Update TRENDS in Ecology and Evolution Vol.22 No.5

Research Focus

A short history of recombination in yeast

Clifford W. Zeyl[1*] and Sarah P. Otto[2*]

[1] Department of Biology, Wake Forest University, PO Box 7325, Winston-Salem, NC 27109, USA
[2] Department of Zoology, University of British Columbia, Vancouver, BC, V6T 1Z4, Canada

Despite it being the darling of fungal genomics, we know little about either the ecology or reproductive biology of the budding yeast, *Saccharomyces cerevisiae*, in nature. A recent study by Ruderfer *et al.* estimated that the ancestors of three *S. cerevisiae* genomes outcrossed approximately once every 50 000 generations, confirming the view that outcrossing is infrequent in natural populations of *S. cerevisiae*. This study also inferred the genomic positions of past recombination events. By comparing past recombination events with present-day recombination rates, this study lays the groundwork for determining whether recombination has improved the long-term survival of descendant lineages by bringing together favorable alleles, a longstanding question in evolutionary genetics.

Recombination

The ubiquity of sex and recombination is one of the most intensively pursued enigmas in evolutionary biology. One explanation, the Fisher–Muller hypothesis [1–4], is that sex and recombination bring adaptive mutations that have arisen in separate lineages together into the same lineage. This advantage can indirectly select for genes that promote sex and recombination. Evidence for this advantage is provided by artificial selection experiments in which recombination rates have increased as a correlated response to selection on other traits [1], as well as by experiments demonstrating that sex and recombination hasten the adaptive process (e.g. [5,6]).

Similarly, recombination is advantageous near a locus where selection maintains a polymorphism ('balancing selection'), because recombination enables beneficial alleles at neighboring loci to spread to chromosomes carrying the different alleles maintained by balancing selection [7]. If recombination brings together a favorable gene combination via either new adaptive alleles or balancing selection, the descendant lineage is more likely to persist, bearing within it the signature of this recombination event. In a recent paper, Ruderfer and colleagues [8] present the results of a study of historical poutcrossing events and infer the genomic positions of previous recombination events in the yeast *Saccharomyces cerevisiae*; up until now, it has been impossible to survey past recombination events to assess whether recombination has helped a lineage to persist.

A mystery in a genetic model

Saccharomyces cerevisiae has been studied for decades, and no eukaryote has yielded more insight into genetics. However, little is known about recombination or outcrossing in natural yeast populations. In the lab, yeast reproduce primarily by budding following a mitotic cell division. Nitrogen starvation and an absence of fermentable sugar are needed to induce meiosis and sporulation in *S. cerevisiae*. The four haploid products of a meiotic division, a tetrad, are enclosed in a sac-like capsule called an ascus. These spores germinate when favorable conditions return.

Mating occurs whenever haploid cells of opposite mating types, MATa and MATα, encounter one another, and is often between closely related cells for two reasons: (i) the close physical proximity of cells of opposite mating type from the same ascus; and (ii) homothallism, the ability of haploid cells of one mating type to produce daughter cells of the opposite mating type. Recombination among inbred yeast leaves no signature, because there is no variation to shuffle between homologous chromosomes. The extent to which yeast lineages have historically outcrossed and recombined in nature has, until now, remained a mystery.

Counting recombination breakpoints

One consequence of outcrossing is that different genes in a recombining genome have different genealogical histories, making phylogenetic inference within sexual genomes problematic. Ruderfer *et al.* [8] turned this problem into an opportunity by searching out those genomic locations where there was a switch in support from one genealogical tree to another, inferring that there must have been a recombination breakpoint in between. The authors aligned the sequences of three *S. cerevisiae* genomes along with a sequence from a closely related species, *S. paradoxus*. Three genealogical relationships are possible among the three *S. cerevisiae* strains (Box 1). In a preliminary analysis, Ruderfer *et al.* used maximum parsimony (Box 1) to determine which tree was supported at over 25 000 nucleotide sites that varied among the three genomes (single nucleotide polymorphisms, or SNPs). The locations of recombination breakpoints were inferred as sites where a consecutive string of SNPs supporting one tree ended and a string supporting a different tree began.

The median distance between recombination breakpoints was estimated to be 943 bp with a broad range (25% quartile: 303 bp; 75% quartile: 2828 bp). This estimate is subject to bias, however, because multiple mutations at a SNP or sequencing errors can occur and violate the assumption of maximum parsimony. To account for such errors and to wring as much information as possible from ambiguous data, the authors then analyzed the DNA sequences using a hidden Markov model (HMM; Box 1).

The median distance between recombination breakpoints estimated by HMM was 2002 bp (25% quartile: 801 bp; 75%

Corresponding author: Zeyl, C.W. (zeylcw@wfu.edu).
* These authors contributed equally to this paper.
Available online 12 February 2007.

quartile: 5412 bp), approximately twice the interval length inferred by maximum parsimony. HMM inferred less recombination because it allows for the possibility that a SNP supports a different genealogical tree than the true tree (e.g. because of multiple mutations or sequencing error) and so is less likely to infer spurious recombination events, especially between closely linked SNPs. The HMM method also enables the probable position of past recombination events to be inferred. Importantly, these are only those past recombination events that generated lineages that were able to survive to the present.

In addition, Ruderfer *et al.* provided a glimpse into the reproductive biology of yeast in nature by estimating the historical rate of outcrossing, O_C. The authors used simulations to generate new genomic sequences, which were then analyzed by HMM in the same manner as the real genomic sequences. (Specifically, they used 'coalescent' simulations, which generate random genealogies under the assumption that offspring are sampled uniformly from individuals in the previous generation. Genomic sequence evolution was then simulated along these random genealogies, using previous estimates of the mutation rate per bp.) Coalescent simulations also allow recombination to occur at rate R, which is the effective number of recombination events expected over all the generations separating two randomly drawn sequences from a population. In mathematical terms, $R = 2rG_TO_C$, where G_T is the expected number of generations separating two sequences from their common ancestor and r is the recombination rate in Morgans per bp. In their simulations, Ruderfer *et al.* observed a median distance between recombination breakpoints that was consistent with the HMM estimate of 2002 bp only when R fell between 0.0021 and 0.0023. Using previous estimates of G_T and r, the authors estimated an outcrossing rate of $O_C = 2 \times 10^{-5}$ generation^{-1}.

Unfortunately, it is difficult to determine the degree of uncertainty in this outcrossing rate. The coalescent simulations required O_C to lie in a narrow range, 1.9–2.1 $\times$ 10^{-5}, to be consistent with the median breakpoint interval inferred by HMM. However, this range ignores uncertainty in the previously estimated parameters (the mutation rate, G_T, and r) and uncertainty in the breakpoint intervals inferred by HMM. Furthermore, coalescent simulations can generate misleading results if the three sequenced genomes of *S. cerevisiae* do not represent a random sample or if past changes in the population size and geographical distribution of *S. cerevisiae* were not accurately modeled.

What does this outcrossing rate mean?

Although the outcrossing rate is low, the total frequency of sex would be substantially higher if we included inbreeding owing to either selfing (within an ascus or following mate switching) or population structure. Inbreeding levels are traditionally measured by F_{IT}, which is the total excess of homozygosity at a locus over that expected under random mating.

In a recent study of 27 strains of *S. cerevisiae* sampled primarily from oak trees and vineyards [9], 25 of the strains were homozygous at all four genes sequenced, suggesting F_{IT} values near one (most matings are inbred). Nevertheless, if we are interested in knowing how often sex and recombination have acted to bring together alleles from different lineages, thus providing the advantages predicted by Fisher and Muller, it is the outcrossing rate, O_C, estimated by Ruderfer *et al.*, that matters most.

Box 1. Inferring recombination breakpoints using parsimony and hidden Markov models

With three strains of *Saccharomyces cerevisiae* and one outgroup species (*S. paradoxus*), there are three possible genealogical tree topologies (Figure Ia–c). One method used to infer the genealogy of a locus is maximum parsimony, the principle that the simplest explanation is the best. In this case, the best tree is the one requiring the fewest possible mutations to explain the sequence data at that locus. Consider a SNP where guanine (G) is present in the outgroup and in lineage 3, and thymine (T) is present in lineages 1 and 2. Only one mutation is required to explain this outcome using tree A, but two mutations are required using trees B and C (see example mutations in Figure Ia–c). Assuming maximum parsimony, this SNP is 'informative' and supports tree A. Occasionally, however, the same mutation will occur more than once in separate lineages (e.g. tree B might be the true genealogy), in which case the assumption of maximum parsimony will be violated and an incorrect tree will be inferred.

Hidden Markov models

To avoid the limitations of parsimony, Ruderfer *et al.* [8] used a hidden Markov model (HMM) to infer the probable genealogical history of each SNP. HMM is a statistical technique that can be used to infer the probability that each site in a sequence belongs to a particular category [10]. In this case, the categories are the genealogical trees shown in Figure Ia–c. For any one SNP, the category to which it belongs is unknown ('hidden'), but the probability that it belongs to each category can be estimated using a model that incorporates the probability of observing each SNP pattern given an underlying tree and the probability that the tree changes from locus to locus (depending on the genetic distance between the loci). These probabilities are unknown, but HMM can 'learn' which set maximizes the overall likelihood of observing the entire data set. Unlike maximum parsimony, HMM is probabilistic: it allows for the possibility that multiple mutations (or even sequencing errors) account for the observed SNP pattern.

Genealogical trees with recombination

Inbreeding among closely related lineages is generally undetectable, because the lineages do not yet differ in sequence. Outcrossing events might be detectable, depending on whether they cause a SNP to support a different tree. Figure Id illustrates the evolutionary history of a set of six SNPs sequenced from *S. paradoxus* and three lineages of *S. cerevisiae*, where each colored letter represents a nucleotide. Over evolutionary time, mutations (circled) have occurred, leading to the observed sequences at the tops of the tree. The second, fourth, and fifth SNPs are uninformative (each tree topology in Figure Ia–c can explain these data with only one mutation per SNP). There is, however, a switch in support between the first SNP, which supports tree A (uniting lineages 1 and 2; Figure Ia), and the third SNP, which supports tree C (uniting lineages 2 and 3, Figure Ic).

Using maximum parsimony, an outcrossing event in the history of these sequences would be inferred with a recombination breakpoint between the first and third SNP. None of the other mating and recombination events are detectable, as they did not cause a switch in which phylogeny was supported. Using HMM, the number of base pairs separating the first and third SNPs would be taken into account in inferring the probability of a recombination event between them. [Recombination events do not appear to be reciprocal on the tree (one-way dashed arrows), because it is more likely that the ancestor of lineage 1, for example, mated with an unsampled lineage similar to lineage 2 than with lineage 2 itself.]

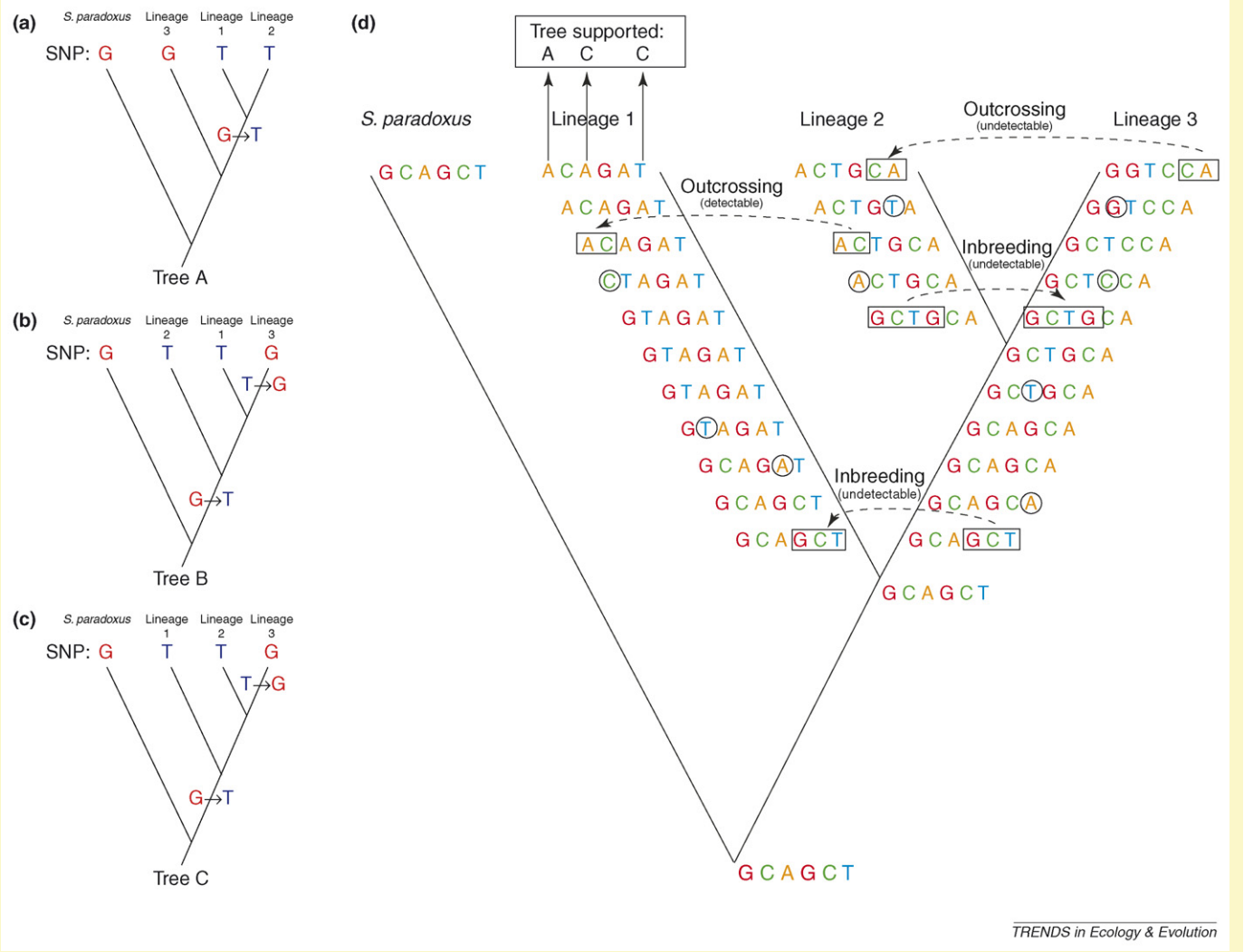

Figure I.

What next?

These data on outcrossing rates in *S. cerevisiae* do not, in and of themselves, prove that sex and recombination have been maintained in yeast because of the advantages of genetic mixing. It is possible that the number of recombination events recorded in these yeast genomes is the same or even less than that expected if we knew the historical rates of outcrossing and recombination.

The Sanger Institute (http://www.sanger.ac.uk) has just made available the genome sequences of 14 strains of *S. cerevisiae* and 21 strains of *S. paradoxus*, with another 22 *S. cerevisiae* and 15 *S. paradoxus* strains to follow. Thus, the potential to explore the history of recombination in yeast has increased substantially, and the results obtained by Ruderfer *et al.* are a first step to determining whether recombination events within certain regions of the genome are more likely to generate lineages that persist over evolutionary time, as predicted by the Fisher–Muller hypothesis. By determining whether the recombination events that have persisted tend to cluster around genes involved in adaptation, we might be able to scan genomes for evidence as to why sex and recombination are so prevalent.

Acknowledgments

S.P.O. is supported by a Discovery grant from the Natural Sciences and Engineering Research Council (Canada) and by a sabbatical grant from the National Evolutionary Synthesis Center (NESCent). C.W.Z. is supported by National Science Foundation grant DEB-0542738 (USA).

References

1 Otto, S.P. and Lenormand, T. (2002) Resolving the paradox of sex and recombination. *Nat. Rev. Genet.* 3, 252–262
2 Fisher, R.A. (1930) *The Genetical Theory of Natural Selection,* Oxford University Press
3 Morgan, T.H. (1913) *Heredity and Sex,* Columbia University Press
4 Muller, H.J. (1932) Some genetic aspects of sex. *Am. Nat.* 66, 118–138
5 Colegrave, N. (2002) Sex releases the speed limit on evolution. *Nature* 420, 664–666
6 Goddard, M.R. *et al.* (2005) Sex increases the efficacy of natural selection in experimental yeast populations. *Nature* 434, 636–640
7 Strobeck, C. *et al.* (1976) The effects of hitchhiking on a gene for recombination. *Genetics* 82, 547–558
8 Ruderfer, D.M. *et al.* (2006) Population genomic analysis of outcrossing and recombination in yeast. *Nat. Genet.* 38, 1077–1081
9 Aa, E. *et al.* (2006) Population structure and gene evolution in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 6, 702–715
10 Eddy, S.R. (2004) What is a hidden Markov model? *Nat. Biotechnol.* 22, 1315–1316

0169-5347/$ – see front matter
doi:10.1016/j.tree.2007.02.005