in escaping agricultural harvesting, succession, or summer drought (*19*). The selective forces driving the molecular evolution of flowering time genes can now be examined in more detail.

There was no clear geographical association of the ecotypes within the groups defined by the *FRI* alleles, or between the early-flowering group 1 ecotypes and the late-flowering group 4 ecotypes from which they may have arisen. Human-induced dispersal has been a major factor in the recent spread of *Arabidopsis* ecotypes and has resulted in little association of geographical and genetic distance (*20, 21*). Human disturbance regularly exposes *Arabidopsis* ecotypes to novel environments, thus maintaining a strong selective pressure for adaptive mutations. This may account for loss-of-function mutations providing the basis for the evolutionary changes in *Arabidopsis* flowering time.

References and Notes

1. G. G. Simpson, A. R. Gendall, C. Dean, *Annu. Rev. Cell Dev. Biol.* **15**, 519 (1999).
2. J. E. Burn, D. R. Smyth, W. J. Peacock, E. S. Dennis, *Genetica* **90**, 145 (1993).
3. J. H. Clarke and C. Dean, *Mol. Gen. Genet.* **242**, 81 (1994).
4. H. Kuittinen, M. J. Sillanpaa, O. Savolainen, *Theor. Appl. Genet.* **95**, 573 (1997).
5. I. Lee, A. Bleecker, R. Amasino, *Mol. Gen. Genet.* **237**, 171 (1993).
6. S. Sanda, M. John, R. Amasino, *J. Hered.* **88**, 69 (1997).
7. K. Napp-Zinn, *Z. Indukt. Abst. Vererb.* **88**, 253 (1957).
8. _____, *Naturwissenschaften* **42**, 650 (1955).
9. I. Lee, S. D. Michaels, A. S. Masshardt, R. M. Amasino, *Plant J.* **6**, 903 (1994).
10. M. Koornneef, H. Blankestijn-de Vries, C. Hanhart, W. Soppe, T. Peeters, *Plant J.* **6**, 911 (1994).
11. For supplementary information on the details of the map-based cloning of *FRIGIDA*, including primer sequences, see *Science* Online (www.sciencemag.org/feature/data/1053859.shl).
12. C. C. Sheldon *et al*., *Plant Cell* **11**, 445 (1999).
13. S. D. Michaels and R. M. Amasino, *Plant Cell* **11**, 949 (1999).
14. B. H. Karlsson, G. R. Sills, J. Nienhuis, *Am. J. Bot.* **80**, 646 (1993).
15. M. Nordborg and J. Bergelson, *Am. J. Bot.* **86**, 470 (1999).
16. J. Hagenblad, U. Johanson, M. Nordborg, unpublished data.
17. C. Alonso-Blanco, S. El-Din El-Assal, G. Coupland, M. Koornneef, *Genetics* **149**, 749 (1998).
18. K. K. Khurmatov, *Arab. Inf. Serv.* **19**, 63 (1982).
19. T. Mitchell-Olds, *Evolution* **50**, 140 (1996).
20. J. Bergelson, E. Stahl, S. Dudek, M. Kreitman, *Genetics* **148**, 1311 (1998).
21. T. F. Sharbel, B. Haubold, T. Mitchell-Olds, *Mol. Ecol.*, in press.
22. R. Schmidt *et al*., *Plant J.* **9**, 755 (1996).

23. We thank M. Smith for excellent care of *Arabidopsis*, K. Torney for construction and transformation of the *35S-FRI* cDNA construct, J. Clarke for ecotype DNA, N. Hartley for sequence analysis, J. Hagenblad and M. Nordborg for sharing unpublished results and seeds, and M. Nordborg, E. Coen, G. Simpson, T. Mitchell-Olds, and D. Stern for invaluable comments. We also thank R. Martienssen and M. Marra for unpublished BAC information. Supported by an EMBO fellowship (U.J.) and a UK Biotechnology and Biological Sciences Research Council Core Strategic Grant to the John Innes Centre.

10 July 2000; accepted 17 August 2000

Selfish DNA in Protein-Coding Genes of *Rickettsia*

Hiroyuki Ogata,[1] Stéphane Audic,[1] Valérie Barbe,[2] François Artiguenave,[2] Pierre-Edouard Fournier,[3] Didier Raoult,[3] Jean-Michel Claverie[1]*

***Rickettsia conorii*, the aetiological agent of Mediterranean spotted fever, is an intracellular bacterium transmitted by ticks. Preliminary analyses of the nearly complete genome sequence of *R. conorii* have revealed 44 occurrences of a previously undescribed palindromic repeat (150 base pairs long) throughout the genome. Unexpectedly, this repeat was found inserted in-frame within 19 different *R. conorii* open reading frames likely to encode functional proteins. We found the same repeat in proteins of other *Rickettsia* species. The finding of a mobile element inserted in many unrelated genes suggests the potential role of selfish DNA in the creation of new protein sequences.**

Selfish DNA (*1*, *2*)—repeated elements without obvious cellular function—is thought to be an important factor in genome evolution. Transposons and other extragenic interspersed repeats are responsible for gene (or exon) shuffling and duplication, as well as regulatory changes (*3*–*5*). However, those mechanisms cannot account for de novo creation of protein domains. The finding, in *Rickettsia conorii*, of a palindromic interspersed repeat inserted in several unrelated protein-coding sequences now suggests that selfish DNA could directly participate in the creation of new protein sequences.

Rickettsiae, the closest extant relatives of mitochondria (*6*), are normally found inside the cells of arthropods, such as lice, fleas, and ticks. They occasionally infect humans and cause serious diseases. The causative agent of Mediterranean spotted fever, also known as boutonneuse fever, is *R. conorii*, an intracellular bacterium of ticks (*Rhipicephalus sanguineus*) (*7*). The complete 1.1-Mb genome of *Rickettsia prowazekii*, a close relative of *R. conorii* transmitted by lice (*Pediculus humanus humanus*), was recently reported (*8*). One salient feature of this genome is the abundance of noncoding sequences, suggesting ongoing genome degradation that parallels the evolutionary scenario presumed to have occurred for mitochondria (*8*, *9*). The sequencing of the entire 1.3-Mb *R. conorii* genome has been undertaken to study the evolution of *Rickettsia* and the molecular basis of different life-styles and pathogenicities (*10*). The palindromic repeat described here was found in 19 open reading frames (ORFs) and in 25 distinct extragenic regions of the genome. The lengths of the 44 repeats identified in *R. conorii* varied from 106 to 150 nucleotides [Web fig. 1 (*11*)]. Because of their well-conserved palindromic nature, most of these sequences are able to form stable hairpin-like mRNA secondary structures (*12*). Hairpins, such as those shown in Fig. 1, are predicted to correspond to the minimum free energy secondary structures for 39 of these repeats.

The amino acid sequences encoded by the 19 repeats found in *R. conorii* ORFs are well conserved and correspond to the same reading frame [Web fig. 2 (*11*)] (*13*). The predicted conformation of the putative peptide chains consists of a central, mildly hydrophobic α helix, flanked by two extended or coil regions including two conserved glycine residues (at positions 14 and 34), a conserved proline (position 40), and numerous conserved charged residues. Insertion sites within the ORFs vary from NH_2-terminal to near COOH-terminal ends (Fig. 2), but secondary-structure prediction and hydropathy analyses suggest that they always occur at the surface of the various proteins. In all cases where the three-dimensional (3D) structures of homologs were available, the insertion sites occurred within NH_2-terminals, loops, or short helices exposed to solvent (*14*).

Of the 19 repeat-containing ORFs in *R. conorii*, 16 have homologs in other species (Table 1) and are thus very likely to encode functional proteins. Most of them are predicted to have important roles in *R. conorii* metabolism. Thirteen have homologs in distant phyla, including four ORFs [guanosine 5′-triphosphate (GTP)–binding protein Era, glutamyl–tRNA synthetase, guanilate kinase, and cell cycle protein MesJ] having homologs within the minimal

[1]Information Génétique & Structurale, CNRS-AVENTIS UMR 1889, 31 Chemin Joseph Aiguier, 13402 Marseille Cedex 20, France. [2]Genoscope, Centre National de Séquençage, 2 rue Gaston Crémieux, 91000 Evry, France. [3]Unité des Rickettsies, Université de la Méditerranée, Faculté de Médecine, CNRS UPRES A 6020, 27 Boulevard Jean Moulin, 13385 Marseille Cedex 05, France.

*To whom correspondence should be addressed. E-mail: jean-michel.claverie@igs.cnrs-mrs.fr

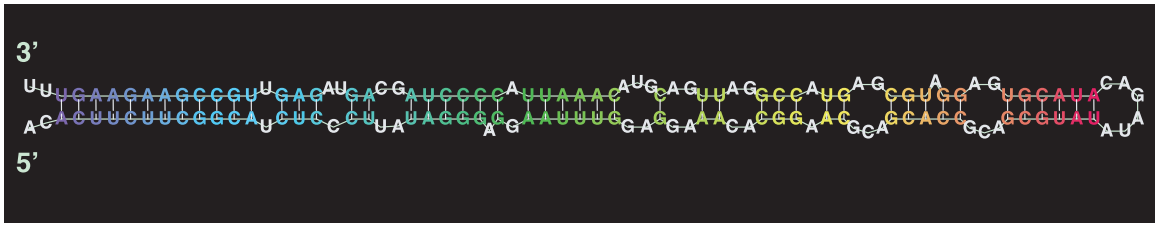

Fig. 1. Predicted secondary structure of the repeat found in the *R. conorii* glutamyl–tRNA synthetase gene. Nucleotides making base pairs in the predicted minimum free-energy structures are shown in the same colors [see also Web fig. 1 (*11*)].

Table 1. ORFs with the repeats inserted. aa, amino acids.

ORF name*	Function	Homologs†	Length (aa)	Location of RPE insertion	Structure/organism‡
		R. conorii			
era (RP118)	GTP-binding protein	P**G**_SYH__	339	4..48	1EGA/*E. coli*
gltX (RP623)	Glutamyl-tRNA synthetase	P**G**CSYHAE	513	336..383	1GLN/*Thermus thermophilus*
gmk (RP765)	Guanylate kinase	P**G**CSY__E	228	6..41	1GKY/ *Saccharomyces cerevisiae*
hemC (RP466)	Porphobilinogen deaminase		350	257..305	1PDA/*E. coli*
kdtA (RP089)	3-deoxy-D-manno-octulosonic-acid transferase	P_C__H__	464	50..96	
mesJ (RP042)	Cell cycle protein	P**G**CSYH__	477	208..255	
mviN (RP590)	Virulence factor	P__S_H__	555	3..50	
pcnB (RP015)	Poly(A) polymerase	PGCS_H_E	435	40..88	
rlpA (RP390)	Rare lipoprotein A precursor	P__SYH__	320	12..57	
ubiG (RP622)	3-Demethylubiquinone-9 3-methyltransferase	P_____E	289	50..97	
ubiH (RP561)	Ubiquinone biosynthesis protein	P___Y__E	430	10..58	
(RP045)	Unknown function	P___YH_E	406	329..376	
(RP167)	Unknown function	_____	92	7..54	
(RP502)	Unknown function	_____	186	103..150	
(RP507)	Unknown function	_____	193	10..57	
(RP545)	Unknown function	PGCSY___	244	116..163	
		R. prowazekii			
alr (RP095)	Alanine racemase	PG_SYH__	404	226..273	2SFP/*Bacillus stearothermophilus*
coxB (RP406)	Cytochrome c oxidase polypeptide II	PG__YHAE	313	5..51	1OCC/*Bos taurus*
kdtA (RP089)	3-Deoxy-D-manno-octulosonic-acid transferase	P_C__H__	461	50..91	
lysC (RP753)	Aspartokinase	PGCSYHAE	446	250..294	
pyrG (RP378)	CTP synthase	PGCSYHAE	586	429..473	
(RP404)	Unknown function	PG_____	281	139..174	
(RP474)	Unknown function	_____	269	152..197	
(RP545)	Unknown function	PGCSY___	236	117..160	
(RP688)	Unknown function	_____	297	9..55	
		R. helvetica			
polA	DNA polymerase I	PGCSYH_E	921	439..485	1D8Y/*E. coli*
		R. felis			
polA	DNA polymerase I	PGCSYH_E	922	440..487	1D8Y/*E. coli*

*Identifiers from the *R. prowazekii* genome are in parentheses. For *R. conorii*, the identifiers indicate their counterparts of *R. prowazekii*. †Organism group abbreviations are as follows. P, Proteobacteria (*E. coli*, *Haemophilus influenzae*, *Helicobacter pylori*); G, Gram-positive bacteria (*Bacillus subtilis*, *Mycoplasma genitalium*, *Mycoplasma pneumoniae*, *Mycobacterium tuberculosis*; homologs of *M. genitalium* are indicated by the abbreviation "G" in boldface.); C, Chlamydia (*Chlamydia trachomatis*, *Chlamydia pneumoniae*); S, Spirochete (*Borrelia burgdorferi*, *Treponema pallidum*); Y, Cyanobacteria (*Synechocystis*); H, Hyperthermoplilic bacteria (*Aquifex aeolicus*); A, Archaea (*Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Archaeoglobus fulgidus*, *Pyrococcus horikoshii*, *Pyrococcus abyssi*, *Aeropyrum pernix*); E, Eukaryotes (*S. cerevisiae*, *Caenorhabditis elegans*). If there is no homolog within an organism group, an underscore (_) replaces the organism abbreviation. ‡Protein Data Bank identifiers and species name of the structure determined.

gene complement of *Mycoplasma genitalium* (*15*). There is no structural correlation among the repeat-containing ORFs (*14*), and they cover a wide spectrum of functional categories (nucleotide metabolism, mRNA processing, translation, and envelope formation). It is unlikely that the 19 repeat-containing ORFs might be pseudogenes for two reasons: (i) they exhibit statistical properties [e.g., fifth-order Markov statistics (*16*)] similar to those of regular genes, and (ii) they do not have any "normal" counterparts in the *R. conorii* genome. Furthermore, the presence of a cognate transcript was verified by reverse transcriptase–polymerase chain reaction (RT-PCR) for *gltX*, *hemC*, and *ubiG*, three of the repeat-containing ORFs in *R. conorii* (*17*–*19*).

All other available microorganism genomes were scanned for similarity to the above *R. conorii* repeats. Twelve homologous sequences were identified within *R. prowazekii* (10), *Rickettsia helvetica* (1), and *Rickettsia felis* (1) [Web fig. 1 (*11*)]. Eleven of these repeats were found inserted in ORFs, most of which correspond to important functions, and have homologs in distant phyla (Table 1). Because of its apparent specificity to *Rickettsia*, this new family of repeats will be referred to as *Rickettsia* palindromic elements (RPEs). The two RPEs identified in *R. helvetica* and *R. felis* are very similar and occur at the same position in the middle of the DNA polymerase I (*polA*) coding region. The nucleotide sequences of the *R. prowazekii* repeats are less conserved than those in *R. conorii*, *R. helvetica*, and *R. felis* [Web fig. 1 (*11*)] (*13*). Their palindromic nature is also less clear, and only 2 out of 10 are predicted to form hairpin structures [Web fig. 1 (*11*)]. Nine of the 10 RPEs identified in *R. prowazekii* occur in ORFs (*8*). Their peptide translations also exhibit an increased sequence divergence (*13*).

There are only three cases of putative orthologs among the RPE-containing ORFs: DNA polymerase I (*polA*) (*R. felis* and *R. helvetica*), 3-deoxy-D-mano-octulosonic-acid transferase (*kdtA*) (*R. prowazekii* and *R. conorii*), and the hypothetical protein RP545 (*R. prowazekii* and *R. conorii*). In all three cases, the regions translated from the RPEs appeared

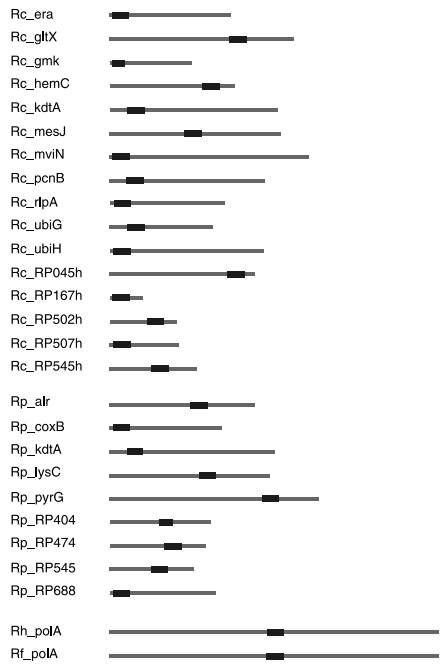

Fig. 2. Locations of the *Rickettsia* RPEs within the ORFs. Gray lines and black boxes represent the entire ORFs and the RPE inserts, respectively. Their widths are proportional to the length of the protein encoded by the corresponding ORFs.

to evolve more rapidly than the rest of the proteins (*20*).

The RPE resembles the 127–base pair (bp) intergenic repeat unit (IRU) (*21*) and the 151-bp-long RSA repeats (*22*) previously described in Enterobacteriaceae, in respect of frequency, size, palindromic structure, and relative sequence conservation. However, IRU and RSA do not share any sequence similarity with RPE and are always found in extragenic regions, except for one documented occasional insertion of RSA at the COOH-terminal of a gene (*23*). They are too small to encode the enzymes found in bacterial insertion sequences and lack their characteristic flanking repeats. A mechanism involving RNA intermediates is thought to be responsible for the amplification of these repeats (*22*).

The surprising occurrences of RPEs within coding regions raise the possibility that they might be excised at the mRNA level. A first argument against this hypothesis, however, is that the RPEs located within coding regions never interrupt the translation frames. Furthermore, direct RT-PCR experiments demonstrated the presence of the RPE sequences on the transcripts of three genes that we tested (*gltX*, *hemC*, and *ubiG*) (*17–19*). RPEs are thus unlikely to be a new type of intron.

Another possibility is that the translated RPE is excised at the protein level like inteins (*24*). Inteins are 360– to 548–amino acid–long peptide intervening sequences, encoding both an endonuclease and a self-splicing activity. Shorter "mini-inteins" lacking the endonuclease domain and only 134 to 198 amino acids long have been described (*24*). Given their much smaller size (~45 residues), translated RPEs are unlikely to encode comparable enzymatic activities. RPEs also lack all consensus motifs associated with protein splicing found in inteins. Furthermore, the analysis of available 3D structures (*25*) indicates that intein insertions often interrupt secondary structure elements and/or occur in buried regions of their target proteins. Inteins must thus be quickly excised to allow the folding of a functional protein. In contrast, RPEs always occur at the surface of a protein (*14*) and appear compatible with the functional fold. Finally, there is preliminary evidence that the GltX and Alr genes (Table 1), when expressed in *Escherichia coli*, produce polypeptides of sizes compatible with the presence of the RPE-derived peptides (*26*).

The previously described IRU and RSA palindromic repeats, similar in size and structure to the RPE, are considered selfish DNA elements occasionally recruited for local regulatory function at the replication, transcription, or translation levels (*22*). The characteristic occurrences of RPEs within coding regions are now raising the possibility that they may have acquired a function at the protein level. The RPE-derived peptide might participate in cellular processes specific to Rickettsiae and related to their obligate intracellular parasitism in arthropods or their pathogenicity in humans. All RPEs are inserted in coding regions by means of the same reading frame, thus producing similar amino acid sequences. This might suggest a positive selection for a RPE-specific function. The alternative is that the RPE-derived peptides are rather neutral, and none of the other reading frames could result in a peptide sequence compatible with the structure and function of the original protein. Two main arguments support the latter hypothesis: (i) The other reading frames would result in stretches of hydrophobic residues, and (ii) the RPE-derived peptide are much less conserved than the rest of the proteins (*20*). The RPE-derived peptide conformation is also predicted to consist of two random-coil regions flanking a central α helix. Such a modular structure is likely to be well tolerated at the surface of the host protein. The insertion of RPEs in many unrelated proteins in *R. conorii* and *R. prowazekii* also argues against a specific functional constraint. In *R. conorii*, the relative frequency of the RPEs in the coding versus the extragenic region is compatible with a model of random insertion followed by counterselection of inadequate RPE reading frames and of inappropriate buried locations in host proteins (*14*). The accumulation of this repeat in Rickettsiae might thus be the result of random genetic drift within a small population [Muller's ratchet (*27*)] associated with the obligate intracellular, slow-growing life-style of these bacteria.

The RPEs appear to represent a previously unknown class of selfish DNA element encoding a peptide insert easily tolerated by arbitrary protein hosts. Over evolutionary times, multiple random insertions of such elements within genes, followed by selection on the resulting peptide sequences within the context of different host proteins, might have contributed to the emergence of new protein sequences, domains, and functions.

References and Notes

1. L. E. Orgel and F. H. C. Crick, *Nature* **284**, 604 (1980).
2. W. F. Doolittle and C. Sapienza, *Nature* **284**, 601 (1980).
3. J. V. Moran, R. J. DeBerardinis, H. H. Kazazian Jr., *Science* **283**, 1530 (1999).
4. R. J. Britten, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 9374 (1996).
5. S. Kato, A. Kanazawa, T. Mikami, Y. Shimamoto, *Curr. Genet.* **34**, 303 (1998).
6. D. Yang, Y. Oyaizu, H. Oyaizu, G. J. Olsen, C. Woese, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4443 (1985).
7. D. Raoult and V. Roux, *Clin. Microbiol. Rev.* **10**, 694 (1997).
8. S. G. E. Andersson *et al.*, *Nature* **396**, 133 (1998).
9. J. O. Andersson and S. G. E. Andersson, *Mol. Biol. Evol.* **16**, 1178 (1999).
10. The shotgun phase of the sequencing project of *R. conorii* is completed, and finishing and annotation are now in progress. The genome sequence data analyzed in this study consists of an average eight-fold shotgun coverage of the ~1.3-Mb genome. The whole-genome sequence of *R. conorii* will be published elsewhere.
11. Supplementary data are available at *Science* Online at www.sciencemag.org/feature/data/1051142.shl.
12. I. L. Hofacker *et al.*, *Monatsh. Chem.* **125**, 167 (1994).
13. Pairwise percent identity among the repeats found in *R. conorii* varies in the range of 30 to 92% (56% on average) at the amino acid sequence level and 53 to 94% (73% on average) at the nucleotide sequence level. Pairwise percent identity among those in *R. prowazekii* varies in the range of 16 to 50% (29% on average) at the amino acid sequence level and 45 to 66% (54% on average) at the nucleotide sequence level. The identity across these two *Rickettsia* is in the range of 17 to 61% (45% on average) at the amino acid sequence level and 35 to 73% (59% on average) at the nucleotide sequence level.
14. Location of the repeat insertions were examined for seven different protein structures in the Protein Data Bank: 1EGA, 1GLN, 1GKY, 1PDA, 2SFP, 1OCC, and 1D8Y. For a view of these structures, see Web fig. 3 (*11*).
15. C. A. Hutchison III *et al.*, *Science* **286**, 2165 (1999).
16. S. Audic and J.-M. Claverie, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 10026 (1998).
17. Briefly, total RNA was extracted with RN easy Midi (Qiagen) and RNA isolation system (Promega) kits from 3-day cultures of a 10^8 inoculum of *R. conorii* Seven (Malish) on Vero cells at 32°C. The presence of a transcript was tested for three ORFs exhibiting RPEs at different relative locations [5′ (*ubiG*), central (*gltX*), and 3′ region (*hemC*)]. These ORFs were also selected on the basis of the high expression levels of the orthologous genes in *E. coli* (*28*). The absence of contaminating DNA was first tested by PCR with primer pairs [1-2, 1-3, and 1-4 (*18*, *19*)], Blue Taq DNA polymerase (Eurbio), annealing

temperature of 51°C, and 40 cycles on a PTC-200 thermocycler (MJ Research). A positive control was then performed under the same PCR condition with genomic DNA as template. All primer pairs generated the expected amplicons. RT-PCR were then performed on the RNA extract with a different thermocycler. All primer pairs generated amplicons of the expected sizes and sequences.

18. For the location and the designation of selected primers, see Web fig. 4 (*11*).
19. Primer sequences are as follows: Rc_gltX1, 5′-TCTTTGGAATCTTGGCTGTAG-3′; Rc_gltX2, 5′-GATTAGTGAGCAGGAAATA-3′; Rc_gltX3, 5′-GGATTGACGTACAAATTAC-3′; Rc_gltX4, 5′-CAAAGACTGTAGAGATATTGG-3′; Rc_hemC1, 5′-TACAGATAGCTTCCAACATC-3′; Rc_hemC2, 5′-CAAACCAATTTTATGCTCGG-3′; Rc_hemC3, 5′-CGGATTGACGTACAAATTAC-3′; Rc_hemC4, 5′-GTATCTAGATGCTAATTGCC-3′; Rc_ubiG1, 5′-CTTGTCACTGTCTAATTCTTC-3′; Rc_ubiG2, 5′-AAATATTAGACGTCGGTTGC-3′; Rc_ubiG3, 5′-CAGATTGACGCACAAATTAC-3′; and Rc_ubiG4, 5′-AATAAACCCTATTCGCCTTG-3′.
20. Amino acid sequence identities for the repeat-derived peptides (and the rest of the proteins) are 43% (85%) for KdtA, 31% (83%) for RP545, and 82% (95%) for PolA.
21. R. G. Lloyd and G. J. Sharples, *Nucleic Acids Res*. **18**, 6503 (1990).
22. S. Bachellier, J.-M. Clément, M. Hofnung, *Res. Microbiol*. **150**, 627 (1999).
23. As revealed by a partial, low-significance match between RSA and 23 amino acid residues at the 3′ end of a *Salmonella typhimurium* hypothetical ORF (SwissProt accession number P16656). A RS3-like inverted repeat is also present in the NH_2-terminal region of the DnaE gene in *R. felis* (9). There is no significant similarity between this RS3-like sequence and our newly described RPE sequence.
24. Reviewed in F. B. Perler, *Nucleic Acids Res*. **27**, 346 (1999).
25. We examined the following structures whose homologs are known to contain inteins: 1TGO, 1AB4, 1ECL, 2REC, 1RLR, and 1SKY.
26. V. Monchois, personal communication.
27. J. J. Muller, *Mutat. Res*. **1**, 2 (1964).
28. H. Tao, C. Bausch, C. Richmond, F. R. Blattner, T. Conway, *J. Bacteriol*. **181**, 6425 (1999).
29. We thank C. Abergel, D. Gautheret, and J. Weissenbach for their help at various stages of this work.

7 April 2000; accepted 24 August 2000

Replaying the Game: Hypnagogic Images in Normals and Amnesics

Robert Stickgold,[1]* April Malia,[1] Denise Maguire,[1] David Roddenberry,[1] Margaret O'Connor[2]

Participants playing the computer game Tetris reported intrusive, stereotypical, visual images of the game at sleep onset. Three amnesic patients with extensive bilateral medial temporal lobe damage produced similar hypnagogic reports despite being unable to recall playing the game, suggesting that such imagery may arise without important contribution from the declarative memory system. In addition, control participants reported images from previously played versions of the game, demonstrating that remote memories can influence the images from recent waking experience.

People who engage in novel physical or mental activities for extended periods of time often experience a hallucinatory replay of the activities as they fall asleep, a phenomenon noted in both literary and scientific sources (*1*–*3*). Although the origin of these images in waking experience appears clear, the source within the brain and the function of such imagery are uncertain. If these images are mediated by declarative memory systems, amnesic patients with bilateral hippocampal lesions should not experience them. Because amnesics are capable of reporting dream experiences (*4*, *5*), the question can be meaningfully addressed.

Most studies of dreaming involve reports of mental activity after forced awakenings from rapid eye movement (REM) sleep (*6*, *7*) or after spontaneous awakenings in the middle of the night or morning (*8*). Hypnagogic mentation represents another source of sleep-state mental imagery (*9*–*12*). Studies of the role of sleep in learning and memory have generally ignored the hypnagogic period, focusing instead on possible roles of REM sleep and deep, slow wave sleep (*13*–*16*).

We asked 27 participants to play 7 hours of the computer game Tetris (*17*) over 3 days, 2 hours at the first exposure and 1 hour each subsequent morning and evening (*18*). Three groups of participants were studied: 12 novices with no prior Tetris experience, 10 experts with considerable Tetris experience, and 5 amnesics with extensive, bilateral medial temporal lobe damage (*19*). Participants were repetitively prompted for mentation reports during the first hour of attempted sleep (*20*).

Even though the five amnesic patients were of average intelligence, they were unable to learn and retain new episodic information regardless of modality of presentation or the nature of the material. The performance of the amnesic patients showed only minimal, albeit significant, improvement over 7 hours of play (Fig. 1). The mean score increased from 537 points on participants' first game to a first session average of 651. Over the next five sessions, the scores increased by 36% to 884 points per game (*21*). Although game playing is often seen as based on procedural learning (which is preserved in amnesia), the amnesic group demonstrated impoverished learning in comparison with normal novices, suggesting that Tetris proficiency may also depend on declarative memory systems. This interpretation is supported by the finding that left and right hippocampal glucose utilization during Tetris play decreases during training in proportion to improvement in performance, suggesting a shift away from hippocampal dependence as learning plateaus (*22*).

Three of the five patients produced a total of eight reports of Tetris imagery across three nights. These reports accounted for 7.4% of all hypnagogic reports collected, similar to that seen in normal participants (7.2%; Fig. 2A) (*23*). In contrast, there was only a single report of a thought about Tetris without clearly associated imagery (<1%; Fig. 2B). As with the normal participants, the visual imagery reported was highly stereotyped (*24*).

Normal controls without prior experience playing Tetris had first game scores averaging 786, significantly higher than that of the amnesics (*25*). They showed considerable improvement in performance over the 3 days and six sessions (Fig. 1) (*26*). The appearance of hypnagogic Tetris images in novices was very similar to that seen with amnesics. Nine of 12 Tetris novices produced a total of 19 reports of Tetris imagery over two nights of recording, a report rate very similar to that of the amnesics (Fig. 2A). The imagery described showed the same highly stereotyped content as seen with the amnesics (*27*). Novices produced 12 reports of thoughts of Tetris without imagery (Fig. 2B). Such reports, clearly linked to declarative memories, were virtually absent in the amnesics.

Ninety percent of the reports with images from these novices (17 out of 19) were obtained on the second night. In contrast, two-thirds of the reports of thoughts alone (8 out of 12) occurred on the first night (Fig. 2). Although this difference in distributions is

[1]Laboratory of Neurophysiology, Department of Psychiatry, Harvard Medical School, 74 Fenwood Road, Boston, MA 02115, USA. [2]Memory Disorders Research Center, Boston University Medical School, Division of Behavioral Neurology, Beth Israel Deaconess Medical Center, Boston, MA 02118 USA.

*To whom correspondence should be addressed. E-mail: rstickgold@hms.harvard.edu