



Collection, Mapping, and Annotation of Over 28,000 cDNA Clones from *japonica* Rice
 Shoshi Kikuchi *et al.*
Science **301**, 376 (2003);
 DOI: 10.1126/science.1081288

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of August 8, 2012):

A correction has been published for this article at:
<http://www.sciencemag.org/content/301/5641/1849.2.full.html>

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
<http://www.sciencemag.org/content/301/5631/376.full.html>

Supporting Online Material can be found at:
<http://www.sciencemag.org/content/suppl/2003/07/16/301.5631.376.DC1.html>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:
<http://www.sciencemag.org/content/301/5631/376.full.html#related>

This article has been **cited by** 429 article(s) on the ISI Web of Science

This article has been **cited by** 100 articles hosted by HighWire Press; see:
<http://www.sciencemag.org/content/301/5631/376.full.html#related-urls>

This article appears in the following **subject collections**:
 Botany
<http://www.sciencemag.org/cgi/collection/botany>

REPORTS

Collection, Mapping, and Annotation of Over 28,000 cDNA Clones from *japonica* Rice

The Rice Full-Length cDNA Consortium

(National Institute of Agrobiological Sciences Rice Full-Length cDNA Project Team: Shoshi Kikuchi,^{1*} Kouji Satoh,¹ Toshifumi Nagata,¹ Nobuyuki Kawagashira,¹ Koji Doi,¹ Naoki Kishimoto,¹ Junshi Yazaki,¹ Masahiro Ishikawa,¹ Hitomi Yamada,¹ Hisako Ooka,¹ Isamu Hotta,¹ Keiichi Kojima,² Takahiro Namiki,² Eisuke Ohneda,² Wataru Yahagi,² Kohji Suzuki,² Chao Jie Li,² Kenji Ohtsuki,² Toru Shishiki²; Foundation of Advancement of International Science Genome Sequencing & Analysis Group: Yasuhiro Otomo,³ Kazuo Murakami,³ Yoshiharu Iida,³ Sumio Sugano,^{3,4} Tatsuto Fujimura,^{3,5} Yutaka Suzuki,^{3,4} Yuki Tsunoda,³ Takashi Kurosaki,³ Takeko Kodama,³ Hiromi Masuda,³ Michie Kobayashi,³ Quihong Xie,³ Min Lu,³ Ryuya Narikawa,³ Akio Sugiyama,³ Kouichi Mizuno,^{3,5} Satoko Yokomizo,³ Junko Niikura,³ Rieko Ikeda,³ Junya Ishibiki,³ Midori Kawamata,³ Akemi Yoshimura,³ Junichirou Miura,³ Takahiro Kusumegi,³ Mitsuru Oka,³ Risa Ryu,³ Mariko Ueda,³ Kenichi Matsubara^{3,6}; RIKEN: Jun Kawai,^{7,8} Piero Carninci,^{7,8} Jun Adachi,⁷ Katsunori Aizawa,⁷ Takahiro Arakawa,⁷ Shiro Fukuda,⁷ Ayako Hara,⁷ Wataru Hashidume,⁷ Norihito Hayatsu,⁷ Koichi Imotani,⁷ Yoshiki Ishii,⁷ Masayoshi Itoh,⁸ Ikuko Kagawa,⁷ Shinji Kondo,⁷ Hedeaki Konno,⁷ Ai Miyazaki,⁷ Naoki Osato,⁷ Yoshimi Ota,⁷ Rintaro Saito,⁷ Daisuke Sasaki,⁷ Kenjiro Sato,⁸ Kazuhiro Shibata,⁸ Akira Shinagawa,⁸ Toshiyuki Shiraki,⁷ Masayasu Yoshino,⁷ Yoshihide Hayashizaki^{7,8})

We collected and completely sequenced 28,469 full-length complementary DNA clones from *Oryza sativa* L. ssp. *japonica* cv. Nipponbare. Through homology searches of publicly available sequence data, we assigned tentative protein functions to 21,596 clones (75.86%). Mapping of the cDNA clones to genomic DNA revealed that there are 19,000 to 20,500 transcription units in the rice genome. Protein informatics analysis against the InterPro database revealed the existence of proteins presented in rice but not in *Arabidopsis*. Sixty-four percent of our cDNAs are homologous to *Arabidopsis* proteins.

Rice (*Oryza sativa*) is an important food crop; it is also a good model for studies of monocot plants because its genome (430 Mb) is small relative to other crop plants of the *Poaceae* species. Draft sequences of the *Oryza sativa* L. ssp. *indica* (1) and *japonica* (2) genomes by the "whole-genome shotgun" sequencing method have been published, as have essentially complete sequences of chromosomes 1 and 4 by a physical mapping ("clone-by-clone") method (3, 4). In addition to genomic data, full-length cDNA clones are necessary to identify exon-intron boundaries and gene-coding regions within genomic sequences and for comprehensive gene-function analyses at the transcriptional (transcriptomic) and translational (protein informatic) (5) levels.

Here we describe the collection, grouping, sequencing (6), mapping, and functional annotation of full-length cDNA (FL-cDNA) clones from ssp. *japonica* (cv. Nipponbare). All the data we present are available to the public through our Web site, the Knowledge-based *Oryza* Molecular Biological Encyclopedia (KOME, <http://cdna01.dna.affrc.go.jp/cDNA>), and the 28,469 complete sequences have been assigned GenBank accession numbers AK058203 through AK074028 and AK98843 through AK111488. (An example of a KOME report is in fig. S1.) Clones will be distributed (www.rgrc.dna.affrc.go.jp/) from the Rice Genome Resource Center (Tsukuba, Japan).

We used the BLASTN and BLASTX programs (both of which use the Basic Local Alignment Search Tool) for the gene homology searches between monocot and dicot plants (6). We found 2603 cDNA clones that were identical to already-known rice genes. We identified 5607 clones as gene products from paralogs to already-known rice genes. Of the rice FL-cDNA clones, 12,527 were homologous to already-known genes of other plants, and 859 were homologous to already-known genes in organisms other than plants. In total, these homology searches enabled us to assign potential functions to 21,596 (75.86%) of our FL-cDNA clones. Homology-based comparative analysis of calcium signal-transduction proteins between plants and animals suggested that some important families, such as voltage-dependent calcium channel proteins, do not exist in plants (7).

We mapped the 28,469 FL-cDNA clones to the rice genome sequences—the *indica* draft genome sequence (1), the *japonica* draft genome sequence (2), and the *japonica* BAC/PAC (bacterial artificial chromosome and P1-derived artificial chromosome) clones (6) (Table 1). More than 94% of our FL-cDNA clones could be mapped to rice genomic (*japonica* and *indica*) sequences. Mapping results of our *japonica*-originated clones to the *indica* genome show that the nucleotide sequences of the gene-coding regions are very similar in these two subspecies. Genome sequence alone could not correctly identify the gene structure, but mapping of cDNA clones and comparison of genome sequences indicate the correct structure of the genes in rice. Rice FL-cDNA clones are useful not only to determine rice gene structure but also to understand gene structure in other *Poaceae* species by the integrative analyses with the expressed sequence tag (EST) clones of these species.

¹Department of Molecular Biology, National Institute of Agrobiological Sciences, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan. ²Hitachi Software Engineering Co., Ltd., 6-81, Onoe-cho, Naka-ku, Yokohama 231-0015, Japan. ³Laboratory of Genome Sequencing & Analysis Group, Foundation of Advancement of International Science (FAIS), 586-9, Akatsuka-Ushigahuchi, Tsukuba, Ibaraki 305-0062, Japan. ⁴Department of Virology, Institute of Medical Science, The University of Tokyo, 4-6-1, Shirogane-dai, Minato-ku, Tokyo 108-8639, Japan. ⁵Institute of Agricultural and Forest Engineering, University of Tsukuba, 1-1-1, Tennodai, Tsukuba, Ibaraki 305-8577, Japan. ⁶Nara Institute of Science and Technology (NAIST), 8916-5, Takayama-cho, Ikoma, Nara 630-0101, Japan. ⁷Laboratory for Genome Exploration Research Group (Institute of Physical and Chemical Research, RIKEN), Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan. ⁸Genome Science Laboratory, RIKEN Wako Main Campus, 2-1, Hiro-sawa, Wako, Saitama 351-0198, Japan.

*To whom correspondence should be addressed. E-mail: skikuchi@nias.affrc.go.jp

Of the 18,933 transcription units (TUs) located on the Syngenta genome sequence (2), 5045 are multi-exon TUs that contain two or more transcripts. We searched for alternative forms in those 5045 loci. By pairwise comparison of transcripts mapped to a given locus, we identified alternative 5' or 3' ends, cryptic exons (exons present in one transcript and entirely absent from another), and those exons flanked by alternative donor/acceptor sites (Table 2). We identified putative alternative transcripts in 2471 loci (45.6% of the redundant TUs and 13.1% of the total TUs). This is notable variability, given that our strategy selects against alternative forms. Alternative initiation sites were shown by 1673 loci (8.8%), whereas variation of termination sites was observed at only 853 loci (4.5%). This imbalance occurred despite the fact that our having used 3' single-pass sequences in clustering clones to select a single representative for the full-length sequencing; this result implies more frequent variation in initiation sites than in termination sites in the rice genome. We detected alternatively spliced internal exons in 94 (0.5%) TUs. Exons were spliced at internal donor and acceptor sites in 180 (1.0%) and 241 (1.3%) loci, respectively. The alternative 5' or 3' ends, cryptic exons, and internal donor/acceptor sites occurred in coding regions [the longest open reading frame

(ORF)] in 1937 (78.4%) loci, and these variations may lead to a functional change in the coded proteins. (This estimate may be affected by our having used the longest ORF as the coding region.)

Antisense RNA inhibits gene expression (8) and genetic imprinting (9). To identify antisense RNA genes, we searched for pairs of transcripts that are transcribed bidirectionally from an overlapping genomic region. We found 902 transcript pairs (1443 cDNA clones), the exons of which overlap at least one nucleotide on opposite strands. The clustering patterns of the variants of the transcripts on the same TU are shown on our Web site. An example is in fig. S2.

From the 5'-end sequences of mRNAs, the promoter sequences can be obtained by comparison with the rice genomic sequences. In a search of the PLACE database, we found cis-acting elements in genomic sequences 1000 base pairs upstream from the 5' termini of each mapped FL-cDNA clone (10). The data are available on our Web site (fig. S1). We identified 20,403 potential promoters.

We searched the InterPro database (6, 11) to compare the profiles of proteins encoded in the rice genome with those in the *Arabidopsis* genome. InterPro domains found in each ORF amino acid sequence can be seen on the KOME report page (fig. S1).

Protein domains found in rice but not in *Arabidopsis* (table S2b) and those more frequent in rice than in *Arabidopsis* (table S2c) could be categorized as those from tissue-specific proteins (e.g., pollen: major pollen allergen Lol pI; seed: Bowman-Birk serine protease inhibitor domain and BURP domain) or related to environmental stresses (e.g., cold; antifreeze protein type I; and drought; ABA/WDS-induced protein). The abundance of these proteins may be characteristic of rice. In contrast, domains found more frequently in *Arabidopsis* than in rice could be categorized into components of signal transduction (e.g., receptor; TIR domain; and diacylglycerol-related signal transduction; DC1 domain) and transposon-related proteins (e.g., putative plant transposon protein, reverse transcriptase, retrotransposon gag protein, integrase, catalytic domain, and plant MuDR transposase). The abundance of domains associated with transposable element-related proteins is a peculiarity of *Arabidopsis*.

Transcription factors are important proteins in rice. Our InterPro search yielded 18 DNA binding domains related to 1336 transcription factors (Table 3). Zinc finger-type transcription factors are most numerous, followed by Myb-type factors; these results are similar to those for *Arabidopsis* (12). To analyze membrane-spanning domains, cellular localization, glycosylation sites, and phosphorylation sites, we used the Membrane Protein Structure And Topology (MEMSAT) (13) and PSORT (14, 15) programs (tables S4 and S5). All the clone-by-clone results of analyses by these programs can be seen on the KOME report page.

To understand the profile and complexity of rice proteins, we compared the 28,444 ORF amino acid sequences from our clones with 27,288 predicted coding sequences (CDS) of the *Arabidopsis* genome (using the same set of data as in the InterPro analysis); (table S4). At an Expect Value (E) $E < 10^{-7}$ (BLASTP), 18,900 FL-cDNA clones (12,996 TUs; 64%) showed homology to *Arabidopsis*-predicted CDS, whereas 9544 (7263 TUs; 36%) did not (Fig. 1). There were 20,473 *Arabidopsis* genes (75%) that had a homolog in our rice FL-

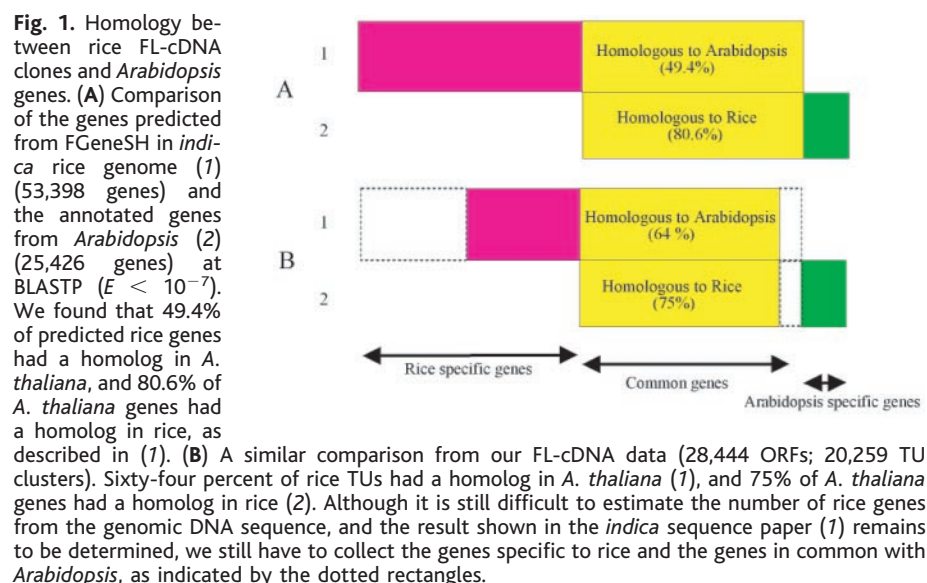


Table 1. Results of mapping FL-cDNA clones to rice genomic sequences. Mbp, million base pairs.

Genomic sequence	Source	Size (Mbp)	No. of mapped clones (%)	No. of nonredundant TUs
japonica draft genome	(2)	390	26,930 (94.6)	18,933
indica draft genome	(1)	363	26,784 (94.1)	19,036
BAC/PAC clones from japonica	IRGSP	368	22,162 (77.8)	15,523

Table 2. Number of transcription units (TUs) with alternative structures. Data represent the 5045 multi-exon TUs that contain at least two transcripts.

Alternative structure	No. (% of total) of TUs with alternative structure
Initiation sites	1673 (8.8%)
Internal exons	94 (0.5%)
Termination sites	853 (4.5%)
Splice donor sites	180 (1.0%)
Splice acceptor sites	241 (1.3%)

REPORTS

Table 3. Transcription factors identified through InterPro search. An InterPro search yielded 18 DNA-binding domains related to 1336 transcription factors, and the proteins with these domains are shown. Zinc finger-type transcription factors are most numerous, followed by Myb-type factors; these results are similar to those for *Arabidopsis* (12).

Category of domain	No. of FL-cDNA clones	Comment
Zn finger	588	Including RING, C2H2, Cx8Cx5C3H, Dof, GATA, CONSTANS, and NF-X1
Myb	158	
ERF	83	Including one ERF
NAM	74	
Homeobox	73	Including ELK, KNOX1, and KNOX2
bZIP	63	Not including Zn finger
AUX/IAA	47	Including four TFB3
WRKY	46	
TFB3	27	Not including ERF or AUX/IAA
GRAS	27	Not including Zn finger
MADS	26	
HSF	25	
Tubby	24	
BRCT	19	Not including Zn finger
Fungal TF	18	
SBP	17	
Jumonji	11	Including jmjC and JmjN; not including Zn finger
TCP	10	
Total	1336	

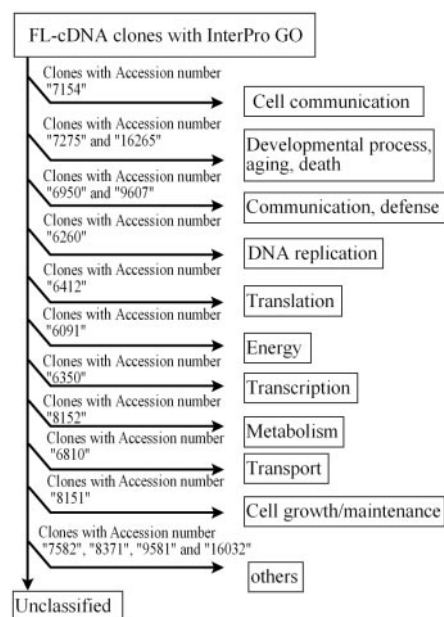


Fig. 2. Flow chart of the classification of FL-cDNA clones by GO term. We classified 21,708 FL-cDNA clones with InterPro GO terms according to the accession number of the associated GO term "biological process." The results of the number of clones in each category are shown in Table 4.

cDNAs, whereas 6815 (25%) did not. For the *indica* genome (1), when both cDNAs and predicted genes were compared against the genomic sequence, 49.4% of predicted rice genes had a homolog in *Arabidopsis*, and 80.6% of *Arabidopsis* genes had a homolog in rice. If we assume that the numbers of rice genes in the *japonica* and *indica* genomes are nearly the

same, these results suggest that more than 5% of *Arabidopsis* homologous genes (the difference between 80.6% and 75%) remain to be collected and that the rice-specific genes to be collected could reduce the fraction of *Arabidopsis*-common TUs to 49.4%. It remains difficult to estimate the correct number of rice genes from the genomic DNA sequence, but this method of comparison with data from other species might help us to estimate the number of genes.

To classify the genes according to their putative function, we used the Gene Ontology (GO) term attached to the InterPro domain names (Fig. 2). The InterPro GO assignment yielded 9734 clones with GO terms associated with "biological processes." The total number of GO terms associated with "biological processes" in our clones is 21,708; these can be divided into 12 categories (Table 4). The InterPro GO assignment yielded 12,346 FL-cDNA clones with GO terms associated with "function." The total number of GO terms associated with "function" was 18,877. The number of terms that shared the description "function" is shown in table S6. Classification of protein kinases was done by GO term (table S7). Finally, the InterPro GO assignment yielded 4025 FL-cDNA clones with GO terms associated with "cellular component," and the total number of GO terms associated with "cellular component" was 4051 (table S8). GO terms attached to each clone can be seen at the KOME page (fig. S1).

We compared the FgenesH-predicted genes for the whole *indica* genome with FL-cDNA clones in terms of functional classifi-

Table 4. GO terms of FL-cDNAs with "biological processes" associated with InterPro domains. The total number of GO terms associated with "biological processes" in our clones is 21,708; they can be exclusively classified into 12 categories. The process of classification is as shown in Fig. 2.

Category	No. of FL-cDNA clones	%
Unclassified	11,974	55.1
Metabolism	5,397	24.9
Transport	1,283	5.9
Translation	884	4.1
Transcription	640	2.9
Cell communication	556	2.6
Communication, defense	249	1.1
Energy	245	1.1
Cell growth/maintenance	207	1.0
Developmental process, aging, death	130	0.6
DNA replication	107	0.5
Others	36	0.2
Total	21,708	100.0

cation. Updated InterPro and GO annotations for the *indica* genome were downloaded from the Bioverse Web site (16). The FgenesH-predicted genes in each category and the FL-cDNA clones in each category are nearly equally distributed, thus demonstrating that there were no functional biases in acquisition of the cDNA data (fig. S4). We also compared the Bioverse annotations of the predicted *Arabidopsis* genes with rice FL-cDNA, and these also show a similar distribution among different functional categories (fig. S5).

Analysis of transcriptional products, as represented by cDNA clones, adds to a greater understanding of genome function and contributes to improving gene prediction tools.

References and Notes

1. J. Yu et al., *Science* **296**, 79 (2002).
2. S. A. Goff et al., *Science* **296**, 92 (2002).
3. T. Sasaki et al., *Nature* **420**, 312 (2002).
4. Q. Feng et al., *Nature* **420**, 316 (2002).
5. M. Seki et al., *Science* **296**, 141 (2002).
6. Starting materials and methods for the construction of cDNA libraries, single-pass sequencing, clustering, and complete sequence analysis are described in materials and methods in supporting online material.
7. T. Nagata, personal communication.
8. P. M. Waterhouse, M. W. Graham, M. B. Wang, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13959 (1998).
9. F. Sleutels, R. Zwart, D.P. Barlow, *Nature* **415**, 810 (2002).
10. K. Higo et al., *Nucleic Acids Res.* **27**, 297 (1999).
11. R. Apweiler et al., *Nucleic Acids Res.* **29**, 37 (2001).
12. J. L. Reichmann et al., *Science* **290**, 2105 (2000).
13. D. T. Jones et al., *Biochemistry* **33**, 3038 (1994).
14. K. Nakai, P. Horton, *Trends Biochem. Sci.* **24**, 34 (1999).
15. PSORT is at <http://psort.nibb.ac.jp>.
16. Bioverse is at <http://bioverse.combio.washington.edu/oriza-sativa>, established by J. McDermott, R. Samudrala and Kane Ka-Shu Wong. We used information from this site in the preparation of figs. S4 and S5.
17. We thank the following people (in alphabetical order in each group) for discussion, encouragement, and technical assistance: K. Kainuma, I. Nonomura, N. Sone, and K. Watanabe [Bio-Oriented Technology Research Advancement Institution (BRAIN)]; K. Onodera, O. Kakui, H. Ishihama, and M. Yokota (FAIS); K. Higo, Y. Otake, I. Hasegawa, S. Iizumi, S. Kimura, M. Okamura, H. Satoh, T. Semba,

K. Takeuchi, and K. Toyoshima (NIAS); R. Ohsugi and Y. Okawa (The Ministry of Agriculture, Forestry, and Fisheries of Japan); T. Akimura, T. Hanagaki, W. Hashizume, K. Hayashida, K. Hiramoto, T. Hiraoka, T. Hirozane, F. Hori, K. Imotani, H. Katoh, Y. Kojima, M. Kouda, S. Koya, C. Kurihara, T. Matsuyama, A. Miyazaki, M. Murata, M. Nakamura, K. Nishi, K. Nomura, R. Numasaki, M. Ohno, C. Sakai, K. Sakai, N. Sakazume, H. Sano, H. Saitoh, Y. Sogabe, M. Tagami, A. Tagawa, F. Takahashi, S. Takaku-Akahira, Y. Takeda, A. Tomaru, H. Torigoe, T. Toya, T. Tanaka, A. Wada, and A. Yasunishi

(RIKEN), and J. Wang, J. McDermott, R. Samudrala, and G. Wong (University of Washington). We also acknowledge the great support of the Yokohama Promotion Division and the Technology Transfer Division of RIKEN. This study was supported by a Rice Genome Full-Length cDNA Library Construction Project grant from BRAIN. It was further supported by a research grant from ACT-JST (Research and Development for Applying Advanced Computational Science and Technology) of the Japan Science and Technology Corporation (JST) to Y.H. This manuscript was reviewed by X. W. Deng (Yale

University), M. Fromm (University of Nebraska), K. Shinozaki (RIKEN Institute), S. Tabata (Kazusa DNA Institute), and G. Wong (University of Washington); we greatly appreciate their input.

Supporting Online Material
www.sciencemag.org/cgi/content/full/301/5631/376/DC1
Materials and Methods
Figs. S1 to S5
Tables S1 to S8
References

9 December 2002; accepted 16 June 2003

Altered Patterns of Sleep and Behavioral Adaptability in NPAS2-Deficient Mice

Carol A. Dudley,¹ Claudia Erbel-Sieler,¹ Sandi Jo Estill,¹ Martin Reick,¹ Paul Franken,² SiNae Pitts,³ Steven L. McKnight^{1*}

Animal behavior is synchronized to the 24-hour light:dark (LD) cycle by regulatory programs that produce circadian fluctuations in gene expression throughout the body. In mammals, the transcription factor CLOCK controls circadian oscillation in the suprachiasmatic nucleus of the brain; its paralog, neuronal PAS domain protein 2 (NPAS2), performs a similar function in other forebrain sites. To investigate the role of NPAS2 in behavioral manifestations of circadian rhythm, we studied locomotor activity, sleep patterns, and adaptability to both light- and restricted food-driven entrainment in NPAS2-deficient mice. Our results indicate that NPAS2 plays a substantive role in maintaining circadian behaviors in normal LD and feeding conditions and that NPAS2 is critical for adaptability to food restriction.

Many and diverse organisms have dedicated regulatory programs that switch genes on and off as a function of the light:dark (LD) cycle (1, 2). Remarkably, for reasons not well understood, these programs continue to operate at or near a 24-hour cycle in the absence of light (3, 4).

In mammals, a specialized region of the brain termed the suprachiasmatic nucleus (SCN) functions as the master pacemaker of circadian rhythm (5, 6). The electrical and metabolic activities of neurons in the SCN fluctuate rhythmically over the LD cycle (7–10). In both diurnal and nocturnal animals, the SCN is “on” during the day and “off” at night. A heterodimeric, activating transcription factor composed of the CLOCK and BMAL1 polypeptides turns SCN gene expression on in the day (11–13). The CLOCK:BMAL1 transcription factor directly induces the expression of genes encoding negative regulators, including the cryptochrome (Cry) proteins, as well as modulatory proteins ex-

emplified by the products of the *Period* (*Per*) genes (12–18). The circadian cycle of the SCN is synchronized with the LD cycle by neuronal input from the retina (19, 20).

Organismal rhythms in mammals are now understood to be executed by the interplay between a master SCN clock and molecular clocks expressed in other parts of the body (21–24). One of these peripheral clocks is specified by a paralog of CLOCK, designated neuronal PAS domain protein 2 (NPAS2). CLOCK and NPAS2 are similar in amino acid sequence, share BMAL1 as an obligate heterodimeric partner, bind to the same DNA recognition element, are suppressed by the cryptochromes, and commonly depend on favorably reducing ratios of nicotinamide adenine dinucleotide (NAD) cofactors (18, 25–27). In contrast, the *Clock* and *Npas2* genes are expressed in regions of the body that mostly do not overlap. The *Clock* gene is expressed in the SCN (28, 29); *Npas2* is not. Instead, *Npas2* is expressed in forebrain cortices that process sensory information, as well as regions of the basal ganglia and limbic system involved in the control of anxiety and emotion (30, 31).

To investigate whether CLOCK and NPAS2 are functionally redundant or serve distinct roles in circadian behavior, we studied NPAS2-deficient (–/–) mice generated by targeted disruption (30) and backcrossed to the

C57BL/6J background over nine generations to create a congenic strain comparable to that used in analysis of CLOCK mice. We used standard assays of wheel running activity to examine circadian periodicity (32). When exposed to a 12-hour:12-hour LD cycle, wild-type mice run at night and rest in the daytime. When released into constant darkness (DD), mice continue to run during the period of subjective darkness and to rest during the period of subjective daylight. Each day, however, mice tend to initiate wheel running slightly earlier, revealing an intrinsic period (τ) a bit shorter than 24 hours. NPAS2 heterozygous (+/–) and NPAS2 (–/–) mice displayed rhythmic locomotor activity largely comparable to that of wild-type (+/+) littermates under normal LD conditions (fig. S1). Rhythmic oscillation of locomotor activity persisted under DD conditions, with wild-type and (+/–) mice both displaying periods of 23.7 hours. NPAS2 (–/–) mice reproducibly displayed a shorter intrinsic period of 23.5 hours that persisted over at least 48 days under conditions of constant darkness. In sharp contrast, CLOCK (–/–) mice studied under similar conditions show a lengthening in period, followed by a rapid deterioration of circadian rhythmicity (28, 33).

Closer inspection of actograms of wild-type and NPAS2 (–/–) mice revealed a substantive difference in wheel running during the active, nocturnal period of the circadian cycle. In contrast to wild-type mice, which displayed a prominent break during the middle of the night, NPAS2 (–/–) mice postponed or completely skipped this break (Fig. 1A). To quantify this difference in activity, we generated aggregate activity profiles and calculated half-maximal nocturnal activity (Fig. 1B). Both wild-type and NPAS2 (–/–) mice exhibited high levels of running activity commencing at lights off. However, at 6.6 hours into the dark period [Zeitgeber time (ZT) 18.6], activity in wild-type and NPAS2 (–/–) animals began to diverge. Wheel running activity in wild-type mice began a steep decline, reaching a half-maximal level at 7.8 hours into the dark period (ZT 19.8). During this same period, wheel running activity in NPAS2 (–/–) mice was sustained at high levels. Not until 9.8 hours into the dark period, or a full 2 hours after activity had subsided to half-maximal in wild-type animals,

¹Department of Biochemistry, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390–9152, USA. ²Department of Biological Sciences, Stanford University, Stanford, CA 94305–5020, USA. ³Department of Psychology, Columbia University, 1190 Amsterdam Avenue, New York, NY 10027, USA.

*To whom correspondence should be addressed. E-mail: smckni@biochem.swmed.edu

ERRATUM

post date 26 September 2003

REPORTS: "Collection, mapping, and annotation of over 28,000 cDNA clones from *japonica* rice" by the Rice Full-Length cDNA Consortium (18 July 2003, p. 376). Under the RIKEN part of the author list, three of authors' names were spelled incorrectly: Wataru Hashizume, not Wataru Hashidume; Yoshiyuki Ishii, not Yoshiki Ishii; and Hideaki Konno, not Hedeaki Konno. Also, one author name was missing: Ayako Yasunishi. In the acknowledgments in reference 17, the following names should not have appeared: W. Hashizume, K. Imotani, A. Miyazaki, and A. Yasunishi.