Characterization of the piRNA Complex from Rat Testes

Nelson C. Lau, *et al.*

Science **313**, 363 (2006);
DOI: 10.1126/science.1130164

which is the start site for replication and transcription (*25*).

Electron micrographs of nucleocapsids from measles virus (*Paramyxoviridae*) and Marburg virus (*Filoviridae*), as well as an RNA-free crystal structure from the Borna virus (*Bornaviridae*) nucleoprotein, suggest that these nucleoproteins also adopt a two-domain structure (*26*, *27*). This suggests that these enveloped viruses use an RNA-sequestering mechanism similar to that observed for the rabies virus N-RNA complex. The N-RNA polymer has thus evolved as the ideal template for the polymerase activity, which exposes the genomic RNA only temporarily to the host cell defense systems during replication. The tight sequestering of RNA observed in the crystal structure suggests further that the closed N-RNA conformation might be stabilized or frozen by small molecules, which could thus act as antiviral agents preventing rabies virus replication.

Note added in proof: This version of the manuscript is slightly changed relative to the version that was published *Science* Express on June 15. Improved refinement of the structure led to a better definition of the following regions, which have been changed accordingly: the N terminus, including the β hairpin leading to helix 1; Tyr^{161}, Arg^{168}, and Arg^{434} contacting the RNA; as well as helices 13 to 16 of the C-terminal domain. None of the changes made influences the overall conformation of the structure or any of the conclusions drawn from the structure.

References and Notes

1. B. Dietzschold, M. Schnell, H. Koprowski, *Curr. Top. Microbiol. Immunol.* **292**, 45 (2005).
2. D. L. Knobel *et al.*, *Bull. World Health Organ.* **83**, 360 (2005).
3. R. A. Lamb, D. Kolakofsky, in *Fields Virology*, D. M. Knipe and P. M. Howley, Eds. (Lippincott, Williams and Wilkins, Philadelphia, ed. 4, 2001), pp. 1305–1340.
4. H. Arnheiter, N. L. Davis, G. Wertz, M. Schubert, R. A. Lazzarini, *Cell* **41**, 259 (1985).
5. D. Kolakofsky, P. Le Mercier, F. Iseni, D. Garcin, *Virology* **318**, 463 (2004).
6. O. Poch, I. Sauvaget, M. Delarue, N. Tordo, *EMBO J.* **8**, 3867 (1989).
7. S. U. Emerson, M. Schubert, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 5655 (1987).
8. D. Kolakofsky, *J. Virol.* **41**, 566 (1982).
9. D. Spehner, R. Drillien, P. M. Howley, *Virology* **232**, 260 (1997).
10. F. Iseni, A. Barge, F. Baudin, D. Blondel, R. W. Ruigrok, *J. Gen. Virol.* **79**, 2909 (1998).
11. B. M. Blumberg, C. Giorgi, D. Kolakofsky, *Cell* **32**, 559 (1983).
12. G. Schoehn, F. Iseni, M. Mavrakis, D. Blondel, R. W. Ruigrok, *J. Virol.* **75**, 490 (2001).
13. Supporting data are available on *Science* Online.
14. D. Thomas *et al.*, *J. Virol.* **54**, 598 (1985).
15. S. Lynch, D. Kolakofsky, *J. Virol.* **28**, 584 (1978).
16. S. Akira, S. Uematsu, O. Takeuchi, *Cell* **124**, 783 (2006).
17. I. Le Blanc *et al.*, *Nat. Cell Biol.* **7**, 653 (2005).
18. L. Espert *et al.*, *J. Biol. Chem.* **278**, 16151 (2003).
19. F. Iseni, F. Baudin, D. Blondel, R. W. Ruigrok, *RNA* **6**, 270 (2000).
20. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
21. B. Dietzschold *et al.*, *Virus Res.* **8**, 103 (1987).
22. H. Toriumi, A. Kawai, *Microbiol. Immunol.* **49**, 757 (2005).
23. S. Longhi *et al.*, *J. Biol. Chem.* **278**, 18638 (2003).
24. A. Kouznetzoff, M. Buckle, N. Tordo, *J. Gen. Virol.* **79**, 1005 (1998).
25. S. U. Emerson, *Cell* **31**, 635 (1982).
26. G. Schoehn *et al.*, *J. Mol. Biol.* **339**, 301 (2004).
27. M. G. Rudolph *et al.*, *Structure* **11**, 1219 (2003).

28. We thank D. Kolakofsky (Geneva), Y. Gaudin, and D. Blondel (Gif-sur-Yvette), for many discussions and ideas; F. Iseni and M. Mavrakis for early contributions to the project; S. Cusack for discussions and comments on the text; and all members of the EMBL European Synchrotron Radiation Facility (ESRF) Joint Structural Biology Group (JSBG) for access to the ESRF beam lines. This work was supported by the EMBL (W.W.), the Deutsche Forschungsgemeinschaft SFB 593 (W.W.), the Université Joseph Fourier (R.W.H.R.), the CNRS (G.S.), an EMBO fellowship (C.R.C.), and a Ph.D. fellowship from the French Ministry for Education, Research, and Technology (A.A.V.A.). Coordinates have been deposited in the Protein Data Bank (accession code no. 2GTT).

Supporting Online Material

www.sciencemag.org/cgi/content/full/1125280/DC1
Materials and Methods
Figs. S1 to S4
Table S1
References

23 January 2006; accepted 17 May 2006
Published online 15 June 2006;
10.1126/science.1125280
Include this information when citing this paper.

Characterization of the piRNA Complex from Rat Testes

Nelson C. Lau,[1]* Anita G. Seto,[1]* Jinkuk Kim,[2,3] Satomi Kuramochi-Miyagawa,[4] Toru Nakano,[4] David P. Bartel,[3,5] Robert E. Kingston[1]†

Small noncoding RNAs regulate processes essential for cell growth and development, including mRNA degradation, translational repression, and transcriptional gene silencing (TGS). During a search for candidate mammalian factors for TGS, we purified a complex that contains small RNAs and Riwi, the rat homolog to human Piwi. The RNAs, frequently 29 to 30 nucleotides in length, are called Piwi-interacting RNAs (piRNAs), 94% of which map to 100 defined (≤101 kb) genomic regions. Within these regions, the piRNAs generally distribute across only one genomic strand or distribute on two strands but in a divergent, nonoverlapping manner. Preparations of piRNA complex (piRC) contain rRecQ1, which is homologous to *qde-3* from *Neurospora*, a gene implicated in silencing pathways. Piwi has been genetically linked to TGS in flies, and slicer activity cofractionates with the purified complex. These results are consistent with a gene-silencing role for piRC in mammals.

Gene-silencing pathways guided by small RNAs, essential for maintaining proper cell growth and differentiation, operate at either the transcriptional or posttranscriptional level (*1*). Posttranscriptional gene silencing acts through mRNA destabilization or inhibition of mRNA translation (*1*), whereas TGS represses gene expression by altering chromatin conformation (*2*). Each pathway uses a core complex containing small RNA associated with a member of the Argonaute (Ago) protein family; however, the different mechanistic needs of each pathway require differences in complex composition. Although RNA-mediated TGS has been studied in fission yeast and other eukaryotes (*2*–*5*), the mechanism of this process in mammals remains elusive.

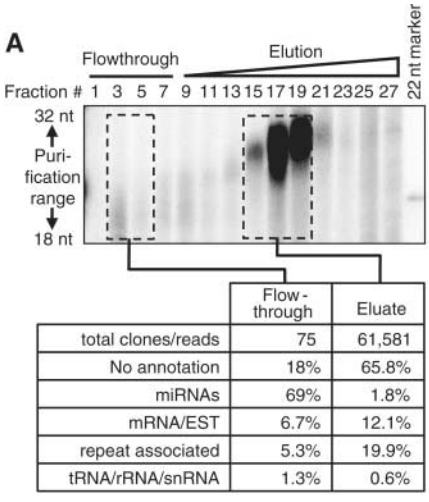
To identify candidate complexes for TGS in mammals, we exploited the previous observations that TGS might use small RNAs longer than the 21- to 23-nucleotide (nt) microRNAs (miRNAs). In *Arabidopsis*, *Tetrahymena*, *Drosophila*, and zebrafish, RNAs that are 24 nt and longer have been associated with TGS and/or genomic repeats, which are often silenced (*6*–*11*). In *Drosophila*, these repeat-associated small interfering RNAs (rasiRNAs) are enriched in the testis (*6*, *12*). Therefore, we prepared extract from rat testes and fractionated it on an ion-exchange Q column, monitoring the small RNAs. A peak of small RNAs longer than a 22-nt marker eluted in mild salt conditions, which suggested the presence of a novel ribonucleoprotein complex (Fig. 1A).

To characterize the small RNAs, we sequenced cDNA libraries made from flowthrough and eluate fractions, obtaining 61,581 reads from the eluate that matched perfectly to the *Rattus norvegicus* genome (*13*). In contrast to the flowthrough RNAs, which were mostly miRNAs (69%), the eluate RNAs derived primarily from regions of the genome not previously thought to be expressed (Fig. 1A). Some eluate reads matched expressed sequence tags (EST) (11%),

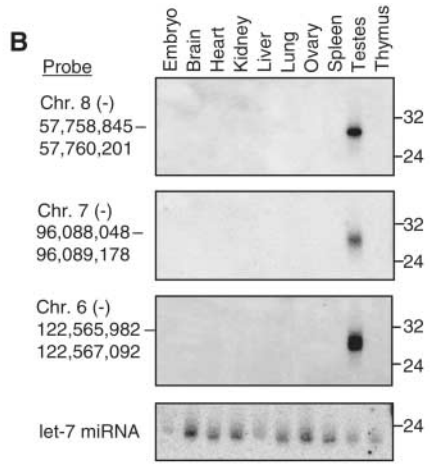
[1]Department of Molecular Biology, Massachusetts General Hospital, 185 Cambridge Street, Boston, MA 02114, USA. [2]Harvard-MIT Division of Health Sciences and Technology, E18-435, 77 Massachusetts Avenue, Cambridge, MA 02139, USA. [3]Howard Hughes Medical Institute and Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA. [4]Department of Molecular Cell Biology, Research Institute for Microbial Diseases, Osaka University, 3-1 Yamada-oka, Suita-shi, Osaka 565-0871, Japan. [5]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: kingston@molbio.mgh.harvard.edu

but only a small fraction matched annotated mRNAs (< 1.1%). Some also matched repeats (20%), but when considering that ~40% of the genome is annotated as repeats (*13*), the eluate reads were depleted in repeat sequences and thus, as a class, did not represent rasiRNAs.



	Flow-through	Eluate
total clones/reads	75	61,581
No annotation	18%	65.8%
miRNAs	69%	1.8%
mRNA/EST	6.7%	12.1%
repeat associated	5.3%	19.9%
tRNA/rRNA/snRNA	1.3%	0.6%



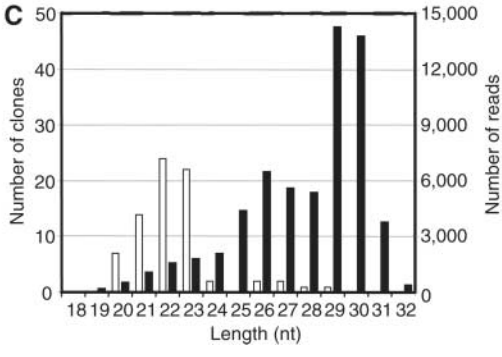


Fig. 1. Testes contain a longer class of small RNAs. (**A**) (Top) Rat testes extract was fractionated on a Q column (0.1 to 1 M potassium acetate gradient). RNA from fractions was end-labeled and resolved on a gel. (Bottom) Small RNAs from column fractions were gel-purified (dashed boxes), converted to cDNAs, and sequenced. (**B**) Rat tissue Northern blot hybridized with body-labeled RNA probes corresponding to small RNA sequences. The blot was stripped before reprobing for the indicated chromosomal regions or let-7 miRNA (loading control). Migration of RNA markers is indicated (right). (**C**) Size distribution of small RNAs from flowthrough (white bars), and eluate (black bars). *Y*-axis scales are different for flowthrough RNAs (left) and eluate RNAs (right).

The eluate RNAs were mostly 25 to 31 nt in length (Fig. 1A), and Northern blot analysis indicated a testis-specific expression pattern (Fig. 1B). Most eluate RNAs began with a 5′ uridine (~84%), but no other sequence features or motifs were detected. A dominant subpopulation at 29 to 30 nt was observed (Fig. 1C); however, these 29- to 30-nt oligomers could not be distinguished from most of the remaining eluate reads by other criteria, including 5′ nucleotide, genomic locus, and annotation. Thus, all the eluate RNAs that did not match annotated noncoding RNAs (miRNA, tRNA, rRNA, and snRNA) were considered together as representing a single newly identified class of small RNAs.

To understand potential functions for these RNAs, we purified the associated proteins. By monitoring the RNAs, we developed a five-step scheme to purify the native complex to near homogeneity (Fig. 2, A and B). Mass spectrometry of the purified complex identified the rat homologs to Piwi (Riwi) and the human RecQ1 protein (Fig. 2 and fig. S1). Western blotting confirmed the copurification of Riwi and rRecQ1 with the small RNAs (Fig. 2C, but see also independent purification described below). We designate RNAs found with rat Piwi to be Piwi-interacting RNAs (piRNAs) and the complex to be the piRNA complex (piRC).

To gain insight into the origins of piRNAs, we examined the genomic loci from which they presumably derived. About two-thirds of the piRNA sequences each perfectly matched a single locus, and in some cases that specific locus was matched by multiple reads (up to 149). For the remaining one-third of the reads, which each mapped to multiple loci

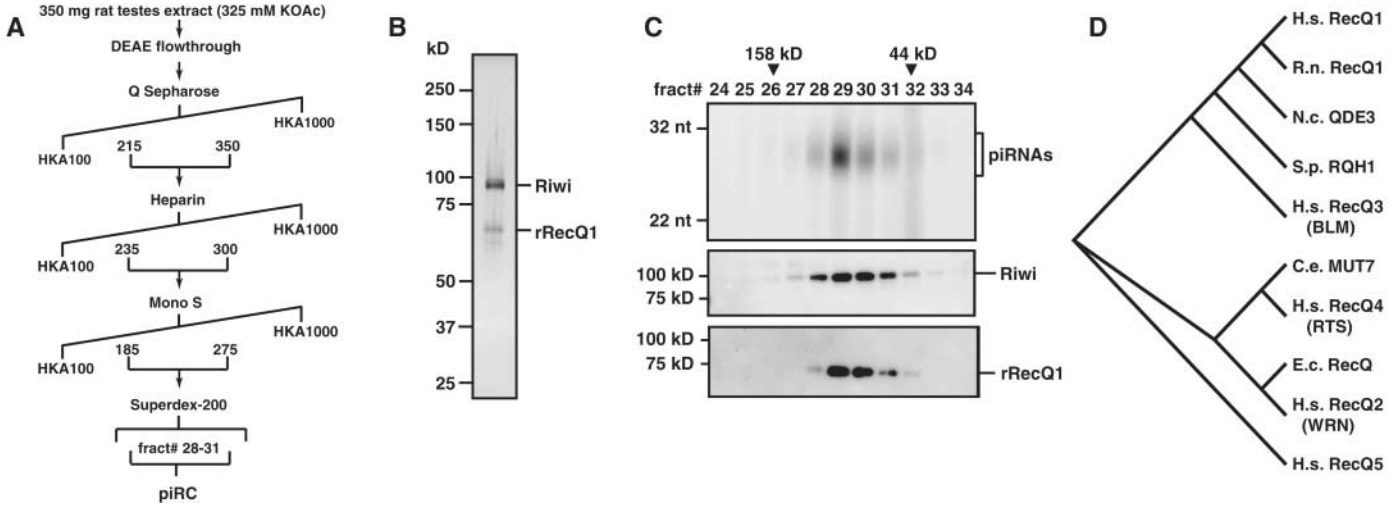


Fig. 2. Purification of a native small RNA-containing complex revealed Riwi and rRecQ1. (**A**) Schematic of piRC fractionation steps from rat testes extract. Numbers represent potassium acetate (KOAc) concentration (mM). (**B**) Proteins from the peak fraction of piRC (Superdex-200 column) were resolved on a gel and silver stained. Bands were excised and identified by mass spectrometry to be Riwi and rRecQ1 (See fig. S1). (**C**) Small RNAs, Riwi, and rRecQ copurify after five steps of chromatographic separation. Final Superdex-200 column fractions were assayed for the presence of small RNAs, Riwi, or rRecQ1. (Top) Small RNAs were end-labeled and resolved on a gel. (Middle and bottom) Western blots probed with antibodies to Miwi (mouse Piwi) and to hRecQ1. Elution profile of protein size markers from Superdex-200 column indicated above. (**D**) Phylogenetic comparison of RecQ DNA helicase family members revealed *Neurospora* QDE-3 to be a close homolog to rRecQ1.

(up to 25,044 loci), we normalized the number of reads by the number of genomic hits and assigned this normalized hit count equally to all the loci; thus, a piRNA read with four perfect genomic hits contributed a quarter of a count to each of its four loci. Counts were integrated into bins and plotted across each chromosome. The majority of counts (94%) fell into 100 genomic clusters that each contained at least 20 uniquely mapping reads (table S1). As exemplified by four clusters on chromosome 20 (Fig. 3A) and illustrated for all 100 clusters (fig. S2), the clusters distributed across the genome; however, some chromosomes were underrepre-

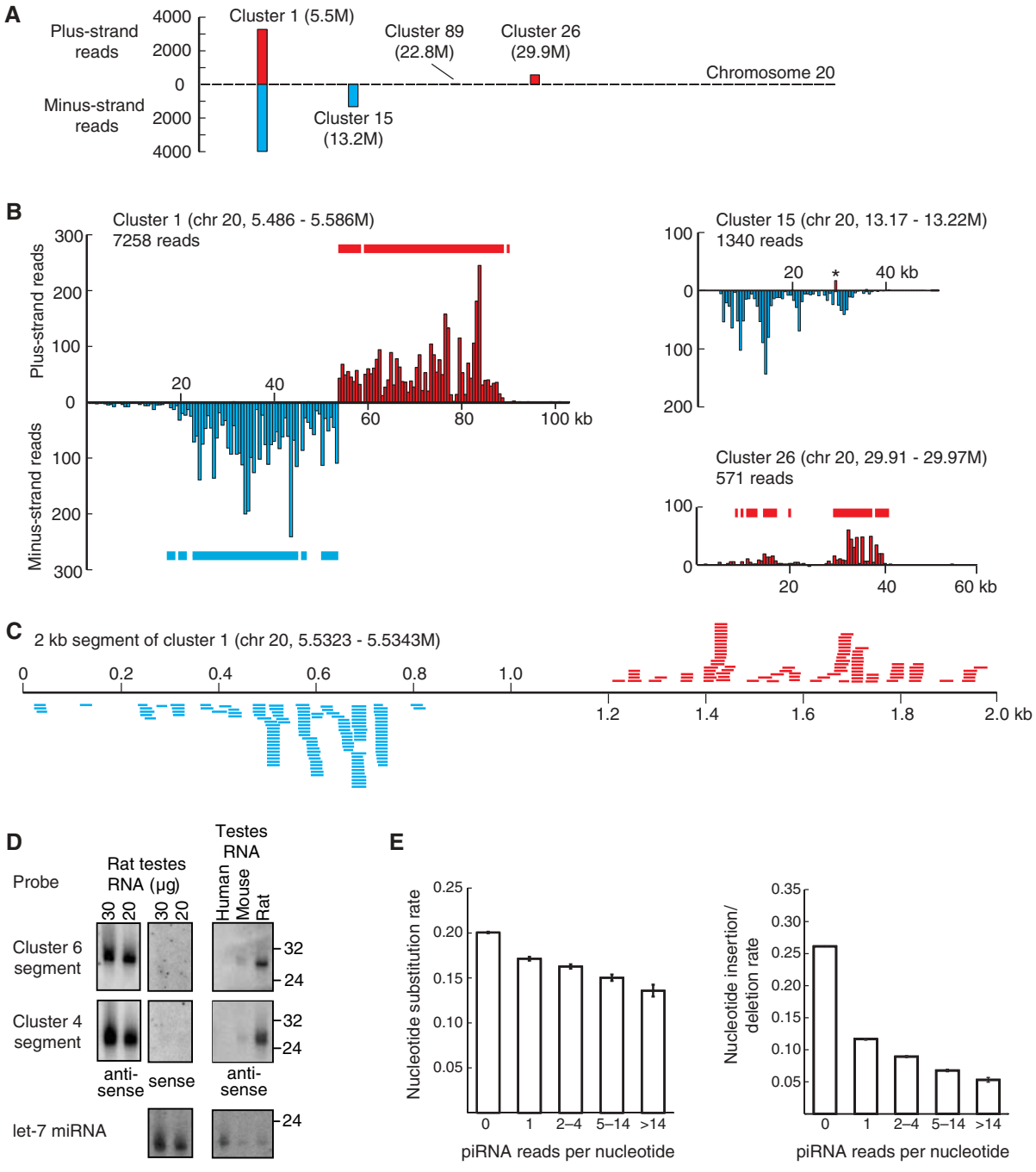


Fig. 3. Genomic characteristics of rat piRNAs. (**A**) Chromosomal view indicating the number of piRNA reads mapping to clusters on chromosome (chr) 20 [(*27*)]. (**B**) Medium-resolution view of clusters 1, 15, and 26. Horizontal bars above and below the histograms indicate regions orthologous to mouse regions that also produce piRNAs (indicated if the bin matches >2 uniquely mapping mouse piRNAs) (tables S1 and S3). An asterisk denotes a group of piRNAs from cluster 15 that perfectly mapped to more than one locus in the genome. (**C**) High-resolution view centered on the gap region that separates minus- and plus-strand piRNA hits in cluster 1. Horizontal bars represent individual piRNAs. (**D**) Northern blot analysis with probes to the indicated clusters, testing strand-specific piRNA expression (left and middle) and cross-hybridization to mouse and human testes RNAs (right). Migration of RNA markers is indicated (far right). Blots were stripped and reprobed to let-7 miRNA (loading control). (**E**) piRNA conservation analysis. Orthologous rat and mouse clusters were identified, and rat residues were binned based on the number of matching rat piRNA reads. The estimated substitution rate per residue (left) and estimated insertion/deletion rate (right), comparing rat to mouse, was calculated for each bin. Error bars indicate 95% confidence intervals of the estimates (*27*).

sented in piRNA hits and clusters (fig. S2B). These clusters spanned 1 to 101 kb (table S1) and in aggregate made up less than 0.1% of the rat genome.

Known silencing RNAs (siRNAs and miRNAs) derive from double-stranded RNA precursors or foldback structures (*1*). In contrast, piRNAs of most clusters mapped exclusively to either the plus or minus genomic strand in irregular, sometimes overlapping, patterns, with no evidence of extensive foldback structures or double-stranded origins (Fig. 3B). Sixteen clusters, such as cluster 1 (Fig. 3, B and C), contained regions of minus- and plus-strand hits that were juxtaposed with each other but separated by a gap of ~100 to 800 base pairs (bps), an orientation that suggests divergent, bidirectional transcription, starting within the gap that separated the two distributions (table S1). Only two clusters had hits that suggested convergent or overlapping transcription (table S1, clusters 31 and 38). Northern blot analysis confirmed that piRNAs derived predominantly from one of the two genomic strands (Fig. 3D). Reverse transcription polymerase chain reaction results suggested that longer transcripts of the same polarity, perhaps piRNA precursors, also derived from these regions (fig. S3).

Analogous production of piRNAs from at least 94 clusters occurred in the mouse, as indicated by the analysis of 68,794 piRNA reads generated in the same manner as those of the rat (Fig. 3B, fig. S4, and tables S3 and S4). Most of the mouse clusters were homologous to rat clusters, with strikingly similar strand specificity and abundance profiles (Fig. 3B and tables S1 and S3). Nonetheless, their sequence conservation was low. Probes against rat clusters 4 and 6 hybridized only weakly to mouse piRNAs, as expected by the numerous point substitutions in the orthologous mouse piRNAs (Fig. 3D and fig. S5). Overall, the single-nucleotide substitution rate of the piRNA clusters was within the 15 to 20% expected for neutral residues (*13*). Nevertheless, residues represented by more reads had lower substitution and insertion/deletion rates, indicating detectable evolutionary pressure to conserve the sequence of the abundant piRNAs (Fig. 3E). We conclude that the production of piRNAs is highly conserved, but the sequence identities of the piRNAs are only weakly conserved. The weak conservation favors models in which piRNAs target the loci/transcripts that correspond to the same loci from which they derive.

We characterized two potential biochemical functions of piRC suggested by activities previously attributed to RecQ and Ago family members. Human RecQ1 is an adenosine triphosphate (ATP)–dependent DNA helicase (*14*). Both adenosine triphosphatase (ATPase) and DNA unwinding activities followed the rRecQ1 protein of piRC (Fig. 4, A and B, and fig. S6). Riwi contains the catalytic residues that other Ago proteins use for RNA-guided cleavage of target RNAs (*15*) (fig. S7A). Using a substrate complementary to a piRNA, we detected cleavage activity, peaking with fractions containing Riwi and piRNAs (Fig. 4D). However, it was not robust, perhaps because of the small representation of the cognate piRNA in the diverse population of piRNAs (<0.2%).

Our purification of piRC uncovered a novel class of small RNAs and identified as copurifying factors Riwi and rRecQ1, two proteins with intriguing functions genetically determined in other species. Piwi represents a subclade of the Ago family of proteins (*16*) and was first discovered to regulate germ stem cell maintenance in *Drosophila* (*17*). Subsequently, mammalian Piwi members were found to regulate germ cell maturation (*18*, *19*). Drosophila *piwi* mutants are also defective in small

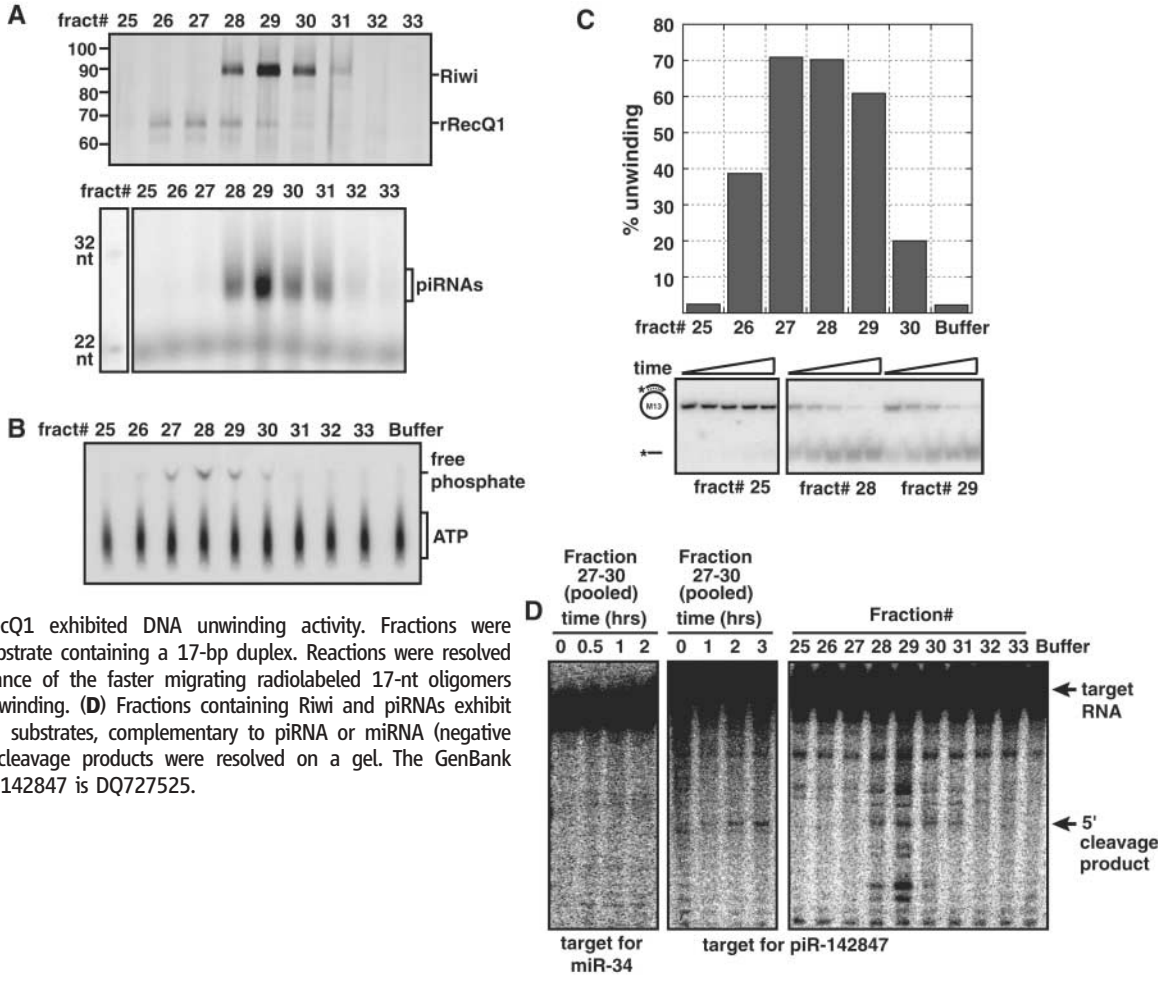


Fig. 4. piRC fractions contained ATP-dependent DNA helicase and slicer activities. (**A**) Visualization of proteins (top, silver stain) and end-labeled small RNAs (bottom) in fractions from final Superdex-200 column (independent purification from that shown in Fig. 2C). Size standards indicated on left. (**B**) Fractions containing rRecQ1 exhibit ATPase activity. Fractions were incubated with radiolabeled ATP. Free phosphate generated by ATPase activity was separated from unhydrolyzed ATP on thin-layer chromatography. (**C**) Fractions containing rRecQ1 exhibited DNA unwinding activity. Fractions were incubated with a DNA substrate containing a 17-bp duplex. Reactions were resolved on a native gel. Appearance of the faster migrating radiolabeled 17-nt oligomers indicated DNA duplex unwinding. (**D**) Fractions containing Riwi and piRNAs exhibit slicer activity. Cap-labeled substrates, complementary to piRNA or miRNA (negative control) sequences, and cleavage products were resolved on a gel. The GenBank accession number for piR-142847 is DQ727525.

RNA-dependent transgene and retrotransposon silencing (*20*, *21*) and lose the inability to localize heterochromatic proteins, including the repressive Polycomb-group proteins (*22*, *23*). *Tetrahymena* Piwi (*TIWI*) is needed for siRNA-mediated DNA elimination (*24*).

In *Neurospora*, a screen for mutants in quelling (gene silencing during vegetative growth) identified both QDE-2, an Ago-family protein, and QDE-3, a RecQ1 homolog (*25*, *26*). When compared with RecQ homology in mammals and other organisms, *Neurospora* QDE-3 resided in the same clade as rRecQ1 (Fig. 2D and fig. S7B). rRecQ1 did not always precisely cofractionate with Riwi and the piRNAs during our final purification step (Fig. 4A). The lack of tight association of rRecQ1 might have reflected conditions specific to this step or might indicate that rRecQ1 is generally less tightly associated with piRNAs than is Riwi. Perhaps rRecQ1 is not critical for piRC function. However, the genetic links between the QDE-2 and QDE-3 silencing factors suggest that the biochemical association between Riwi and rRecQ1 has biological importance and, furthermore, implies a gene-silencing function for piRC. Addressing the functions of piRC and the biogenesis and localization of the piRNAs will be important questions for elucidating the potential for piRC to regulate the genome.

References and Notes

1. P. D. Zamore, B. Haley, *Science* **309**, 1519 (2005).
2. M. A. Matzke, J. A. Birchler, *Nat. Rev. Genet.* **6**, 24 (2005).
3. I. M. Hall *et al.*, *Science* **297**, 2232 (2002).
4. A. Verdel *et al.*, *Science* **303**, 672 (2004).
5. T. A. Volpe *et al.*, *Science* **297**, 1833 (2002).
6. A. A. Aravin *et al.*, *Dev. Cell* **5**, 337 (2003).
7. P. Y. Chen *et al.*, *Genes Dev.* **19**, 1288 (2005).
8. A. Hamilton, O. Voinnet, L. Chappell, D. Baulcombe, *EMBO J.* **21**, 4671 (2002).
9. S. R. Lee, K. Collins, *Genes Dev.* **20**, 28 (2006).
10. K. Mochizuki, M. A. Gorovsky, *Genes Dev.* **18**, 2068 (2004).
11. D. Zilberman, X. Cao, S. E. Jacobsen, *Science* **299**, 716 (2003).
12. A. A. Aravin *et al.*, *Mol. Cell. Biol.* **24**, 6742 (2004).
13. R. A. Gibbs *et al.*, *Nature* **428**, 493 (2004).
14. S. Cui *et al.*, *J. Biol. Chem.* **278**, 1424 (2003).
15. F. V. Rivas *et al.*, *Nat. Struct. Mol. Biol.* **12**, 340 (2005).
16. M. A. Carmell, Z. Xuan, M. Q. Zhang, G. J. Hannon, *Genes Dev.* **16**, 2733 (2002).
17. D. N. Cox *et al.*, *Genes Dev.* **12**, 3715 (1998).
18. W. Deng, H. Lin, *Dev. Cell* **2**, 819 (2002).
19. S. Kuramochi-Miyagawa *et al.*, *Development* **131**, 839 (2004).
20. A. I. Kalmykova, M. S. Klenov, V. A. Gvozdev, *Nucleic Acids Res.* **33**, 2052 (2005).
21. M. Pal-Bhadra, U. Bhadra, J. A. Birchler, *Mol. Cell* **9**, 315 (2002).
22. C. Grimaud *et al.*, *Cell* **124**, 957 (2006).
23. M. Pal-Bhadra *et al.*, *Science* **303**, 669 (2004).
24. K. Mochizuki, N. A. Fine, T. Fujisawa, M. A. Gorovsky, *Cell* **110**, 689 (2002).
25. C. Catalanotto, G. Azzalin, G. Macino, C. Cogoni, *Genes Dev.* **16**, 790 (2002).
26. C. Cogoni, G. Macino, *Science* **286**, 2342 (1999).
27. Materials and methods are available as supporting material on *Science* Online.

28. The order of listing of the first two authors is arbitrary. We thank N. Francis for initial work on this project; W. Johnston for technical assistance; L. Davidow, J. Morris, L. Lim, and J. Ruby for bioinformatics assistance; Z. Zhang and J. Goldman for performing preliminary assays; D. Schwarz for advice on slicer assays; and C. Woo, E. Troemel, J. Song, and S. Aigner for comments on the manuscript. Supported by Helen Hay Whitney (N.L.) and Damon Runyon (A.S.) postdoctoral fellowships, a Korea Foundation for Advanced Studies predoctoral fellowship (J.K.), and grants from NIH (D.B. and R.K.). D.B. is a Howard Hughes Medical Institute investigator.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1130164/DC1
Materials and Methods
SOM Text
Figs. S1 to S7
Tables S1 to S6

18 May 2006; accepted 7 June 2006
Published online 15 June 2006;
10.1126/science.1130164
Include this information when citing this paper.

Aneuploidy and Isochromosome Formation in Drug-Resistant *Candida albicans*

Anna Selmecki,[1] Anja Forche,[1] Judith Berman[1,2]*

Resistance to the limited number of available antifungal drugs is a serious problem in the treatment of *Candida albicans*. We found that aneuploidy in general and a specific segmental aneuploidy, consisting of an isochromosome composed of the two left arms of chromosome 5, were associated with azole resistance. The isochromosome forms around a single centromere flanked by an inverted repeat and was found as an independent chromosome or fused at the telomere to a full-length homolog of chromosome 5. Increases and decreases in drug resistance were strongly associated with gain and loss of this isochromosome, which bears genes expressing the enzyme in the ergosterol pathway targeted by azole drugs, efflux pumps, and a transcription factor that positively regulates a subset of efflux pump genes.

Candida albicans is the most prevalent human fungal pathogen and is especially problematic in immune-compromised individuals. Azoles are antifungal drugs used extensively in the therapy of *C. albicans* infections because they cause few side effects. Resistance to azoles arises during long-term low-level prophylactic treatment regimes (*1*). The evolution of azole-resistance can occur via different pathways, e.g., increased activity of transcription factors that regulate drug pumps (*2*) or mutations in the ergosterol biosynthetic pathway (*3*), and may be facilitated by the heat shock protein Hsp90 (*4*). Although *C. albicans* does not have a complete sexual cycle, surveys of clinical strains suggest that it tolerates genome flexibility that generates a moderate level of selectable genetic diversity (*5*–*11*).

We adapted comparative genome hybridization (CGH) arrays for the analysis of gene copy number at all loci (*10*) for 70 azole-resistant and azole-sensitive strains from clinical and laboratory sources (table S1). We found 37 aneuploid chromosomes in 23 strains (Fig. 1A, black and white bars). Aneuploidy was seven times as prevalent in fluconazole-resistant [Flu^R, minimum inhibitory concentration (MIC) $\geq$ 4] (21 out of 42 strains) as in fluconazole-sensitive (Flu^S, MIC < 4) (2 out of 28 strains) isolates. Aneuploidy, primarily trisomy, was most prevalent on chr5 (15 events), which also exhibited a high level of segmental aneuploidy (8 events) (Fig. 1A, gray bars) (*10*).

All eight Flu^R strains with a chr5 segmental aneuploidy displayed distinctive features by CGH: increased gene copy number on the left arm of the chromosome; one or two copies of genes on the right arm of the chromosome; and, most strikingly, the breakpoint between the two arms was identical (Fig. 1B). This common breakpoint was located between open reading frames orf19.3161 and orf19.4219, which flank a gap between contigs 19.10170 and 19.10202 in the version 19 assembly of the genome sequence. We sequenced the region between these two contigs (fig. S1) from strains carrying either one of the two chr5 homologs, which allowed us to distinguish alleles found in this partially heterozygous region of the genome. The sequence gap was within one arm of an inverted repeat flanking the binding site for Cse4p, the histone H3 variant that is necessary for centromere function (*12*) (Fig. 1B, diagram).

The combination of CGH data and Southern blot analysis revealed an isochromosome (two identical chromosome arms flanking a centromere) in the eight strains with extra copies of chr5L. An isochromosome composed of two chr5L arms should include a unique fragment,

[1]Department of Genetics, Cell Biology, and Development, [2]Department of Microbiology, University of Minnesota, 6-160 Jackson Hall, 321 Church Street SE, Minneapolis, MN 55455, USA.

*To whom correspondence should be addressed. E-mail: jberman@umn.edu