On being the right size

Edward C Holmes

The first replicating molecules were probably composed of RNA and undoubtedly small, limited in size by a self-destructing error rate. A new study shows that a relatively minor increase in replication fidelity may have had a large effect on the size, and hence complexity, of early replicators.

© 2005 Nature Publishing Group http://www.nature.com/naturegenetics

1953 was a landmark year for biological science. To many, it is best known as the beginning of a new age in molecular genetics, starting with the publication of Watson and Crick's structure of DNA[1]. Yet 1953 also saw the dawn of modern studies on the origin of life. By simulating probable conditions of the earth's early atmosphere and using 'spark discharge' as an energy source, Stanley Miller achieved the prebiotic synthesis of a variety of organic compounds including amino acids, thereby showing that an endogenous origin of life on earth was feasible[2]. In stark contrast to the developments made since Watson and Crick's work, retracing the steps in the origin of life remains an elusive goal, hampered by a multitude of interconnected problems. On page 1008 of this issue, Kun *et al.*[3] shed new light on one of the key hurdles associated with the development of life on earth: how to evolve genetic complexity.

The error threshold

Understanding the origin of life requires us to solve some of the most fundamental problems in evolution: how to produce the first polymers, how to generate the first replicating molecule, how to make the first cell. Perhaps a more generic problem is how a simple replicating system can evolve increased genotypic, and hence phenotypic, complexity. Since the discovery of ribozymes, self-replicating RNA molecules with catalytic properties, it has been generally accepted that early life at one point existed as an 'RNA world', in which the only replicators were RNA molecules[4], perhaps functioning together as 'bags of genes'[5]. Although this is an interesting theory, reliance on RNA replication comes at a cost; compared with DNA, the copying of RNA is highly error-prone. This, in turn, means that there is an upper limit on the size of primitive RNA replicators, as those larger than this intrinsic 'error threshold' will be unable to copy themselves with sufficient fidelity, and the system will degenerate. Hence, many of the ribozymes we know today are only a few hundred nucleotides in length. To create more genetic complexity, it is therefore necessary to encode more information in longer genes by using a replication system with greater fidelity. But there's a catch: to replicate with greater fidelity requires a more accurate and hence complex replication enzyme, but such an enzyme cannot be created because this will itself require a longer gene, and longer genes will breach the error threshold. This evolutionary chicken-and-egg dilemma has been dubbed Eigen's paradox (following Manfred Eigen's seminal work on the nature of early replicators[6]) and is one of the most intractable puzzles in the origin of life. Kun *et al.*[3] now go a long way to providing an answer.

Edward C. Holmes is in the Department of Biology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA. e-mail: ech15@psu.edu

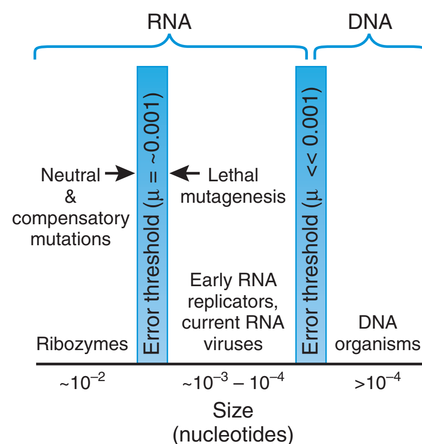

Figure 1 Error thresholds and the origin of life. For evolution to proceed in the RNA world, a reduction in the error rate (μ) to ~0.001 mutations per nucleotide per replication must be achieved. For more complex genomes to evolve, with sizes $>10^4$ nucleotides, a second threshold needs to be crossed in which $\mu << 0.001$. This necessitates the evolution of DNA replication, which provides a higher copying fidelity than RNA. Neutral and compensatory changes tend to dampen the effects of deleterious mutations, allowing a relaxed error threshold. In contrast, antiviral drugs that increase the error rate tend to push contemporary RNA viruses over the error threshold and into lethal mutagenesis.

A distinctive feature of RNA molecules is that they generally form complex secondary structures, complete with loops, hairpins and bulges. This is central to understanding the evolution of early replicators, because such structures, and the complex fitness landscapes they enforce, means that many of the mutations that arise either will be neutral, and hence have no effect on fitness, or will interact epistatically. In sum, the effect of RNA secondary structure is to remove the one-to-one relationship between genotype and phenotype, producing what Kun *et al.* call a relaxed error threshold[3], buffering early replicators against mutational meltdown. Crucially, if the fitness landscapes of present-day ribozymes can be calculated, as Kun *et al.* attempt to do[3], then it should also be possible to estimate, albeit roughly, what sizes early RNA replicators would have been able to achieve under given mutation rates. As raw data, Kun *et al.* examined the fitness landscapes of two small ribozymes, the hairpin and *Neurospora* VS ribozymes[3]. The secondary structures of these ribozymes can be predicted, and extensive mutagenesis studies can be used to estimate fitness. Fitting data to model suggested that a decrease in the error rate from 0.1–0.01 to 0.001 mutations per nucleotide per replication would result in many replicators in the size range of 7–8 kb, equivalent to that of contemporary tRNAs and RNA viruses (**Fig. 1**).

From RNA to DNA

Despite the results of Kun *et al.*[3], some key questions remain unanswered. In particular, although an increase in replicator size from several hundred nucleotides to >7 kb represents a large enhancement of genetic complexity and is undoubtedly sufficient for an RNA world, it is very much smaller than either the tiniest cellular life forms known today or the estimated size of the genome of the last universal common ancestor[7]. More fundamentally, the most abundant RNA life forms, RNA viruses, usually have genomes <12 kb in length. Hence, although most RNA viruses are larger and more complex than ribozymes and have greater copying fidelity, they are also at the mercy of an error threshold (**Fig. 1**). This

recognition is the theoretical basis of a new and promising form of viral treatment that involves forcing RNA viruses over the error threshold through the application of antiviral agents like ribavirin that act as mutagens, thereby producing a "lethal mutagenesis"[8]. How, then, to move from the small genomes of RNA viruses to those of more complex life forms? Given that the largest replicating RNA molecules are the coronaviruses, at 30 kb, the answer must involve the evolution of DNA replication that comes equipped with proof-reading. That is another intricate evolutionary story, particularly given the growing evidence for RNA proof-reading[9]. Explaining the full history of the origin of life will continue to perplex evolutionary biologists for many generations to come.

1. Watson, J.D. & Crick, F.H.C. *Nature* **171**, 737–738 (1953).
2. Miller, S.L. *Science* **117**, 528 (1953).
3. Kun, A., Santos, M. & Szathmáry, E. *Nat. Genet.* **37**, 1008– 011 (2005).
4. Gilbert, W. *Nature* **319**, 618 (1986).
5. Szathmáry, E. & Demeter, L. *J. Theor. Biol.* **128**, 463–486 (1987).
6. Eigen, M. *Naturwissenschaften* **58**, 465–523 (1971).
7. Szathmáry, E. *Nature* **433**, 469–470 (2005).
8. Crotty, S., Cameron, C.E. & Andino, R. *Proc. Natl. Acad. Sci. USA* **98**, 6895–6900 (2001).
9. Poole, A.M. & Logan, D.R. *Mol. Biol. Evol.* **22**, 1444–1455 (2005).

Elucidating mouse transmission ratio distortion

Mary F Lyon

A proportion of wild mice carry a variant region of chromosome 17 that results in severe transmission ratio distortion in males. The genetic basis of this distortion has long been enigmatic, but a recent study begins to disentangle it.

© 2005 Nature Publishing Group http://www.nature.com/naturegenetics

On page 969 of this issue, Bernhard Herrmann and colleagues report the cloning of a distorter gene in the transmission ratio distortion system of the mouse t complex[1]. Why is this so exciting? The t complex has been baffling geneticists for more than 70 years. It is a variant form of the proximal region of chromosome 17 found in a proportion of wild mice. Male mice heterozygous with respect to a t haplotype (as the variant forms are called) transmit it to an abnormally high proportion of their young, up to 99%. Males carrying two t haplotypes are sterile. Females show normal transmission and fertility. The t complex occupies roughly one-third of chromosome 17 but is inherited as a unit, owing to strong crossover suppression caused by four nonoverlapping inversions (**Fig. 1**). Study of rare crossovers that give rise to partial haplotypes has shown that the transmission ratio distortion depends on the action of several factors located in or between the inversions. The current view is that there are three or more distorters, Tcd1–Tcd3, that act on a responder, Tcr[2]. The distorters have a harmful effect and act on any chromosome, but the t form of the responder, Tcr^t, is resistant to this effect and acts only in *cis*. Thus, sperm carrying Tcr^t are protected from the harmful distorters and are able to fertilize eggs, giving rise to the ratio distortion. The distorters are postulated to act cumulatively, and when they are homozygous, the resistance of Tcr^t is overcome so that no sperm can fertilize eggs, and sterility results. The sperm are not killed, but their swimming is much impaired and few reach the oviduct[3] (**Fig. 1**).

A key obstacle to cloning the responder and distorters is that they cannot be localized precisely, owing to the crossover-suppressing effect of the inversions. There may even be more than one distorter in a particular inversion, but they will be inherited as one. The responder was recently cloned by Bernhard Herrmann and colleagues[4], who found in the appropriate region a hybrid gene, formed by the fusion of two kinases. The 5′ end of a ribosome S6 kinase was fused to a previously unknown kinase related to the microtubule-associated protein family, called sperm motility kinase or *Smok1*. The fused gene was called $Smok1^{Tcr}$. When this gene was introduced as a transgene at a random location, if a partial t haplotype with distorters was also present, then the chromosome carrying Tcr was transmitted at a high ratio. Thus, Tcr showed behavior typical of the responder, and Herrmann and colleagues concluded that it was the responder. They considered Tcr to be part of a signaling cascade and surmised that the distorters acted upstream of Tcr in this cascade.

Mary F. Lyon is in the MRC Mammalian Genetics Unit, Harwell, Didcot, Oxfordshire, OX11 0RD, UK.
e-mail: m.lyon@har.mrc.ac.uk

Search for distorters

They began a search for distorters in the region in which *Tcd1* was postulated to lie and isolated a gene called T-cell activation Rho GTPase-activating protein, *Tagap1*, which differed between t and wild-type strains. The t haplotypes carried four *Tagap1* loci, whereas the wild type carried only one. The t haplotype cDNAs showed several nucleotide differences from wild-type, including one that resulted in a truncated protein with an intact N-terminal RhoGAP domain. *Tagap1* is transcribed in the testes and at a higher level in t haplotypes than in the wild type.

Thus, *Tagap1* fulfilled two necessary criteria for a distorter gene: it was expressed in the testes and differed between t haplotypes and the wild type. The crucial question was whether it affected ratio distortion. In view of the higher expression of the t form of *Tagap1*, Herrmann and colleagues assessed the effect of overexpressing the wild-type allele by means of a transgene. They compared the transmission ratio of a haplotype lacking *Tcd1* in males with and without the transgene, and males with the transgene had a higher ratio (**Fig. 1**). Conversely, *Tagap1*-knockout males had significantly reduced transmission. This means that *Tagap1* possesses the essential property of distorting the transmission of a t haplotype and fulfills the criteria to be considered the first distorter gene to be identified in the t complex. It also fits with Herrmann's earlier prediction that the distorters would act upstream of *Smok1* in a signaling cascade. The effect of the cascade on microtubule function could affect flagellar movements. Thus, at long last we have a picture of the mechanism of transmission ratio distortion in the t complex.

Puzzles and questions

But this is not the end of the story. *Tagap1* in its distorter role acts as a hypermorph.