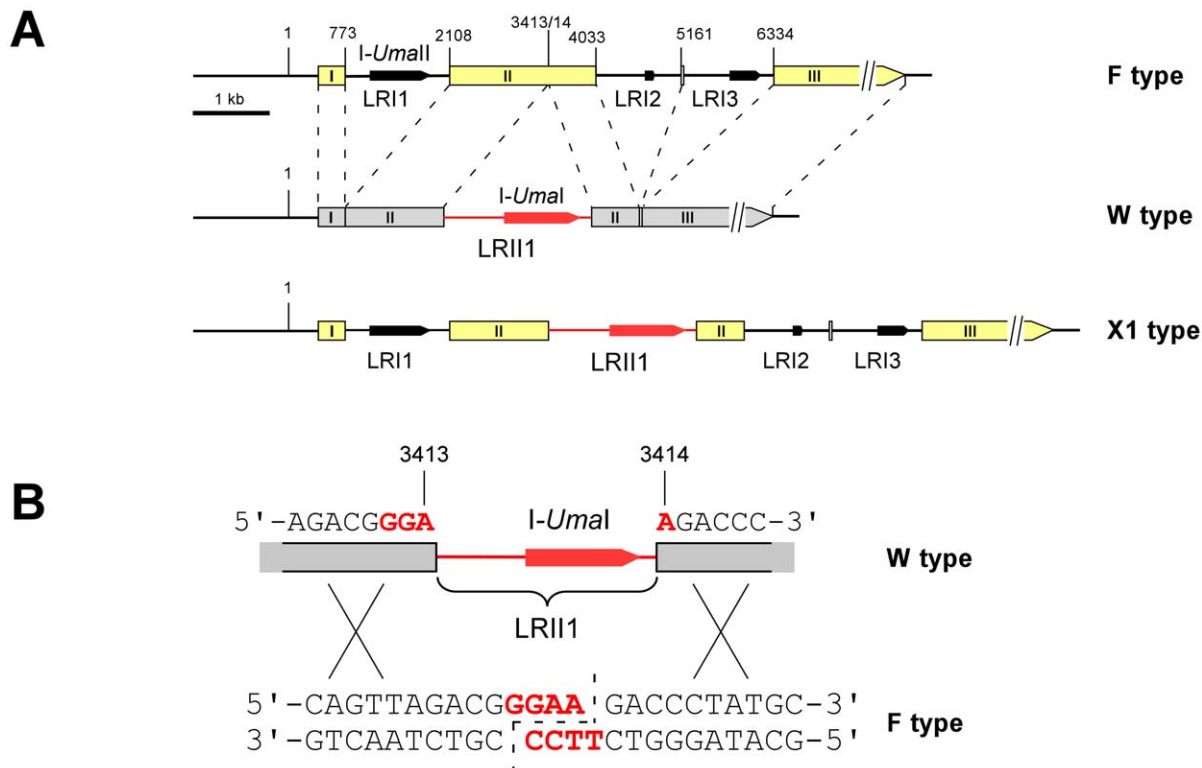


Figure 1. Determination of the I-Umal cleavage site. (A) Schematic of the polymorphic region within the LSU rDNA of *U. maydis*. Depicted are the parental W and F types and the recombinant X1 type. Boxes refer to deduced exon sequences (the three major ones are numbered from I to III). Stippled lines connect homologous sequences. All numbers refer to the complete mitochondrial genome sequence of *U. maydis* strain 521. Black arrows mark predicted HEGs within intronic regions (termed LRI1, LRI2, LRI3 and LRII1). LRII1 and I-Umal are marked in red. The schematic (drawn to scale) is adapted from Fedler *et al.* [18] with permission of the Genetics Society of America. (B) The schematic shows the target cleavage site (red, bold face type) within the recipient F type sequence (indicated positions refers to the mitochondrial genome sequence of strain 521; see part A). The staggered line (stippled) marks produced 3' overhangs. The upper part shows the donor sequence of the W type interrupted by intron LRII1. Crosses refer to potential homologous recombination events.
doi:10.1371/journal.pone.0049551.g001

frame-shifted region (starting at nucleotide position 532 in I-Umal_{mut}), it did not affect the native protein chain. Immunoblot analysis confirmed that AP10 was correctly expressed as His-tagged fusion (Figure 2B). Time-course analysis revealed that the activity of the mutant form was strongly diminished compared to the wild-type form as judged from the appearance of a faint distinct cleavage product after longer incubation (Figure 2B).

To corroborate this finding, plasmids pAP1DK and pAP2DK providing for an internal in frame deletion within the coding region of the second LAGLIDADG domain (amino acid positions 214–249) were constructed. Consistent with a requirement for both LAGLIDADG domains, extracts from corresponding *E. coli* strains lacked homing endonuclease activity (Figure 2C).

Determination of the I-Umal Target Site Specificity

Based on the DNA sequences bordering the observed cleavage site, we generated a series of potential target site fragments to delineate the substrate specificity of I-Umal (Table S3). For this purpose, the enzyme assay was modified in that reaction products were subsequently cleaved within the plasmid backbone (see schematic in Figure 3). First, we determined the minimum target site length. This showed that sequences flanking the central-four base pairs 5'-GGAA-3' could be shortened to 6 bp on either side (substrate plasmid pUC19-N) without strongly affecting cleavage efficiency, whereas further truncation to 5 bp abolished cleavage (pUC19-I; Table 1). However, 4 or 5 bp fragments on the left side

still provided for cleavage in the presence of a 7 bp fragment on the right side (pUC19-P,R; Figure 3), but not the other way around (pUC19-Q) (Table 1). Taken together, this showed that a 15 bp sequence was sufficient for cleavage and implicated that 4 bp on the left and 6 bp on the right of the central cleavage site define the minimum target site. This was explicitly verified, although the corresponding substrate (pUC19-Y) was clearly less efficiently cleaved than the one with 7 bp on the right side (pUC19-R), indicating that position +9 contributed to the target site specificity (Figure 3). Further shortening of the left side (pUC19-Z) ultimately confirmed the minimum target site length of 14 bp *in vitro* (Figure 3). To assess further target site requirements, plasmid pUC19-M, in which position -6 was substituted within a longer region, was constructed. This showed that cleavage was only weakly affected providing evidence for contacts beyond position -6 on the left side (Figure 3). In favor of contact points beyond the 14-base pairs target site motif, a substitution of position +1 abolished cleavage within pUC19-Y (pUC19-Y*), while the same substitution in pUC19-T only partially reduced cleavage (Figure 3).

Substrate requirements were further analyzed based on substitutions within the central-four base pairs and adjacent positions. As expected, substitution of the GGAA motif abolished cleavage (pUC19-E in Figure 3). In particular, substitution of the two central adenine bases exhibited a strong effect (pUC19-L), whereas the two guanine bases on the left were not critical

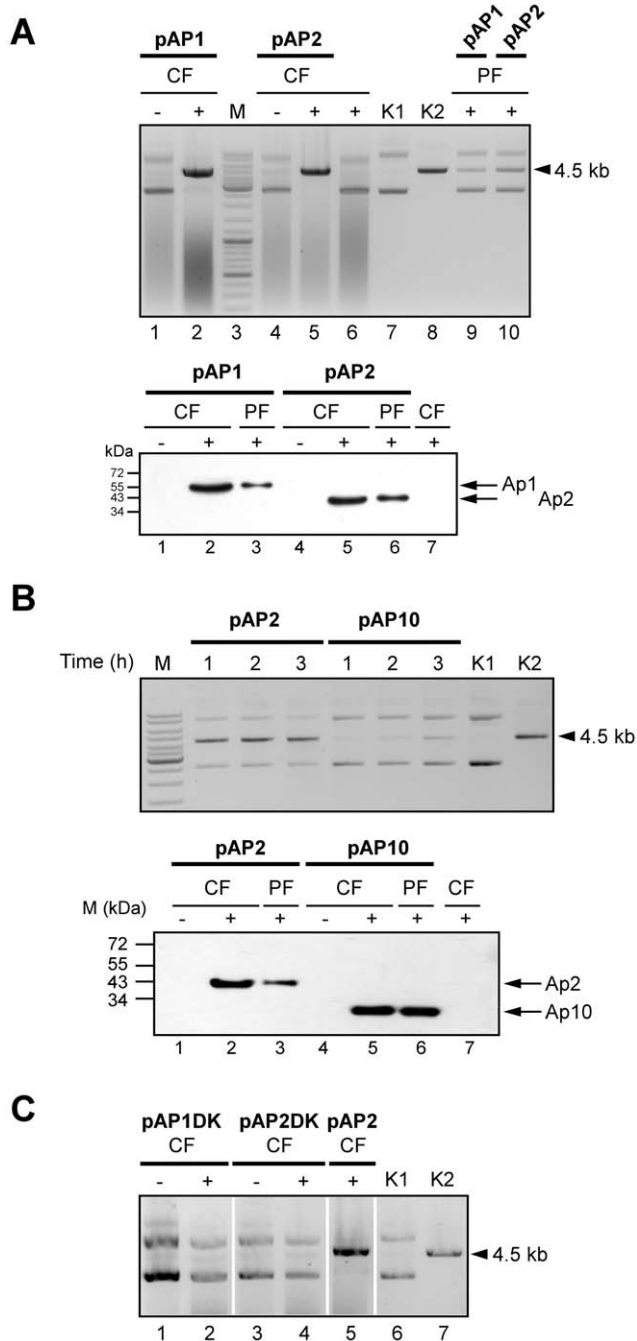


Figure 2. Enzymatic analysis of *I-Umal* in dependence on the two LAGLIDADG domains. (A) Cleavage analysis with crude fractions (CF) from either non-induced (–) or induced (+) pAP1 (1,2) and pAP2 (4,5) *E. coli* cells in the presence of substrate plasmid pSL521. CF from non-transformed *E. coli* cells incubated under inducing conditions served as a negative control (6). Cleavage analysis with purified fractions (PF) from induced pAP1 (9) and pAP2 (10) cells. DNA ladder: thick bands correspond to 500, 1000 and 3000 bp (3). The bands directly above 3 kb are 3.5, 4 and 5 kb. pSL521 uncut (K1) or linearized (K2) served as markers. The size of pSL521 is 4503 bp (arrowhead). Bottom panel: verification of *I-Umal* expression. Protein fractions from either non-induced (–) or induced (+) cells were applied to SDS-PAGE for subsequent immunoblot analysis to detect His-tagged *I-Umal* expressed in pAP1 (1–3; additional THX domain) and pAP2 (4–6) cells. CF from non-transformed *E. coli* cells incubated under inducing conditions served as a negative control (7). All lanes are from the same blot. The predicted molecular masses for AP1 and AP2 are 55.8 and 42.3 kDa,

respectively. (B) Comparison between the enzymatic activities of *I-Umal* (pAP2) and *I-Umal*_{mut} (pAP10). PF from induced pAP2 or pAP10 cells was incubated with substrate plasmid pSL521 for 1, 2 and 3 h, respectively. The thick marker (M) band corresponds to 3 kb (see part A). Controls K1, K2 as in (A). Bottom panel: verification of *I-Umal*_{mut} expression. Protein fractions from either non-induced (–) or induced (+) cells were applied to SDS-PAGE for subsequent immunoblot analysis to detect His-tagged *I-Umal*_{mut} expressed in pAP10 cells (4–6). The corresponding fractions from pAP2 (1–3) and CF from induced, non-transformed *E. coli* served for controls (7). The predicted molecular mass for AP10 is 27.0 kDa. (A, B) Loaded protein amounts were 16–18 μg for CF and 0.23–0.34 μg for PF. (C) Analysis of AP1DK and AP2DK activities. CF from non-induced (–) or induced (+) pAP1DK (1,2) and pAP2DK (3,4) cells were incubated with substrate plasmid pSL521. The reaction with CF from induced pAP2 cells served for control (5). Controls K1, K2 as in (A). All lanes are from the same blot.
doi:10.1371/journal.pone.0049551.g002

(pUC19-O in Figure 3). On the other side, individual substitutions of the two critical base pairs on the right only weakly affected cleavage (pUC19-T,U). Likewise, individual substitutions at the adjacent position –3 (pUC19-W) or +3 (pUC19-V) largely remained without effect, whereas the substitution at –3 in combination with the substitution at –2 (pUC19-F) strongly impaired cleavage (pUC19-X in Figure 3). In conclusion, this analysis provided evidence for asymmetric recognition of the two halves of the DNA target site and has shown interdependence between adjacent base pair substitutions.

Transcriptional Regulation of *I-Umal*

Under natural conditions, *I-Umal* only may find a target in matings with a recipient strain that lacks the LRI11 intron. We therefore wondered whether expression of *I-Umal* is regulated during sexual development. To test this, we applied a quantitative real-time PCR (qRT-PCR) using RNA from parental F (FB1) and W (GF5) type strains as well as from dikaryotic cells resulting from mating of these strains under standard conditions. As expected, the developmentally-regulated *a2* mating type-specific gene *lga2* was highly expressed under mating conditions, only weakly in the *a2* partner and not in the *a1* partner due to absence of the gene [21,22]. However, the W type-associated *I-Umal* gene was highly expressed irrespective of mating, showing a similar expression profile as two other mitochondrial genes analyzed (*cox1*, *nad6*). This indicated that *I-Umal* is constitutively expressed and not subject of developmental regulation (Figure 4).

Analysis of *I-Umal*

Of further interest was the investigation of a potential HE activity encoded by the *I-Umal* gene, which is positioned within the LRI1 intron (see Figure 1A). Like *I-Umal*, the putative *I-Umal*II protein contains two HEG domains. According to the determined codon usage (see Table S1), no sequence adaptation was necessary for expression in *E. coli*. Specifically, among the very rarely used codons, only a single TTG triplet was found, which lies immediately next to the stop codon and therefore was not considered critical. Again, *I-Umal*II remained soluble as C-terminal His-tagged fusion in the absence of a THX domain (protein AP8) as verified from immunoblot analysis (Figure S2A). For enzymatic analysis, we used a substrate plasmid (pSLMF34) comprising the putative target site region from W type mtDNA, which flanks the LRI1 insertion site (see Figure 1A). Explicitly, the substrate plasmid contained a 531 bp fragment with the predicted cleavage site at position 315. However, enzymatic analysis using either crude or affinity-purified fractions showed no cleavage activity (Figure S2B). To exclude the possibility that products remained associated with the enzyme, reaction products were amended with

Table 1. Survey of tested target sites.

Identity: pUC19-	Length ¹	Sequence ²	Cleavage efficiency ³
A	-12/+24	CAGTTAGAC GGAA GACCCTATGCAGCTTTACTGTA	+++
B	-24/+24	GCGGTTTACCTTCAGTTAGAC GGAA GACCCTATGCAGCTTTACTGTA	+++
C	-24/+12	GCGGTTTACCTTCAGTTAGAC GGAA GACCCTATGC	+++
D	-12/+12	CAGTTAGAC GGAA GACCCTATGC	+++
E	-24/+24	GCGGTTTACCTTCAGTTAGAC CTCT GACCCTATGCAGCTTTACTGTA	-
F	-12/+12	CAGTTAGAC CGAA GACCCTATGC	+/+++
G	-11/+11	AGTTAGAC GGAA GACCCTATG	+/+++
H	-9/+9	TTAGAC GGAA GACCCTA	+/+++
I	-7/+7	AGAC GGAA GACCC	+/-
K	-12/+12	CAGTTAGAC GA AAGACCCTATGC	+/+++
L	-12/+12	CAGTTAGAC GGCT GACCCTATGC	+/-
M	-12/+12	CAGTTA TAC GGAA GACCCTATGC	+/+++
N	-8/+8	TAGAC GGAA GACCCT	+/+++
O	-12/+12	CAGTTAGAC CTA AAGACCCTATGC	+/+++
P	-7/+9	AGAC GGAA GACCCTA	+/+++
Q	-9/+7	TTAGAC GGAA GACCC	+/-
R	-6/+9	GAC GGAA GACCCTA	+/+++
T	-12/+12	CAGTTAGAC GGCA GACCCTATGC	++
U	-12/+12	CAGTTAGAC GGAT GACCCTATGC	++
V	-12/+12	CAGTTAGAC GGAA CACCCTATGC	+++
W	-12/+12	CAGTTAGAC TGAA GACCCTATGC	+/+++
X	-12/+12	CAGTTAGAC TCGA AACCCTATGC	+
Y	-6/+8	GAC GGAA GACCCT	+
Y*	-6/+8	GAC GGCA GACCCT	-
Z	-5/+9	AC GGAA GACCCTA	-

¹The numbering refers to the central-four base pairs GG₋₁A₊₁A.

²The central-four nucleotides (or mutated derivatives) of the cleavage site are in bold face type. Nucleotide substitutions are underlined and in bold face type.

³See legend of Figure 3 for the assignment of + and - symbols.

doi:10.1371/journal.pone.0049551.t001

SDS (0.25% w/v)/EDTA (10 mM) prior to gel electrophoresis as reported previously for I-CreI [23]. Despite this, no cleavage products were detected (data not shown).

Identification of Additional I-UmaI Target Sites and of Putatively Functional I-UmaI Homologs

Based on the identification of the defined I-UmaI target site we investigated the recurrence of this site elsewhere in *U. maydis* mitochondrial as well as genomic DNA. For a corresponding BLASTN search, we used the minimum target site -6/+8 (contained in pUC19-Y; see Figure 3). Within the mitochondrial genome of *U. maydis* (F type strain 521) this site was confined to exon II of the LSU rDNA. Furthermore, this site was absent from the entire genomic sequence of *U. maydis*. Only one potential I-UmaI target site was detected in the *um11087* ORF, namely 5'-AGACGGCA₊₁AGACCCT-3' (data not shown). This sequence corresponds to the -7/+8 target sequence with one mismatch at position -1, and therefore, it remains to be shown whether this site is cleaved (see Table 1). To analyze the occurrence of the I-UmaI target site within genomes of other species we applied a BLASTN search against the (nr/nt) nucleotide collection. Interestingly, this revealed that sequence motifs matching the -7/+8 target site requirement were preferentially associated with LSU rRNA genes

of members of the Ustilaginomycetes and Agaricomycetes, but almost absent from ascomycetous species (Table S4).

Next, we analyzed the occurrence of I-UmaI homologs applying a TBLASTN search. This revealed more than 50 candidates mainly associated with mitochondrial small subunit (SSU) or LSU rRNA genes (E-values <1e-18). Consistently, these sequences were not found positive for I-UmaI target sites (data not shown). To sort out candidates representing putatively functional I-UmaI homologs, we surveyed the highest 10 scores more closely, using the criteria applied in Table 2. This revealed interrupted I-UmaI target sites at the predicted exon/intron junction in only the top four candidates (Figure S3 and Table 2). Strikingly, using Rfam the corresponding introns were all classified as group II introns, as previously determined for LRII (Table 2; [18]). In addition, all these introns were inserted into homologous positions in LSU rDNA of these species (Figure S3 and Table 2). In further support of putatively functional homologs, an amino acid sequence alignment revealed distinct regions exclusively shared with I-UmaI (red bars in Figure S4). Interestingly, apart from the homolog in *Sporisorium reilianum*, representing the closest known relative to *U. maydis*, the remaining three all belong to the same order, namely the Agaricales (see Discussion). However, three out of the four corresponding ORF sequences had premature stop codons and are presumably degenerate (Table 2). Hence, only the homolog of

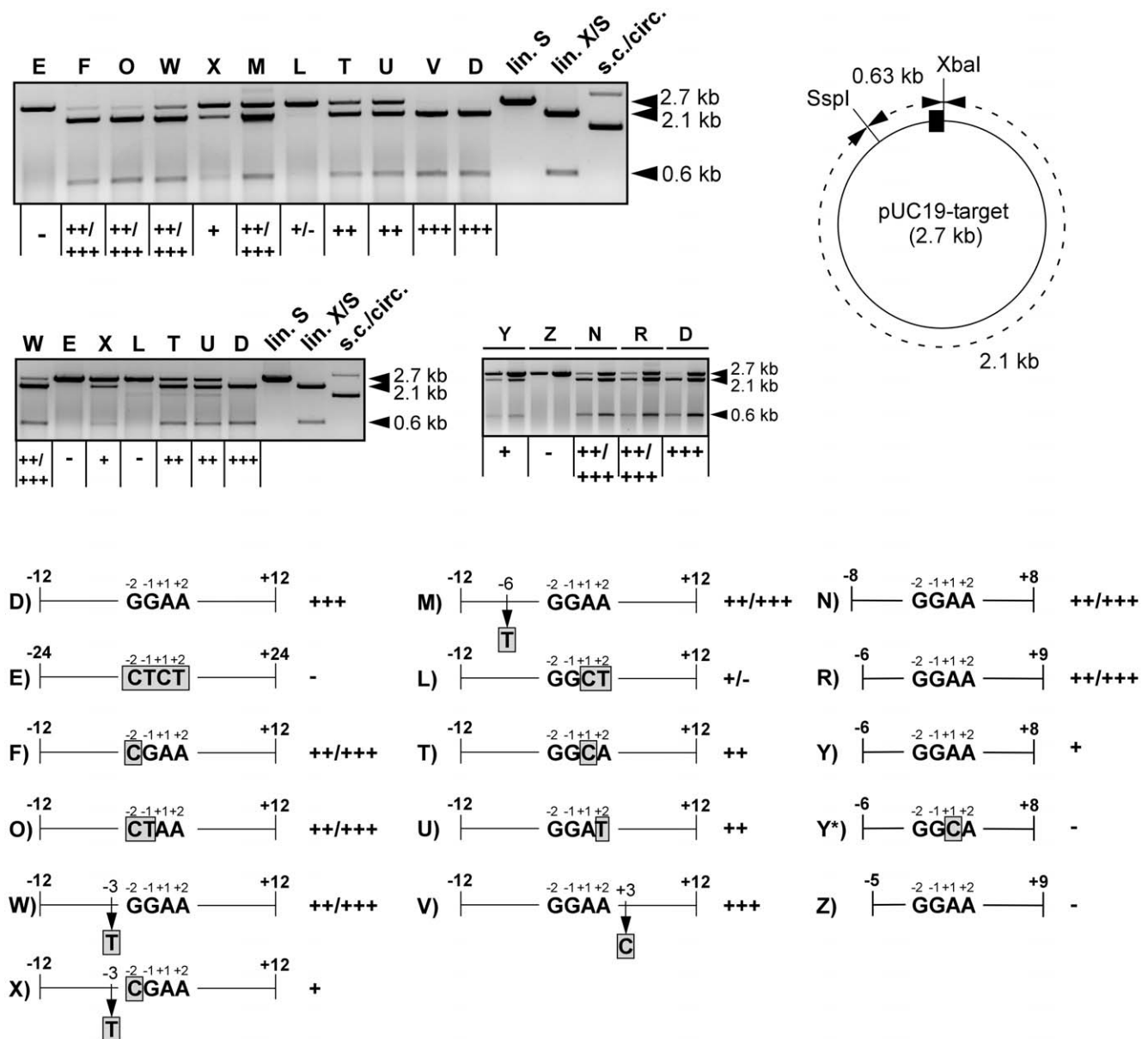


Figure 3. Analysis of the I-UmaI target site specificity. CF from induced pAP2 cells was incubated with various substrate plasmids under standard conditions for analysis of cleavage efficiencies (see Materials and Methods). Letters denote the individual substrate plasmids (e.g. D for pUC19-D). Reaction products were additionally cleaved with *SspI* to yield 2.1 and 0.63 kb fragments in case of cleavage (schematic on the right: the I-UmaI target site fragment is indicated as black box. *XbaI* cleaves at the right border and in combination with *SspI* produces fragments similar in size to those produced by I-UmaI/*SspI*). Marker lanes: lin. S, pUC19-D cleaved with *SspI*; lin. X/S, pUC19-D cleaved with *XbaI*/*SspI*; s.c./circ., uncleaved pUC19-D showing the supercoiled and circular forms. The + and - symbols refer to the cleavage efficiencies. +++, complete cleavage; +++/+, >50% cleavage; ++, ~50% cleavage; +, <50% cleavage; +/-, faint cleavage bands were detected; -, no detectable cleavage. The double lanes in the lower panel on the right refer to two different substrate concentrations used: for the right lane, the concentration was 3.3-fold higher than normally used. Note the inefficient cleavage of pUC19-Y. The schematic (bottom) shows the lengths and substitutions (boxed, shaded gray) of the tested constructs. Numbers refer to the positions within the target site as indicated (see Table 1).
doi:10.1371/journal.pone.0049551.g003

the shiitake mushroom *Lentinula edodes* (NCBI accession no. YP_006576298) appears to be preserved. In this regard, its cognate target site might differ from the one of I-UmaI based on three consecutive mismatches adjacent to the minimum I-UmaI target site (see Figure S3). In conclusion, HEGs encoding putatively functional I-UmaI homologs appear to be confined to group II introns associated with a few members of the Ustilaginales and Agaricales, consistent with the widespread occurrence of

predicted I-UmaI target sites in mitochondrial LSU rRNA genes of members of the corresponding classes.

Discussion

This study has demonstrated HE activity of I-UmaI along with the identification of the cognate target site. I-UmaI is a Mg^{2+} -dependent HE requiring both LAGLIDADG domains for activity. The I-UmaI target site precisely borders the LRIII insertion site

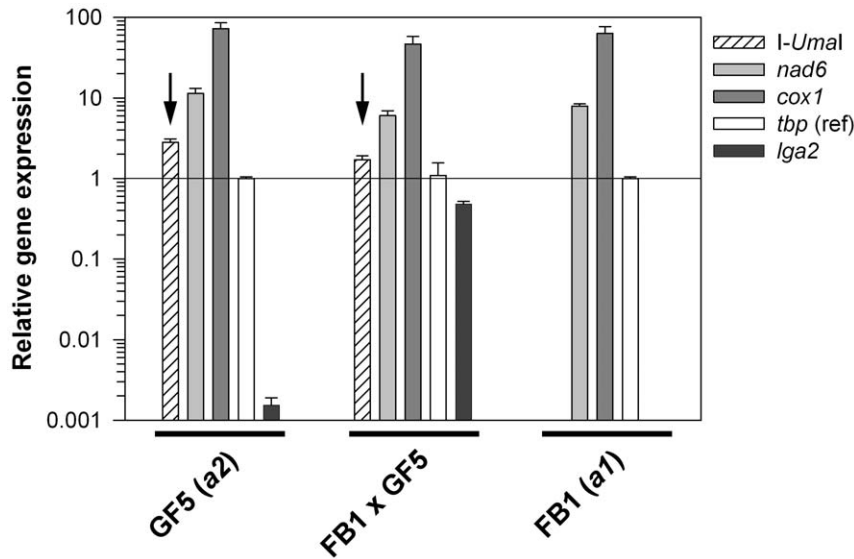


Figure 4. qRT-PCR analysis to assess transcription of I-Umal. Comparison of transcript levels of I-Umal, *nad6*, *cox1*, *tbp*, and *lga2* in dependence on sexual development. RNA was isolated from either FB1, GF5 or the mating of FB1/GF5 strains cultivated for 48 h on solid charcoal-containing complete medium. For each strain, results were normalized against the expression of the probable TATA-box binding factor gene *tbp* (*um10143*). Means and standard errors refer to four technical replicates. Note the absence of *lga2* (control for mating-dependent induction) and I-Umal transcripts in strain FB1 (*a1*, F type). The detection threshold is set to 10^{-3} (relative units). doi:10.1371/journal.pone.0049551.g004

within exon II of the LSU rRNA gene. This substantiates a role of I-Umal in the mobilization of the LR111 intron under conditions of biparental inheritance. The minimum length of the I-Umal target site is 14 bp as determined *in vitro* and thus belongs to the shortest ones reported for homing endonucleases. To this end, specific homing sites with lengths of 14 bp are known from the His-Cys box endonuclease I-PpoI from *Physarum polycephalum* and the

monomeric archaeal LHE I-DmoI [3,5,24,25]. Homing endonucleases of the LAGLIDADG family are typically associated with group I self-splicing introns [3,5,26], but also occur in group II introns [27] and references therein, [28,29]. Together with recent investigation [29], this study reinforces enzymatic activity of group II intron-associated LHEs.

Table 2. Identification of I-Umal homologs.

NCBI accession no. ¹ /species (phylum) ²	Similarity: E-value ¹	Predicted length of HE ³	LAGLIDADG domains ⁴	Intron type/ (E-value) ⁵	Host gene	I-Umal target site ⁶
ACL27287 (I-Umal)/ <i>Ustilago maydis</i> (B)	–	336 (0)	2	Group II/(1.23e-14)	LSU	yes
FQ311469/ <i>Sporisorium reilianum</i> (B)	3e-101	nd (12)	2	Group II/(1.23e-14)	LSU	yes
AB697988/ <i>Lentinula edodes</i> (B)	5e-63	365 (0)	2	Group II/(2.75e-12)	LSU	yes
HQ540074/ <i>Limacella glischra</i> (B)	7e-56	329 (1)	2	Group II/(2.80e-12)	LSU	yes
AF087656/ <i>Agrocybe aegerita</i> (B)	4e-48	319 (4)	2	Group II/(1.37e-12)	LSU	yes
FN377860/ <i>Glomus coronatum</i> (G)	3e-40	309 (1)	2	no hit	LSU	no
AF404306/ <i>Rhizophyidium sp. 136</i> (C)	3e-29	352 (3)	2	no hit	<i>cox1</i>	no
HQ259115/ <i>Moniliophthora roreri</i> (B)	4e-27	327 (0)	2	no hit	<i>nad5</i>	no
DQ364632/ <i>Gibberella zeae</i> (A)	1e-25	372 (0)	2	no hit	<i>cox1</i>	no
JN007486/ <i>Chaetomium thermophilum</i> (A)	2e-25	348 (0)	2	no hit	<i>cox1</i>	no
JQ015302/ <i>Madurella mycetomatis</i> (A)	8e-25	366 (0)	2	no hit	<i>cox2</i>	no

¹NCBI TBLASTN 2.2.26+ search (July 2012) against the nr data base with the I-Umal protein sequence as query. Shown are the top 10 hits.

²B: Basidiomycota; G: Glomeromycota; C: Chytridiomycota; A: Ascomycota.

³Length in amino acids. See Figure S4 for the ORF regions derived from the indicated accession numbers. In brackets: the number of premature stop codons (UAA, UAG) within the predicted ORFs. UGA has been converted to Trp according to the mitochondrial codon usage; for FN377860 and AF404306, each one nucleotide (positions 465 and 301, respectively) has been removed to provide for a continuous ORF; nd: not determined for FQ311469 due to 7 frameshift mutations within the predicted ORF region (data not shown).

⁴According to Pfam; see alignment in Figure S4.

⁵According to Rfam; see reference [18] for I-Umal.

⁶See Figure S3; for HQ540074, the minimum target sequence on the left of the intron insertion site is fully conserved, with positions 348–401 in HQ540074 corresponding to 3360–3413 in the *U. maydis* 521 mtDNA sequence. The right border could not be aligned due to the limited sequence available in HQ540074.

doi:10.1371/journal.pone.0049551.t002

Cleavage Specificity of I-UmaI

Homing endonucleases differ from classical restriction enzymes in that longer target sites are recognized albeit with an increased tolerance, which correlates inversely to the number of specific protein-DNA contacts [3,6,30,31]. Our analysis has shown that I-UmaI provides for cleavage of a target site with a minimum length of 14 bp, although cleavage of longer sites is clearly favored, implying that additional DNA contacts exist in flanking regions (see Table 1). This was explicitly corroborated by mutational analysis demonstrating that critical substitutions within the minimum target site were tolerated within a longer target site region. This analysis has further provided evidence for asymmetric recognition of the target site. This is exemplified by different length requirements of the two halves flanking the site of strand cleavage as well as by the finding of unequal contributions of the central-four base pairs (see Figure 3). In agreement, I-UmaI has two distinct LAGLIDADG domains both being required for cleavage. Unlike homodimers, monomers are not constrained to highly symmetrical DNA targets and their homing sites tend to be less palindromic [3]. In addition, monomeric LAGLIDADG enzymes are known to interact differently with the two DNA strands in their target sequences [32–34]. In accordance with cleavage by other homing endonucleases [3], I-UmaI generates a four-nucleotide 3'-overhang (5'-GGAA-3'). By contrast, the LR111 intron lies in the sequence context 5'-GGA-intron-A-3', which likely reflects the initial invasion in recipient DNA. The same arrangement was seen in X1 type mtDNA resulting from recombination between parental W and F types (see Figure 1A; [18]). This makes it unlikely that the LR111 intron is directly copied into the cleavage site, but instead might be transferred along with co-conversion of homologous flanking exon sequences, as explicitly reported for intron homing [35,36].

Evidence for Intron Homing in *U. maydis*

Spreading of HEGs can occur with frequencies up to 100% [5,13–15]. Previously, we detected that under conditions of biparental inheritance with combinations of F/W (W/F) mitotypes, the parental F type was almost completely lost in favor of the recombinant X1 type. Thus far, under conditions of biparental inheritance, only combinations with the W type strain MF18 provided for marked inheritance of the F type, while little X1 type was produced [18]. The I-UmaI gene of this strain carries a frameshift mutation leading to loss of the C-terminal LAGLIDADG domain. In this study, we have shown that both domains are required for cleavage at the expected intron insertion site. This clearly correlates homing efficiency of LR111 with the functionality of the encoded HE.

Previously, we detected additional recombinant forms of the LSU rRNA gene, which suggested homing of F type-associated introns into W type mtDNA. For example, the recombinant X2 type contains the F type-derived intron LR11 within W type mtDNA, while the introns LR12 and LR13 are lacking (see Figure 1A). Furthermore, the X2 type was efficiently produced in combinations with strain MF18 under conditions of biparental inheritance [18], suggesting conversion of the LR11 intron. We therefore attempted to determine the potential activity of the LR11-encoded HE I-UmaII using the expected target site region at the splice junction between exons I and II (see Figure 1A). The corresponding sequence was shown fully conserved in four independent W type strains analyzed [18]. However, despite successful expression in *E. coli*, no cleavage activity was detected under the experimental conditions applied. It is conceivable that I-UmaII exhibited a strongly reduced activity *in vitro*, which escaped detection. Additionally, the possibility remains that the target site

has diverged. In case of the LHE I-AniI it has been shown that its optimal target site deviates in two positions from its physiological target site in the host gene of *Aspergillus nidulans* [31].

Significance of I-UmaI and its Propagation

Although dependence of mitochondrial intron homing on endonuclease function has been intensively studied in yeast species [3,5], [37] and reference therein, this study has underscored the potential of intron homing in a fungal species that normally undergoes uniparental mtDNA inheritance. Consequently, if this process would be unconstrained it would lead to inactivation of all target sites, a situation that is not encountered in *U. maydis* [10,18]. On the other side, if mtDNA inheritance would be strictly uniparental like in animals [16], HEGs might be subject of degeneration unless their products are additionally charged with beneficial functions, such as promoting splicing of host introns through a maturase activity [6,10,38]. The present finding of a functional copy of I-UmaI therefore confirms that mtDNA exchange is well-balanced between these two extremes. To this end, a dual function, which would justify the high expression level detected for I-UmaI irrespective of sexual development (see Figure 4), cannot be excluded for I-UmaI.

Based on the current database, putatively functional or former functional homologs of I-UmaI are confined to members of the Ustilaginales and Agaricales representing phylogenetic distant orders within the Basidiomycota [39]. This suggests that a common ancestor intron has been acquired independently, in line with reported evidence for horizontal spread of the ω -associated HEG in Saccharomycete yeasts and the current view of horizontal transfer of mobile introns between distantly related species [5,40]. In this regard, it is not excluded that I-UmaI-associated introns are more spread than anticipated from available genomic data, as suggested from the wide occurrence of predicted I-UmaI target sites in LSU rRNA genes of members of the Ustilaginomycetes and Agaricomycetes (see Table S4) as well as from the identification of *U. maydis* strains, which do not carry LR111 (see Figure 1A; [18]). Therefore, it may be worthwhile to await additional genomic data to trace the history of invasion and transmission of this mobile element.

In this context, it will be exciting to examine whether maintenance of I-UmaI target sites was due to uniparental mtDNA inheritance or simply reflects the absence of I-UmaI homologs in populations of corresponding species. The finding that I-UmaI and homologous genes are subject of degeneration (see Table 2) implies that underlying intron homing is highly dynamic giving rise to fixation of cognate target sites [10,40]. To this end, the only intact putatively functional I-UmaI homolog is from *L. edodes*. Its deduced target site is overlapping with that of I-UmaI such that the minimum target site is preserved, while adjacent positions on the left side differ (see Figure S3). This supports our assumption that base pair positions extending the I-UmaI minimum target site contribute to the cleavage specificity. To test this, further work elucidating the target site specificity of the putative *L. edodes* enzyme will be necessary. In this context, it might be particularly interesting to identify contacting amino acids within the presumed LAGLIDADG domains of the *U. maydis* and *L. edodes* proteins (see Figure S4). This may yield insight into mechanisms underlying the acquisition of new target sites.

LHEs are generally regarded as tools for genomic engineering based on their exceptional cleavage specificity. For example, they serve to introduce strand breaks to promote homologous recombination for the purpose of targeted gene disruption. In this matter, monomeric enzymes like I-UmaI might be more favorable for heterologous expression than their homodimeric

cousins considering the smaller sizes and the fact that they do not rely upon dimerization [6].

Conclusion

This study has provided for the functional characterization of a novel group II-associated homing endonuclease. Together with the characterization of a naturally occurring mutant variant it strongly supports the concept of rigorous intron homing under conditions of biparental inheritance. This in turn emphasizes the need for underlying control by uniparental mtDNA inheritance. Bioinformatic analyses suggest that I-UmaI represents a descendant of an archetype HEG that has invaded two phylogenetically distant orders consistent with the wide occurrence of predicted cognate target sites in members of the respective classes. Based on its relatively short target site length, I-UmaI provides a potentially valuable framework for developing variant enzymes with altered target specificities.

Materials and Methods

Strains, Growth Conditions and Chemicals

Growth conditions of the *U. maydis* wild-type strains 521 (*alb1*), FB1 (*alb1*), GF5 (*a2b13*), MF18 (*alb17*) and mating performance were as described [18]. *E. coli* K12 strain TOP10 (Life Technologies, Karlsruhe, Germany) was used as host for plasmid amplifications and enzyme expressions. *E. coli* strains were cultivated in dYT/Ap (ampicillin; 100 µg/ml). If not further specified, all chemicals were of analytical grade and obtained from Sigma (Taufkirchen, Germany) or Roth (Karlsruhe, Germany).

DNA and RNA Procedures

Isolation of *U. maydis* DNA and nucleic-acid procedures were done as described [41]. RNA was isolated from *U. maydis* cultivated for 2 days on solid CM charcoal plates using the TRIzol reagent (Life Technologies). Plasmids were isolated from *E. coli* using the lysis by boiling method [42]. Restriction enzymes were from NEB (New England Biolabs, Frankfurt a.M., Germany), oligonucleotides from MWG (Ebersberg, Germany). The correctness of all plasmid constructs was verified by sequencing (Sequencing Service, Faculty of Biology, LMU Munich: <http://www.gi.bio.lmu.de/sequencing/>).

Expression Constructs

For pAP1, the I-UmaI ORF (NCBI accession no. EU921805; positions 2892–3899) was amplified (Phusion High-Fidelity DNA Polymerase, NEB) from genomic DNA of strain GF5 for cloning into pBAD102/D-TOPO (Life Technologies), using the primer combination 5'-CACCATGGATACAACCTTATGATTCTAC-3' (LRIf2; *NcoI* site underlined)/5'-TTTACGATAACGATT-CATCGTCG-3'. pAP2 was derived from pAP1 by removing the internal 377 bp *NcoI* fragment, which encodes the N-terminal thioredoxin (THX) domain. pAP1DK and pAP2DK were generated from pAP1 and pAP2, respectively, by cleavage with *XbaI/NheI* and subsequent religation of the larger 4986 bp fragment. pAP10 carries the mutant I-UmaI ORF of strain MF18 (NCBI accession no. EU921803; region 2884–3866) up to the native stop codon (U₆₁₀AG). To generate pAP10, a PCR-based mutagenesis approach was applied using genomic DNA of strain MF18 to yield an ORF fragment carrying the TGA₅₃₄ to TGG₅₃₄ and ATA₅₄₉ to ATG₅₄₉ substitutions within the frame-shifted 3' region. Overlapping primers 5'-GTAGTTTAATAA-CATCAGGAAGATCATCCAGGTAATTCCACC-3' (reverse) and 5'-GGAATTACCTGGATGATCTTCGTGATGTTAT-TAAACTACACCG-3' (forward), comprising the two substitu-

tions (underlined), were used in combination with outer primers LRIf2 (see above) and 5'-GCCGAAATTGAAATGATCCTTC-3' (reverse), respectively. Gel-purified fragments were annealed prior to a second PCR using the outer primers. The resulting PCR product was ligated into pBAD102/D-TOPO and the internal 377 bp *NcoI* fragment was removed subsequently. Due to an internal deletion within the region of the inner primers, the same experimental approach was applied, using the generated plasmid as template, with the outer reverse primer 5'-GATGACCGG-TACGCGTAGAATCG-3', which is antiparallel to the 780–802 region in pBAD102/D-TOPO. The PCR product was cleaved with *BglII/HindIII* to correct the corresponding region in pAP10.

Substrate Plasmids

A 1075 bp DNA fragment comprising the predicted I-UmaI target site was amplified from genomic DNA of *U. maydis* strain 521 (F type) using the primer combination 5'-GGAATTCCA-TATGCTCCTCGCCGAATACGAGAGG-3')/5'-GGAATTC-CATATGTCCTCCAGTCAAACCTGACCACC-3' (*NdeI* sites underlined) and inserted into the *NdeI* site of pSL1180 (GE Healthcare, Darmstadt, Germany) to yield pSL521. For the collection of short I-UmaI target site plasmids, termed pUC19-A to Z, matching oligonucleotides (3.4 µM each; see Table S3) were combined in restriction enzyme buffer containing bovine serum albumine (BSA; 1 mg/ml) and protruding ends were filled in the presence of Klenow enzyme (NEB). After heat inactivation (10 min, 75°C), double-stranded DNA was restricted with *BamHI/XbaI*, followed by heat inactivation (20 min, 85°C) and purification on a matrix (JetSorb; GENOMED, Löhne, Germany). Cleaved fragments were ligated in the presence of T4 ligase (NEB) into the compatible sites of pUC19 (GE Healthcare).

Enzyme Preparations

Overnight cultures (200 rpm, 37°C) of the various *E. coli* strains were transferred to 50 ml fresh dYT/Ap medium at starting optical densities at 600 nm (OD₆₀₀) of 0.1 and incubated to densities of 0.5. Cultures were subsequently supplemented with or without 0.01% (w/v) arabinose (Ara) and incubated for additional 90 min. For control, the non-transformed *E. coli* TOP10 strain treated with 0.01% (w/v) Ara in dYT was used. Cells were collected by centrifugation (1700 g; 4°C), washed in 10 ml ice-cold extraction buffer (EB: 50 mM Tris-HCl, pH 8.0; 50 mM NaCl; 4% glycerol; v/v), resuspended in 3 ml ice-cold EB+ (EB supplemented with 0.1 mM EDTA, 2 mM DTT, 1x Complete Protease Inhibitor; Roche, Mannheim, Germany) and stored frozen at -80°C. For enzyme extraction, pellets were subjected to four cycles of freeze-thawing, followed by sonification, using a Bandelin Sonopulse HD3080 sonifier (running at 50%, 5 sec/10 sec pulse/pause intervals for 10 min on ice). Homogenates were centrifuged (15 min, 8500 g; 4°C) and supernatants of the crude fraction (CF) were stored frozen (-80°C) in aliquots. For affinity purification, 550 µl of CF were applied to Ni-NTA spin columns (Qjagen, Hildesheim, Germany) as specified by the manufacturer. Columns were equilibrated in EB+ supplemented with 8 mM imidazole prior to application of the CF samples. Columns were washed with three volumes (500 µl) of EB/20 mM imidazole, followed by elution with two volumes (230 µl) of EB/250 mM imidazole supplemented with 1x Complete Protease Inhibitor to yield the purified fraction (PF). The combined eluates were used for enzyme analysis. Protein concentrations were determined using the Bio-Rad Protein Assay (Bio-Rad, Munich, Germany) with BSA as standard. Protein concentrations of the enzyme fractions were 1749±58 µg/ml for CF and 27.2±5.3 µg/ml for PF.

Enzyme Assay

The applied enzymatic conditions basically followed the protocol of Monteilhet *et al.* [37]. If not further specified, enzyme preparations (10 to 20 μ l) were incubated in the presence of substrate plasmid (50–100 ng) in 60 to 100 μ l reaction buffer (EB supplemented with 0.1 mg/ml BSA, 10 mM MgCl₂) for 30 min at 30°C. Reaction mixtures were either preceded by phenol/chloroform extraction, followed by ethanol precipitation, or by exposure to 85°C (5 min), followed by purification of plasmid-containing supernatants on a matrix (JetSorb; GENOMED). Both treatments gave the same product yields (data not shown). In either case, DNA was dissolved in 0.5x TE (Tris-EDTA)/RNase (10 μ g/ml) and applied to agarose (1%)/ethidium bromide gel electrophoresis. Alternatively, reaction products were digested with *SspI* prior to gel documentation. Substrate plasmid pUC19-D either non-digested or cleaved with either *SspI* or *SspI/XbaI*, and pSL521 either non-digested or cleaved with *BamHI* were used for controls. Size markers were phage λ DNA digested with *PstI* (0.25 μ g) or GeneRuler™ DNA Ladder Mix (1 μ g; Fermentas, St. Leon-Rot, Germany).

Determination of the Target Cleavage Site

Linearized target plasmid pSL521 (ca. 1 μ g) received from digestion with AP1 and AP2, respectively, was extracted from agarose gels and treated with T4 DNA polymerase (0.5 units; NEB) for 10 min at 4°C in the absence of dNTPs, followed by 30 min incubation at 12°C in the presence of 0.2 mM dNTPs. Blunt-ended plasmids were religated either in the absence or presence of the 692 bp *DraI* fragment isolated from pUC19. The cleavage sites of three independent clones (two linearized by AP1 and AP2, respectively, and ligated to the *DraI*-fragment, and one linearized by AP2 and directly religated) were analyzed by sequencing.

Quantitative Real-time PCR (qRT-PCR) Analysis

Total RNA (2 μ g) was treated with DNase (Ambion, Life Technologies). qRT-PCR analysis was applied using the SuperScript III First-Strand Synthesis SuperMix (Life Technologies) in the presence of hexamer primers and the Platinum SYBR qPCR Supermix-UDG kit (Life Technologies) in the presence of 10 nM fluorescein (Bio-Rad, Munich, Germany). Gene specific primers for *I-UmaI* were: 5'-GGAATTACCTGAATGATCTTCGTGA-3'/5'-CCGAAATTGAAATGATCCTTCTCCA-3'. Specific primers against *nad6*, *cox1*, *lga2* and *thp* have been described [43].

Protein Gels and Immunodetection

Protein amounts as specified in the text were applied to SDS-PAGE (10%) for immunodetection as described [41]. A monoclonal anti-His(C-term) antibody (Life Technologies; 1:5000) and a goat anti-mouse immunoglobulin G-horseradish peroxidase HRP conjugate (1:10000; Promega, Mannheim, Germany) were used as primary and secondary antibodies, respectively. Protein markers peqGOLD IV or V (Peqlab, Erlangen, Germany) were used as size markers.

Databases

Protein and nucleotide sequences were compared using the NCBI BLAST database (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Mitochondrial sequences of *U. maydis* were retrieved from NCBI under accession no. DQ157700 if not further specified in the text. Protein subdomains were predicted by Pfam (<http://pfam.sanger.ac.uk/search>). Analysis of intronic regions was done with the software provided by Rfam (<http://rfam.sanger.ac.uk/search>).

ClustalX was used for comparative sequence alignments (<http://www.clustal.org/>). The BROAD database (http://www.broadinstitute.org/annotation/genome/ustilago_maydis/Home.html) was used to search *U. maydis* genomic sequences for *I-UmaI* sites. *U. maydis* gene sequences (um numbers) were retrieved from the MIPS (Munich Information Center for Protein Sequences) *Ustilago maydis* Genome Database (MUMDB; <http://mips.helmholtz-muenchen.de/genre/proj/ustilago>).

Supporting Information

Figure S1 Cleavage conditions for I-UmaI. (A) Test of different temperature conditions. CF from induced pAP2 cells was incubated with pUC19-B (see Table 1) under standard conditions at either 30°C or 37°C using different amounts of CF (18, 9, 3.6 μ g from left to right). Marker lanes: s.c./circ., uncleaved pUC19-B showing the supercoiled and circular forms; lin. X/S, pUC19-B cleaved with *XbaI/SspI*; lin. S, pUC19-B cleaved with *SspI*. Panel on the right: test of different pH conditions. Replicate samples for pH 8.0. (B) Cleavage of different substrate plasmids (see Table 1) either supercoiled (1–4) or precleaved with *SspI* (5–8). Marker lanes: lin. S, pUC19-D cleaved with *SspI*; lin. X/S, pUC19-D cleaved with *XbaI/SspI*. Note that the cleavage efficiencies of the different substrate plasmids were maintained under the two conditions. (C) Metal ion requirement for *I-UmaI*. CF from induced pAP2 cells was incubated with substrate plasmid pUC19-D in the presence of different cations under standard conditions. MgCl₂, MnCl₂, ZnCl₂, CoCl₂, CaCl₂, CuCl₂: each 10 mM (1.3 mM for CoCl₂). Panel on the right: test of different MgCl₂ concentrations: lanes 1–6:0, 0.5, 1, 2, 5, 10 mM. (A–C) Reaction products (1–4 in part B) were cleaved with *SspI* resulting in 2.1 and 0.63 kb fragments in case of cleavage by *I-UmaI*. All lanes per image are from the same gel. The efficiency of the cleavage is seen from the disappearance of the 2.7 kb band in favor of the 2.1 and 0.63 kb bands (see schematic in Figure 3). (TIF)

Figure S2 Enzymatic analysis of I-UmaII. (A) Verification of *I-UmaII* expression. CF (1,2) or PF (3) from either non-induced (–) or induced (+) pAP8 cells were applied to SDS-PAGE for subsequent immunoblot analysis to detect His-tagged *I-UmaII*. CF from non-transformed *E. coli* cells incubated under inducing conditions served as a negative control (4). Approximately 18 μ g protein were loaded for CF and 0.25 μ g for PF. The upper band corresponds to the predicted molecular mass of AP8 (34.7 kDa). (B) Enzyme assay with CF (1, 2) and PF (4) from either non-induced (–) or induced (+) pAP8 cells. CF from non-transformed *E. coli* cells incubated under inducing conditions served as a negative control (3). The used substrate plasmid pSLMF34 (3959 bp) uncut (K1) or linearized (K2) served as marker. See Figure 2A for the used DNA ladder. All lanes are from the same gel. (TIF)

Figure S3 Insertion sites of introns containing predicted I-UmaI homologs. Shown are the insertion sites of intronic regions containing the predicted *I-UmaI* homologs in mtDNA of *S. reilianum*, *L. edodes*, and *A. aegerita* (NCBI accession numbers are written nearby). The *I-UmaI* target site (–6/+9; see Table 1) is typed in bold face and shaded gray. The central-four bases are typed in red. Numbers refer to positions in the corresponding NCBI sequences. Intronic regions are depicted by black horizontal lines, with the thick blue arrows marking the positions of the corresponding HEGs (drawn to scale; start and end positions are indicated). Dots mark identical bases in the sequence alignments,

with the upper sequence corresponding to the *U. maydis* F type (NCBI accession no. DQ157700). In case of *A. aegerita*, the intron start and end positions have been shifted upstream by 2 bp relative to the sequence in AF087656 to match the intron/exon borders experimentally confirmed for *U. maydis* [18].

(TIF)

Figure S4 Sequence alignment of predicted I-UmaI homologs. The amino acid alignment includes I-UmaI and predicted homologs (see Table 2). Identities of $\geq 40\%$ are shaded gray. Gaps have been inserted to maximize the alignment. Amino acids exclusively conserved between I-UmaI and its three closest homologs shown in this alignment are typed in red. Corresponding regions (interruptions by max. one amino acid, with the conserved residue occurring in max. one additional row) are further denoted by red bars above the I-UmaI sequence. The sequence of *S. reilianum* has been omitted due to the accumulation of multiple frameshift mutations (see Table 2). All letters and regions marked in red also exist in the *S. reilianum* sequence except for R₁₀₅ (referred to I-UmaI), E₂₃₁, K₂₅₅, which map to regions of predicted frameshifts. LAGLIDADG motifs are marked in blue. The alignment does not match the predicted first LAGLIDADG motifs deduced from the HQ540074 and AF087656 sequences, which by Pfam lie within amino acid positions 53–135 and 8–101, respectively. Premature stops in the predicted amino acid sequences are indicated by asterisks (see Table 2). The corresponding ORF regions from which the protein sequences were deduced are: FQ311469:78810–79834, AB697988:55865–56962, HQ540074:984–1973, AF087656:9784–10743, FN377860:2520–3450, AF404306:33766–34825, HQ259115:28041–29024, DQ364632:67072–68190, JN007486:77114–78160,

References

- Lambowitz AM, Belfort M (1993) Introns as mobile genetic elements. *Ann Rev Biochem* 62: 587–622.
- Bonen L, Vogel J (2001) The ins and outs of group II introns. *Trends Genet* 17: 322–331.
- Chevalier BS, Stoddard BL (2001) Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res* 29: 3757–3774.
- Bonocora R, Shub DA (2001) A novel group I intron-encoded endonuclease specific for the anticodon region of tRNA-Fmet genes. *Mol Microbiol* 39: 1299–1306.
- Stoddard BL (2005) Homing endonuclease structure and function. *Q Rev Biophys* 38: 49–95.
- Stoddard BL (2011) Homing endonucleases: from microbial genetic invaders to reagents for targeted DNA modification. *Structure* 19: 7–15.
- Zhao L, Bonocora RP, Shub DA, Stoddard BL (2007) The restriction fold turns to the dark side: a bacterial homing endonuclease with a PD-(D/E)-XK motif. *EMBO J* 26: 2432–2442.
- Jin Y, Binkowski G, Simon LD, Norris D (1997) Ho endonuclease cleaves MAT DNA *in vitro* by an inefficient stoichiometric reaction mechanism. *J Biol Chem* 272: 7352–7359.
- Bakhrat A, Jurica MS, Stoddard BL, Raveh D (2004) Homology modeling and mutational analysis of Ho endonuclease of yeast. *Genetics* 166: 721–728.
- Burt A, Koufopanou V (2004) Homing endonuclease genes: the rise and fall and rise again of a selfish element. *Curr Opin Genet Dev* 14: 609–615.
- Belfort M, Roberts RJ (1997) Homing endonucleases - keeping the house in order. *Nucleic Acids Res* 25: 3379–3388.
- Haugen P, Simon DM, Bhattacharya D (2005) The natural history of group I introns. *Trends Genet* 21: 111–119.
- Netter P, Petrochilo E, Slonimski PP, Bolotin-Fukuhara M, Coen D, et al. (1974) Mitochondrial genetics. VII. Allelism and mapping studies of ribosomal mutants resistant to chloramphenicol, erythromycin and spiramycin in *S. cerevisiae*. *Genetics* 78: 1063–1100.
- Jacquier A, Dujon B (1985) An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. *Cell* 41: 383–394.
- Colleaux L, d'Auriol L, Betermier M, Cottarel G, Jacquier A, et al. (1986) Universal code equivalent of a yeast mitochondrial intron reading frame is expressed into *E. coli* as a specific double strand endonuclease. *Cell* 44: 521–533.
- Barr CM, Neiman M, Taylor DR (2005) Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. *New Phytol* 168: 39–50.
- Basse CW (2010) Mitochondrial inheritance in fungi. *Curr Opin Microbiol* 13: 712–719.
- Fedler M, Luh KS, Stelter K, Nieto-Jacobo F, Basse CW (2009) The *a2* mating-type locus genes *lga2* and *rga2* direct uniparental mitochondrial DNA (mtDNA) inheritance and constrain mtDNA recombination during sexual development of *Ustilago maydis*. *Genetics* 181: 847–860.
- Roberts RJ, Belfort M, Bestor T, Bhagwat AS, Bickle TA, et al. (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res* 31: 1805–1812.
- Chevalier B, Sussman D, Otis C, Noël AJ, Turmel M, et al. (2004) Metal-dependent DNA cleavage mechanism of the I-CreI LAGLIDADG homing endonuclease. *Biochemistry* 43: 4015–4026.
- Urban M, Kahmann R, Bölker R (1996) Identification of the pheromone response element in *Ustilago maydis*. *Mol Gen Genet* 251: 31–37.
- Bortfeld M, Auffarth K, Kahmann R, Basse CW (2004) The *Ustilago maydis a2* mating-type locus genes *lga2* and *rga2* compromise pathogenicity in the absence of the mitochondrial p32 family protein Mrb1. *Plant Cell* 16: 2233–2248.
- Seligan LM, Stephens KM, Savage JH, Monnat RJ Jr (1997) Genetic analysis of the *Chlamydomonas reinhardtii* I-CreI mobile intron homing system in *Escherichia coli*. *Genetics* 147: 1653–1664.
- Lowery R, Hung L, Knoche K, Bandziulis R (1992) Properties of I-PpoI: a rare-cutting intron encoded endonuclease. *Promega Notes* 38: 8–12.
- Silva GH, Dalggaard JZ, Belfort M, Van Roey P (1999) Crystal structure of the thermostable archaeal intron-encoded endonuclease I-DmaI. *J Mol Biol* 286: 1123–1136.
- Haugen P, Bhattacharya D (2004) The spread of LAGLIDADG homing endonuclease genes in rDNA. *Nucleic Acids Res* 32: 2049–2057.
- Toor N, Zimmerly S (2002) Identification of a family of group II introns encoding LAGLIDADG ORFs typical of group I introns. *RNA* 8: 1373–1377.
- Franz Lang B, Laforest M-J, Burger G (2007) Mitochondrial introns: a critical view. *Trends in Genetics* 23: 119–125.
- Mullineux ST, Costa M, Bassi GS, Michel F, Hausner G (2010) A group II intron encodes a functional LAGLIDADG homing endonuclease and self-splices under moderate temperature and ionic conditions. *RNA* 16: 1818–1831.
- Gimble FS, Moure CM, Posey KL (2003) Assessing the plasticity of DNA target site recognition of the Pl-SceI homing endonuclease using a bacterial two-hybrid selection system. *J Mol Biol* 334: 993–1008.
- Scalley-Kim M, McConnell-Smith A, Stoddard BL (2007) Coevolution of homing endonuclease specificity and its host target sequence. *J Mol Biol* 372: 1305–1319.

JQ015302:16802–17902. See also Table 2 (footnote 3) for the conversion of ORF regions.

(TIF)

Table S1 Codon usage of the mitochondrial genome of *U. maydis*.

(DOC)

Table S2 Differences in the codon usage between *U. maydis* and *S. cerevisiae*.

(DOC)

Table S3 Primers for target site analysis.

(DOC)

Table S4 Occurrence of potential I-UmaI target sites.

(DOC)

M&M S1 Enzyme assay under different pH conditions.

(DOC)

M&M S2 Construction of pAP8 and substrate plasmid pSLMF34.

(DOC)

Acknowledgments

We would like to thank our student R. Frank (KIT, Karlsruhe) for supporting cleavage studies.

Author Contributions

Conceived and designed the experiments: CB AP. Performed the experiments: AP CB BM. Analyzed the data: CB AP JK. Contributed reagents/materials/analysis tools: JK CB. Wrote the paper: CB. Bioinformatic analyses: CB.

32. Perrin A, Buckle M, Dujon B (1993) Asymmetrical recognition and activity of the I-SceI endonuclease on its site and on intron-exon junctions. *EMBO J* 12: 2939–2947.
33. Christ F, Schoettler S, Wende W, Steuer S, Pingoud A, et al. (1999) The monomeric homing endonuclease PI-SceI has two catalytic centres for cleavage of the two strands of its DNA substrate. *EMBO J* 18: 6908–6916.
34. Moure CM, Gimble FS, Quijcho FA (2008) Crystal structures of I-SceI complexed to nicked DNA substrates: snapshots of intermediates along the DNA cleavage reaction pathway. *Nucleic Acids Res* 36: 3287–3296.
35. Lazowska A, Meunier B, Macadre C (1994) Homing of a group II intron in yeast mitochondrial DNA is accompanied by unidirectional co-conversion of upstream-located markers. *EMBO J* 13: 4963–4972.
36. Eskes R, Liu L, Ma H, Chao MY, Dickson L, et al. (2000) Multiple homing pathways used by yeast mitochondrial group II introns. *Mol Cell Biol* 20: 8432–8446.
37. Monteilhet C, Dziadkowiec D, Szczepanek T, Lazowska J (2000) Purification and characterization of the DNA cleavage and recognition site of I-SceI mitochondrial group I intron encoded endonuclease produced in *Escherichia coli*. *Nucleic Acids Res* 28: 1245–1251.
38. Belfort M (2003) Two for the price of one: a bifunctional intron-encoded DNA endonuclease-RNA maturase. *Genes Dev* 17: 2860–2863.
39. Hibbett DS, Binder M, Bischoff JF, Blackwell M, Cannon PF, et al. (2007) A higher-level phylogenetic classification of the Fungi. *Mycol Res* 111: 509–547.
40. Goddard MR, Burt A (1999) Recurrent invasion and extinction of a selfish gene. *Proc Natl Acad Sci USA* 96: 13880–13885.
41. Basse CW, Stumpferl S, Kahmann R (2000) Characterization of a *Ustilago maydis* gene specifically induced during the biotrophic phase: evidence for negative as well as positive regulation. *Mol Cell Biol* 20: 329–339.
42. Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular cloning: a laboratory manual*. 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
43. Nieto-Jacobo F, Pasch D, Basse CW (2012) The mitochondrial Dnm1-like fission component is required for *lga2*-induced mitophagy but dispensable for starvation-induced mitophagy in *Ustilago maydis*. *Eukaryot Cell* 11: 1154–1166.