Interim Report IR-09-074

Evolutionary maintenance of selfish homing endonuclease genes in the absence of horizontal transfer

Koji Yahara (omori@bio-math10.biology.kyushu-u.ac.jp)
Masaki Fukuyo (ikobaya@ims.u-tokyo.ac.jp)
Akira Sasaki (sasaki_akira@soken.ac.jp)
Ichizo Kobayashi (ikobaya@ims.u-tokyo.ac.jp)

Approved by

Ulf Dieckmann
Leader, Evolution and Ecology Program

June 2010

Contents

Evolutionary maintenance of selfish homing endonuclease genes

in the absence of horizontal transfer

Koji Yahara[*‡#], Masaki Fukuyo[**], Akira Sasaki[§++], and Ichizo Kobayashi[**]

[*] Division of Biostatistics, Kurume University School of Medicine, Fukuoka 830-0011, Japan

[‡] Division of Infectious Diseases, Kurume University School of Medicine, Fukuoka 830-0011, Japan

[#] Life Science Systems Department, Fujitsu Kyushu Systems Ltd., Fukuoka 814-8589, Japan

[§] Department of Evolutionary Studies of Biosystems (Sokendai-Hayama), The Graduate University for Advanced Studies (Sokendai), Hayama, Kanagawa 240-0193, JAPAN

[+] Evolution and Ecology Program, International Institute for Applied Systems Analysis, A-2361 Laxenburg, Austria

[†] PRESTO, Japan Science and Technology Agency, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, JAPAN

[**] Laboratory of Social Genome Sciences, Department of Medical Genome Sciences, Graduate School of Frontier Science & Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan

Author for correspondence: Ichizo Kobayashi

Address: Laboratory of Social Genome Sciences, Department of Medical Genome Sciences,

Graduate School of Frontier Science & Institute of Medical Science, University of Tokyo,

4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan

Phone: +81-3-5449-5326   Facsimile: +81-3-5449-5422

E-mail: ikobaya@ims.u-tokyo.ac.jp

Manuscript information: 23 pages, 3 figures, 3 tables

ABSTRACT

Homing endonuclease genes are 'selfish' mobile genetic elements whose endonuclease promotes the spread of its own gene by creating a break at a specific target site and using the host machinery to repair the break by copying and inserting the gene at this site. Horizontal transfer across the boundary of a species or population within which mating takes place has been thought to be necessary for their evolutionary persistence. This is based on the assumption that they will become fixed in a host population, where opportunities of homing will disappear, and become susceptible to degeneration. To test this hypothesis, we modeled behavior of a homing endonuclease gene that moves during meiosis through double-strand break repair. We mathematically explored conditions for persistence of the homing endonuclease gene and elucidated their parameter dependence as phase diagrams. We found that, if the cost of the pseudogene is lower than that of the homing endonuclease gene, the two forms can persist in a population through autonomous periodic oscillation. If the cost of the pseudogene is higher, two types of dynamics appear that enable evolutionary persistence: bistability dependent on initial frequency, or fixation irrespective of initial frequency. The prediction of long persistence in the absence of horizontal transfer was confirmed by stochastic simulations in finite populations. The average time to extinction of the endonuclease gene was found to be thousands of meiotic generations or more using realistic parameter values. These results provide a solid theoretical basis for understanding these and other extremely selfish elements.

INTRODUCTION

Homing endonucleases are a large class of proteins that are common in simple eukaryotes and in prokaryotes (1). Their coding regions (called genes here for convenience), are inserted into a specific DNA sequence and spliced out of either the transcribed mRNA as an intron, or out of the translated protein as an intein. Homing endonuclease genes act as "selfish" mobile genetic elements, because their gene product introduces a break at an 'empty' target site, and forces the host machinery to repair the break by copying the endonuclease gene. These genes are not essential to their hosts. Currently, much is known about the molecular structure and action of homing endonucleases and their genes, but little information is available on their population biology (2).

From a population genetics standpoint, a homing endonuclease allele could easily become fixed in a host population and susceptible to degeneration because homing opportunities disappear. Therefore, the long-term survival of a homing endonuclease allele at a specific site is proposed to depend on horizontal transfer across the boundary of a species or population, within which mating takes place, to another species or population that do not yet contain the homing endonuclease allele at that particular site (3, 4). Specifically, they have been hypothesized to persist over a long evolutionary period only if they can transmit to other species before extinction. For this reason, active homing endonuclease genes are often used as an indicator of DNA transfer events. This interpretation is consistent with phylogenetic analysis of homing endonuclease genes such as ω and PI-SceI (VDE) of *Saccharomyces cerevisiae* (5), the group I intron of the mitochondrial cox1 gene (6), and the group I intron of the 23S rRNA genes of hyperthermophilic bacteria.

Although these interpretations are widely accepted, a functional homing intein endonuclease in the PRP8 gene of euascomycetes (*Pezizomycota*) has been found that can survive within a species for several hundred million years, in the absence of interspecies horizontal transfer (4). Since the conditions that allow the evolutionary persistence of homing endonuclease genes without horizontal transfer are unknown, this process must be explored. In this work, we investigate if horizontal gene transfer across the boundary of a species or population within which mating takes place is necessary for the evolutionary maintenance of homing endonuclease genes. We built a model of a hypothetical homing endonuclease gene that spreads through meiotic double-strand break repair, similar to VDE (7). We analyzed its allele frequency over evolutionary periods in which an alternation of meiotic generations was interspersed with stages of haploid and diploid mitosis, followed by degenerate mutation, resulting in a pseudogene, which would be lost. We mathematically explored conditions that allowed persistence of the homing endonuclease gene, and numerically confirmed the evolutionary dynamics.

We discovered conditions for persistence of the homing endonuclease gene, elucidated their dependence on parameters such as its cost, and generated phase diagrams. Persistence was possible through autonomous periodic oscillation, or through two other types of evolutionary dynamics, even in the absence of horizontal transfer. These results were confirmed by stochastic simulations in finite populations.

MODEL

We constructed a model of a haploid-diploid cycle in a Mendelian population of unicellular eukaryotes, in which the haploid cells correspond to gametes in higher eukaryotes. We focused on a homing endonuclease locus that moves by homing in the early phase of meiosis. At each round of meiosis, which was designated as a meiotic generation, the survival of each allele depended on the frequencies of the interacting haploid genotypes in the population. The model is illustrated in Figure 1A, and symbols are explained in Table 1.

Each haploid cell carried one of three homing endonuclease alleles: a gene encoding the homing endonuclease (H^+), a pseudogene formed by its degenerate mutation (H^d), or its empty target site (H^-). A zygote was formed by random mating of the haploid cells. All six combinations of haploid cells could be produced by syngamy with probabilities given by the product of the frequency (x, y or z) of each allele. For example, the probability of the formation of diploid H^+/H^- was *2xy*, and that of H^+/H^d was *2yz*. We assumed that the haploid cells could freely switch their mating type (8) or that the homing endonuclease alleles were randomly associated with the mating types.

During meiosis of a H^+/H^- diploid cell, the homing endonuclease produced from the H^+ allele cleaves the unique empty target site (H^-) of the homologous chromosome. Homing endonucleases are known to have a single target site in a haploid genome (9). While they can evolve new target specificities (10, 11), they also tolerate individual base pair variation at their recognition sequences (12). We therefore ignored the possibility that mutations at their target prevented recognition. The DNA double-strand break generated by endonuclease cleavage is efficiently repaired by the recombination repair system of the host cell that is induced at the meiosis stage. Repair occurs with no deleterious effects (13) through the copying of the homologous H^+ allele, which results in gene conversion from H^- to H^+. This homing process takes place at the rate r. The homing endonuclease will not cleave the H^+ or H^d allele because of insertion at the recognition site. The pseudogene H^d and the empty target site H^- do not produce the functional homing endonuclease, so the homing process and the accompanying change in allele frequency can occur only in H^+/H^- diploids.

After the homing step, each diploid cell that results from production of a zygote was assumed to undergo meiosis to regenerate haploid cells. The frequency of each allele was calculated at this stage. We assumed a cost to the host cell of carrying a homing endonuclease gene, c_1, manifested as a reduced growth rate. The relative fecundity of a cell carrying an H^+ allele to one carrying only an H^- empty target site depended on the cost expressed as $\alpha = e^{-c_1}$. Also assumed was the cost of carrying a pseudogene, c_2, which was also represented by a reduced growth rate. The relative fecundity of a cell carrying H^d allele to one carrying only H^- alleles similarly depended on the cost expressed as $\beta = e^{-c_2}$. Assuming an independent contribution of H^+ or H^d to the cost, the relative fecundity of a diploid cell was given by the product of the fecundities of two haploid cells Two contrasting cases for the relative amounts of costs c_1 and c_2 (α and β) were assumed. If the cost of carrying an intact homing gene was more than the cost of carrying a pseudogene ($\beta > \alpha$), we set $\beta = \sqrt{\alpha}$ ($c_2 = c_1/2$). In contrast, if carrying a pseudogene gene was more costly ($\alpha > \beta$), then $\beta = \alpha^2 (c_2 = 2c_1)$. Because the cost to the host of carrying some homing endonuclease genes is very small (2), we assumed in our numerical analysis that c_1 was not large, and therefore α was high (for example, $c_1 = 0.01, \alpha = 0.99$).

The above calculation is summarized in Table 2 as a mating table containing all patterns of syngamy, with probabilities of occurrence and the number of progeny for each pattern. The relative number for a particular haploid progeny from one type of diploid is given by the probability of occurrence of syngamy, the rate of homing, and the relative fecundity of the diploid cell. Note that β can be replaced according to the definition of Table 1.

From the mating table (Table 2), we arrived at the following equations. The fitness W, which is the expected number of progeny per haploid individual, was calculated for each allele as follows:

$$W(x) = x + (1-r)y\alpha + z\beta$$

$$W(y) = \alpha\left\{ y\alpha + 2x\left(\frac{1}{2}(1-r) + r\alpha\right) + z\beta \right\}$$ (1)

$$W(z) = \beta(x + y\alpha + z\beta)$$

The allele frequency in the next meiotic generation can be calculated as the expected number of haploid cells harboring each allele, for example, $W(y)y$, divided by the total number of haploid cells in the population, or the mean fitness $\bar{W} = W(x)x + W(y)y + W(z)z$:

$$y^* = \frac{\alpha y\left\{ \alpha y + 2x\left(\frac{1}{2}(1-r) + r\alpha\right) + \beta z \right\}}{(x + \alpha y + \beta z)^2 - 2r\alpha(1-\alpha)xy}$$ (2)

$$z^* = \frac{\beta z(x + \alpha y + \beta z)}{(x + \alpha y + \beta z)^2 - 2r\alpha(1-\alpha)xy}$$

where $y_{next} = y^*$ and $z_{next} = z^*$ are the frequencies of H^+ and H^d in the next meiotic generation. Note that x can be replaced by $x = 1 - y - z$. The frequency of H^- in the next meiotic generation was calculated using $x^* = 1 - y^* - z^*$.

When the effect of mutation was introduced, equations (2) changed as follows:

$$y_{next} = (1-u)y^*$$

$$z_{next} = uy^* + (1-v)z^*$$ (3)

As illustrated in Figure 1B, only a limited case of mutations was taken into account. That is, only degeneration of the H^+ allele to H^d at rate u, and precise loss of the pseudogene creating an empty target site H^- at rate v was assumed. Since v is conceivably smaller than u, we assumed $v = \dfrac{u}{10}$ in our numerical analysis. All other mechanisms with the potential to change allele frequency, such as gene conversion caused by mechanisms other than homing, horizontal gene transfer, or migration were not assumed here. We defined horizontal gene transfer as transfer of a gene across the boundary of a species or population within which mating takes place, that is a Mendelian population. An empty target site H^- could be converted to H^+ only through the homing process, resulting in an increase in the frequency of the homing endonuclease gene in the population.

We mathematically explored conditions for persistence of the H^+ allele based on the above equations. Stability of three marginal equilibria in which each of the three alleles predominates in the population was analyzed. Stability of the internal equilibrium in which the three alleles could coexist was then analyzed, together with the conditions for its existence. We summarized the parameter dependence of the conditions that allowed the evolutionary maintenance of the homing endonuclease gene with phase diagrams, and numerically confirmed the evolutionary dynamics. (See SI for details on the stability analyses and derivation of the phase diagrams.)

We also carried out computer simulations in finite populations with an effective size $N \geq 10^6$ (14) to consider the possibility of stochastic extinction, and examined the validity of our mathematical analysis. Because we did not assume that any of the mechanisms described above recovered the H^+ allele, it would inevitably become extinct during a long evolutionary time span because of degeneration caused by accumulation of mutations. The average time until the extinction of the H^+ allele was calculated by 50 runs of simulation for each parameter set.

RESULTS

The conditions for evolutionary maintenance of the homing endonuclease gene were classified by the magnitude of α and β. If $\beta > \alpha$, i.e., the cost of carrying the homing endonuclease gene H$^+$ is larger than the cost of carrying the homing endonuclease pseudogene H^d, the homing endonuclease gene can be maintained through periodic oscillation of the three alleles. In contrast, if $\beta < \alpha$, the homing endonuclease gene can predominate in the population. Phase diagrams and the evolutionary dynamics are as follows.

Phase diagrams and evolutionary dynamics for $\beta > \alpha$: If $\beta > \alpha$, which is the case when the pseudogene has lost its homing endonuclease activity but retains splicing activity, thus imposing only a minor additional cost on the host cell, the pseudogene H^d is advantageous over H$^+$, if there is no H$^-$ allele in the equilibrium. In this case, the phase diagram shows two regions: the stable H$^-$ monomorphism and the limit cycle of all three homing alleles (Figure 2).

The black line in the phase diagram shows the boundary of the area of H$^-$ predominance obtained from equation (4),

$$r < \frac{1 - \alpha + u\alpha}{(1-u)\alpha(2\alpha - 1)} \tag{4}$$

indicating that H$^-$ is stable if the rate of homing is smaller than a threshold. Specifically, the empty allele H$^-$ predominates if the cost of homing is considerably high compared to the rate of homing.

In the region to the right of the black line, where equation (4) is not satisfied, however, none of the three alleles can permanently predominate in the population. All equilibria, including internal equilibrium, are unstable, and limit or heteroclinic cycles occur, depending on the presence or absence of mutation. Accordingly, the H$^+$ allele can be maintained through a limit cycle in the presence of mutation, if the homing rate r is considerably high compared to the cost of carrying a homing endonuclease gene.

The gray line in the phase diagram of Figure 2 is the boundary for the area of the existence of the internal equilibrium. This is obtained from the equation (S15) of SI, indicating that an internal equilibrium exists on its right. In the space between the black and gray lines, H$^-$ is stable, irrespective of its initial frequency, and its evolutionary dynamics move away from the internal equilibrium towards marginal equilibrium where H$^-$ dominates. The enlargement in Figure 2 shows how these two boundaries shift with mutation. In the presence of mutation, the black boundary line shifts upward by the mutation rate u, while the gray boundary line shifts slightly downward.

The evolutionary dynamics of the limit cycle of all three homing alleles is numerically illustrated in Figure 3. The H$^+$ allele is maintained through periodic oscillation of the three homing endonuclease alleles. At the beginning, the majority of the allele population is H$^-$, providing ample opportunity for homing, which enables the spread of the H$^+$ allele. After the apparent near-fixation, however, the H$^+$ allele is susceptible to degeneration because there is little opportunity for homing. This, in turn, enables the spread of the H^d allele. The first 800 meiotic generations in the upper right of Figure 3 show this fixation-to-degeneration process.

H^d has a relative advantage over H^+ while it still imposes cost c_2 (assumed to be $c_1/2$ and thus $\beta = \sqrt{\alpha}$) on the host, and is thus disadvantageous over H^- when the threat by H^+ becomes small. This, in turn, increases H^- frequency, which again generates the homing opportunity for the H^+ allele. The above process is repeated, and the period is approximately 2000 generations using the parameter set of Figure 3. The period decreases monotonically as the relative fecundity of a cell carrying H^+ decreases, as is shown in SI (Figure S1). The limit cycle enables the H^+ allele to be maintained over thousands of meiotic generations in the absence of horizontal transfer. A schematic diagram of the limit cycle of homing endonuclease alleles in a population is shown in SI (Figure S2). The result is consistent with cyclical models of the gain, degeneration and loss of homing endonuclease genes that are formulated based on the presence of horizontal transfer (3, 5). Results under different cost and mutation parameter settings are summarized in SI (Figure S3). In all the parameter spaces explored, the limit cycle appeared and the homing endonuclease gene was maintained.

Stochastic simulations for $\beta > \alpha$: In order to consider stochastic extinction, and examine the validity of our mathematical analysis, we also carried out simulations in finite populations. Because we did not assume any mechanism for recovery of the H^+ allele, such as horizontal transfer, migration or reverse mutation, the homing endonuclease allele would inevitably become extinct at some point over a long evolutionary period. The average time until extinction is expected to depend on the cost and mutation rate. The results summarized in Table 3 indicate that the average time is more than thousands of meiotic generations, unless the cost is very large. An example of the long-term trajectory under the same parameter set as in Figure 3, with $N \geq 10^6$ is illustrated in SI (Figure S4).

Phase diagrams and evolutionary dynamics when $\beta < \alpha$: If $\beta < \alpha$, for example, if the intron or intein has lost the ability to be spliced out and consequently has a deleterious effect on its host, the pseudogene H^d is disadvantageous compared to the H^+ allele. In this case, the phase diagram shows three regions: H^- stable, bistable, and H^+ stable (Figure S5 in SI). In this phase diagram, the black line, obtained from equation (4), represents the boundary of the area of H^- predominance as in Figure 2. As above, H^- is stable on its left. In contrast, the gray line in the phase diagram is the boundary of the area of H^+ predominance in a population. In the absence of mutation, this is obtained from equation (5),

$$r > 1 - \alpha \tag{5}$$

which indicates that H^+ is stable if the rate of homing r is larger than the threshold. In the presence of mutation, an H^+-dominating marginal equilibrium exists if $\beta < \alpha$:

$$\left(y, z \right) = \left(1 - \frac{u\alpha}{\alpha - \beta}, \frac{u\alpha}{\alpha - \beta} \right) \tag{6}$$

where higher terms of u are disregarded. The condition for the local stability of H^+-predominant equilibrium (6) is again approximated by (5). Therefore, in the deep gray "H^+ stable" region where the equation (5) is satisfied, H^+ predominates irrespective of its initial frequency. In the light gray "bistable" region, where both equation (4) and (5) are satisfied, the result depends on the initial frequency. In the white "H^- stable" region, H^- predominates irrespective of its initial frequency. The enlargement in Figure S5 in SI shows how these two boundaries shift with mutation. Similar to Figure 2, the black boundary line shifts upward by the mutation rate u, while the gray boundary line shifts slightly downward.

In contrast to the case of $\alpha < \beta$, no internal equilibrium exists when $\alpha > \beta$.

The evolutionary dynamics of the "bistable" and "H$^+$ stable" regions are numerically illustrated in Figures S6A and S6B in SI, respectively. The former dynamics were confirmed by randomly assigning 100 patterns of initial frequencies to the three alleles, showing that the homing endonuclease gene H$^+$ predominates if its initial frequency is relatively high. The latter shows that the homing endonuclease gene H$^+$ predominates irrespective of its initial frequency.

Stochastic simulations for $\beta < \alpha$: Under $N \geq 10^6$, H$^+$ does not become extinct within 10^5 meiotic generations for any parameter set used in Table 3. An example of unrealistically disadvantageous parameter sets for the H$^+$ allele, and its trajectory to extinction, is illustrated in SI (Figure S7).

DISCUSSION

Our mathematical analyses revealed that evolutionary persistence is indeed possible for a homing endonuclease gene in the absence of gene transfer across the boundary of a species or population within which mating takes place. The dynamics enabling the persistence were classified by the cost to the host cell of carrying the homing endonuclease gene H$^+$ or its pseudogene H^d, and two phase diagrams were constructed.

If the cost of the pseudogene was lower than that of the homing endonuclease gene ($\beta > \alpha$), persistence of the homing endonuclease gene was possible through autonomous periodic oscillation in the population, in the presence of mutation (Figure 2, Figure 3). An example of this case would be if the pseudogene lost homing endonuclease activity but retained splicing activity. The endonuclease domain of VDE is known to degenerate after becoming fixed in a population. This type of pseudogene is common in *Saccharomyces* (7).

Conversely, if the cost of the pseudogene was higher than that of the intact homing endonuclease gene ($\beta < \alpha$), persistence of the homing endonuclease gene was possible through the evolutionary dynamics of bistability. Furthermore, if the homing rate became sufficiently high, persistence of the homing endonuclease gene was possible irrespective of its initial frequency (Figure S5, Figure S6B). An example of this case would be the presence of a pseudogene whose intron or intein has a reduced ability to be spliced out (15), which could have a deleterious effect on the host cell.

In the former case, opportunities of homing must be regularly created, even after fixation of the homing endonuclease gene, for the evolutionary maintenance of homing endonuclease genes. This is because the pseudogene is advantageous over the homing endonuclease gene and can always invade the population. Horizontal gene transfer is believed to be the most effective mechanism for this. Although we do not deny the effectiveness of horizontal gene transfer, our results suggest that horizontal gene transfer is not the only mechanism that enables persistence of homing endonuclease genes over a long evolutionary period. This was confirmed by stochastic simulations on finite populations with effective population sizes $\geq 10^6$ which was consistent with estimates for yeast and other small eukaryotes (14, 16). The average time to extinction of allele H$^+$ is thousands of meiotic generations, unless its cost is high. Because meiosis in *Saccharomyces cerevisiae* is initiated only upon nutritional starvation of diploid cells, its frequency is expected to be much smaller than that of mitosis of either haploid cells or diploid cells. Although the exact ratio of meiosis to mitosis in natural environments has not been examined for *Saccharomyces cerevisiae* (17), the ratio in *Saccharomyces paradoxus,* the closest wild relative of *S.cerevisiae*, is estimated to be 1:1000 in Europe and 1:3000 in the Far East (14). Therefore

the ratio could range from 1:100 to 1:10000 for *S. cerevisiae*. The thousands of meiotic generations would then correspond to 10^5 - 10^7 mitotic generations. In addition, the doubling time for unicellular organisms in natural environments is likely to be much longer than in laboratory conditions. *Escherichia coli* reproduces at a doubling time of 40 h in the human intestine (18), but exhibits a doubling time of 0.5 h under laboratory conditions. We could not find comparable estimates for *Saccharomyces cerevisiae*, which can be regarded as an opportunistic pathogen on fruits and other plants (17), and might assume a similar 10^2 fold difference. In addition, microbes are likely to be in a resting stage more often than a dividing stage. We found no estimate of how often yeast are in a dividing stage in natural environments, but it could be 1 to 100. Taken together, the average time for extinction estimated above appears to be, in practice, very long. Any mechanism increasing H^+ frequency such as reverse mutation or migration would further promote its persistence.

In experiments with VDE of *Saccharomyces cerevisiae*, the cost of carrying a homing endonuclease gene to the host cell seems very small because effect of VDE on mitotic replication rates is less than 1% in inbred populations (19). The observation corresponds to the regions in our phase diagrams where the H^+ allele is maintained under the relatively low cost of carrying a homing endonuclease gene.

While cyclical models of the gain, degeneration and loss of homing endonuclease genes based on their horizontal transfer, have been formulated (3, 5), this work is, to our knowledge, the first published theoretical analysis to successfully summarize the parameter dependence of the conditions for the cycle in the absence of horizontal gene transfer. Whether the evolutionary maintenance of homing endonuclease genes through periodic oscillation without horizontal gene transfer occurs in nature is worth exploring.

Other than mutation, migration and horizontal gene transfer, other mechanisms might affect endonuclease gene frequency. One possible mechanism is gene conversion of H^d/H^- during meiosis, producing H^d/H^d or H^-/H^-, and resulting in a haploid cell with an empty site allele. Mitotic gene conversion in H^+/H^- diploids would be very low for VDE, because its meiosis-specific, karyopherin-mediated nuclear import is required for its action on the chromosomal site (13, 20). Another possible mechanism is reverse transcription of spliced mRNAs that do not contain introns, followed by homologous recombination of the cDNA with the genomic DNA, resulting in intron loss (21). These mechanisms would help create homing opportunities, and would be advantageous for the homing endonuclease allele.

In each meiotic generation, the survival of each homing endonuclease allele depends on the frequencies of the interacting haploid cells. Our haploid-diploid cycle model was constructed to analyze the population dynamics of homing endonuclease alleles that involved interdependence among several alleles. The same framework could be applied to other homing endonuclease genes that spread through comparable chromosome cycles. Homing endonucleases of bacteriophage T4 (I-*Tev* I, I-*Tev* II, F-*Tev* I, and F-*Tev* II) are one example (1). Bacteriophage double-strand break repair mechanisms involving homologous recombination (22) would provide an opportunity for homing during co-infection of a single host cell. Homing endonuclease ω of yeast mitochondria are another example. Mitochondria are biparentally inherited in *S. cerevisiae* and recombine through a haploid-diploid-like cycle in which a double-strand break induces homologous recombination and gene conversion (5).

Another model of the evolutionary trajectory of a selfish invasion, followed by slow decay to complete loss, has been proposed for vertically transmitted bacteria of the genus *Wolbachia* (23). In this case, crosses between a male infected with *Wolbachia* and an uninfected female produce only few progeny because of cytoplasmic incompatibility. The infected males thus sacrifice themselves to reduce the frequency of uninfected females. Invasion and temporary predominance of the infected cytotype is followed by invasion of the resistant cytotype, allowing the eventual spread of the uninfected cytotype. However, the

dynamics are frequency dependent and do not allow the uninfected cytotype to spread irrespective of its initial frequency. This characteristic of dynamics, and the asymmetric relationship between the males and females, distinguishes their behavior from that of the homing endonuclease genes.

Other selfish genetic elements that cause a meiotic drive, such as Segregation Distorter in *Drosophila melanogaster* (24) or the spore killer in ascomycetes (25), also have a three-allele polymorphism, i.e., killer, sensitive and resistant To be precise, their "alleles" are a combination of a killer/non-killer, and another allele. Segregation Distorter is different from the homing endonuclease genes, in that the polymorphism is maintained by a strong directional selection against the killer alleles in diploid organisms. In contrast, the life cycle and mechanism for maintaining polymorphism in spore killer seems similar to those of the homing endonuclease genes. The published model (23), however, allowed sensitive to killer mutations, and did not analytically find conditions for the limit cycle and bistability, as were found here. None of the above models, including the one presented here, consider spatial structure, meaning the spatial localization of interactions between organisms. Because spatial structure can have a signficant effect on the qualitative differences in the direction of evolution (26), inclusion of this effect into the models is worth exploring. Another study on a selfish genetic element with three alleles (toxin-producing, sensitive, resistant) revealed that spatial structure promotes coexistence of the three alleles in a population (27).

This study is the first theoretical analysis of the evolutionary persistence of homing endonuclease genes in the absence of horizontal transfer to successfully summarize parameter dependence as phase diagrams. The results will provide a solid theoretical basis for our understanding of the evolution of homing endonuclease genes and other selfish genetic elements that spread through a comparable cycle.

REFERENCES

1. Belfort, M. & Roberts, R. J. (1997) Homing endonucleases: keeping the house in order *Nucleic Acids Res* **25,** 3379-88.
2. Burt, A. & Trivers, R. (2006) *Genes in Conflict: The Biology of Selfish Genetic Elements* (Harvard University Press.
3. Burt, A. & Koufopanou, V. (2004) Homing endonuclease genes: the rise and fall and rise again of a selfish element *Curr Opin Genet Dev* **14,** 609-15.
4. Butler, M. I., Gray, J., Goodwin, T. J. & Poulter, R. T. (2006) The distribution and evolutionary history of the PRP8 intein *BMC Evol Biol* **6,** 42.
5. Goddard, M. R. & Burt, A. (1999) Recurrent invasion and extinction of a selfish gene *Proc Natl Acad Sci USA* **96,** 13880-5.
6. Cho, Y. & Palmer, J. D. (1999) Multiple acquisitions via horizontal transfer of a group I intron in the mitochondrial cox1 gene during evolution of the Araceae family *Mol Biol Evol* **16,** 1155-65.
7. Okuda, Y., Sasaki, D., Nogami, S., Kaneko, Y., Ohya, Y. & Anraku, Y. (2003) Occurrence, horizontal transfer and degeneration of VDE intein family in Saccharomycete yeasts *Yeast* **20,** 563-73.
8. Haber, J. E. (1998) Mating-type gene switching in Saccharomyces cerevisiae *Annu Rev Genet* **32,** 561-99.
9. Posey, K. L., Koufopanou, V., Burt, A. & Gimble, F. S. (2004) Evolution of divergent DNA recognition specificities in VDE homing endonucleases from two yeast species *Nucleic Acids Res* **32,** 3947-56.
10. Rosen, L. E., Morrison, H. A., Masri, S., Brown, M. J., Springstubb, B., Sussman, D., Stoddard, B. L. & Seligman, L. M. (2006) Homing endonuclease I-CreI derivatives with novel DNA target specificities *Nucleic Acids Res* **34,** 4791-800.
11. Scalley-Kim, M., McConnell-Smith, A. & Stoddard, B. L. (2007) Coevolution of a homing endonuclease and its host target sequence *Journal of Molecular Biology* **372,** 1305-1319.
12. Chevalier, B. S. & Stoddard, B. L. (2001) Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility *Nucleic Acids Res* **29,** 3757-74.
13. Fukuda, T., Nogami, S. & Ohya, Y. (2003) VDE-initiated intein homing in Saccharomyces cerevisiae proceeds in a meiotic recombination-like manner *Genes Cells* **8,** 587-602.
14. Tsai, I. J., Bensasson, D., Burt, A. & Koufopanou, V. (2008) Population genomics of the wild yeast Saccharomyces paradoxus: Quantifying the life cycle *Proc Natl Acad Sci USA* **105,** 4957-62.
15. Kawasaki, M., Nogami, S., Satow, Y., Ohya, Y. & Anraku, Y. (1997) Identification of three core regions essential for protein splicing of the yeast Vma1 protozyme. A random mutagenesis study of the entire Vma1-derived endonuclease sequence *J Biol Chem* **272,** 15668-74.
16. Lynch, M. & Conery, J. S. (2003) The origins of genome complexity *Science* **302,** 1401-4.
17. Landry, C. R., Townsend, J. P., Hartl, D. L. & Cavalieri, D. (2006) Ecological and evolutionary genomics of Saccharomyces cerevisiae *Mol Ecol* **15,** 575-91.
18. Hartl, D. L. & Dykhuizen, D. E. (1984) The population genetics of Escherichia coli *Annu Rev Genet* **18,** 31-68.
19. Goddard, M. R., Greig, D. & Burt, A. (2001) Outcrossed sex allows a selfish gene to invade yeast populations *Proc R Soc Lond Ser B* **268,** 2537-42.
20. Nagai, Y., Nogami, S., Kumagai-Sano, F. & Ohya, Y. (2003) Karyopherin-mediated nuclear import of the homing endonuclease VMA1-derived endonuclease is required for self-propagation of the coding region *Mol Cell Biol* **23,** 1726-36.
21. Jeffares, D. C., Mourier, T. & Penny, D. (2006) The biology of intron gain and loss

Trends Genet **22,** 16-22.

22. Takahashi, N. & Kobayashi, I. (1990) Evidence for the double-strand break repair model of bacteriophage lambda recombination *Proc Natl Acad Sci USA* **87,** 2790-4.

23. Hurst, L. D. & McVean, G. T. (1996) Clade selection, reversible evolution and the persistence of selfish elements:the evolutionary dynamics of cytoplasmic incompatibility *Proc R Soc Lond Ser B* **263,** 97-104.

24. Charlesworth, B. & Hartl, D. L. (1978) Population Dynamics of the Segregation Distorter Polymorphism of DROSOPHILA MELANOGASTER *Genetics* **89,** 171-192.

25. Nauta, M. J. & Hoekstra, R. F. (1993) Evolutionary dynamics of spore killers *Genetics* **135,** 923-30.

26. Mochizuki, A., Yahara, K., Kobayashi, I. & Iwasa, Y. (2006) Genetic addiction: selfish gene's strategy for symbiosis in the genome *Genetics* **172,** 1309-23.

27. Kerr, B., Riley, M. A., Feldman, M. W. & Bohannan, B. J. (2002) Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors *Nature* **418,** 171-4.

FIGURE LEGENDS

Figure 1. Model of homing endonucleases. (A) Life cycle of homing endonuclease alleles. Allele H^+ produces a homing endonuclease that cleaves the empty allele, H^-, on its homologous chromosome, in a diploid cell during meiosis. The DNA double-strand break is repaired by the host with the H^+ allele as the template. This process of homing results in gene conversion from H^- to H^+. The homing endonuclease does not cleave H^d, which represents a pseudogene degenerated from H^+. Diploid cells are formed by syngamy from all pairwise combinations of haploid cells, whose probabilities depend on the frequency of each haploid cell type. **(B) Mutation.** The H^+ allele becomes pseudogene H^d by degenerate mutation at the rate u. Pseudogene H^d is occasionally lost from its target site with the rate v, which is smaller than u.

Figure 2. Phase diagram for $\alpha < \beta$ indicating the region in which the H^+ allele is maintained through periodic oscillation. The black line is obtained from equation (4). In the gray region on its right, a limit or heteroclinic cycle occurs. H^- is stable in the white region on the left side of the black boundary irrespective of its initial frequency. The gray line is obtained from equation (S15) of SI, indicating that an internal equilibrium exists on its right. The enlargement in the upper right shows how these two boundaries shift with mutation. The black boundary line shifts upward by the mutation rate u. In contrast, the gray boundary line shifts downward. The two boundaries intersect at $(\alpha, r) = (1, 0)$. These diagrams are illustrated with mutation rates $u = 10^{-4}, v = 10^{-5}$, which are higher than expected, for clarity.

Figure 3. Evolutionary dynamics of periodic oscillation of homing endonuclease alleles for $\alpha < \beta$. Results obtained numerically for a parameter set in the "Cycle" region of Figure 2 with mutations: $\alpha = 0.99$ ($c_1 = 0.01$), $r = 0.2$, $u = 6 \times 10^{-6}, v = 6 \times 10^{-7}$. The initial frequencies are 99% for H^- and 1% for H^+.

Table 1. Definitions

Variable and parameter	Symbol	Value
Homing endonuclease absent allele	H^-	
Homing endonuclease positive allele	H^+	
Homing endonuclease defective (pseudogene) allele	H^d	
Frequency of H^-	x	
Frequency of H^+	y	
Frequency of H^d	z	
Rate of homing	r	
(DNA double-strand breakage by the homing endonuclease and succeeding repair by a host cell)		
Cost of carrying a homing endonuclease gene H^+ on a host cell	c_1	
Cost of carrying a homing endonuclease pseudogene H^d on a host cell	c_2	
Relative fecundity of a cell carrying an H^+ allele to that carrying only H^- empty sites	α	e^{-c_1}
Relative fecundity of a cell carrying an H^d allele to that carrying only H^- empty sites	β	e^{-c_2}
Rate of degeneration (pseudogene formation)	u	
Rate of losing pseudogene (precise excision)	v	$u/10$
Effective population size	N	

If $\alpha > \beta$, then $\beta = \alpha^2 \, (c_2 = 2c_1)$. In contrast, if $\beta > \alpha$, then $\beta = \sqrt{\alpha} \; (c_2 = c_1 / 2)$.

Table 2. Mating table

Zygote genotypes	Probability	Progeny		
		H^-	H^+	H^d
H^-/H^-	x^2	1	0	0
H^+/H^-	$2xy$	$a(1-r)/2$	$a^2r + a(1-r)/2$	0
H^+/H^+	y^2	0	a^2	0
H^d/H^-	$2xz$	$\beta/2$	0	$\beta/2$
H^d/H^d	z^2	0	0	β^2
H^d/H^+	$2yz$	0	$\alpha\beta/2$	$\alpha\beta/2$

Table 3. Average logarithmic time to extinction
of H^+ allele in stochastic simulations for $\alpha < \beta$

$c_1 = 0.01$ ($\alpha \approx 0.990$)		
N 10^6	10^7	10^8
u		
6×10^{-6} 3.8 ± 0.4	> 5	> 5
6×10^{-5} > 5	> 5	> 5
6×10^{-4} > 5	> 5	> 5

$c_1 = 0.03$ ($\alpha \approx 0.970$)		
N 10^6	10^7	10^8
u		
6×10^{-6} 3.4 ± 0.3	> 5	> 5
6×10^{-5} > 5	> 5	> 5
6×10^{-4} > 5	> 5	> 5

$c_1 = 0.05$ ($\alpha \approx 0.951$)		
N 10^6	10^7	10^8
u		
6×10^{-6} 3.1 ± 0.2	> 5	> 5
6×10^{-5} > 5	> 5	> 5
6×10^{-4} > 5	> 5	> 5

The average logarithmic time ($\pm$ SD) in terms of

meiotic generation was calculated from 50 runs

of simulation. The rate of homing, *r*, was 0.2. All logarithms are to base 10.

Figure 1

A.

syngamy

$2xy$ ... $2xz$ ...other 4 patterns

r

endonuclease

mutation

meiosis

B.

H$^+$ u H$^\mathbf{d}$ v H$^-$

Degeneration
(pseudogene formation)

Loss of pseudogene

Figure 2

Figure 3

A.

B.
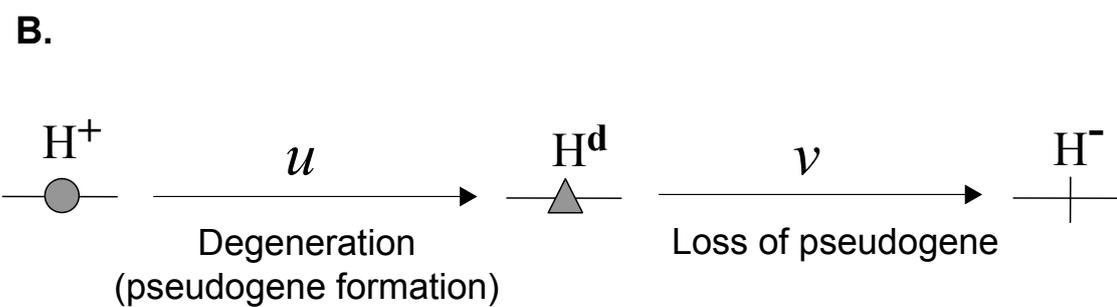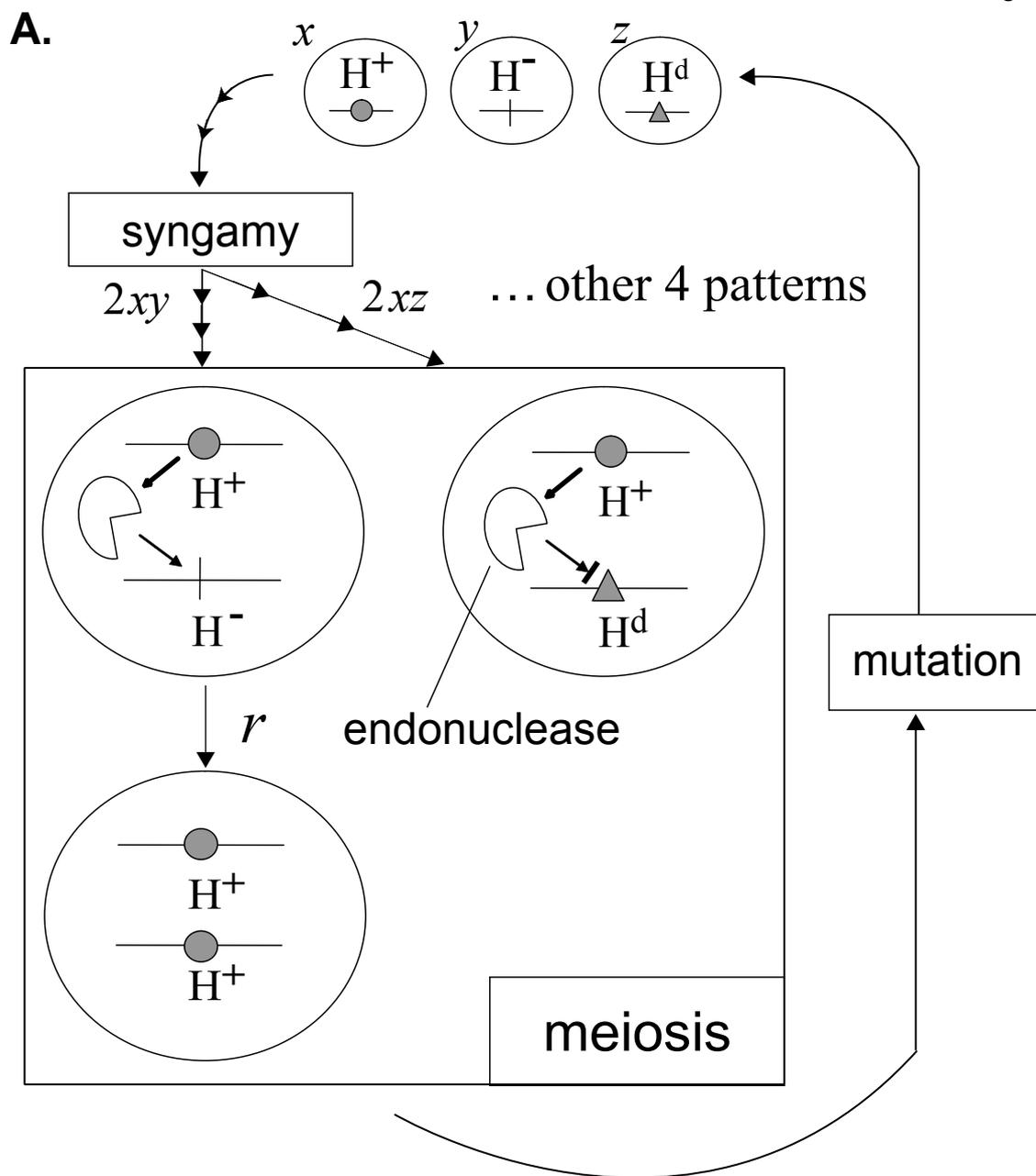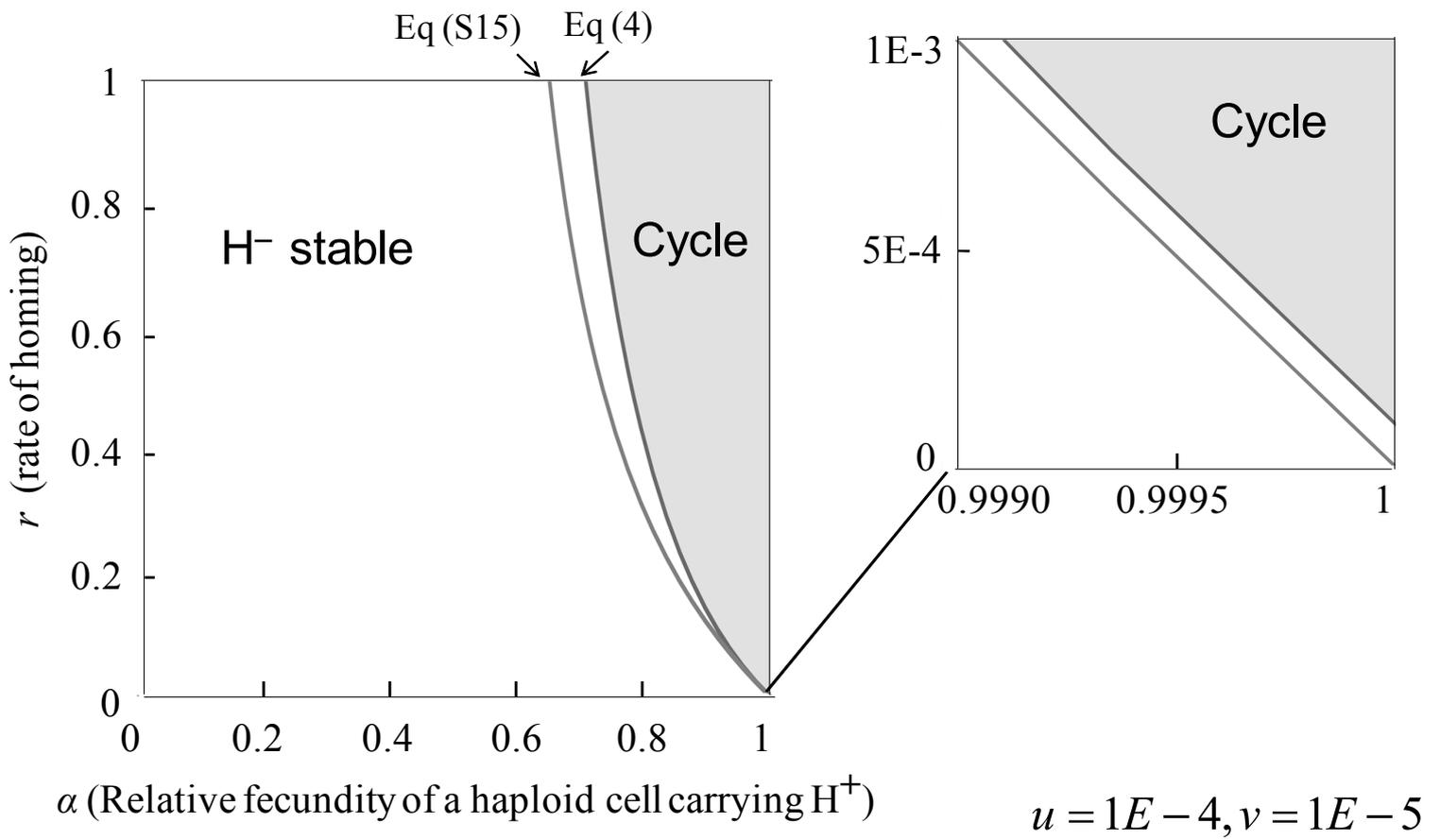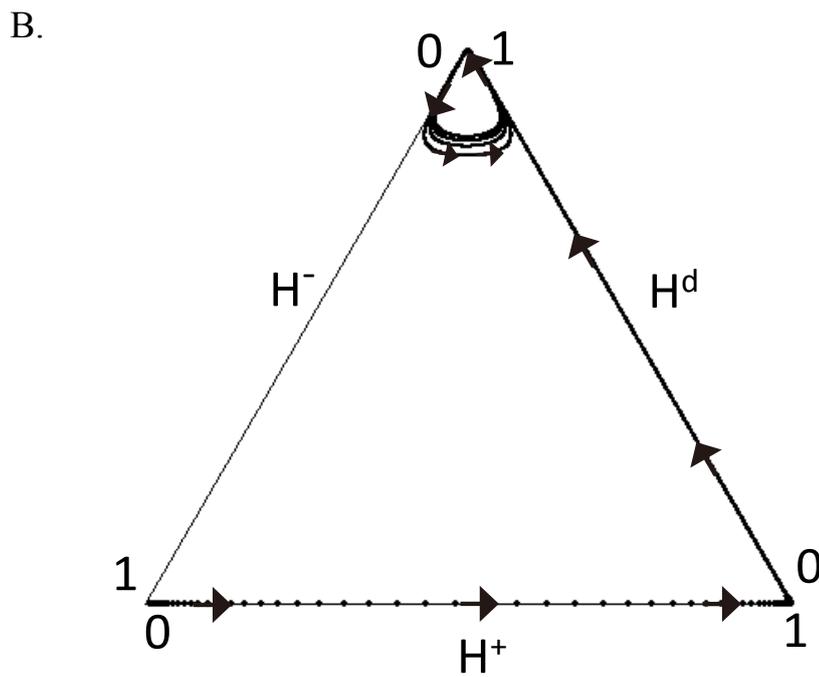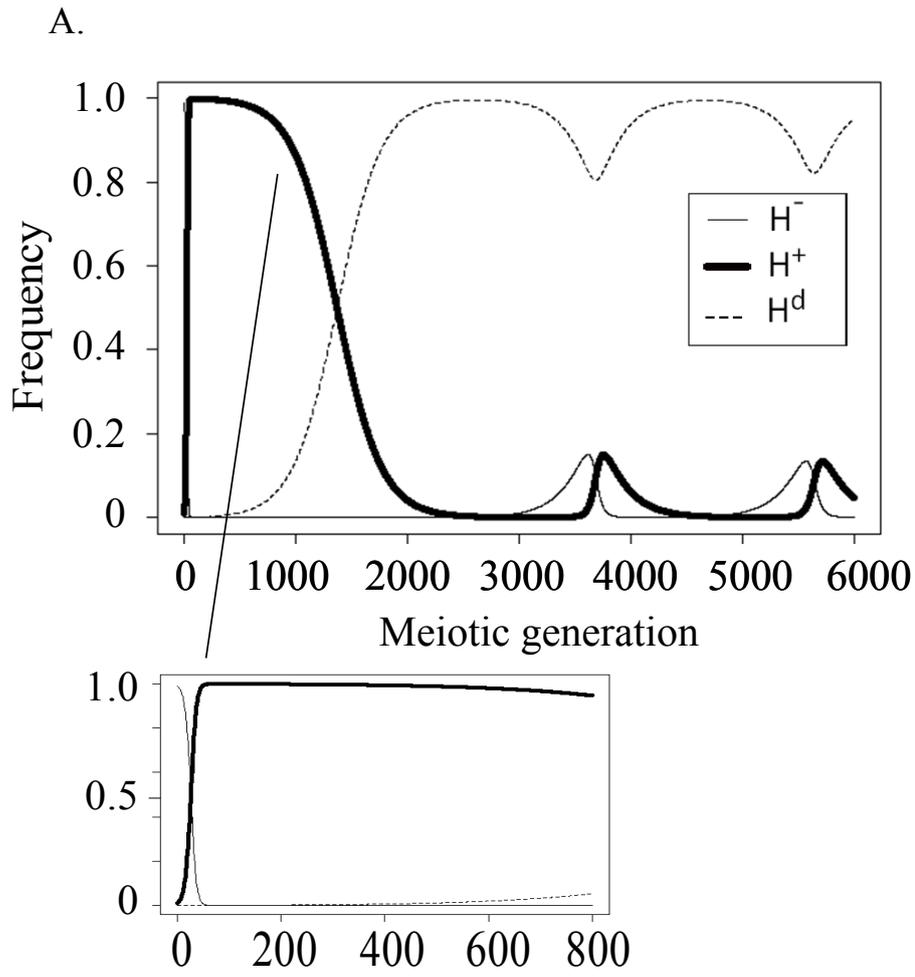
SUPPORTING INFORMATION

Stability of marginal equilibria: Based on the equations (2) and (3) in the text, Jacobian matrix for a marginal equilibrium where the empty allele H⁻ predominates (i.e., $x=1, y=0, z=0$) is calculated as follows

$$\begin{pmatrix} (1-u)\alpha(r(2\alpha-1)+1) & 0 \\ u\alpha(r(2\alpha-1)+1) & (1-v)\beta \end{pmatrix} \tag{S1}$$

indicating that it is locally stable if the condition below is satisfied:

$$r < \frac{1-\alpha+u\alpha}{(1-u)\alpha(2\alpha-1)} \tag{S2}$$

This gives the inequality (4) in the text. The black line of Figure 2 or Figure S5 is obtained from the equation (4).

For another marginal equilibrium where the homing endonuclease gene H⁺ predominates (i.e., $x=0, y=1, z=0$), we approximately consider the effect of mutation which causes slight deviation from the equilibrium.

By introducing ε as the effect of mutation and dy, dz as deviation from the equilibrium, we replace the mutation rate u, v by $\varepsilon U, \varepsilon V$ and frequency y, z by $1-\varepsilon dy$, εdz, respectively. Substituting the equilibrium condition, $y_{next} = y$ and $z_{next} = z$ of (2)-(3) in the text, and expanding the resultant in the Taylor series yields

$$(dy, dz) = \left(\frac{\varepsilon U\alpha}{\alpha-\beta}, \frac{\varepsilon U\alpha}{\alpha-\beta} \right)$$

at equilibrium. Therefore the equilibrium frequencies with mutation is

$$(y, z) = \left(1 - \frac{u\alpha}{\alpha-\beta}, \frac{u\alpha}{\alpha-\beta} \right) \tag{S3}$$

if u^2 terms are neglected. This equilibrium is in a feasible region only if $\alpha > \beta$. Jacobian matrix for $\varepsilon = 0$ is

$$\begin{pmatrix} \dfrac{1-r}{\alpha} & \dfrac{1-r-\beta}{\alpha} \\ 0 & \dfrac{\beta}{\alpha} \end{pmatrix} \tag{S4}$$

indicating that it is locally stable only if the conditions below are satisfied from eigenvalue of the Jacobian matrix:

$$\beta < \alpha \tag{S5a}$$

$$r > 1 - \alpha \qquad\qquad\qquad\qquad \text{(S5b)}$$

This gives the condition (5) in the text for the stability of H$^+$ equilibrium in the case of $\beta < \alpha$ and in the absence of mutation. Namely, only if $\beta < \alpha$ and the rate of homing r is larger than a threshold, this marginal equilibrium is stable. Considering the effect of mutation and expanding the equation (S1) in Taylor series yields the condition for the local stability of equilibrium (S3)

$$r > 1 - \alpha + \varepsilon \frac{U(\alpha^2 + \beta - 2\alpha\beta)}{\alpha - \beta} + O[\varepsilon^2] \qquad\qquad \text{(S6)}$$

The gray line of Figure S5 is obtained from the equation (S6).

For the third marginal equilibrium where the pseudogene H^d predominates (i.e., $x=0$, $y=0$, $z=1$), we similarly need to consider the effect of mutation which causes slight deviation from the equilibrium. By introducing ε as the effect of mutation and dz as deviation from the equilibrium, we replace the mutation rate u, v by $\varepsilon U, \varepsilon V$ and frequency y, z by 0, $1 - \varepsilon dz$, respectively. Expanding in Taylor series yields (dy, dz) = (0, 0) in the equilibrium. Therefore the equilibrium is $(y, z) = (0, 1)$. Jacobian matrix for this marginal equilibrium can thus be expressed as follows by regarding $\varepsilon = 0$:

$$\begin{pmatrix} \dfrac{\alpha}{\beta} & 0 \\[3mm] \dfrac{1-\alpha}{\beta} & \dfrac{1}{\beta} \end{pmatrix} \qquad\qquad\qquad \text{(S7)}$$

indicating that this equilibrium is always unstable because β is less than 1 by definition.

Therefore, if $\beta > \alpha$ and if the equation (4) in the text is not satisfied, all of these marginal equilibria are unstable. The marginal equilibrium where the homing endonuclease gene H$^+$ predominates (i.e., $x=0$, $y=1$, $z=0$) can be stable only if $\beta < \alpha$.

Internal equilibrium: We then calculate an internal equilibrium from $W(x) = W(y) = W(z) = \phi$ and $x + y + z = 1$. In the absence of mutation, the equilibrium frequency of y is obtained from the equation (1) by

$$W(x) - \frac{W(z)}{\beta} = -ry\alpha = \phi - \frac{\phi}{\beta}$$

yielding

$$\hat{y} = \frac{(1-\beta)\phi}{r\alpha\beta} \qquad\qquad\qquad \text{(S8)}$$

The equilibrium frequency of x is similarly obtained from the equation (1) by

$$\frac{W(y)}{\alpha} - \frac{W(z)}{\beta} = \frac{\phi}{\alpha} - \frac{\phi}{\beta} = -rx + 2rx\alpha$$

yielding

$$\hat{x} = -\frac{(\alpha - \beta)\phi}{r\alpha(2\alpha - 1)\beta} \tag{S9}$$

The equilibrium frequency of z is also calculated by substituting the equation (S8) and (S9) into $W(z)$ of the equation (1).

$$\hat{z} = \frac{\left(2\alpha^2(r + \beta - 1) - \alpha(r + \beta - 2) - \beta\right)\phi}{r\alpha(2\alpha - 1)\beta^2} \tag{S10}$$

The unknown ϕ is determined from the equation (S8), (S9), (S10) and $x + y + z = 1$.

$$\phi = \frac{r\alpha(2\alpha - 1)\beta^2}{2(\alpha - \beta)(1 - \alpha)(1 - \beta) + r\alpha(2\alpha - 1)} \tag{S11}$$

By substituting the equation (S11) into equation (S8)-(S10), we obtain the internal equilibrium.

$$\hat{x} = \beta(\beta - \alpha) / K$$
$$\hat{y} = \beta(1 - \beta)(2\alpha - 1) / K \tag{S12}$$
$$\hat{z} = \left[(1 - \alpha)\{2\alpha - \beta(1 + 2\alpha)\} + r\alpha(2\alpha - 1)\right] / K$$

where $K = 2(\alpha - \beta)(1 - \alpha)(1 - \beta) + r\alpha(2\alpha - 1)$. Now we show that if $\beta < \alpha$, there is no internal equilibrium. That $\beta < \alpha$ and $\hat{x} > 0$, $\hat{y} > 0$ implies that $K < 0$, or

$$r > \frac{(1 - \alpha)\{2\alpha - \beta(1 + 2\alpha)\}}{\alpha(1 - 2\alpha)} \tag{S13}$$

Applying $\beta < \alpha$ in the right hand side yields $r > 1$, which is simply not allowed by definition − the homing rate must be less than 1. Thus there should be no internal equilibrium if $\beta < \alpha$. On the other hand, if $\beta > \alpha$, the internal equilibrium can exist as shown in the following.

Condition for existence of internal equilibrium for β>α: Suppose now that $\beta > \alpha$. For $\hat{x}$ and $\hat{y}$ to be positive in (S12), $\alpha > 1/2$ and $K > 0$ follows. The positivity of the denominator K is equivalent to

$$r > \frac{2(\beta - \alpha)(1 - \alpha)(1 - \beta)}{\alpha(2\alpha - 1)} \equiv r^{**} \tag{S14}$$

For the positivity of $\hat{z}$ in (S12), we also have

$$r > \frac{(1-\alpha)\left\{\beta(1+2\alpha)-2\alpha\right\}}{\alpha(2\alpha-1)} \equiv r^* \qquad (S15)$$

Because $\beta > \alpha > 1/2$ in the region we are considering here, $r^* > r^{**}$ must always follow, and hence (S15) is more stringent condition for the existence of internal equilibrium than (S14). To summarize, the internal equilibrium exists if the following 3 conditions are met: i) $\beta > \alpha$ (intact homing allele is more costly than defective allele), ii) $\alpha > 1/2$ (homing allele is not too costly), and iii) the homing rate r is greater than a threshold r^* defined in (S15).

Instability of internal equilibrium and period of the limit cycle for β>α: From the eigenvalue and determinant of the Jacobian matrix of the internal equilibrium, phase diagram for its instability can be drawn. If we neglect the mutation, the determinant of Jacobian at the internal equilibrium is larger than 1, and hence the internal equilibrium is unstable, if:

$$\frac{\left\{r(2\alpha-1)+2(1-\alpha)\right\}(\beta-\alpha)(1-\beta)}{r\beta^2(2\alpha-1)} > 0 \qquad (S15)$$

As $\beta > \alpha > 1/2$ must follow for the existence of the internal equilibrium, this inequality is always satisfied, implying that the internal equilibrium is always unstable, by having a pair complex eigenvalues with their modulus greater than 1. Therefore, in the "Cycle" region of Figure 2, all equilibria including the internal equilibrium are unstable and the limit cycle occurs in the presence of mutation. The parameter dependence of the period of the limit cycle is analyzed from the eigenvalues of the Jacobian matrix of the internal equilibrium. Let us denote the right hand sides of (2), with $x = 1 - y - z$, as functions of y and z: $y^* = f(y,z)$, $z^* = g(y,z)$. Let T and D be the trace and the determinant of the Jacobian matrix at the internal equilibrium:

$$T = \left(\frac{\partial f}{\partial y} + \frac{\partial g}{\partial z}\right)(\hat{y},\hat{z}), \quad D = \left(\frac{\partial f}{\partial y}\frac{\partial g}{\partial z} - \frac{\partial f}{\partial z}\frac{\partial g}{\partial y}\right)(\hat{y},\hat{z}), \qquad (S16)$$

where $\hat{y}$ and $\hat{z}$ are defined in (S12). It can be shown, by substituting (S12) into (S16), that

$$T = \frac{(1-\alpha)(1-\beta)(2\alpha+1)(\beta-\alpha)+2r\alpha\beta(2\alpha-1)}{r\alpha\beta(2\alpha-1)}$$

$$D = \frac{2(1-\alpha)(1-\beta)(\beta-\alpha)+r(2\alpha-1)(2\alpha\beta-\alpha+\beta)}{r\beta^2(2\alpha-1)} \qquad (S17)$$

The eigenvalue λ of the Jacobian is obtained from the characteristic equation $\lambda^2 - T\lambda + D = 0$. If the eigenvalues are complex, they can be written in the polar coordinate $\lambda = \rho e^{\pm i\theta}$, with radius $\rho = \sqrt{D}$ and the angular frequency:

$$\theta = \arctan\sqrt{\frac{4D}{T^2}-1}$$
$$= \sqrt{(\beta-\alpha)(1-\beta)}+O(\varepsilon^2)$$
(S18)

In the last equation, we expanded θ in Taylor series with respect to small costs $1-\alpha$ and $1-\beta$ of H^+ and H^d, respectively, by assuming that they are of the order of ε, and neglected the terms whose orders are equal to or higher than ε^2. The approximate period $2\pi/\theta$ for the case of $\beta=\sqrt{\alpha}$ is illustrated as a function of α and r in Figure S1. The period is insensitive to the homing rate r of the homing endonuclease, and monotonically decreases with decrease of the relative fecundity of a cell carrying H^+, as long as the costs of H^+ and H^d are not very large (see the last equation of S18),

Figure S1. Parameter dependence of the period of the limit cycle. The approximate period $2\pi/\theta$ for the case of $\beta=\sqrt{\alpha}$ is illustrated as a function of α and r (the lower equation of S18). The period monotonically decreases with decrease of the relative fecundity of a cell carrying H^+.

Figure S2. Schematic diagram of limit cycle of homing endonuclease alleles in a population. The majority of the allele population at the beginning is H^- providing ample opportunity of homing, which enables spread of the H^+ allele. H^+ allele is then susceptible to degeneration while there is little opportunity for homing, which, in turn, enables spread of H^d allele. When the majority in the population becomes H^d allele, H^- is advantageous and H^d decreases in frequency. This all happens in a population without any interaction with other populations/species.

Figure S3. Parameter dependence of the limit cycle. Results obtained numerically for different parameter sets in the "Cycle" region of Figure 2: r=0.2, $0.01 \le c_1 \le 0.05$, $6E-6 \le u \le 6E-4$. The initial frequency of H^- is 99% while that of H^+ is 1%, The larger the cost, the larger the amplitude. On the other hand, the higher the mutation rate, the smaller the amplitude.

Figure S4. Evolutionary dynamics of stochastic extinction of a homing endonuclease allele after thousands of meiotic generations in a finite population for $\beta>\alpha$. An example obtained numerically for a parameter set in the "Cycle" region of Figure 2 with mutations: $\alpha=0.990$ ($c_1=0.01$), r=0.2, $u=6\times10^{-6}, v=6\times10^{-7}$, $N=10^6$. The initial frequency of H^- is 99% while that of H^+ is 1%.

Figure S5. **Phase diagram for** $\alpha > \beta$ **indicating two regions in which the H⁺ allele can dominate.** The black line obtained from equation (4) indicates that H⁻ is stable on its left. The gray line obtained from the equation (S6) of SI indicates that the H⁺ allele is stable on its right. In the light gray space between the black and gray lines, the dynamics are bistable as shown in Figure S5A. The enlargement in the upper right shows how the two boundaries shift with mutation as in Figure 3. These diagrams are illustrated with mutation rates $u = 10^{-4}, v = 10^{-5}$, which are higher than expected, for clarity.

Figure S6. **Evolutionary dynamics in which H⁺ allele can dominate when** $\alpha > \beta$. **(A) Bistability.** Results obtained numerically for a parameter set in the "bistable" region of Figure S4 with mutations: $\alpha = 0.84, \beta = \alpha^2 = 0.71$ $r=0.2$, $u = 6\times10^{-6}, v = 6\times10^{-7}$. One hundred patterns were generated from random initial frequencies of the three alleles. The black lines show the dynamics of fixation of the H⁺ allele, while the gray lines show those of fixation of H⁻. **(B) H⁺ stability.** Results obtained numerically for a parameter set in the "H⁺ stable" region of Figure S4 with mutations: $\alpha = 0.98, \beta = \alpha^2 = 0.96$,$r=0.2$, $u = 6\times10^{-6}, v = 6\times10^{-7}$. The initial frequencies of the two evolutionary dynamics illustrated here are as follows: $(x,y,z) =$ (0.1, 0.8, 0.1) and (0.45, 0.05, 0.5). The dynamics result in H⁺ predominance, irrespective of its initial frequency.

Figure S7. **Evolutionary dynamics of stochastic extinction of a homing endonuclease allele after thousands of meiotic generations in a finite population for** $\beta < \alpha$. An example obtained numerically for a parameter set: $\alpha = 0.909$ ($c_1 = 0.095$), $r=0.2$, $u = 3\times10^{-3}, v = 3\times10^{-4}$, $N = 10^3$. The initial frequency of H⁻ is 20% while that of H⁺ is 80%. After almost fixation, the homing endonuclease allele could become extinct under an unrealistically large cost, high mutation rate and small effective population size.

H$^+$
(Empty target sites)

H$^-$

Homing

Loss of
pseudogene

Cycle in a population

H^d H$^+$
(Homing endonuclease
positive allele)

H$^-$
H^d
(Pseudogene)

Degeneration
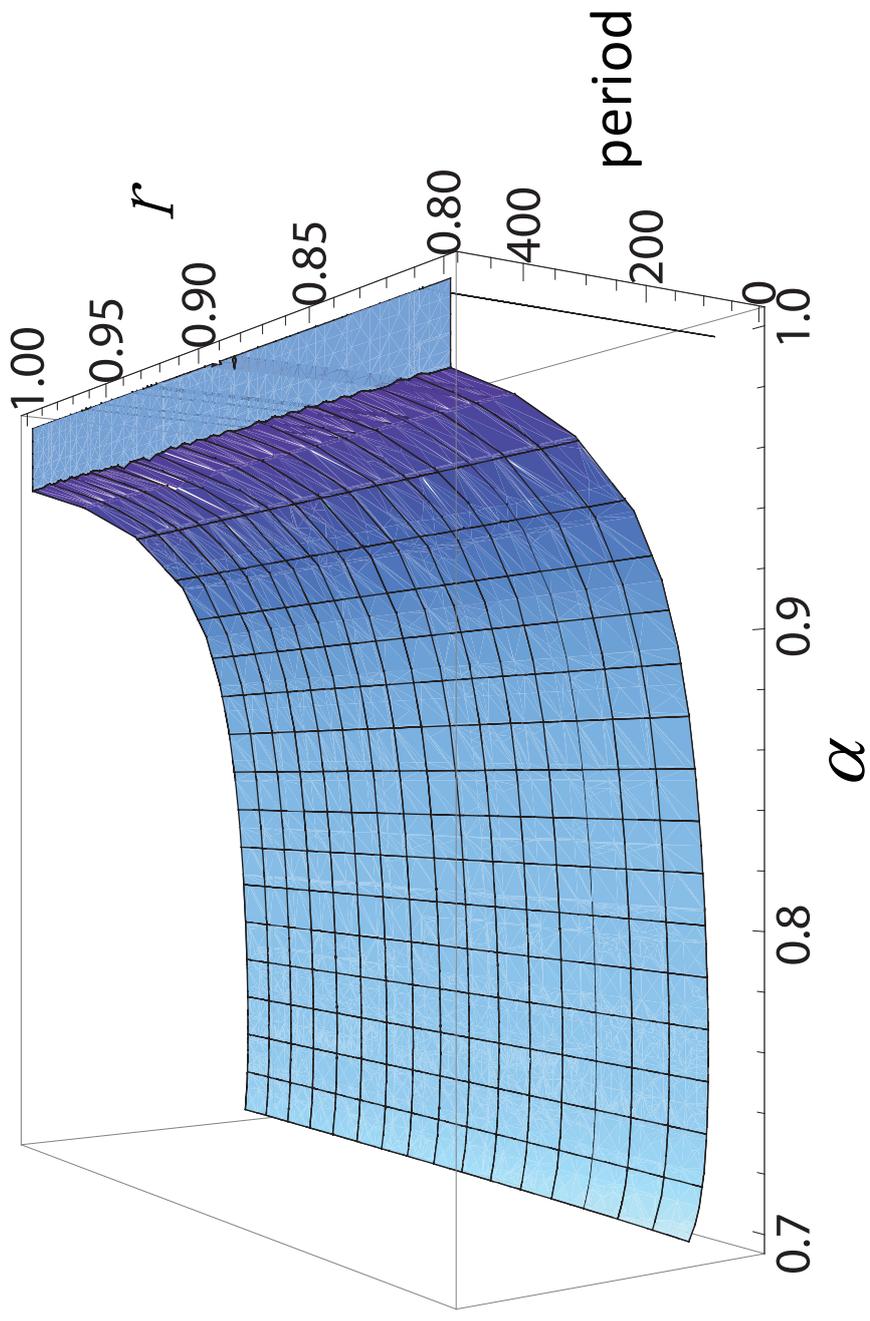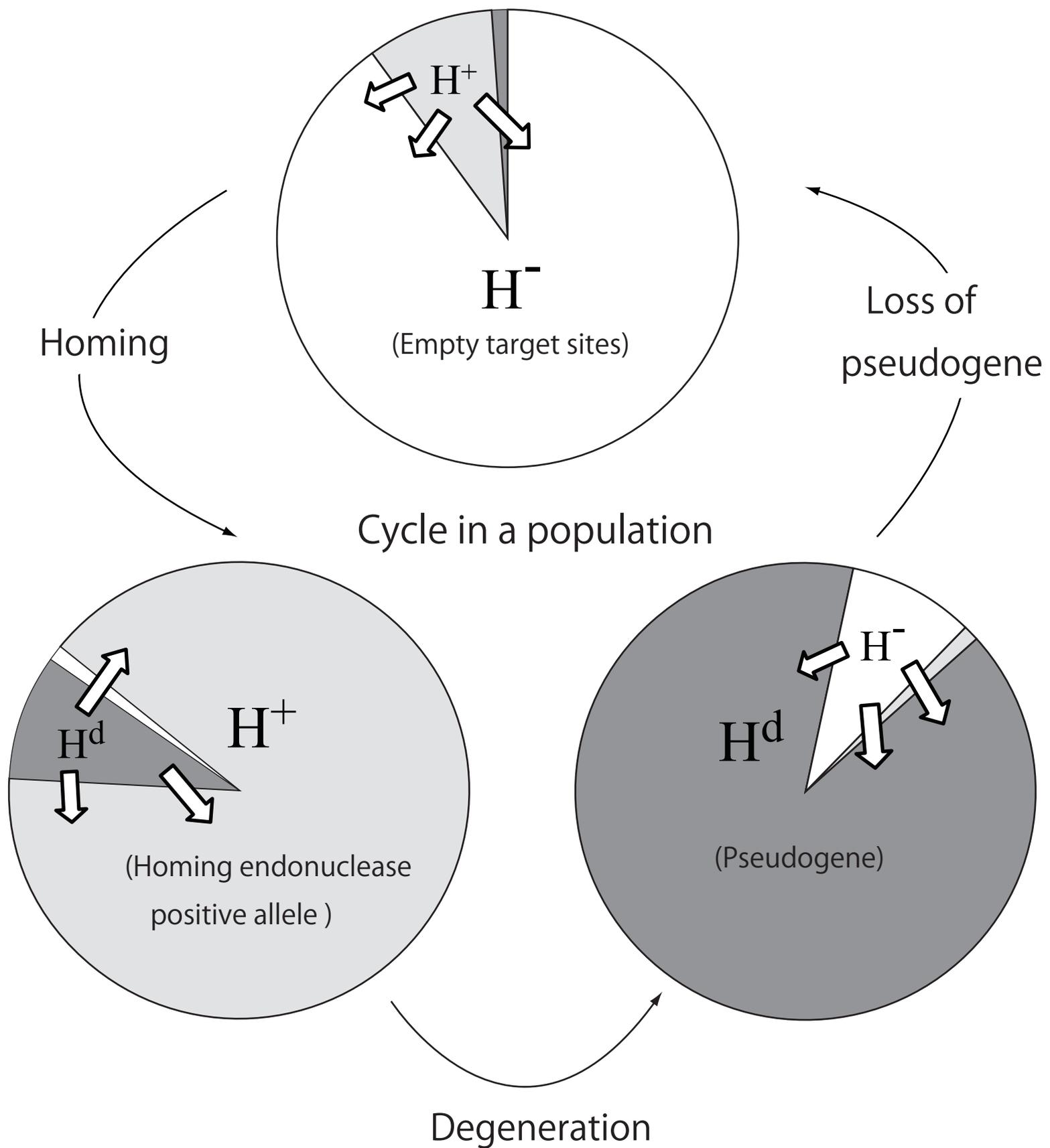
Higher mutation rate

Larger cost

Legend: H⁻, H⁺, Hᵈ

$c_1 = 0.01, u = 6E - 6$

$c_1 = 0.01, u = 6E - 5$

$c_1 = 0.01, u = 6E - 4$

$c_1 = 0.03, u = 6E - 6$

$c_1 = 0.03, u = 6E - 5$

$c_1 = 0.03, u = 6E - 4$

$c_1 = 0.05, u = 6E - 6$

$c_1 = 0.05, u = 6E - 5$

$c_1 = 0.05, u = 6E - 4$
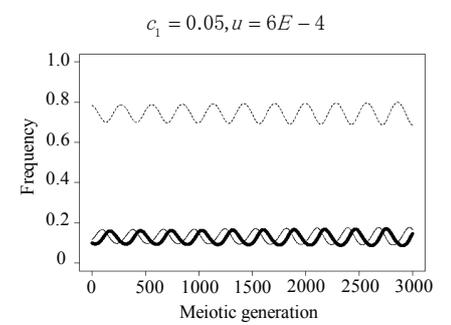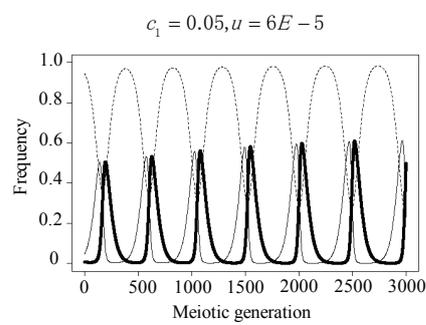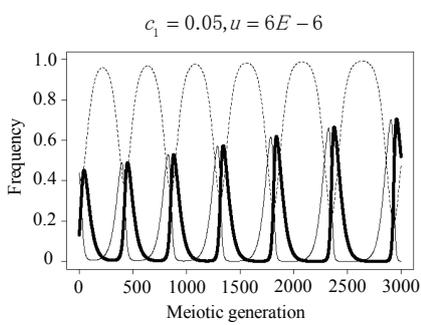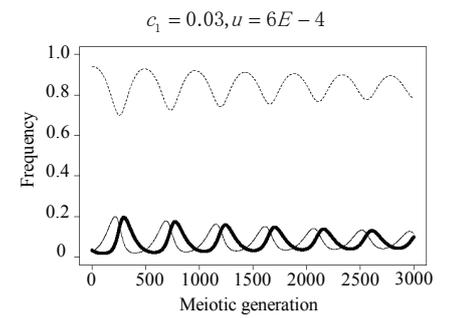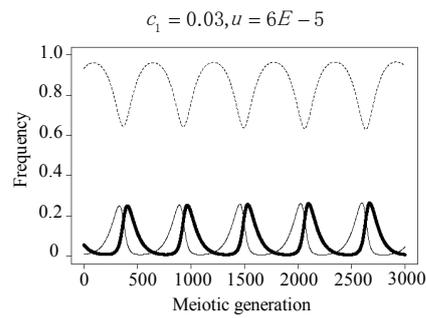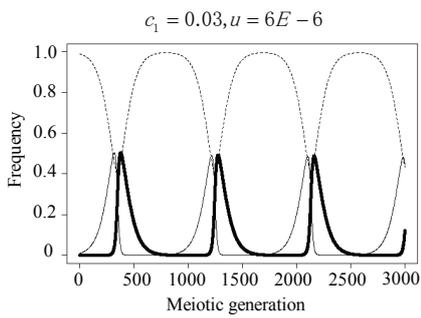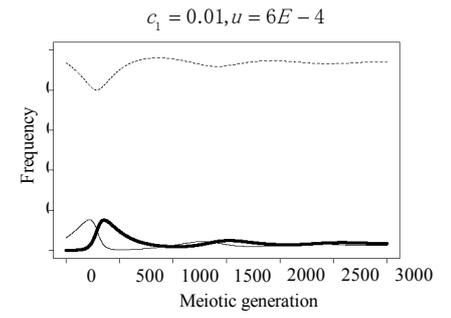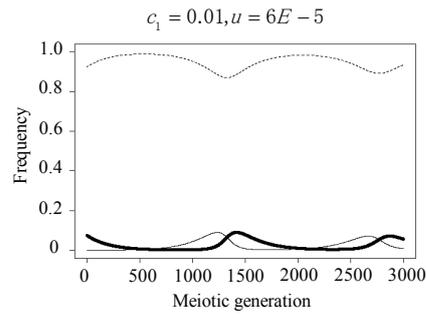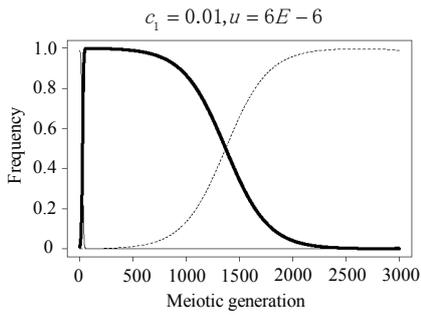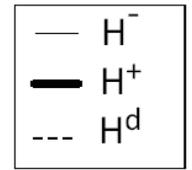
A.
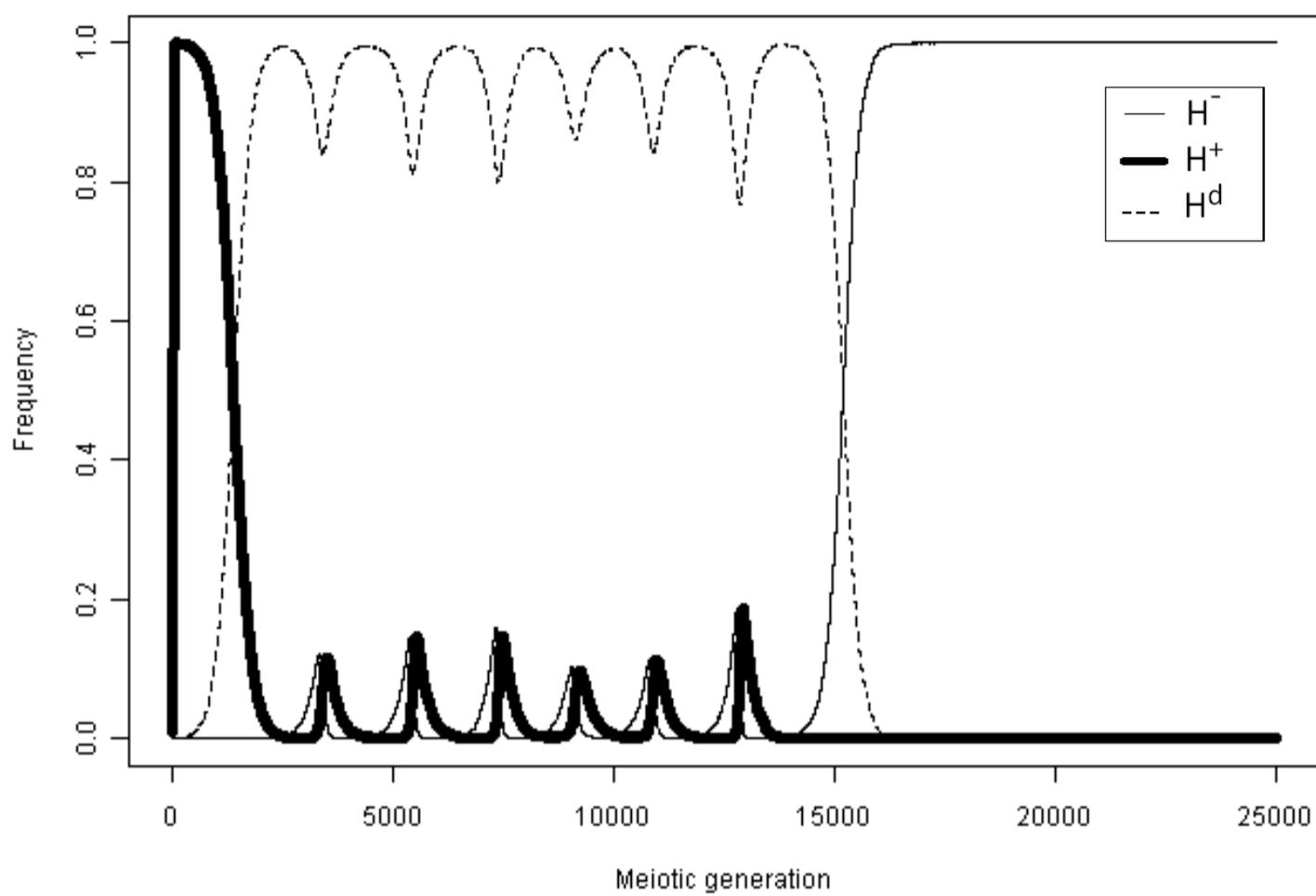
B.
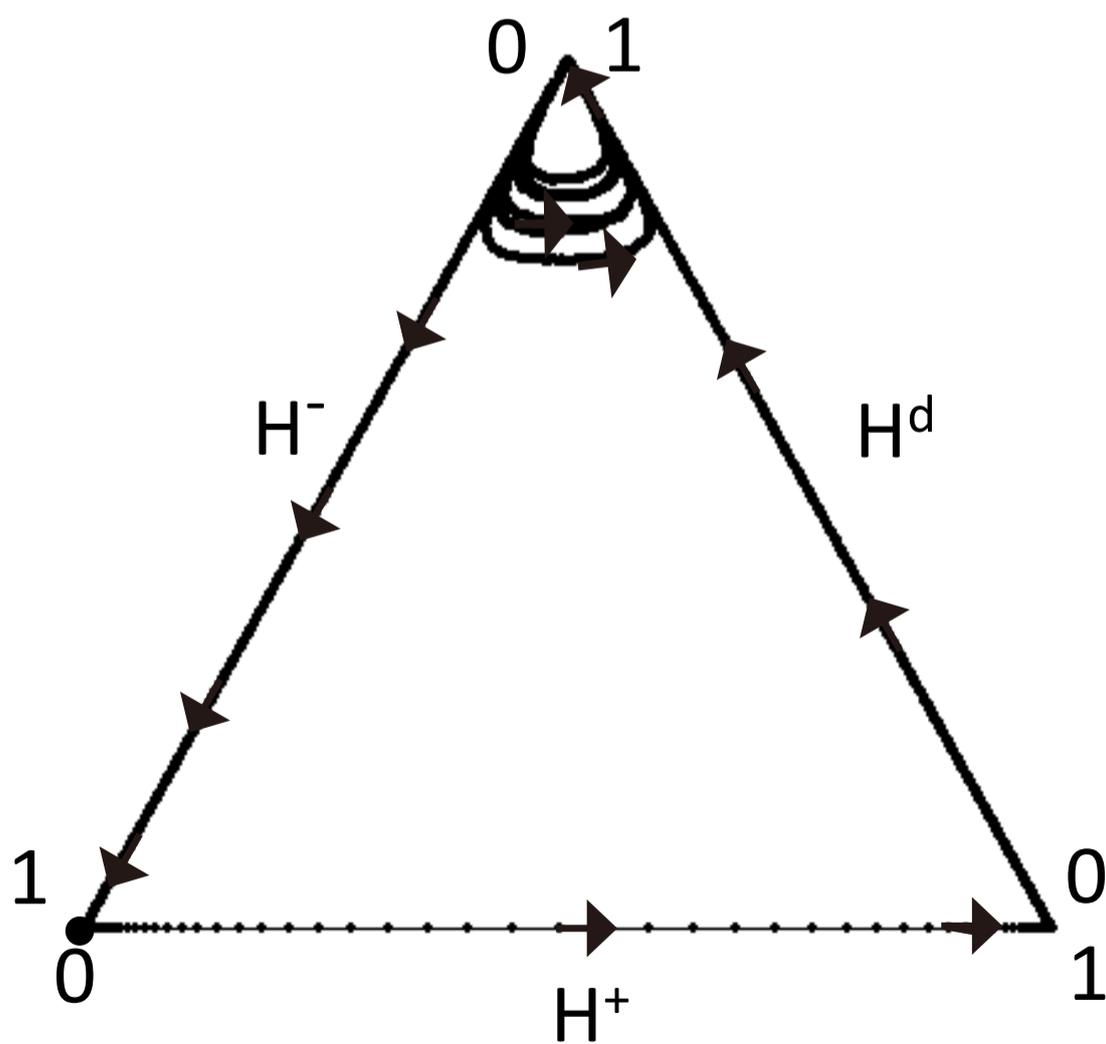
$$c_1 = 0.01, \; c_2 = \frac{c_1}{2}, u = 6E-6, \; N = 10^6$$

Figure S5

$u = 1E - 4, v = 1E - 5$

A.

B.

$$c_1 = 0.095, \ c_2 = 2c_1, \ u = 3E-3, \ N = 10^3$$