

Recurrent invasion and extinction of a selfish gene

Matthew R. Goddard* and Austin Burt

Department of Biology, Imperial College, Silwood Park, Ascot, Berkshire, SL5 7PY, United Kingdom

Edited by John Maynard Smith, University of Sussex, Brighton, United Kingdom, and approved September 9, 1999 (received for review July 6, 1999)

Homing endonuclease genes show super-Mendelian inheritance, which allows them to spread in populations even when they are of no benefit to the host organism. To test the idea that regular horizontal transmission is necessary for the long-term persistence of these genes, we surveyed 20 species of yeasts for the ω -homing endonuclease gene and associated group I intron. The status of ω could be categorized into three states (functional, nonfunctional, or absent), and status was not clustered on the host phylogeny. Moreover, the phylogeny of ω differed significantly from that of the host, strong evidence of horizontal transmission. Further analyses indicate that horizontal transmission is more common than transposition, and that it occurs preferentially between closely related species. Parsimony analysis and coalescent theory suggest that there have been 15 horizontal transmission events in the ancestry of our yeast species, through simulations indicate that this value is probably an underestimate. Overall, the data support a cyclical model of invasion, degeneration, and loss, followed by reinvasion, and each of these transitions is estimated to occur about once every 2 million years. The data are thus consistent with the idea that frequent horizontal transmission is necessary for the long-term persistence of homing endonuclease genes, and further, that this requirement limits these genes to organisms with easily accessible germ lines. The data also show that mitochondrial DNA sequences are transferred intact between yeast species; if other genes do not show such high levels of horizontal transmission, it would be due to lack of selection, rather than lack of opportunity.

Homing endonuclease genes (HEGs) are optional or nonessential genes widely distributed in fungi, protists, bacteria, and viruses (1, 2). At least among eukaryotes, where they are often found in organelles associated with group I self-splicing introns, they have no known host function. Rather, they are thought to be selfish or parasitic genes that spread in populations because their catalytic activity results in self-propagation, as they code for a sequence-specific endonuclease (i.e., a protein that cleaves DNA at a particular recognition sequence, in this case usually 12–40 bp long). In heterozygous or heteroplasmic individuals, in which there are both HEG⁺ and HEG⁻ chromosomes, the protein recognizes and cuts the HEG⁻ chromosomes; the HEG⁺ chromosomes are protected because the presence of the gene interrupts the recognition sequence. A cut chromosome turns on the cell's recombinational repair system (3), which uses the homologous chromosome (in this case HEG⁺) as a template. After repair, the HEG is found on both chromosomes, and the cell is then homozygous HEG⁺. Consequently, these genes show strong transmission ratio distortion; they are often inherited by more than 95% of progeny, rather than the Mendelian 50% (4, 5).

This process of “homing” easily explains how such a gene can increase in frequency and become fixed within a population. For example, with a transmission rate of 95%, a gene would increase its frequency from 0.001 to 0.999 in about 15 outcrossed generations. However, what then? Once fixed in the population, there will be little selection against nonsense or frameshift mutations which destroy the enzyme's ability to cut DNA; only the production of deletion mutants, in which the element is completely lost (reconstituting an intact recognition sequence), will maintain selection for endonuclease function. Moreover, if there is a cost to the host cell for producing a functional endonuclease, then natural selection will also increase the

frequency of nonfunctional elements. It is therefore not easy to see how functional HEGs can be maintained at detectable frequencies over long evolutionary time periods (assuming no host benefit). How, then, to explain the ever-increasing list (see refs. 2 and 6) of such genes?

One possible explanation is that high frequencies of functional elements are indeed transient states, but they recur because functional elements are occasionally copied to a new location with the appropriate recognition sequence, either in the same species (transposition), or in a different one (horizontal transmission). If this were to happen, the element would again go to high frequencies, before giving way to nonfunctional elements in the face of mutation pressure and/or natural selection. If such movements are sufficiently frequent (i.e., if they occur, on average, at least once before degeneration), then the element can persist over long evolutionary time scales. Comparative sequence analysis has shown that such transfers do occur at least occasionally (e.g., see refs. 7–9), but the data thus far are too fragmentary to allow one to estimate the frequency of movement and determine whether it is common enough to allow persistence. To address this question directly, we have analyzed a particular element, ω (omega, also known as r1 and Sc LSU.1), in a series of closely related host species, the saccharomycete yeasts.

An optional genetic element, ω was first described in *Saccharomyces cerevisiae* (10). It consists of an HEG inserted into a group I self-splicing intron and is found interrupting the large subunit (LSU) rRNA gene of the mitochondrion (Fig. 1A). In crosses between ω^+ and ω^- genotypes, fully 99% of the segregant progeny are ω^+ (4) (note that mitochondria in *S. cerevisiae* are biparentally inherited and recombine). A very similar element has been sequenced from the homologous position in the LSU rDNA of *Kluyveromyces thermotolerans* (11). Curiously, *Saccharomyces paradoxus* and *Kluyveromyces lactis* have only a group I intron at this site, without an HEG (12, 13). Although it has yet to be confirmed directly, these HEG⁻ introns presumably would not show biased inheritance. *Kluyveromyces marxianus* also has an HEG⁻ intron quite similar in sequence to ω in its mitochondrial ATPase subunit 9 gene (ref. 14; see GenBank accession no. U75348 for a similar element in the same gene of *K. lactis*). Yet other species are negative when probed with labeled ω sequences, indicating that either they do not contain the element or it is very different in sequence (11, 13, 15). However, there has been no attempt to determine whether this variable distribution is due solely to multiple independent losses, or to losses and gains by horizontal transmission. In this study, we have surveyed 20 species of yeasts by PCR, using primers matching the flanking LSU rRNA gene. The size of the resulting amplicon indicates whether an intron is present, and if so, whether it contains an HEG. All intron-containing amplicons were then sequenced for

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: HEG, homing endonuclease gene; LSU, large subunit.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AJ229045–AJ229076 and AJ229176).

*To whom reprint requests should be addressed. E-mail: m.goddard@ic.ac.uk.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

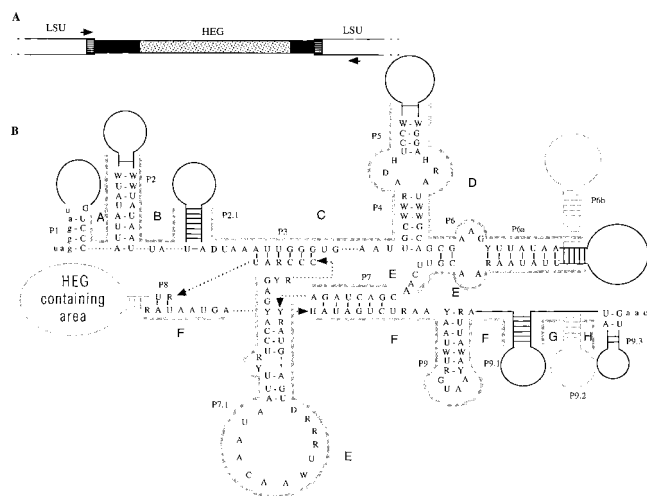


Fig. 1. (A) Structure of ω . White boxes represent the mitochondrial LSU rDNA, black boxes represent the ω group I intron, and the stippled box represents the ω HEG. The two boxes with horizontal lines represent the endonuclease recognition site, which is interrupted by the presence of ω . Also shown are the PCR primer-binding sites (not to scale). (B) Consensus primary and secondary structure of the group I introns, showing base-paired regions P1 to P9.3. If present, HEGs are found in region P8. Solid black lines represent areas of secondary structure that are present in all introns, but are not sequentially conserved. Dashed arrows connect nucleotides that have been separated for ease of display. Stem-loop structures in gray are optional or unalignable (P6b is found only in *Kluyveromyces dobzhanskii*, *K. lactis*, *Zygosaccharomyces bisporus* and *Zygosaccharomyces rouxii*; P9.2 is found only in *Saccharomyces castellii* and *S. sp.-Japan*) and were not included in the phylogenetic analysis. Areas A–H, indicated by thick gray lines, are homologous areas that were aligned independently and used in phylogenetic analysis. Exon sequences are in lowercase letters.

phylogenetic analysis, comparing the phylogeny of the introns and HEGs with that of the host yeast. Such comparisons of phylogenetic trees allow very sensitive tests for horizontal transmission (e.g., see ref. 16).

Materials and Methods

Strains. Eighteen species of yeast from four genera in the Saccharomycetaceae were obtained from the Centraalbureau Voor Schimmelcultures (CBS; Delft, The Netherlands) (see Fig. 2). In addition, two as-yet-unnamed isolates from Japan and Brazil were included in the survey. These isolates have been placed tentatively in *Saccharomyces sensu stricto*, based on karyotype analysis and their ability to form hybrids (17, 18).

Molecular Methods. DNA was extracted from overnight cultures of yeast by following the procedure of Strathern and Higgins (19). Presence or absence of the intron was assayed by PCR, with primers designed from an alignment of *S. cerevisiae* and *Pichia canadensis* mitochondrial LSU rRNA genes (Fig. 1A) (ω 01: 5'-GATAACGAATAAAAGTTACGCTAGGG-3', positions 2660–2685 in the *S. cerevisiae* gene, GenBank accession no. V00699; and ω 02: 5'-CTTCAGCAGATAGGAACCATACTG-3', positions 3983–4006). Amplicons containing introns were sequenced directly with an ABI 373 automated sequencer. *S. cerevisiae* and *K. thermotolerans* ω sequences were taken from Jacquier and Dujon (11). To reconstruct the host phylogeny, we amplified and sequenced the internal transcribed spacer region (ITS1–5.8S–ITS2) for all 20 yeast species, using primers ITS1 and ITS4 of White *et al.* (20), and supplemented them with 18S rDNA sequences taken from GenBank (21). To screen for ω elements elsewhere in the genome, total DNA from the 20 species was digested with *Hae*III, electrophoresed, transferred to

a nylon membrane, and probed with the ω intron sequences from *S. paradoxus* and *K. lactis*. The resulting autoradiograms were examined and compared with a positive control that had been probed with 200 bp of the 3' flanking mitochondrial LSU gene.

Analyses. DNA sequences were aligned by using CLUSTALW (22). Initial alignments of complete intron sequences were poor, and so secondary structures were obtained by using those of Cech (23) as a guide and by analyzing the introns in sections with the *m-fold* (v2.3) server (<http://www.ibt.wustl.edu/~zucker/rna>). Folding temperatures were set to 25°C and all other parameters were kept as default. These secondary structures were used to define eight homologous regions (A–H in Fig. 1B), each of which was aligned independently. These areas exclude unalignable loops and optional stem-loop regions. We also excluded one side of each base-paired region from the final data matrix, because changes on one side are unlikely to be independent of changes on the other.

For the ITS1–5.8S–ITS2 sequences, only the 5.8S rRNA gene and a small part of the ITS2 were alignable across all 20 species. Complete 18S rDNA sequences were available for all species except the unnamed ones from Brazil and Japan; in most cases (12/18), the 18S data were from the same strain as that included in our study. Separate analysis of the 5.8S and 18S data showed no conflict between the two (i.e., no branch with a greater than 50% bootstrap score for one gene was contradicted by the other); thus, they were combined. This combined analysis left the closely related *Saccharomyces sensu stricto* group unresolved, because there were too few variable sites; therefore, we analyzed these species independently, using the whole of the ITS1–5.8S–ITS2 region. Finally, HEGs were aligned by using amino acid translations from the yeast mitochondrial genetic code, with start codons inferred from those of *S. cerevisiae* and *K. thermotolerans* (11). All alignments are available on request from the authors. Phylogenetic analyses were performed with MACCLADE (v. 3.07; ref. 24), PAUP* (v. 4.0d64; ref. 25), and MATHEMATICA (v. 3.0.1; ref. 26).

Results and Discussion

Of the 20 species surveyed, 14 had an intron; of these, only 5 had an HEG (Fig. 2). Two of the HEGs had insertions disrupting the reading frame and are presumably nonfunctional, whereas the other 3 have no such obvious defect and are presumed to be functional. Thus, the most common state is a nonfunctional element [either nonfunctional HEG or no HEG (11 species)]; the next most common state is no element (6 species); and the least common is a putatively functional element (3 species). To test whether ω occurs in other genomic locations in these species, genomic DNA for the 20 species was probed with the intron sequences from *S. paradoxus* and *K. lactis*. All 14 species that were expected to show a strong hybridization signal did so; extra bands were not observed for any species except *K. lactis* and the closely related *K. dobzhanskii*, which presumably corresponds to the HEG⁻ introns in their ATPase genes (see Introduction). Thus, ω usually exists at only one site in the genome, and transposition to a new location does not appear to be a common event.

When the different intron states are mapped onto the host phylogeny, they are found to be intermingled (Fig. 2), which suggests a history of horizontal transmission. The alternative hypothesis is that a complete ω element was present in the common ancestor, and then lost many times. Once lost from a lineage, it would be absent from all descendent taxa; therefore, this hypothesis predicts that species without an HEG or without an intron would tend to be clumped on the host phylogeny. However, no evidence of that was found: the inferred minimum number of losses (at least nine losses of the HEG sequence (including losses of the whole intron) and four losses of the

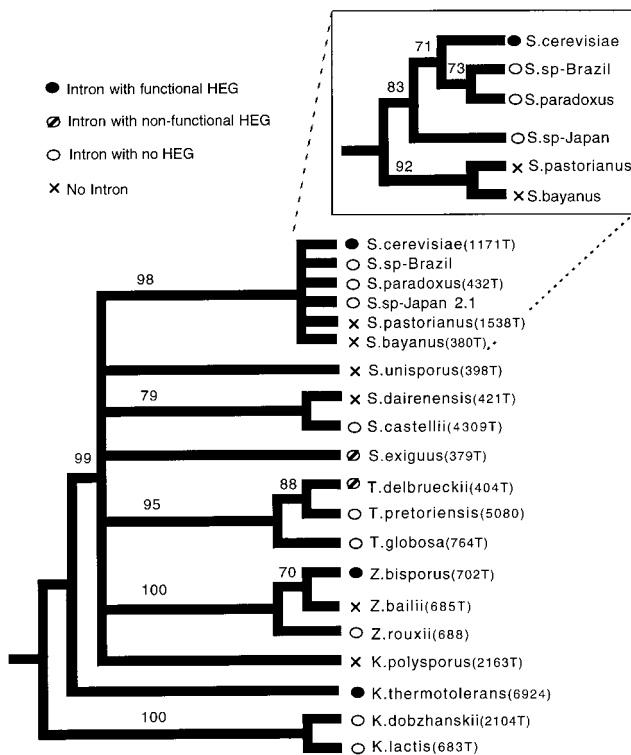


Fig. 2. Intron status and phylogenetic relationships of yeasts. Bootstrap consensus tree (1,000 replicates) obtained from a maximum parsimony branch-and-bound analysis of 18S + 5.8S data, with gaps scored as fifth bases. The bootstrap score for each branch is shown. Likelihood analyses gave the same topology. The root is inferred from a larger analysis of 18S rDNA of hemiascomycete yeasts (ref. 20 and unpublished observations). (Inset) A separate analysis of the *Saccharomyces sensu stricto*, using complete internal transcribed spacer sequence data and a parsimony branch-and-bound bootstrap analysis (1,000 replicates). Numbers in parentheses identify the Centraalbureau Voor Schimmelcultures cultures; the genus "T." is *Torulaspota*. T indicates the type strain.

intron) is no less than that observed when HEG and intron status is randomized on the phylogeny ($P = 0.82$ and $P = 0.19$ respectively, $n = 100$ randomizations).

A much more robust assay for horizontal transmission is to test whether ω has a different phylogenetic history than that of the host. An appropriate statistical test for differences in phylogenetic history is found in the partition homogeneity test (PHT; see Farris *et al.*, ref. 27), in which the sum of the lengths of the most parsimonious trees fitted to the two datasets independently is compared with the sum of the lengths of trees fitted to random partitions of the same data. First, we compared the two components of ω , the HEG and the intron. They do not appear to have different histories; the intron phylogeny is simply a less well resolved version of the HEG phylogeny (Fig. 3A). Moreover, the HEG and intron datasets do not differ significantly by the PHT ($P = 0.38$, Fig. 3A). Both of them show substantial differences in topology from that of the host tree, however, and the datasets differ significantly by the PHT ($P = 0.004$ and $P < 0.001$ respectively, Fig. 3B and C). Such phylogenetic incongruence is strong evidence for horizontal transmission.

Such analyses alone do not indicate the phylogenetic relationships of donor and recipient species; for example, it is not clear from these results whether the transfers have been among different members of the Saccharomycetaceae or whether the family has been repeatedly exchanging elements with more distantly related taxa. To address this issue, we searched Gen-

Bank with BLASTN (v. 2.0.8; ref. 28) and each of our five HEGs after they had been deposited (release 111.0). In each case we retrieved only our own sequences plus an HEG found at the homologous site of a more distantly related saccharomycete yeast, *P. canadensis*. All other HEGs [including many at other locations in our yeasts (6)], are too divergent at the nucleotide level to be recovered by BLASTN. That is, despite the horizontal transmissions, the saccharomycete LSU rDNA HEGs appear to be monophyletic with respect to elements in more distantly related species, suggesting that transfers of HEGs among the LSU rRNA genes of saccharomycete yeasts have been more common than transfers involving either more distantly related species or other genomic sites within these species. Even within the Saccharomycetaceae, there are suggestions that horizontal transmission is more likely between closely related species, as the *P. canadensis* HEG sequence retrieved from GenBank is the most divergent of all the saccharomycete sequences. Moreover, the deepest branch in our host tree (separating *K. lactis* and *K. dozhanskii* from all others) is also present in the intron tree (Fig. 3C), though without a root we cannot say whether it is also the deepest branch. Similar BLASTN searches with our 14 intron sequences recovered group I introns from a number of genes and taxa, but none scored higher than our saccharomycete LSU introns, with the occasional exception of the intron in the *K. lactis* ATPase gene. Interestingly, the insertion site for this latter element is quite different from the 18-bp recognition sequence used by ω in *S. cerevisiae* (29). Perhaps the transposition event was associated with a change of HEG.

Having demonstrated that horizontal transmission has occurred among yeasts, we then wanted to estimate its frequency. One approach, calculating how many horizontal transmission events would be necessary to account for the difference between the host and intron phylogenies, is equivalent to asking how many branches of the host tree would have to be cut off and reconnected in a new position to turn the host tree into the intron tree (30). Using the subtree-pruning-reconnection (SPR) routine of PAUP*, we generated all of the 462 one-step rearrangements of the 14-taxon host tree (Fig. 3C) and tested, by likelihood analysis, how well each one fit the intron data. (In general, for n taxa and a binary tree, PAUP* will generate $4n^2 - 26n + 42$ rearrangements, although these must be filtered by using information about the root to exclude impossible transfers between ancestor and descendent taxa.) The best tree and all other trees that were not significantly worse by the Kishino-Hasegawa (31) test at $P = 0.5$ were then used as the starting trees for another round of rearrangement. This procedure was continued until one of the three most likely intron trees was recovered. Five rearrangements were necessary, although after three rearrangements we had arrived at a tree that was not significantly worse than the most likely trees (Table 1). Thus, our best estimate from this analysis is five horizontal transfer events, with a lower bound of three. Berbee and Taylor (32) estimate, from 18S rDNA sequence divergence calibrated with the fossil record, that *S. cerevisiae* and *K. lactis* diverged approximately 70 million year ago. From this date and the maximum-parsimony host phylogeny with maximum-likelihood branch lengths, we estimate that the total amount of time in the host phylogeny is about 420 million years, giving a horizontal transmission rate of about 0.01 per million years.

This rate is likely to be an underestimate of the true value for two reasons. First, not all horizontal transmission events will be phylogenetically detectable (33). For example, when the source and recipient species are closest relatives, the event will not change the topology (although in principle it might still be detectable by changes in branch length). Coalescent modeling suggests that for 14 taxa, we might miss 2/3 of the horizontal transfers (calculated by using equation 17 of ref. 33), raising our best estimate to 15 transfers and a rate of about 0.04 per million

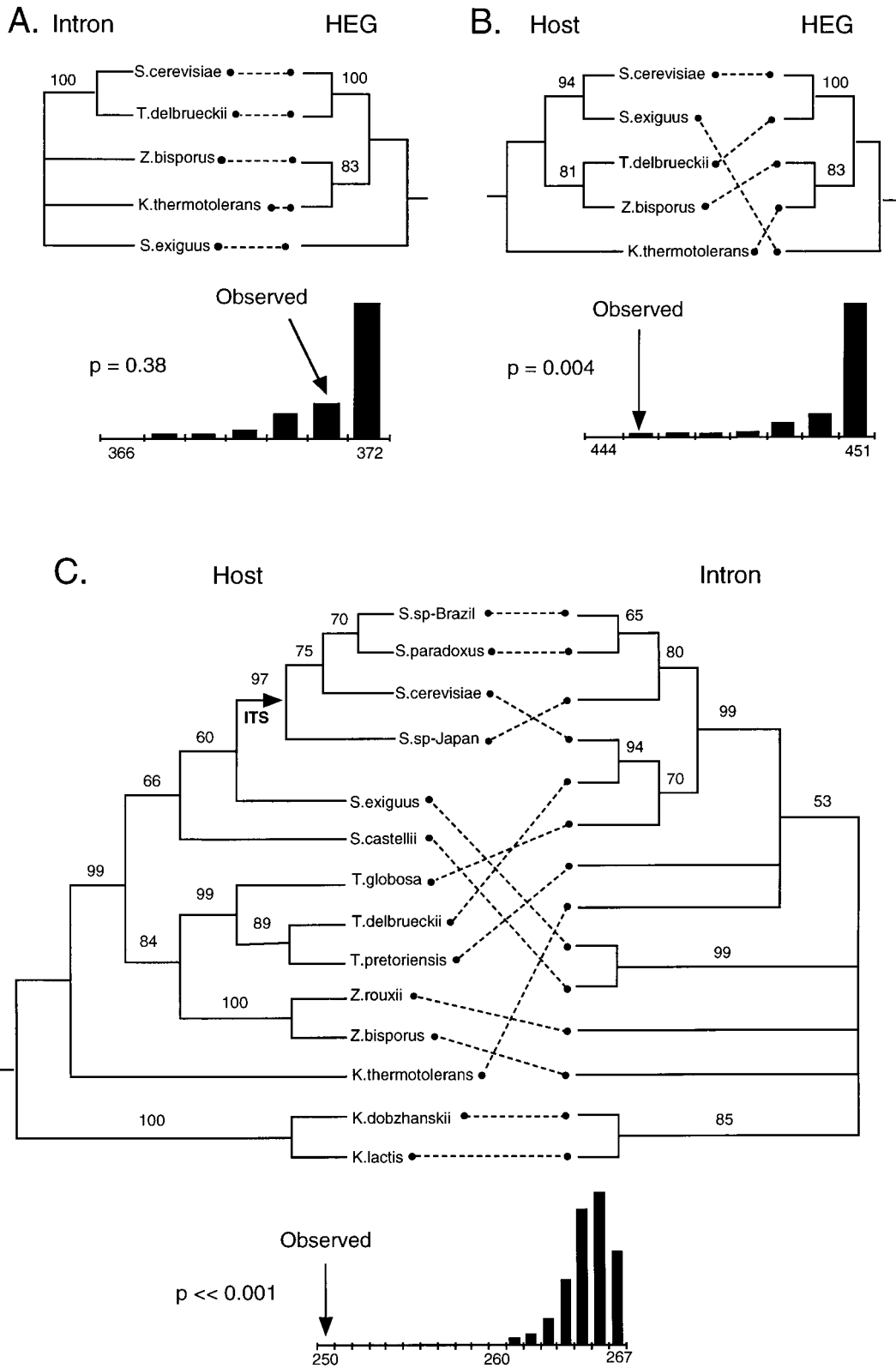


Fig. 3. Phylogenetic comparisons between intron, HEG, and host. The histogram associated with each tanglegram shows the distribution of summed tree lengths for random partitions of the data; arrows show the length of the actual partition.

years. This correction is likely to be conservative, because it presupposes that the probability of transfer is independent of phylogenetic distance; if, as seems likely, transfers are more

common between closely related species (above), the proportion of transfers that are missed can be much greater. Second, our estimate is from the minimum number of rearrangements nec-

Table 1. Results of a heuristic search for the minimum number of horizontal transmissions necessary to reconcile the host and intron trees

No. horizontal transfers	Ln likelihood*	Significantly worse than intron tree†	No. trees‡
0 (Host tree)	-458.8	Yes ($P < 0.0001$)	NA
1	-429.3	Yes ($P < 0.0001$)	4
2	-405.5	Yes ($P = 0.04$)	3
3	-397.6	No ($P = 0.22$)	18
4	-391.9	No ($P = 0.83$)	16
5 (Intron tree)	-390.8	No ($P = 1$)	170

Results of an algorithm to find the minimum number of rearrangements necessary to change the host phylogeny (Fig. 3C) into one of the three most likely intron phylogenies. The latter are given by the three possible resolutions of the following consensus tree: (((((((S. paradoxus, Brazil), Japan), ((S. cerevisiae, T. delbrueckii), T. globosa)), K. thermotolerans, (S. exiguus, S. castellii)), T. pretoriensis), Z. bisporus), Z. rouxii), (K. dobzhanskii, K. lactis)). Likelihoods were calculated by using the HKY model with the transition-to-transversion ratio and proportion of invariant sites estimated from the data. *Ln likelihood of the most likely tree.

†By the Kishino–Hasegawa test (see ref. 30).

‡Including most likely tree and all others not significantly worse (at $P = 0.5$); used as starting trees in the next iteration. NA, not applicable.

essary to transform one phylogeny into another, and if there has been the equivalent of reversals or parallelisms, this minimum estimate will underestimate the true number (just as parsimony analysis is likely to underestimate the true number of character-state changes in a phylogeny). Two rooted trees cannot differ by more than $n - 2$ rearrangements ($n - 3$ for unrooted trees), meaning that even if horizontal transmission events were infinitely frequent, we could not estimate more than 12 events by this method. Moreover, we have simulated an infinite frequency of horizontal transfers by randomizing the association between host and intron, and in four analyses, the values obtained were 7(3), 6(4), 6(5), and 7(4) events (lower bounds in parentheses). These results suggest that our observed value of 5(3) is indeed close to saturation. We also performed an alternative test for correlation of phylogenetic signal between host and intron datasets, which compares the length of the shortest tree fitted to the combined dataset to the length of trees fitted to combined datasets in which the host–intron association has been randomized (34). This test showed a highly significant correlation between the two datasets ($P < 0.001$). Such correlations can arise either because horizontal transmission is not frequent enough to completely randomize the host–intron association, or because it tends to be local (i.e., more common among closely related hosts; see above).

An alternative approach to quantifying horizontal transmission rates is to combine our demonstration of horizontal transmission with the laboratory observations and population genetic reasoning outlined earlier into a cyclical evolutionary model. In this model there are three character states of interest (functional element, nonfunctional element, and no element), and only three of the six conceivable character transitions are evolutionarily possible: empty \rightarrow functional (horizontal transmission); functional \rightarrow nonfunctional (degeneration by mutation pressure and/or costs of enzyme production); and nonfunctional \rightarrow empty. This last step requires a precise excision of the element to reconstitute the recognition sequence; such mutations have been observed for some mitochondrial HEGs and introns and are thought to arise by reverse transcription of spliced RNAs (10). Armed with this model, we can estimate the maximum likelihood transition rates among the three states by following Pagel (35), using the observed character states and the host tree

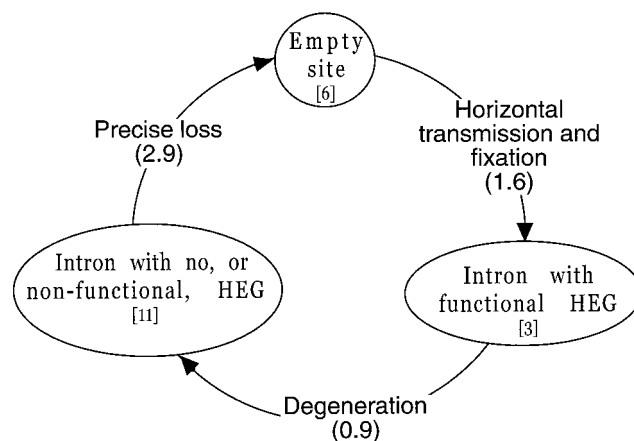


Fig. 4. Cyclical model of ω gain and loss. Numbers in brackets indicate the number of taxa, of the 20 surveyed, with that intron state; numbers in parentheses are maximum likelihood average waiting times for changing from one state to the next (millions of years), calculated by mapping the character states onto the following most parsimonious tree, with branch lengths estimated by maximum likelihood, forcing a molecular clock: (((((((Brazil: 0.00007, S. paradoxus: 0.00007): 0.00034, S. cerevisiae: 0.00041): 0.00042, Japan: 0.00083): 0.00015, (S. bayanus: 0.00009, S. pastorianus: 0.00009): 0.00089): 0.010178, (S. unisporus: 0.009518, S. exiguus: 0.009518): 0.001636): 0.000697, (S. dairensis: 0.004508, S. castellii: 0.004508): 0.007343): 0.001739, (((T. pretoriensis: 0.000591, T. delbrueckii: 0.000591): 0.001229, T. globosa: 0.001820): 0.009593, (Z. rouxii: 0.007138, (Z. baillii: 0.003529, Z. bisporus: 0.003529): 0.003609): 0.004275): 0.001462, K. polysporus: 0.012875): 0.000716): 0.004425, K. thermotolerans: 0.018016): 0.009386, (K. dobzhanskii: 0.002081, K. lactis: 0.002081): 0.025320): 0.000000.

for all 20 species with branch lengths estimated by maximum likelihood (Fig. 4; MATHEMATICA file available on request). The fit of this three-parameter model is much better than a two-parameter model with no horizontal transmission ($\Delta \ln \text{lik} = 13.2$; $\text{A}df = 1$), again supporting our conclusion that horizontal transmission does indeed occur. Moreover, the fit of the three-parameter model is no better than that of a one-parameter model in which the three transition rates are set equal to one another ($\Delta \ln \text{lik} = 2.0$; $\text{A}df = 2$), indicating no significant differences among the three rates. Under this latter model, the maximum likelihood estimate for any one of the transition rates is 0.53 per million years, with two-unit support limits of 0.07 and infinity (assuming the host tree is correct). That is, a complete circuit around the cycle is expected to take 5.7 million years, with a lower bound of 0 and an upper bound of 42 million years.

Conclusions

Our survey and phylogenetic analyses indicate that the evolutionary biology of ω is highly dynamic. Within any particular host lineage, there appears to be a perpetual cycle of invasion by horizontal transmission, degeneration, and eventual loss, followed by reinvasion. This picture matches expectations from population genetic reasoning, as selection for HEG function will be minimal or nonexistent once it has gone to fixation (assuming there is no benefit to the host). This pattern of rapid spread through a population followed by slow degeneration may be a common property of selfish genetic elements, and in such cases, horizontal transmission may be necessary for their long-term persistence (36, 37). Indeed, we speculate that a requirement for a minimum frequency of horizontal transmission may be a critical factor determining the distribution and abundance of HEGs, perhaps explaining why, among eukaryotes, they are mostly found in fungi and protists. We suggest they are largely absent from animals because access to the germ line is too limited, at least for chromosomally integrated DNA.

Our attempts to quantify the frequency of horizontal transmission suggest that it is occurring at least on a time scale of 10^6 to 10^7 years, and perhaps much faster. The methods used to derive this quantitative estimate of the frequency of horizontal transmission should be widely applicable. These analyses also indicate that ω is more likely to move by horizontal transmission to the mitochondrial LSU of a new species than it is to transpose to a new site in the same species. Presumably the latter is rare because, for homing to be effective, a simultaneous change in the enzyme recognition sequence would often be required. In addition, horizontal transmission is apparently more common between closely related species than between more distantly related ones. The relatively high frequency of horizontal transmission observed for ω is re-

markable, because it is a selfish gene without an extrachromosomal step in its life cycle (unlike, say, transposable elements). It indicates that DNA sequences do get moved, intact, from one species to another. To the extent that normal host-benefiting genes do not show horizontal transmission (still an open question in yeasts), the cause would be a lack of positive selection, rather than a lack of opportunity.

We are grateful to E. Louis for supplying the Japanese and Brazilian isolates; we thank D. Swofford for allowing us to use test versions of PAUP*, and T. Barraclough, V. Koufopanou, M. Thomas, and M. Tristem for useful comments on a previous draft. This work was funded by a Natural Environment Research Council (London) studentship to M.G. and grant (GR3/10626) to A.B.

1. Mueller, J., Bryk, M., Loizos, N. & Belfort, M. (1993) in *Nucleases*, eds. Linn, S. M., Lloyd, R. S. & Roberts, R. J. (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed., Vol. 2, pp. 111–143.
2. Belfort, M. & Roberts, R. J. (1997) *Nucleic Acids Res.* **25**, 3379–3388.
3. Orr-Weaver, T. L. & Szostak, J. W. (1985) *Microbiol. Rev.* **49**, 33–58.
4. Jacquier, A. & Dujon, B. (1985) *Cell* **41**, 383–394.
5. Gimble, F. S. & Thorner, J. (1992) *Nature (London)* **357**, 301–305.
6. Dalgaard, J. Z., Klar, A. J., Moser, M. J., Holley, W. R., Chatterjee, A. & Mian, I. S. (1997) *Nucleic Acids Res.* **25**, 4626–4638.
7. Paquin, B., Laforest, M.-J. & Lang, B. F. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 11807–11810.
8. Turmel, M., Côté, V., Otis, C., Mercier, J.-P., Gray, M. W., Lonergan, K. M. & Lemieux, C. (1995) *Mol. Biol. Evol.* **12**, 533–545.
9. Vaughn, J., Mason, M., Sper-Whitis, G., Kuhlman, P. & Palmer, J. (1995) *J. Mol. Evol.* **41**, 563–572.
10. Dujon (1989) *Gene* **82**, 91–114.
11. Jacquier, A. & Dujon, B. (1983) *Mol. Gen. Genet.* **192**, 487–499.
12. Ragnini, A., Grisanti, P., Rinaldi, T., Frontali, L. & Palleschi, C. (1991) *Curr. Genet.* **19**, 169–174.
13. Wilson, C. & Fukuhara, H. (1991) *Curr. Genet.* **19**, 163–167.
14. Dujon, B., Colleaux, L., Jacquier, A., Michel, F. & Monteilhet, C. (1986) in *Extrachromosomal Elements in Lower Eukaryotes*, eds. Wickner, R. B., Hinnebusch, A., Lambowitz, A. M., Gunsalus, I. C. & Hollaender, A. (Plenum, New York).
15. Skelly, P. J. & Maleszka, R. (1991) *Curr. Genet.* **19**, 89–94.
16. Page, R. (1994) *Cladistics* **10**, 155–173.
17. Naumov, G. I., Naumova, E. S. & Louis, E. J. (1995) *J. Gen. Appl. Microbiol.* **41**, 499–505.
18. Naumov, G. I., Naumova, E. S., Hagler, A. N., Mendonca-Hagler, L. C. & Louis, E. (1995) *Antonie Van Leeuwenhoek* **67**, 351–355.
19. Strathern, J. & Higgins, D. (1991) in *Guide to Yeast Genetics and Molecular Biology*, eds. Guthrie, C. & Fink, G. (Academic, San Diego), pp. 322–323.
20. White, T. J., Bruns, T., Lee, S. & Taylor, J. (1990) in *PCR Protocols: A Guide to Methods and Applications*, eds. Innes, M. A., Gelfand, D. H., Sninsky, J. J. & White, T. J. (Academic, San Diego), pp. 315–322.
21. James, S. A., Cai, J., Roberts, I. N. & Collins, M. (1997) *Int. J. Syst. Bacteriol.* **47**, 453–460.
22. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673–4680.
23. Cech, T. R. (1988) *Gene* **73**, 259–271.
24. Maddison, W. P. & Maddison, D. R. (1992) *MACCLADE (Sinauer, Sunderland, MA)*, Version 3.07.
25. Swofford, D. L. (1999) *PAUP* Phylogenetic Analysis Using Parsimony (Sinauer, Sunderland, MA)*, Version 4.
26. Wolfram, S. (1996) *The Mathematica Book* (Wolfram Media, Champaign, IL).
27. Farris, J. S., Källersjö, M., Kluge, A. G. & Bult, C. (1995) *Cladistics* **10**, 315–319.
28. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Miller, W. & Lipman, D. J. (1997) *Nucleic Acids Res.* **25**, 3389–3402.
29. Colleaux, L., D'Auriol, L., Glabert, F. & Dujon, B. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6022–6026.
30. Hein, J. (1993) *J. Mol. Evol.* **36**, 396–405.
31. Kishino, H. & Hasegawa, M. (1989) *J. Mol. Evol.* **29**, 170–179.
32. Berbee, M. L. & Taylor, J. W. (1993) *Can. J. Bot.* **71**, 1114–1127.
33. Hudson, R. R. & Kaplan, N. L. (1985) *Genetics* **111**, 147–164.
34. Koufopanou, V., Burt, A. & Taylor, J. W. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 5478–5482.
35. Pagel, M. (1994) *Proc. R. Soc. London B* **255**, 37–45.
36. Lohe, A. R., Moriyama, E. N., Lidholm, D.-A. & Hartl, D. L. (1995) *Mol. Biol. Evol.* **12**, 62–72.
37. Hurst, L. D. & McVean, G. T. (1996) *Proc. R. Soc. London B* **263**, 97–104.