Probing Gene Expression in Live Cells, One Protein Molecule at a Time

Ji Yu,[1]* Jie Xiao,[1]* Xiaojia Ren,[1] Kaiqin Lao,[2] X. Sunney Xie[1]†

We directly observed real-time production of single protein molecules in individual *Escherichia coli* cells. A fusion protein of a fast-maturing yellow fluorescent protein (YFP) and a membrane-targeting peptide was expressed under a repressed condition. The membrane-localized YFP can be detected with single-molecule sensitivity. We found that the protein molecules are produced in bursts, with each burst originating from a stochastically transcribed single messenger RNA molecule, and that protein copy numbers in the bursts follow a geometric distribution. The quantitative study of low-level gene expression demonstrates the potential of single-molecule experiments in elucidating the workings of fundamental biological processes in living cells.

The central dogma of molecular biology states that DNA is transcribed into mRNA, which is then translated into protein. Ever since the pioneering work on the *lac* operon (1), our knowledge of gene expression has come primarily from genetic and biochemical studies (2–4) conducted with large populations of cells and molecules. Recently, many in vitro single-molecule experiments have probed real-time dynamics and yielded valuable mechanistic insights into macromolecules (5–8), including transcriptional (9) and translational (10) machineries. In order to understand the workings of these machineries in their physiological contexts, we set out to probe gene expression at the single-molecule level by real-time monitoring of protein production in live cells.

Gene expression is often stochastic (11–14), because most genes exist at single or low copy numbers in a cell. Some genes are expressed at high levels and others at low levels. The mRNA expression can now be tracked in a single cell with single-molecule sensitivity (15, 16). The protein expression has been traditionally characterized by averages of cell populations, in which stochasticity is masked. More information is available from both the distribution of expression levels among a cell population (17–19) and the temporal evolution of a single cell by using fluorescent reporters (20). However, these studies have been restricted to high expression levels because of the low sensitivity for protein detection, yet many important proteins are produced at small copy numbers (21, 22). Here, we demonstrate probing protein expression in individual *Escherichia coli* cells under the control of a repressed *lac* promotor, one molecule at a time (23).

The most popular reporters for monitoring gene expression in live cells are green fluorescent protein (GFP) and its derivatives, such as yellow fluorescent protein (YFP) (24–26). We use a YFP variant, Venus, as the reporter be-

cause of its short maturation time (27). However, it is difficult to image a single GFP or YFP molecule in cytoplasm, because its fluorescence signal spreads to the entire cytoplasm by fast diffusion during the image acquisition time and is overwhelmed by cellular autofluorescence. On the other hand, single YFP fusion protein molecules on cell membranes can be detected (28, 29) because their diffusion is slowed. Therefore, we designed a fusion protein consisting of Venus and a membrane protein, Tsr, as the reporter for monitoring *lac* promoter activity. A well-studied methylation-dependent chemotaxis receptor protein (MCP) (30), Tsr contains two transmembrane domains and is fused to the N terminus of Venus.

We constructed an *E. coli* strain SX4 in which a single copy of the chimeric gene *tsr-venus* was incorporated into the *E. coli* chromosome, replacing the native *lacZ* gene. The endogenous *tsr* gene of *E. coli* was left intact.

Because the *tsr* gene is highly expressed (30), a small amount of exogenous Tsr-Venus poses minimal perturbation to cells' normal functions. Western assay of induced SX4 cells showed the presence of Venus only in the membrane fraction and not in the cytoplasmic fraction, suggesting efficient membrane localization of Tsr-Venus. We also compared the levels of induced expression of Tsr-Venus and Venus in two strains, both under the control of the *lac* promoter [Supporting Online Material (SOM) Text and fig. S1]. No notable difference was observed, indicating that the introduction of the *tsr* sequence does not change the yield of Venus production, which is not the case for many other membrane-targeting sequences that we tested.

We first show the ability to detect single Tsr-Venus fluorescent protein molecules expressed in SX4 cells (Fig. 1). Figure 1A shows two diffraction-limited fluorescent spots [full width at half maximum (FWHM) ~ 300 nm] in the left cell. A line cross section of the fluorescence image along the cells' long axes shows the signal distinctly above the cells' autofluorescence background (Fig. 1C). We attribute each fluorescent peak to an individual Tsr-Venus molecule on the basis of abrupt disappearance of the signal upon photobleaching, which is characteristic of single molecules. Figure 1D shows such a photobleaching time trace. Had the signal arisen from multiple molecules, its disappearance would be in multiple steps. In addition, the fluorescence intensity of each peak is consistent with in vitro measurements of purified single Venus molecules (fig. S2).

A sketch of our live-cell experiment is shown in Fig. 2. Upon an infrequent and spontaneous
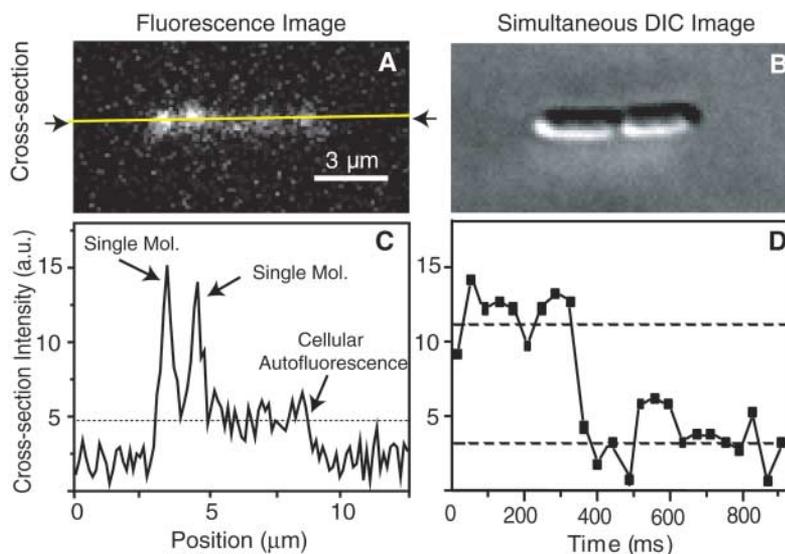

Fig. 1. Single-molecule detection of a fluorescent fusion protein, Tsr-Venus, in live *E. coli* cells. (**A**) Fluorescence and (**B**) DIC images of two *E. coli* cells (strain SX4) expressing Tsr-Venus. Two single fusion protein molecules were detected as diffraction-limited fluorescent spots (FWHM at ~300 nm) in the left cell. The fluorescence image is taken with 514-nm laser excitation and a 100-ms exposure time at 0.3 kW/cm². (**C**) Line cross section of the fluorescence signal along long axes of the two *E. coli* cells. a.u., arbitrary units. (**D**) Fluorescence time trace of a single Tsr-Venus molecule in an *E. coli* cell, showing abrupt photobleaching (40-ms exposure at 0.5 kW/cm²).

[1]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA. [2]Applied Biosystems, Foster City, CA 94404, USA.

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

dissociation event of the repressor from the operator region of DNA, transcription by RNA polymerase is initiated, generating one mRNA molecule. A few ribosome molecules bind to the mRNA, producing a burst of fusion protein molecules. These molecules can be detected after the completion of their assembly process, including protein folding, incorporation onto the inner cell membrane, and maturation of the Venus fluorophore. Meanwhile, the repressor quickly rebinds to the operator under the highly repressing condition until the next event of protein production.

An epifluorescence microscope and a charge-coupled device (CCD) camera were used to image Venus with 514-nm laser excitation, while differential interference contrast (DIC) images were taken simultaneously to record the cell contours during growth. To count the fusion protein molecules as they were continuously generated, we photobleached the Venus fluorophores after their detection. Specifically, we applied a 1200-ms laser exposure every 3 min. The laser power used was 0.3 kW/cm^2, at which the sample photobleaching time constant is $\sim$250 ms. Fluorescence images were recorded only in the first 100 ms, during which photobleaching is minimal, and were discarded in the following 1100 ms in order to avoid variation in the integrated signal and reduction of the signal-to-background ratio due to photobleaching. The 3-min dwell time, which defines the temporal resolution, was chosen to avoid photodamage to the cells. Cells grown under such laser illumination in a temperature-controlled sample chamber have an average cell division time of $\tau_{cell} = 55$ min, the same τ_{cell} as in shaking M9 liquid culture without laser illumination. In each image, one cell usually produces no more than five fluorescent protein molecules, which can be spatially resolved. In principle, higher expression levels can be quantified with integrated fluorescence signals.

The activity of the *lac* promoter in the SX4 strain under the highly repressing condition was monitored for cells immobilized by an agarose gel pad of M9 media maintained at 37°C through several cell cycles, and time-lapse movies were recorded (movies S1 and S2). A sequence of images from one of them is shown in Fig. 3A. In each fluorescence image, the fluorescent spots correspond to newly synthesized fluorescent molecules during the last 3 min. Although Tsr is known to cluster at one of the cell poles (SOM Text and fig. S6), we found that Tsr-Venus protein molecules initially land on random positions in the membrane and migrate to and cluster at the cell poles at a longer time scale (SOM Text and fig. S6). Time traces of the fluorescent protein molecules along cell lineages were extracted from the time-lapse fluorescence-DIC movies (Fig. 3B). More than 60 time traces have been collected for statistical analyses.

Several qualitative features are evident from these time traces. First, protein molecules are generated in bursts. Second, the number of protein molecules in each burst varies. Third, the bursts exhibit particular temporal spreads. Analysis of the data allows us to address the following four questions: Do these gene expression bursts occur randomly in time? How many mRNA molecules are responsible for each gene expression burst under the repressed condition? What is the distribution of the number of protein molecules in each burst? And what is the origin of the temporal spread of the individual bursts?

To address the first question, we show (Fig. 4A) the distribution of the number of gene expression bursts per cell cycle for all cells. The histogram is well fit with a Poisson distribution, which suggests that gene expression bursts occur randomly and are uncorrelated in time. We also observed a weak cell cycle dependence of the burst frequency (fig. S3), which might arise from an increase of gene copy number associated with

DNA replication during cell growth and does not change the Poissonian distribution. The average number of bursts is $n_{burst} = 1.2$ per cell cycle, yielding an average time of 46 min between two adjacent bursts. This time is comparable to in vitro dissociation times of *lac* repressor from *lac* operator O_1 (20 to 50 min) (*31*, *32*); albeit the dissociation time in live cells can be different, and each repressor dissociation event may not lead to successful transcription because of either temporary unavailability of RNA polymerase or failed transcription.

In order to answer whether the bursts arise from one copy or multiple copies of mRNA, we determined the average number of mRNA molecules per burst (*m*) according to $m = n_{mRNA}\tau_{cell}/(n_{burst}\tau_{mRNA})$, where n_{mRNA} is the steady-state abundance of *tsr-venus* mRNA molecules averaged over a cell population, τ_{cell} is the average cell division time, n_{burst} is the average number of expression bursts per cell division cycle, and τ_{mRNA} is the cellular lifetime of the *tsr-venus* mRNA. By using real-time reverse transcription polymerase chain reaction (RT-PCR), we obtained $n_{mRNA} = 0.037 \pm 0.013$ for SX4 cells (table S1 and fig. S5). We also measured the cellular lifetime of *tsr-venus* mRNA to be $\tau_{mRNA} = 1.5 \pm 0.2$ min (fig. S4) by a real-time RT-PCR assay after a pulse induction of the mRNA production. It follows that $m = 1.14 \pm 0.42$ (a 95% confidence interval; see SOM Text for more details). We thus conclude that under the repressed condition each gene expression burst results from one mRNA molecule, implying that Lac repressor quickly rebinds the exposed operator region of DNA, allowing transcription initiation of one mRNA molecule.

Next, we show the histogram of the number of protein molecules (*n*) produced per mRNA molecule (Fig. 4B). The distribution fits well with a single exponential decay, which is termed the geometric distribution for integer *n*. This distribution arises from the stochastic cellular lifetime of an mRNA molecule with mean $\tau_{mRNA} = 1.5$ min (fig. S4) due to degradation of mRNA by ribonuclease (RNase) E, which competes with ribosomes for mRNA binding (*33*). It was shown theoretically that the probability of generating *n* protein molecules from one mRNA follows a geometric distribution (*11*, *12*, *34*):

$$P(n) = \rho^n(1 - \rho) \qquad (1)$$

where ρ is the probability of the ribosome binding and $1 - \rho$ is the probability of RNase E binding to the overlapping site on mRNA (*33*). We proved that this model is consistent with our single-molecule measurement. Data fitting of Fig. 4B to Eq. 1 yields $\rho = 0.8 \pm 0.1$ and an average of 4.2 ± 0.5 molecules produced per burst. This number multiplied by the number of bursts per cell cycle (1.2) results in a steady-state protein abundance of 5.0 ± 0.8 molecules per cell in a cell population. Consistently, we ex-
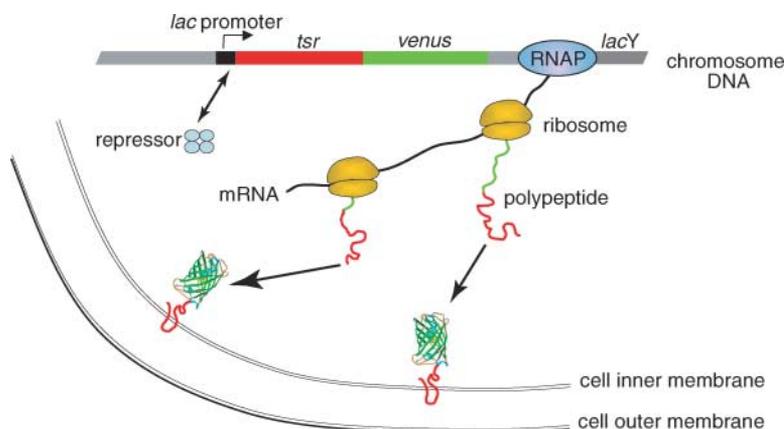

Fig. 2. Experimental design for live-cell observations of gene expression. Tsr-Venus is expressed under the control of *lac* repressor, which binds tightly to the *lac* operator on DNA. Transcription of one mRNA by an RNA polymerase results from an infrequent and transient dissociation event of repressor from DNA. Multiple copies of protein molecules are translated from the mRNA by ribosomes. Upon being assembled into *E. coli*'s inner membrane, Tsr-Venus protein molecules can be detected individually by a fluorescence microscope.
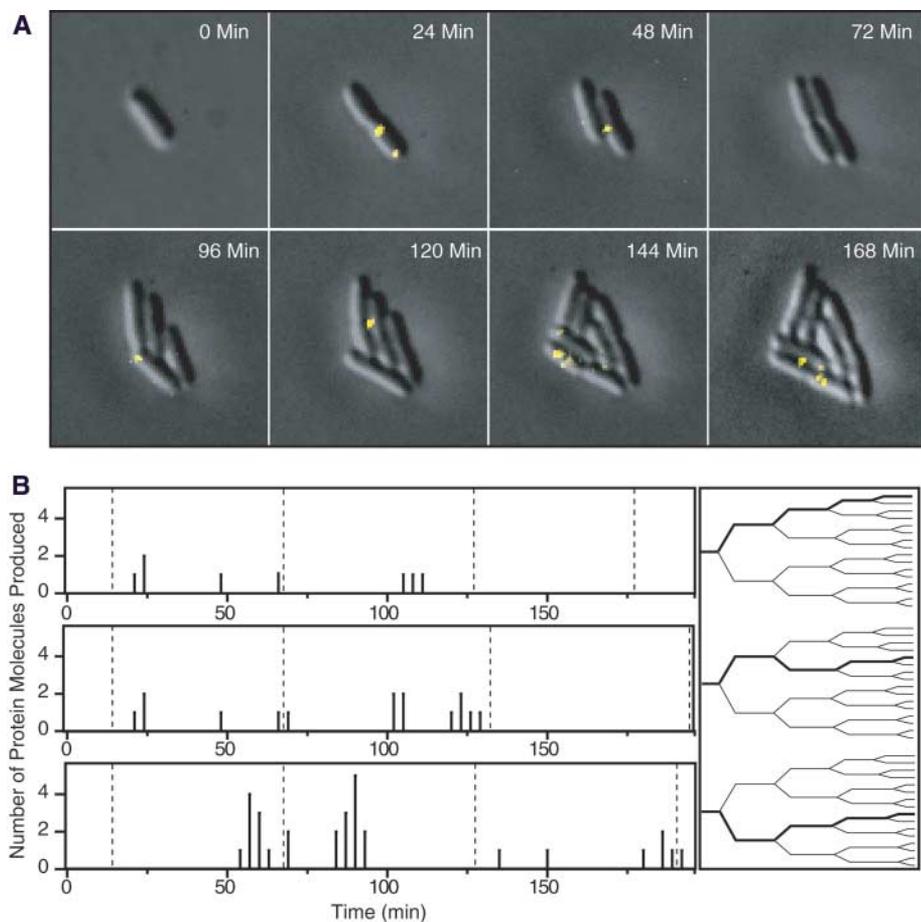
Fig. 3. Real-time monitoring of the expression of *tsr-venus* under the control of repressed *lac* promoter. (**A**) Sequence of fluorescent images (yellow) overlaid with simultaneous DIC images (gray) of *E. coli* cells expressing Tsr-Venus on agarose gel pad of M9 medium. The cell cycle is 55 ± 10 min in a temperature-controlled chamber on a microscope stage. The eight frames are from time-lapse fluorescence movie S1 taken over 195 min with 100-ms laser exposures (0.3 kW/cm²) every 3 min. An 1100-ms exposure is applied after each image collection to photobleach the Venus fluorophores. (**B**) Time traces of the expression of Tsr-Venus protein molecules (left) along three particular cell lineages (right) extracted from the time-lapse fluorescence-DIC movie of (A). The time resolution is 3 min. The vertical axis is the number of protein molecules newly synthesized during the last three minutes. The dotted lines mark the cell division times. The time traces show that protein production occurs in random bursts, within which variable numbers of protein molecules are generated. Each gene expression burst lasts ~3 to 15 min.
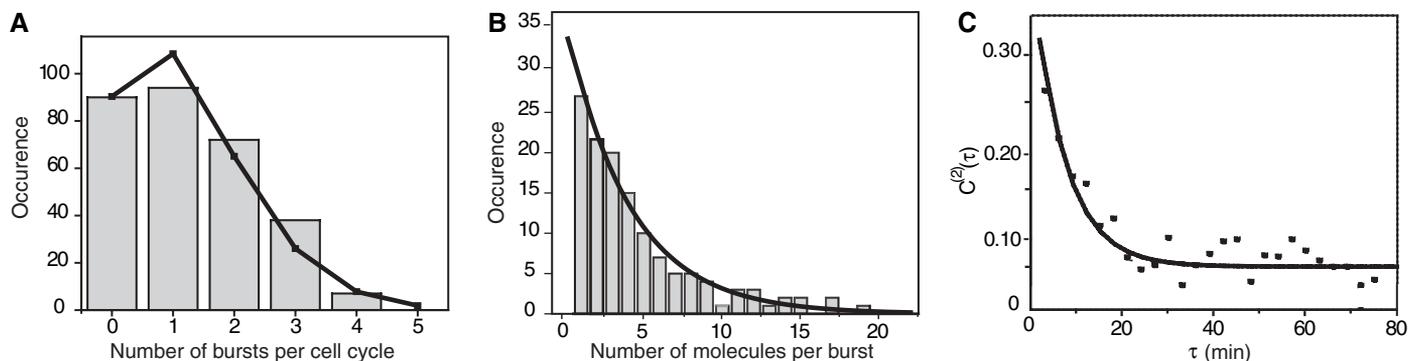
perimentally measured this abundance to be 4.1 ± 1.8 molecules per cell by counting the molecules in ~300 individual cells under the microscope at the same time.

Lastly, the temporal spread of the expression bursts can be characterized from the auto-correlation function of the fluctuation in protein expression, $C^{(2)}(\tau)$ (Fig. 4C), averaged from 30 different cell lineages from 15 different movies. The single exponential fit of $C^{(2)}(\tau)$ gives a decay time constant of 7.0 ± 2.5 min, corresponding to the average spread of the stochastic arrival times of fluorescent reporter proteins within a burst, despite the fact that the polypeptides are generated within the short lifetime of an mRNA ($\tau_{mRNA} = 1.5$ min). We show (SOM Text) that, under the condition that there is one rate-limiting step for the posttranslation assembly of the fusion protein,

$$C^{(2)}(\tau) = \left(\frac{s\rho}{1-\rho} \right)^2 \left[1 + \frac{\kappa}{s} \exp(-\kappa\tau) \right]$$
(2)

where s is the average rate of the expression burst and κ is the rate constant of Tsr-Venus assembly process, consisting of transcription, translation, folding, and chromophore maturation. The fitting of Fig. 4C with Eq. 2 gives $s = (29 \pm 8 \text{ min})^{-1}$, in agreement with the average number of expression bursts per cell cycle of 1.2 ± 0.3 (Fig. 4A); $\rho = 0.7 \pm 0.1$, consistent with the value of 0.8 ± 0.1 determined from Fig. 4B; and $1/\kappa = 7.0 \pm 2.5$ min, corresponding to the rate-limiting step of the protein assembly process. Considering the fast transcription (~45 bases/s) and translation (~15 residues/s) rates, we tentatively assign $1/\kappa$ to the fluorophore maturation process (SOM Text). Although we can only spatially resolve a few molecules within an *E. coli* cell because of the diffraction limit, the long spread of the stochastic arrival times of Venus allows many more protein molecules per expression burst to be counted in several consecutive images.

Fig. 4. Statistical analyses of the protein production time traces. (**A**) Histogram (gray bars) of the number of expression events per cell cycle. The data fit well to a Poisson distribution (solid line) with an average of 1.2 gene expression burst per cell cycle. (**B**) Distribution of the number of fluorescent protein molecules detected in each gene expression burst, which follows a geometric distribution (solid line), giving a probability of ribosome binding of 0.81 ± 0.05 and an average number of molecules per burst of 4.2. (**C**) Autocorrelation function of the protein production time traces calculated according to Eq. S9. The result is averaged from 30 individual cell lineages because of the insufficient statistics of a single time trace. The fitting to Eq. 2 (solid line) gives $1/\kappa = 7.0 \pm 2.5$ min, which is attributed to posttranslational assembly of the fluorescent fusion protein.

Gene expression, central to life processes, is intrinsically a stochastic process involving low copy number of biomolecules. Our real-time assay allows probing of low copy number proteins in single live cells not accessible by current technologies. This approach, together with other emerging single-molecule techniques (35), will yield further insight into not only gene expression but also other fundamental biological processes.

References and Notes

1. F. Jacob, J. Monod, Cold Spring Harbor Symp. Quant. Biol. 26, 193 (1961).
2. J. Beckwith, Science 156, 597 (1967).
3. W. Gilbert, B. Muller-Hill, Proc. Natl. Acad. Sci. U.S.A. 58, 2415 (1967).
4. M. Ptashne, A. Gann, Genes and Signals (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2002).
5. X. S. Xie, H. P. Lu, J. Biol. Chem. 274, 15967 (1999).
6. C. Bustamante, Z. Bryant, S. Smith, Nature 421, 423 (2003).
7. A. Ishijima, T. Yanagida, Trends Biochem. Sci. 26, 438 (2001).
8. S. Weiss, Science 283, 1676 (1999).
9. M. D. Wang et al., Science 282, 902 (1998).
10. S. Blanchard, R. Gonzalez, H. Kim, S. Chu, J. Puglisi, Nat. Struct. Mol. Biol. 11, 1008 (2004).
11. O. Berg, J. Theor. Biol. 71, 587 (1978).
12. D. Rigney, J. Theor. Biol. 79, 247 (1979).
13. J. Paulsson, Nature 427, 415 (2004).
14. J. Paulsson, Phys. Life Rev. 2, 157 (2005).
15. Y. Shav-Tal et al., Science 304, 1797 (2004).
16. I. Golding, E. Cox, Proc. Natl. Acad. Sci. U.S.A. 101, 11310 (2004).
17. N. Barkai, S. Leibler, Nature 403, 267 (2000).
18. E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, A. van Oudenaarden, Nat. Genet. 31, 69 (2002).
19. M. B. Elowitz, A. J. Levine, E. D. Siggia, P. S. Swain, Science 297, 1183 (2002).
20. N. Rosenfeld, J. W. Young, U. Alon, P. S. Swain, M. B. Elowitz, Science 307, 1962 (2005).
21. S. Ghaemmaghami et al., Nature 425, 737 (2003).
22. P. Guptasarma, Bioessays 17, 987 (1995).
23. Materials and methods are available as supporting material on Science Online.
24. R. J. Bongaerts, I. Hautefort, J. M. Sidebotham, J. C. Hinton, Methods Enzymol. 358, 43 (2002).
25. R. Y. Tsien, Annu. Rev. Biochem. 67, 509 (1998).
26. M. Chalfie, Y. Tu, G. Euskirchen, W. W. Ward, D. C. Prasher, Science 263, 802 (1994).
27. T. Nagai et al., Nat. Biotechnol. 20, 87 (2002).
28. P. H. Lommerse et al., Biophys. J. 86, 609 (2004).
29. J. Deich, E. M. Judd, H. H. McAdams, W. E. Moerner, Proc. Natl. Acad. Sci. U.S.A. 101, 15921 (2004).
30. S. L. Mowbray, M. O. J. Sandgren, J. Struct. Biol. 124, 257 (1998).
31. K. Bondeson, A. Frostellkarlsson, L. Fagerstam, G. Magnusson, Anal. Biochem. 214, 245 (1993).
32. W. Hsieh, P. Whitson, K. Matthews, R. Wells, J. Biol. Chem. 262, 14583 (1987).
33. O. Yarchuk, N. Jacques, J. Guillerez, M. Dreyfus, J. Mol. Biol. 226, 581 (1992).
34. H. H. McAdams, A. Arkin, Proc. Natl. Acad. Sci. U.S.A. 94, 814 (1997).
35. L. Cai, N. Friedman, X. S. Xie, Nature 440, 358 (2006).
36. We thank P. Choi, L. Cai, N. Friedman, J. Elf, L. Xun, and R. Losick for helpful discussions and critical reading of the manuscript, J. Hearn for technical assistance, and A. Miyawaki (RIKEN) for the generous gift of the venus gene. This work was supported by NIH Director's Pioneer Award program, a NIH R21 grant, and in part by U.S. Department of Energy, Office of Science, and Applied Biosystems' Exploratory Fund. J.Y. acknowledges a Genome-Related Research Award from Merck.

Late Colonization of Easter Island

Terry L. Hunt[1]* and Carl P. Lipo[2]

Easter Island (Rapa Nui) provides a model of human-induced environmental degradation. A reliable chronology is central to understanding the cultural, ecological, and demographic processes involved. Radiocarbon dates for the earliest stratigraphic layers at Anakena, Easter Island, and analysis of previous radiocarbon dates imply that the island was colonized late, about 1200 A.D. Substantial ecological impacts and major cultural investments in monumental architecture and statuary thus began soon after initial settlement.

With an empty landscape containing gigantic statues and other cultural achievements, Easter Island, or Rapa Nui, symbolizes an isolated civilization that once flourished but suffered ecological catastrophe. Central to understanding the cultural, ecological, and demographic processes that shaped Rapa Nui's prehistory is the establishment of a reliable chronology for the island. Here, we provide radiocarbon dates from deposits likely to represent the earliest occupation on the island and evaluate previous 14C dates to show that Rapa Nui's prehistoric chronology is later than has been commonly assumed.

Early Polynesians colonized Fiji, Tonga, and Samoa in the central South Pacific about 2800 years ago. Accumulating evidence shows that continued expansion from Samoa-Tonga into eastern Polynesia (e.g., Cooks, Societies, Marquesas, and Hawaii) did not occur until after 800 A.D., although notably longer chronologies have been suggested and debated (1–3). Many eastern Polynesian sites once considered centuries older have been shown to be consistently younger (4, 5). For example, it is now thought that New Zealand was colonized after 1200 A.D., about 400 years after the date that has long been assumed (6). This date has been confirmed by radiocarbon dating of seeds gnawed by the Polynesian rat (Rattus exulans) (7), clear evidence of human presence because rats are commensal species.

Smith (8) obtained initial radiocarbon dates from Rapa Nui with the Norwegian expedition of Heyerdahl in the 1950s. His suite of 19 radiocarbon dates included one from a burn horizon at Poike Ditch of 400 A.D. [1570 ± 80 years before the present (yr B.P.), 384 to 664 calibrated (cal) A.D.]. Subsequent researchers cited the

[1]Department of Anthropology, University of Hawai'i Manoa, 2424 Maile Way, Honolulu, HI 96822, USA. [2]Department of Anthropology and Institute for Integrated Research in Materials, Environments, and Society, California State University Long Beach, 1250 Bellflower Boulevard, Long Beach, CA 90840, USA.

*To whom correspondence should be addressed. E-mail: thunt@hawaii.edu

Table 1. Radiocarbon dates from recent excavations at the Anakena Dune site.

Sample Beta-	Material/layer	Radiocarbon age (yr B.P.)	13C/12C	Conventional C14 age (yr B.P.)	2σ calibration (cal A.D.) (19, 20)	Probability
196711	Charcoal/Unit 1 Layer 8	660 ± 40	−24.9	660 ± 40	1294–1403	1.00
196712	Charcoal/Unit 1 Layer 5	690 ± 60	−25.6	680 ± 60	1274–1414	1.00
196713	Charcoal/Unit 1 Layer 8	670 ± 40	−24.8	670 ± 60	1291–1400	1.00
196714	Charcoal/Unit 1 Layer 9	600 ± 60	−26.0	590 ± 60	1300–1368	0.369341
					1373–1450	0.630659
196715	Charcoal/Unit 1 Layer 11	670 ± 40	−22.5	710 ± 40	1279–1391	1.00
196716	Charcoal/Unit 1 Layer 12	720 ± 60	−24.7	720 ± 60	1229–1251	0.046194
					1260–1400	0.953806
209903	Charcoal/Unit 5 Base	870 ± 80	−23.8	870 ± 80	1029–1300	0.996544
					1368–1372	0.003456
209904	Charcoal/Unit 5 Base	870 ± 40	−23.8	870 ± 40	1055–1058	0.003916
					1151–1278	0.996084