Interactions between RNA polymerase and the "core recognition element" counteract pausing

Irina O. Vvedenskaya *et al.*

Science **344**, 1285 (2014);
DOI: 10.1126/science.1253458

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.

stratification of genetic variation, negative results and lack of replication are likely to dominate the outcome of genetic studies in uncharacterized populations. Here we have demonstrated a high degree of fine-scale genomic structure across Mexico, shaped by pre-Columbian population dynamics and affecting the present-day genomes of Mexican mestizos, which is of both anthropological and biomedical relevance. Studies such as this one are crucial for enabling precision medicine, providing novel data resources, empowering the next generation of genetic studies, and demonstrating the importance of understanding and measuring fine-scale population structure and its associations with biomedical traits.

REFERENCES AND NOTES

1. S. Gravel *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 11983–11988 (2011).
2. S. Wang *et al.*, *PLOS Genet.* **3**, e185 (2007).
3. V. Acuña-Alonzo *et al.*, *Hum. Mol. Genet.* **19**, 2877–2885 (2010).
4. A. L. Williams *et al.*, *Nature* **506**, 97–101 (2014).
5. R. Lisker, E. Ramírez, V. Babinsky, *Hum. Biol.* **68**, 395–404 (1996).
6. K. Sandoval *et al.*, *Am. J. Phys. Anthropol.* **148**, 395–405 (2012).
7. A. Gorostiza *et al.*, *PLOS ONE* **7**, e44666 (2012).
8. D. Reich *et al.*, *Nature* **488**, 370–374 (2012).
9. See supplementary materials on *Science* Online.
10. J. Novembre *et al.*, *Nature* **456**, 98–101 (2008).
11. D. M. Altshuler *et al.*, *Nature* **467**, 52–58 (2010).
12. B. M. Henn *et al.*, *PLOS ONE* **7**, e34267 (2012).
13. J. Hey, *PLOS Biol.* **3**, e193 (2005).
14. J. K. Pickrell, J. K. Pritchard, *PLOS Genet.* **8**, e1002967 (2012).
15. A. Pascual Soto, *El Tajín. En Busca de los Orígenes de una Civilización* (UNAM-INAH, Mexico, 2006).
16. L. Campbell, T. Kaufman, *Annu. Rev. Anthropol.* **14**, 187–198 (1985).
17. M. R. Nelson *et al.*, *Am. J. Hum. Genet.* **83**, 347–358 (2008).
18. D. H. Alexander, J. Novembre, K. Lange, *Genome Res.* **19**, 1655–1664 (2009).
19. A. Moreno-Estrada *et al.*, *PLOS Genet.* **9**, e1003925 (2013).
20. A. Brisbin *et al.*, *Hum. Biol.* **84**, 343–364 (2012).
21. N. A. Johnson *et al.*, *PLOS Genet.* **7**, e1002410 (2011).
22. L. R. Botigué *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 11791–11796 (2013).
23. S. Wang *et al.*, *PLOS Genet.* **4**, e1000037 (2008).
24. I. Silva-Zolezzi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **21**, 8611–8616 (2009).
25. M. A. Nalls *et al.*, *Am. J. Hum. Genet.* **82**, 81–87 (2008).
26. C. A. Peralta *et al.*, *Am. J. Nephrol.* **31**, 202–208 (2010).
27. L. Fejerman *et al.*, *Cancer Res.* **68**, 9723–9728 (2008).
28. J. L. Hankinson, J. R. Odencrantz, K. B. Fedan, *Am. J. Respir. Crit. Care Med.* **159**, 179–187 (1999).
29. R. Kumar *et al.*, *N. Engl. J. Med.* **363**, 321–330 (2010).
30. K. Salari *et al.*, *Genet. Epidemiol.* **29**, 76–86 (2005).
31. D. B. Hancock *et al.*, *PLOS Genet.* **5**, e1000623 (2009).
32. D. G. Torgerson *et al.*, *J. Allergy Clin. Immunol.* **130**, 76, e12 (2012).

ACKNOWLEDGMENTS

We thank all volunteers for generously donating DNA samples and participating in the study. This project was possible with the joint support from multiple Institutions in Mexico and the United States. Stanford University supported C.D.B. with funding from the Department of Genetics. INMEGEN received support from the Federal Government of Mexico, particularly the Ministry of Health, the Mexican Health Foundation (FUNSALUD), and the Gonzalo Río Arronte Foundation. State governments and universities of Durango, Campeche, Guanajuato, Guerrero, Oaxaca, Sonora, Tamaulipas, Veracruz, Yucatan, and Zacatecas contributed significantly to this work. This research was also supported by the George Rosenkranz Prize for Health Care Research in Developing Countries awarded to A.M.-E.; University of California San Francisco (UCSF) Chancellor's Research Fellowship, Dissertation Year Fellowship, and NIH Training Grants T32GM007175 and T32HG000044 (to C.R.G.); the Robert Wood Johnson Foundation Amos Medical Faculty Development Award; the Sandler Foundation; the American Asthma Foundation (to E.G.B.); CONACYT grant 129693 (to H.R.-V.); BBSRC grant BB/I021213/1 (to A.R.-L.); the National Institutes of Health (NIH) (grants R01GM090087, R01HG003229, ES015794, GM007546, GM061390, HL004464, HL078885, HL088133, HL111636, RR000083, P60MD006902, and ZIA ES49019); and National Science Foundation award DMS-1201234. This work was supported in part by the Intramural Research Program of NIH, National Institute of Environmental Health Sciences (to S.J.L.). Some computations were performed using the UCSF Biostatistics High Performance Computing System. We also thank B. Henn, S. Gravel, and J. Byrnes for helpful discussions; C. Gunter and M. Carpenter for editing the manuscript; and M. Morales for informatics and programming support. C.D.B. is on the advisory board of a project at 23andMe; and on the scientific advisory boards of Personalis, Inc.; InVitae; Etalon, Inc.; and Ancestry.com. The collections and methods for the Population Reference Sample (POPRES) are described by Nelson *et al.* (2008). The POPRES data sets used for the analyses described here were obtained from dbGaP through accession number phs000145.v1.p1. Access to the MCCAS data set may be obtained under the terms of a data transfer agreement with the National Institute of Environmental Health Sciences; the contact is S.J.L.. Individual-level genotypes for new data presented in this study are available, through a data access agreement to respect the privacy of the participants for the transfer of genetic data, by contacting C.D.B., A.M.-E., and INMEGEN (http://www.inmegen.gob.mx/).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/344/6189/1280/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S20
Tables S1 to S6
References (*33*–*64*)

3 February 2014; accepted 21 May 2014
10.1126/science.1251688

TRANSCRIPTION

Interactions between RNA polymerase and the "core recognition element" counteract pausing

Irina O. Vvedenskaya,[1]* Hanif Vahedian-Movahed,[2]* Jeremy G. Bird,[1,2]* Jared G. Knoblauch,[1] Seth R. Goldman,[1] Yu Zhang,[2] Richard H. Ebright,[2]† Bryce E. Nickels[1]†

Transcription elongation is interrupted by sequences that inhibit nucleotide addition and cause RNA polymerase (RNAP) to pause. Here, by use of native elongating transcript sequencing (NET-seq) and a variant of NET-seq that enables analysis of mutant RNAP derivatives in merodiploid cells (mNET-seq), we analyze transcriptional pausing genome-wide in vivo in *Escherichia coli*. We identify a consensus pause-inducing sequence element, $G_{-10}Y_{-1}G_{+1}$ (where –1 corresponds to the position of the RNA 3′ end). We demonstrate that sequence-specific interactions between RNAP core enzyme and a core recognition element (CRE) that stabilize transcription initiation complexes also occur in transcription elongation complexes and facilitate pause read-through by stabilizing RNAP in a posttranslocated register. Our findings identify key sequence determinants of transcriptional pausing and establish that RNAP-CRE interactions modulate pausing.

Regulation of gene expression during transcription elongation often involves sequences in DNA that cause the transcription elongation complex (TEC) to pause. Pausing can affect gene expression by facilitating engagement of regulatory factors, influencing formation of RNA secondary structures, and enabling synchronization of transcription and translation.

Several lines of evidence suggest that pausing involves specific sequence signals that inhibit nucleotide addition (*1*–*11*). To define key sequence determinants for pausing, we used native elongating transcript sequencing (NET-seq), which permits occupancies of TECs to be mapped genome-wide with base-pair resolution (*12*, *13*) (fig. S1). The occupancy of the TEC at a given position is correlated with the tendency of the TEC to pause at the position. Accordingly, NET-seq analysis enables identification of pause sites. To perform NET-seq in *Escherichia coli*, cells carrying a chromosomal *rpoC-3xFLAG* gene, encoding RNAP β′ subunit with a C-terminal 3xFLAG tag were grown to midexponential phase; cells were flash-frozen and lysed; 3xFLAG-tagged TECs were immunoprecipitated with an antibody against FLAG; RNAs were extracted from TECs; and RNA 3′ ends were converted to cDNAs and analyzed using high-throughput sequencing. We defined pause sites as positions where TEC occupancy exceeded TEC occupancy at each position 25 base pairs (bp) upstream and downstream. We identified 15,553 pause sites, which corresponds to ~19,800 total pause sites, given the estimated ~78% saturation of the analysis (tables S1 to S7). Alignment of pause-site sequences revealed a clear consensus pause element (PE): $G_{-10}Y_{-1}G_{+1}$, where position –1 corresponds to the position of the RNA 3′ end (Fig. 1A and fig. S2). Of the

[1]Department of Genetics and Waksman Institute, Rutgers University, Piscataway, NJ 08854, USA. [2]Department of Chemistry and Waksman Institute, Rutgers University, Piscataway, NJ 08854, USA.
*These authors contributed equally. **†Corresponding author. E-mail: bnickels@waksman.rutgers.edu (B.E.N.); ebright@waksman.rutgers.edu (R.H.E.)**

identified pause sites, ~35% exhibited a 3-of-3 match to the consensus PE, and ~42% exhibited a 2-of-3 match to the consensus PE (tables S4 and S5). Comparing the total number of pause sites with a 3-of-3 match (~6900) to the total number of sequences in the transcribed portion of the genome with a 3-of-3 match (~43,500 to 58,000) (*14*) indicates that, under these conditions, functional pausing occurs at ~12 to 16% of sequences with a 3-of-3 match.

To validate the NET-seq results, we selected two consensus PEs for analysis in vitro. One, located in *yrbL*, has –1C, and the other, located in *gltP*, has –1T (Fig. 1B and fig. S3). Transcription assays show that the *yrbL* PE and *gltP* PE induce pausing in vitro (Fig. 1B and fig. S4) and that consensus base pairs at positions +1, –1, and –10 are required for efficient pause capture in vitro (Fig. 1B).

We next determined the contribution of individual bases of the consensus PE to pausing at the *yrbL* PE (Fig. 2A). The results indicate that introduction of each of the three nonconsensus base pairs at position +1 (C, A, and T) reduces pausing; introduction of each of the two nonconsensus base pairs at position –1 (A and G) reduces pausing; and introduction of two of three nonconsensus base pairs at position –10 (A and T) reduces pausing (Fig. 2A). In each case, the effect on pausing is manifest at the level of pause capture efficiency.

The consensus PE comprises sequence determinants that previously have been implicated as important for pausing: G at +1 (*10*, *15*), T or C at –1 (*3*, *5*, *6*, *10*), and G at –10 (*4*, *10*). The consensus PE is especially similar to a consensus pause–inducing sequence identified by Herbert *et al.* in single-molecule studies (*10*).

The sequence determinants in the consensus PE ($G_{-10}Y_{-1}G_{+1}$) can be understood in terms of the structure and mechanism of the TEC (Fig. 2B). In each nucleotide-addition cycle in transcription elongation, RNAP translocates between a "pretranslocated state," in which the RNAP active center "*i*" and "*i* + 1" sites interact with RNA positions –2 and –1, and a "posttranslocated state,"

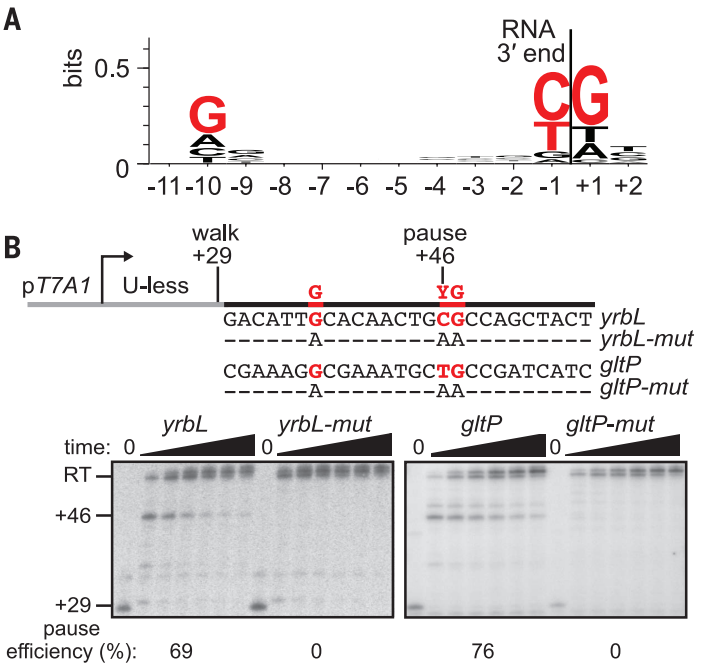
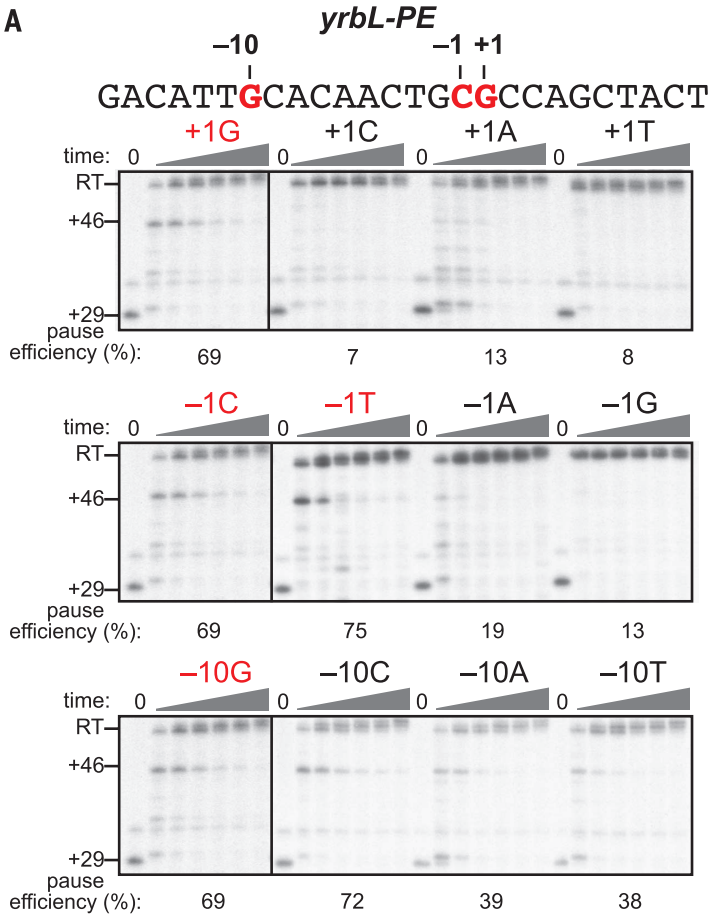


Fig. 1. Identification of consensus PE. (**A**) Sequence logo for consensus PE from NET-seq. Red, bases with ≥0.2 bit sequence-information content. (**B**) In vitro transcription assays with consensus PEs and mutant PEs. (Top) Templates; (bottom) results. +29, RNA before addition of uridine 5′-triphosphate (UTP); +46, RNA in TEC at PE; RT, read-through RNA; red, consensus PE bases.



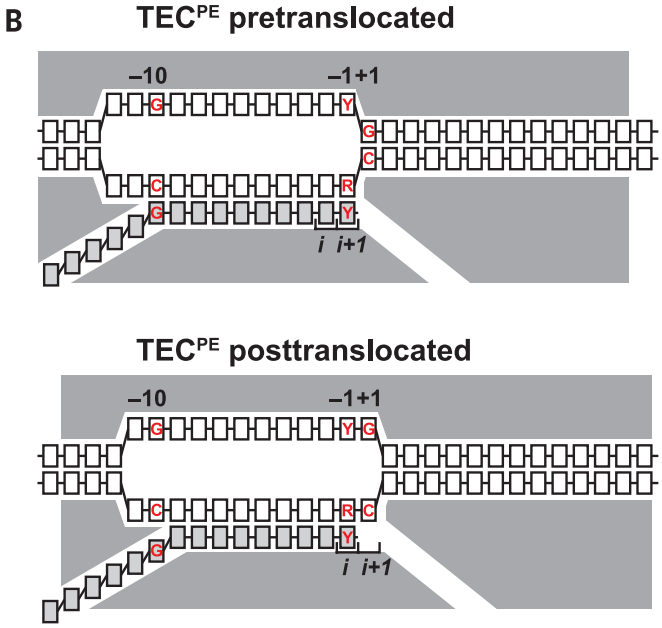


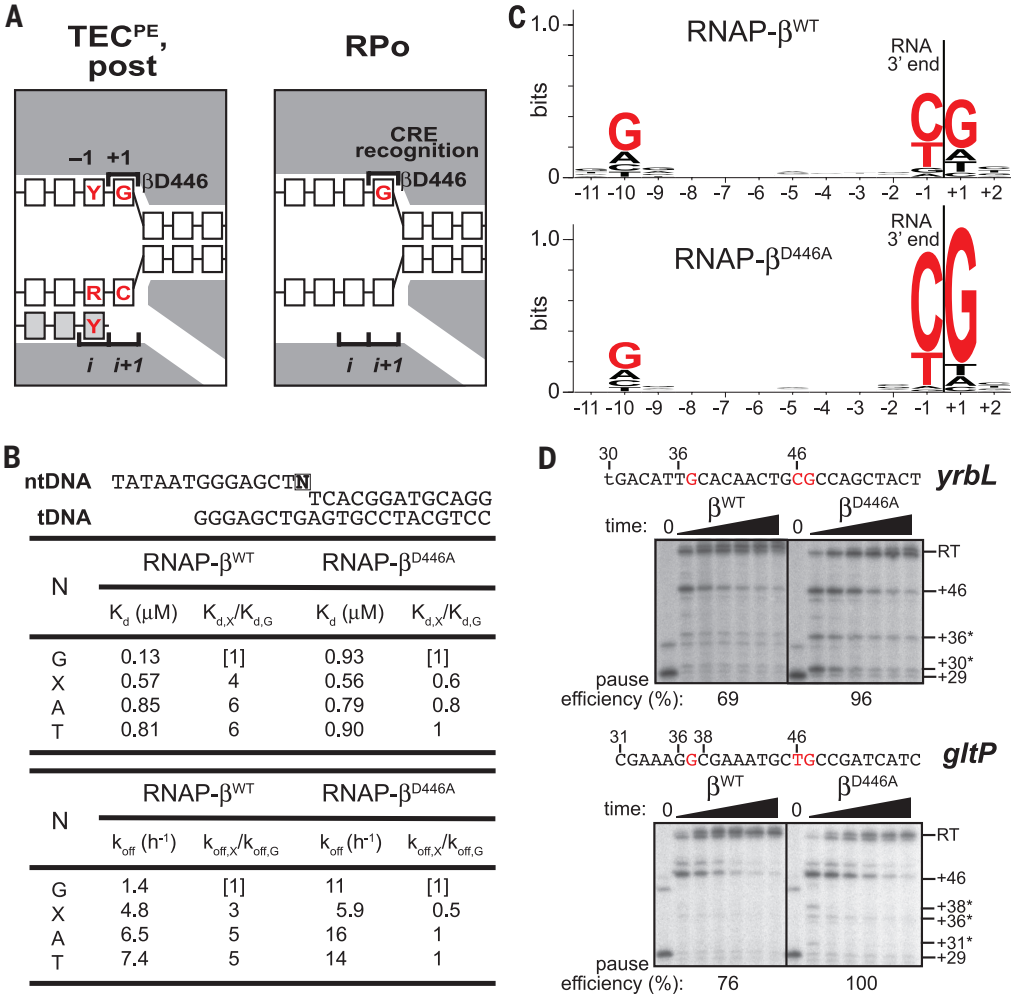
Fig. 2. Contributions of individual base pairs of consensus PE. (**A**) In vitro transcription assays with *yrbL*-PE derivatives. Red, consensus PE bases. (**B**) Schematic representation of TEC in pretranslocated state at consensus PE (top) and TEC in posttranslocated state at consensus PE (bottom). White boxes, DNA; gray boxes, RNA; gray shading, RNAP; red, consensus PE bases; *i* and *i* + 1, RNAP active center *i* and *i* + 1 sites.

in which the RNAP *i* site interacts with RNA position –1 and the RNAP *i* + 1 site is unoccupied and available for binding of a nucleoside triphosphate (NTP) (*1*, *2*, *16*). Translocation requires breaking the DNA base pair at position +1 and breaking the RNA-DNA base pair at position –10 (*17*, *18*). Because the DNA base pair at position +1 must be broken for forward translocation, G/C at +1 will disfavor forward translocation relative to the less stable A/T. Similarly, because the RNA-DNA base pair at position –10 must be broken for forward translocation, G/C at –10 will disfavor forward translocation relative to the less stable A/T. Furthermore, because 5′-rG-rN-3′/3′-dC-dN-5′ is more stable than 5′-rC-rN-3′/3′-dG-dN-5′ (*19*), rG:dC at position –10 also will disfavor translocation over rC:dG at –10. Translocation requires movement of the template-strand DNA nucleotide and base-paired RNA nucleotide at position –1 from the RNAP active center *i* + 1 site to the *i* site and movement of the template-strand DNA nucleotide at position +1 to the *i* + 1 site. Because available evidence indicates that the RNAP active center *i* and *i* + 1 sites preferentially interact with 5′-rR-rY-3′/3′-dY-dR-5′ (*3*), a nontemplate-strand Y at position –1 and nontemplate-strand G at position +1 will disfavor forward translocation relative to all other sequences (the former by stabilizing the pretranslocated state; the latter by destabilizing the posttranslocated state). Thus, each position of the consensus sequence is predicted to favor the pretranslocated state over the posttranslocated state (–10 through effects on duplex stability, –1 through effects on active-center binding, and +1 through both). Accordingly, each position of the consensus PE would be predicted to increase the opportunity for the TEC to enter an "elemental pause" state (a state that, according to one view, is accessed from the pretranslocated state and serves as an obligatory intermediate for pausing) (*7*, *8*, *10*, *11*, *20*) or a "backtracked" state (a state that, according to another view, is accessed from the pretranslocated state and serves as the primary state for pausing) (*2*, *16*).

A TEC in a posttranslocated state at a PE will contain an unpaired G at the downstream end of the nontemplate strand of the unwound "transcription bubble" (Fig. 3A). In transcription initiation complexes, RNAP core enzyme makes sequence-specific interactions with an unpaired G at the downstream end of the transcription bubble ("core recognition element" CRE) (*21*) (Fig. 3A). In transcription initiation, interaction between RNAP and the unpaired G of the CRE (G_{CRE}) facilitates promoter unwinding to form a stable, catalytically competent, RNAP-promoter open complex (RPo) (*21*). It has been proposed that RNAP-G_{CRE} interaction may be functionally important in transcription elongation as well as in transcription initiation (*21*), but this has not been previously documented.

The observation that the consensus PE contains the sequence feature required for establishment of RNAP-G_{CRE} interaction raises the possibility that RNAP-G_{CRE} interaction may mediate or modulate pausing. To test this possibility, we constructed and analyzed a mutant RNAP defective in sequence-specific recognition of G_{CRE}. The crystal structure of RPo shows that RNAP βD446 makes H bonds with Watson-Crick atoms of G_{CRE} and suggests that D446 reads the identity of G_{CRE} (*21*). Replacement of βD446 by alanine results in the loss of the ability to distinguish G, A, T, or an abasic site at the position corresponding to G_{CRE} (Fig. 3B and fig. S5), confirming that βD446 reads the identify of G_{CRE} and providing a reagent to assess functional significance of sequence-specific RNAP-G_{CRE} interactions.

To establish whether RNAP-CRE interactions affect pausing, we used a variant of NET-seq, "merodiploid NET-seq" (mNET-seq), that enables analysis of mutant RNAP derivatives, including mutant RNAP derivatives that do not support viability in haploid (fig. S1). mNET-seq uses merodiploid cells containing a plasmid-encoded, epitope-tagged RNAP and a chromosome-encoded,



N	RNAP-β^{WT}		RNAP-β^{D446A}	
	K_d (μM)	$K_{d,X}/K_{d,G}$	K_d (μM)	$K_{d,X}/K_{d,G}$
G	0.13	[1]	0.93	[1]
X	0.57	4	0.56	0.6
A	0.85	6	0.79	0.8
T	0.81	6	0.90	1

N	RNAP-β^{WT}		RNAP-β^{D446A}	
	k_{off} (h^{-1})	$k_{off,X}/k_{off,G}$	k_{off} (h^{-1})	$k_{off,X}/k_{off,G}$
G	1.4	[1]	11	[1]
X	4.8	3	5.9	0.5
A	6.5	5	16	1
T	7.4	5	14	1

Fig. 3. Sequence-specific RNAP-G_{CRE} interactions modulate pausing. (**A**) Structural organization of TEC in posttranslocated state at consensus PE (left) and RPo in transcription initiation at promoter containing consensus CRE (right) (*21*). The presence in each case of an unpaired G at downstream end of nontemplate strand of transcription bubble suggests the possibility of equivalent sequence-specific interactions between RNAP core and the G. Red at left, bases of the consensus PE; red at right, G_{CRE}. (**B**) Results of fluorescence-detection equilibrium assays (top) and kinetic (bottom) assays of interactions of RNAP derivatives with nucleic acid scaffolds containing G, A, T, or an abasic site (X) at position corresponding to G_{CRE}. ntDNA, nontemplate-strand DNA; tDNA template-strand DNA. (**C**) Sequence logos for consensus PE with RNAP-β^{WT} [(top) 0.4 bit sequence-information content at G_{CRE}] and RNAP-β^{D446A} [(bottom) 1 bit of sequence-information content at G_{CRE}], as defined by mNET-seq. Red, bases with ≥0.2 bit sequence information content. (**D**) Pause-capture efficiencies of RNAP β^{WT} (left panels) and RNAP β^{D446A} (right panels) at *yrbL* PE and *gltP* PE in vitro. +29, RNA before addition of UTP; +46, RNA in TEC at PE; asterisks, RNA in TECs at additional sites where RNAP β^{D446A} exhibits higher pause-capture efficiency than RNAP β^{WT}; RT, read-through RNA; red, consensus PE bases.

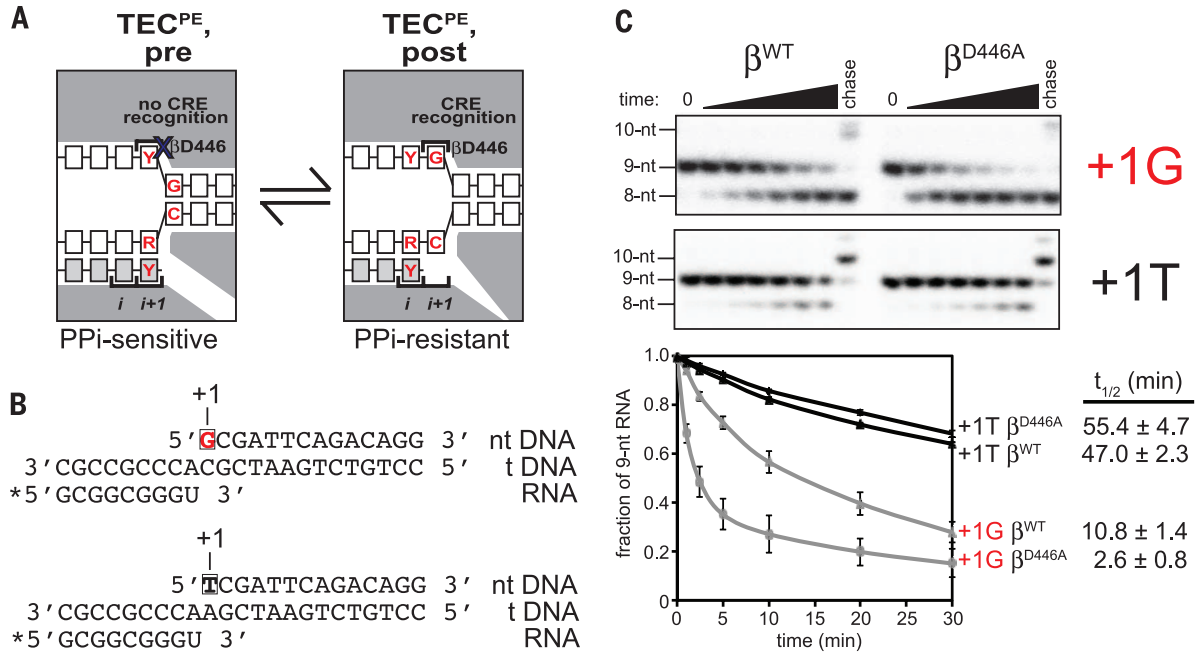


Fig. 4. Sequence-specific RNAP-G_{CRE} interactions modulate translocation bias. (**A**) Structural organization of TEC in pretranslocated state at consensus PE (left; unfavorable RNAP-CRE interaction) and TEC in posttranslocated state at consensus PE (right; favorable RNAP-CRE interaction). PPi, pyrophosphate. (**B**) Nucleic-acid scaffolds used for translocation-bias assays. Asterisk, radiolabel on RNA 5′ end; box, position corresponding to G_{CRE}; red, consensus PE base. (**C**) Translocational bias for RNAP β^{WT} and RNAP β^{D446A} on nucleic-acid scaffolds containing G or T at position corresponding to G_{CRE}. Gel images show pyrophosphorolysis reaction progress from 0 to 30 min. Labels: 9-nucleotide (nt), scaffold; 8-nt, product of pyrophosphorolysis; 10-nt, product of "chase" reaction with GTP (+1G template) or UTP (+1T template). Graph shows fraction of unaltered scaffold (means ± SEM; three measurements) as function of time.

untagged RNAP, and it involves selective analysis of transcripts associated with epitope-tagged RNAP in the presence of a mixed population of epitope-tagged RNAP and untagged RNAP. We introduced into cells a plasmid encoding 3xFLAG-tagged wild-type RNAP β subunit (β^{WT}) or 3xFLAG-tagged RNAP β subunit containing D446A (β^{D446A}), we isolated RNAs associated with RNAP-β^{WT} or RNAP-β^{D446A} by immunoprecipitation, and we identified pause sites. For RNAP-β^{WT}, alignment of pause sites revealed a consensus sequence matching the consensus PE, which validated mNET-seq as an effective system for analysis of pausing (Figs. 1A and 3C; figs. S2 and S6; and tables S4 to S9). For RNAP-β^{D446A} we identified ~60 to 90% more pause sites than with RNAP-β^{WT} (tables S4 to S10). Alignment of the pause sites revealed that ~30% more of the pause sites carried a G at position +1 (Fig. 3C, fig. S6, and table S6). We conclude that RNAP-β^{D446A} is more susceptible than RNAP-β^{WT} to pausing at sites with G at position +1.

We next compared pausing properties of RNAP-β^{D446A} and RNAP-β^{WT} in vitro, using templates carrying the *yrbL* PE and *gltP* PE (Fig. 3D and fig. S7). RNAP-β^{D446A} enhances pausing at *yrbL* PE and *gltP* PE. RNAP-β^{D446A} also enhances pausing at other positions where the next nucleotide to be added to the transcript is G (positions with asterisks in Fig. 3D). The results indicate that a substitution that disrupts sequence-specific RNAP-CRE interaction increases pausing, both in vivo and in vitro, at positions where the posttranslocated state contains G_{CRE}. We conclude that sequence-specific RNAP-G_{CRE} interaction occurs during elongation and counteracts pausing.

To explore why disruption of sequence-specific RNAP-G_{CRE} interaction enhances pausing, we assessed whether RNAP-G_{CRE} interactions affect the translocational register of the TEC by assessing sensitivity of TECs to pyrophosphorolysis (Fig. 4). Sensitivity to pyrophosphorolysis provides a measure of TEC translocation because a TEC in a pretranslocated state is sensitive to pyrophosphorolysis but a TEC in a posttranslocated state is resistant (Fig. 4A) (*3*). We performed assays with RNAP-β^{WT} or RNAP-β^{D446A} on templates containing G or T at position +1 (Fig. 4B). We found that TECs with RNAP-β^{WT} were ~5 times as sensitive to pyrophosphorolysis when the template contained +1G as when the template contained +1T, which indicated that a greater proportion of TECs on templates containing +1G were in a pretranslocated state than of TECs on templates containing +1T (Fig. 4C). The results directly demonstrate the effect of G at position +1 on translocation bias. TECs with RNAP-β^{D446A} were ~4 times as sensitive to pyrophosphorolysis as TECs with RNAP-β^{WT} when the template contained +1G (Fig. 4C), whereas, in contrast, TECs with RNAP-β^{D446A} and RNAP-β^{WT} exhibited identical sensitivities to pyrophosphorolysis when the template contained +1T (Fig. 4C). The results indicate that, on templates containing +1G, a greater proportion of TECs with RNAP-β^{D446A} are in a pretranslocated state as compared to TECs with RNAP-β^{WT}. We conclude that sequence-specific RNAP-G_{CRE} interactions stabilize the TEC posttranslocated state, which provides a mechanistic explanation for the finding that RNAP-G_{CRE} interactions counteract pausing.

Our findings define the key sequence determinants of transcriptional pausing as $G_{-10}Y_{-1}G_{+1}$ (Figs. 1 and 2). The consensus PE promotes pausing by disfavoring translocation of the TEC to the posttranslocated state (Fig. 2B), which increases the opportunity for the TEC to enter an "elemental pause" state and/or a "backtracked" state. We further show that sequence-specific RNAP-G_{CRE} interactions counteract pausing by stabilizing the TEC in a posttranslocated state (Figs. 3 and 4). Because residues of RNAP core that mediate sequence-specific RNAP-CRE interaction are conserved in RNAP from all living organisms, we suggest that RNAP-CRE interaction counteracts pausing in all multisubunit RNAPs. The consensus PE will, on average, be encountered by RNAP every ~32 bp during transcription elongation for organisms, such as *E. coli*, with ~50% G/C content and will be encountered even more frequently for organisms with higher G/C content. RNAP-G_{CRE} interactions may help overcome a barrier to forward translocation that occurs each time RNAP encounters a $G_{-10}Y_{-1}G_{+1}$ sequence during transcription elongation. A major function of RNAP-G_{CRE} interactions may be to suppress noise during transcription elongation by smoothing the sequence-dependent energy landscape for transcription elongation.

REFERENCES AND NOTES

1. M. H. Larson, R. Landick, S. M. Block, *Mol. Cell* **41**, 249–262 (2011).
2. M. Dangkulwanich, T. Ishibashi, L. Bintu, C. Bustamante, *Chem. Rev.* **114**, 3203–3223 (2014).
3. P. P. Hein, M. Palangat, R. Landick, *Biochemistry* **50**, 7002–7014 (2011).
4. S. Kyzer, K. S. Ha, R. Landick, M. Palangat, *J. Biol. Chem.* **282**, 19020–19028 (2007).
5. C. L. Chan, R. Landick, *J. Mol. Biol.* **233**, 25–42 (1993).
6. V. A. Aivazashvili, R. Sh. Bibilashvili, R. M. Vartikian, T. A. Kutateladze, *Mol. Biol. (Moscow)* **15**, 915–929 (1981).
7. A. Weixlbaumer, K. Leon, R. Landick, S. A. Darst, *Cell* **152**, 431–441 (2013).
8. R. Landick, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 8797–8798 (2009).
9. M. L. Kireeva, M. Kashlev, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 8900–8905 (2009).
10. K. M. Herbert *et al.*, *Cell* **125**, 1083–1094 (2006).
11. K. C. Neuman, E. A. Abbondanzieri, R. Landick, J. Gelles, S. M. Block, *Cell* **115**, 437–447 (2003).
12. L. S. Churchman, J. S. Weissman, *Nature* **469**, 368–373 (2011).
13. Materials and methods are available as supplementary materials on *Science* Online.
14. B. J. Haas, M. Chin, C. Nusbaum, B. W. Birren, J. Livny, *BMC Genomics* **13**, 734 (2012).

15. D. N. Lee, L. Phung, J. Stewart, R. Landick, *J. Biol. Chem.* **265**, 15145–15153 (1990).
16. M. Dangkulwanich *et al.*, *eLife* **2**, e00971 (2013).
17. L. Bai, A. Shundrovsky, M. D. Wang, *J. Mol. Biol.* **344**, 335–349 (2004).
18. V. R. Tadigotla *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 4439–4444 (2006).
19. N. Sugimoto *et al.*, *Biochemistry* **34**, 11211–11216 (1995).
20. J. Zhou, K. S. Ha, A. La Porta, R. Landick, S. M. Block, *Mol. Cell* **44**, 635–646 (2011).
21. Y. Zhang *et al.*, *Science* **338**, 1076–1080 (2012).

ACKNOWLEDGMENTS

This work was supported by NIH grants GM041376 (R.H.E.), GM088343 (B.E.N.), and GM096454 (B.E.N.). Reads are deposited in the Sequence Read Archive (accession SRP039384).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/344/6189/1285/suppl/DC1
Materials and Methods
Figs. S1 to S9
Tables S1 to S12
References (*22*–*32*)

17 March 2014; accepted 2 May 2014
10.1126/science.1253458

DISEASE ECOLOGY

Ecological and evolutionary effects of fragmentation on infectious disease dynamics

Jussi Jousimo,[1]* Ayco J. M. Tack,[1]* Otso Ovaskainen,[1] Tommi Mononen,[1,2] Hanna Susi,[1] Charlotte Tollenaere,[1] Anna-Liisa Laine[1]†

Ecological theory predicts that disease incidence increases with increasing density of host networks, yet evolutionary theory suggests that host resistance increases accordingly. To test the combined effects of ecological and evolutionary forces on host-pathogen systems, we analyzed the spatiotemporal dynamics of a plant (*Plantago lanceolata*)–fungal pathogen (*Podosphaera plantaginis*) relationship for 12 years in over 4000 host populations. Disease prevalence at the metapopulation level was low, with high annual pathogen extinction rates balanced by frequent (re-)colonizations. Highly connected host populations experienced less pathogen colonization and higher pathogen extinction rates than expected; a laboratory assay confirmed that this phenomenon was caused by higher levels of disease resistance in highly connected host populations.

Although infectious diseases pose serious threats to human health and food security, pathogens are also a fundamental component of natural biodiversity (*1*, *2*). Infection prevalence fluctuates through space and time in natural plant-pathogen associations (*3*–*5*), but there is little evidence of devastating epidemics in native interactions. This is in stark contrast to the boom-and-bust dynamics of many fungal pathogens attacking commercially important crops under agricultural settings (*6*, *7*) and epidemics caused by introduced species (*8*). With most of the focus in epidemiology targeted toward understanding conditions generating infectious disease outbreaks (*9*–*11*), the mechanisms that enable long-term disease persistence in wild communities remain elusive.

Spatially explicit metapopulation theory has identified habitat size and connectivity to other habitat patches as the key parameters affecting species distributions (*12*). The limited data available have confirmed metapopulation theory to be relevant also for disease dynamics, with colonizations and extinctions of local host populations generating the observed patterns of infection (*13*). Far less is known about how evolutionary dynamics vary through space and what the resulting effects on disease dynamics are (*14*). There is substantial theory demonstrating that the spatial scale of host-pathogen interactions and resulting rates of gene flow are critical for how coevolutionary dynamics play out between hosts and their pathogens (*15*, *16*). Hence, although host availability for pathogens increases with increasing density of host networks, the evolutionary potential of host populations may also change as a function of connectivity. Increased host gene flow into the local populations generates higher variation in resistance, allowing resistance to be selected for (*17*, *18*). A comparison across species found the relative rate of gene flow in hosts versus parasites to be the strongest predictor of pathogen local adaptation to host resistance. When host dispersal rates exceeded those of their parasites, the potential for parasite maladaptation emerged (*19*). However, such species-level approaches provide little insight into how coevolutionary dynamics and resulting infectious disease dynamics are played out across the spatially heterogeneous landscapes that most species inhabit. Indeed, the spatial and the genetic complexities that natural host-pathogen interactions support have rendered epidemiological predictions challenging in wild communities, yet identifying the effect of such complexity on epidemiological dynamics may provide key insights into the battle against diseases.

To understand how spatial and evolutionary processes interact to shape disease dynamics, we analyzed the spatial and environmental factors that drive metapopulation dynamics of an obligate fungal pathogen occupying a highly fragmented host population network. In our analyses, we targeted the three key steps of metapopulation dynamics: presence-absence of disease within local host populations (i.e., occupancy); colonization by the pathogen of previously unoccupied host populations; and extinction, whereby a pathogen population previously found infecting a local host population is deemed lost. We also analyzed factors affecting within-population abundance of disease. Our spatial variables included host population size, presence of road, host population connectivity (S^{H}), and pathogen population connectivity at the beginning (S^{P}) and the end (S^{M}) of epidemics. The environmental variables were percentage of dry host plants within populations, habitat boundary type, light availability, and precipitation in July and August. We used a hierarchical spatiotemporal Bayesian logistic regression model that allows estimation of the residual spatial and temporal autocorrelation.

Our data were collected from *Podosphaera plantaginis*, a specialist powdery mildew naturally infecting *Plantago lanceolata* in the Åland archipelago, southwestern Finland. *P. lanceolata* is an obligate outcrossing perennial herb that reproduces both sexually and clonally via side rosettes. The pollen of *P. lanceolata* is wind-dispersed, and seeds typically drop close to the mother plant. In Åland, the host populations are highly fragmented, forming a network of several thousand populations spanning an area of about 50 km by 70 km. Locations of the local *P. lanceolata* populations in the Åland Islands have been annually mapped since the 1990s. Because of clonal reproduction and a seed bank, host populations rarely go extinct, and hence the spatial configuration of the host populations remains relatively constant (*12*). Visible signs of infection by *P. plantaginis* begin to appear in late June in those host populations where the pathogen has successfully survived the winter as resting spores. The resting spores remain attached to the senescent host leaves throughout the winter, and in the spring the spores may reinfect the same host plant or new host plants in close vicinity (*20*, *21*). The epidemic starts from these initial disease foci when the pathogen is transmitted among hosts by clonally produced dispersal spores that are passively carried by wind (*5*, *22*). Some six to eight clonally produced generations follow one another in quick

[1]Metapopulation Research Group, Department of Biosciences, University of Helsinki, Post Office Box 65 (Viikinkaari 1), FI-00014 University of Helsinki, Helsinki, Finland. [2]Department of Biomedical Engineering and Computational Science, Aalto University School of Science, FI-00076 Aalto, Finland. *These authors contributed equally to this work. **†Corresponding author. E-mail: anna-liisa.laine@helsinki.fi**