phytoplankton communities of oligotrophic ocean systems, and still less so for terrestrial plants. We have described phenotypes likely to result from changes in CO_2 concentrations alone; how changes in other variables, such as temperature, pH and nutrients modify these phenotypes remain to be seen. Selection experiments in more realistic systems will be necessary to validate the evolutionary response to global atmospheric change. □

Methods

Selection experiment

Ten replicate lines were founded from a single clone of *C. reinhardtii* M566B (laboratory isolate), and ten replicate lines were founded from a single clone of CC2344 (Chlamydomonas Genetics Center, Duke University). Five replicates from each clone were grown in an increasing CO_2 environment and five replicates from each clone were grown in an ambient CO_2 environment. The ambient CO_2 environment consisted of flasks being bubbled with air containing 430 p.p.m. CO_2 for the entire experiment. Lines in the high CO_2 treatment were initially grown in flasks being bubbled with air containing 430 p.p.m. CO_2, and CO_2 levels were raised steadily to 1,050 p.p.m. over the first 600 generations of the experiment. These lines were then grown at 1,050 p.p.m. CO_2 for a further 400 generations. Lines were propagated by batch culture grown in bubbled flasks containing 300 ml of Suoka high salt medium[28] (HSM) in a phytotron chamber under constant light ($800 \pm 20\,\mu mol\,m^{-2}\,s^{-1}$) at 25 °C. One millilitre (about 10^5 cells) was transferred every 3–4 days for approximately 1,000 generations for each replicate line.

Pure culture growth rates and limiting densities

Pure culture growth rates were measured in 384-well plates containing 90 μl HSM per well. Cultures were first acclimated (3–6 days), then diluted and transferred to assay plates. For the two evolved lines that often failed to grow at ambient CO_2 concentrations, several extra acclimation cultures were inoculated, and the surviving cultures were used for growth assays. The plates were grown in the same phytotron chamber as above at either 430 p.p.m. or 1,050 p.p.m. CO_2. Absorbance of each culture was measured every 24 h. Limiting densities were calculated from the maximum absorbance maintained by a culture. Values given are means of three independent replicates.

Competitive fitness assay

Competitive fitness was measured by inoculating 300 ml of HSM with equal volumes (approximately equal numbers) of acclimated selection line and a marked strain CC48 arg^- (from Chlamydomonas Genetics Center, Duke University). The flasks were grown in the same phytotron chamber as above, bubbled with either 430 p.p.m. or 1,050 p.p.m. CO_2. The cultures were sampled every 3 days and plated on HSM plus arginine plates. After colony growth, the plates were counted and then replica-plated onto HSM-only plates. The relative frequencies of marker and selection lines were calculated by difference. Dead (arginine-requiring) colonies were usually visible on the HSM-only plates. In cases where the selection lines were absent on plates, they were assumed to be present just below the detection limit of the assay, and were entered into the analysis as having a frequency of 0.005. Values given are means from three independent replicates.

Photosynthesis and respiration assays

Photosynthetic oxygen evolution and respiration (oxygen uptake in the dark) were measured in a Clark-type oxygen electrode illuminated at $800\,\mu mol\,m^{-2}\,s^{-1}$. Cultures were depleted of oxygen by bubbling with N_2/CO_2 at either 430 p.p.m. or 1,050 p.p.m. CO_2. Net photosynthesis (oxygen electrode output from illuminated cells) was used for analysis. Respiration was calculated from oxygen uptake in the dark immediately after a light period. Values given are means of two independent replicate measurements. Chlorophyll was determined by acetone extraction[28]. Values given are means of two replicate measurements from the same culture.

Cell measurements

Cells from acclimated liquid cultures were fixed and 200 (high lines) or 170 (ambient lines) cells measured under a microscope. Cell volume was calculated based on the shape of cells[29].

Received 26 May; accepted 17 August 2004; doi:10.1038/nature02945.

1. Urban, O. Physiological impacts of elevated CO_2 concentration ranging from molecular to whole plant responses. *Photosynthetica* **41,** 9–20 (2003).
2. Griffin, K. L. *et al.* Plant growth in elevated CO_2 alters mitochondrial number and chloroplast fine structure. *Proc. Natl Acad. Sci. USA* **98,** 2473–2478 (2001).
3. Watson, R., Houghton, J. & Yihui, D. (eds) *Climate Change 2001: The Scientific Basis* (Intergovernmental Panel on Climate Change, Geneva, 2001).
4. Saxe, H., Ellsworth, D. S. & Heath, J. Tree and forest functioning in an enriched CO_2 atmosphere. *New Phytol.* **139,** 395–436 (1998).
5. Lee, T. D., Tjoelker, M. G., Ellsworth, D. S. & Reich, P. B. Leaf gas exchange responses of 13 prairie grassland species to elevated CO_2 and increased nitrogen supply. *New Phytol.* **150,** 405–418 (2001).
6. Fuhrer, J. Agroecosystem responses to combinations of elevated CO_2, ozone and global climate change. *Agric. Ecosyst. Environ.* **97,** 1–20 (2003).
7. Luescher, A., Hendrey, G. R. & Noesburger, J. Long-term responsiveness to free air CO_2 enrichment of functional types, species and genotypes of plants from fertile permanent grassland. *Oecologia* **113,** 37–45 (1998).
8. Tilman, D. & Lehman, C. Human-caused environmental change: Impacts on plant diversity and evolution. *Proc. Natl Acad. Sci. USA* **98,** 5433–5440 (2001).
9. Bazzaz, F. A., Jasienski, M., Thomas, S. C. & Wayne, P. Microevolutionary responses in experimental populations of plants to CO_2-enriched environments. *Proc. Natl Acad. Sci. USA* **92,** 8161–8165 (1995).
10. Ward, J. K., Antonovics, J., Thomas, R. B. & Strain, B. R. Is atmospheric CO_2 a selective agent on model C3 annuals? *Oecologia* **123,** 330–341 (2000).
11. Andalo, C., Goldringer, I. & Godelle, B. Inter- and intragenotypic competition under elevated carbon dioxide in *Arabidopsis thaliana*. *Ecology* **82,** 157–164 (2001).
12. Potvin, C. & Tousignant, D. Evolutionary consequences of simulated global change: Genetic versus plastic responses. *Oecologia* **108,** 683–693 (1997).
13. Raven, J. A. & Girard-Bascou, J. Algal model systems and the elucidation of photosynthetic metabolism. *J. Phycol.* **37,** 943–950 (2001).
14. Colman, B., Huertas, I. E., Bhatti, S. & Dason, J. S. The diversity of inorganic carbon acquisition mechanisms in eukaryotic microalgae. *Funct. Plant Biol.* **29,** 261–270 (2002).
15. Falconer, D. S. Selection in different environments: effects on environmental sensitivity (reaction norm) and on mean performance. *Genet. Res.* **56,** 57–70 (1990).
16. Sültemeyer, D. Carbonic anhydrase in eukaryotic algae: characterization, regulation, and possible function during photosynthesis. *Can. J. Bot.* **76,** 962–972 (1998).
17. Badger, M. R. *et al.* The diversity and coevolution of Rubisco, plastids, pyrenoids, and chloroplast-based CO_2 concentrating mechanisms in algae. *Can. J. Bot.* **76,** 1052–1071 (1998).
18. Badger, M. R. & Spalding, M. H. *Photosynthesis; Physiology and Metabolism* 369–397 (Kluwer, New York, 2000).
19. Thyssen, C., Hermes, M. & Sültemeyer, D. Isolation and characterization of *Chlamydomonas reinhardtii* mutants with an impaired CO_2-concentrating mechanism. *Planta* **217,** 102–112 (2003).
20. Suzuki, K. & Spalding, M. H. Adaptation of *Chlamydomonas reinhardtii* high-CO_2-requiring mutants to limiting CO_2. *Plant Physiol.* **90,** 1195–1200 (1988).
21. Spalding, M. H., Van, K., Wang, Y. & Nakamura, Y. Acclimation of *Chlamydomonas* to changing carbon availability. *Funct. Plant Biol.* **29,** 221–230 (2002).
22. Maclean, R. C. & Bell, G. Experimental adaptive radiation in *Pseudomonas*. *Am. Nat.* **160,** 569–581 (2002).
23. Turner, P. E. & Elena, S. F. Cost of host radiation in an RNA virus. *Genetics* **156,** 1465–1470 (2000).
24. Travisano, M. & Lenski, R. E. Long-term experimental evolution in *Escherichia coli*. IV. Targets of selection and the specificity of adaptation. *Genetics* **143,** 15–26 (1996).
25. Notley-McRobb, L. & Ferenci, T. Adaptive *mgl*-regulatory mutations and genetic diversity evolving in glucose-limited *Escherichia coli* populations. *Environ. Microbiol.* **1,** 33–43 (1999).
26. Yedid, G. & Bell, G. Macroevolution simulated with autonomously replicating computer programmes. *Nature* **420,** 810–812 (2002).
27. Xue, X. *et al.* Interactions between photosynthesis and respiration in the green alga *Chlamydomonas reinhardtii*. *Plant Physiol.* **112,** 1005–1014 (1996).
28. Harris, E. H. *The Chlamydomonas Sourcebook: A Comprehensive Guide to Biology and Laboratory Use* (Academic, San Diego, California, 1989).
29. Hillebrand, H., Dürselen, C.-D., Kirschtel, D., Pollingher, U. & Zohary, T. Biovolume calculation for pelagic and benthic microalgae. *J. Phycol.* **35,** 403–424 (1999).

Acknowledgements This research was supported by a Discovery Grant from the Natural Sciences and Engineering Council of Canada to G.B.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.B. (graham.bell@mcgill.ca).

Pack-MULE transposable elements mediate gene evolution in plants

Ning Jiang[1]*†, Zhirong Bao[2]*†, Xiaoyu Zhang[1]†, Sean R. Eddy[2] & Susan R. Wessler[1]

[1]*Department of Plant Biology, University of Georgia, Athens, Georgia 30602, USA*
[2]*Howard Hughes Medical Institute and Department of Genetics, Washington University, St Louis, Missouri 63108, USA*

* These authors contributed equally to this study
† Present addresses: Department of Horticulture, Michigan State University, East Lansing, Michigan 48824, USA (N.J.); Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA (Z.B.); Department of Molecular, Cell, and Developmental Biology, University of California, Los Angeles, California 90095, USA (X.Z.)

***Mutator*-like transposable elements (MULEs) are found in many eukaryotic genomes and are especially prevalent in higher plants[1–3]. In maize, rice and *Arabidopsis* a few MULEs were shown to carry fragments of cellular genes[4–6]. These chimaeric elements are called Pack-MULEs in this study. The abundance of MULEs in rice and the availability of most of the genome sequence permitted a systematic analysis of the prevalence and**

nature of Pack-MULEs in an entire genome. Here we report that there are over 3,000 Pack-MULEs in rice containing fragments derived from more than 1,000 cellular genes. Pack-MULEs frequently contain fragments from multiple chromosomal loci that are fused to form new open reading frames, some of which are expressed as chimaeric transcripts. About 5% of the Pack-MULEs are represented in collections of complementary DNA. Functional analysis of amino acid sequences and proteomic data indicate that some captured gene fragments might be functional. Comparison of the cellular genes and Pack-MULE counterparts indicates that fragments of genomic DNA have been captured, rearranged and amplified over millions of years. Given the abundance of Pack-MULEs in rice and the widespread occurrence of MULEs in all characterized plant genomes, gene fragment acquisition by Pack-MULEs might represent an important new mechanism for the evolution of genes in higher plants.

MULEs are DNA transposons that can be classified as either autonomous (transposase-encoding) or non-autonomous (not encoding but requiring transposase). Unlike non-autonomous DNA elements from most transposon families, non-autonomous MULEs are usually not deletion derivatives of autonomous MULEs. Instead, a variety of sequences are found between MULE terminal inverted repeats (TIRs) including fragments from host genes. This phenomenon was first reported for the maize *Mu1* element, which contains part of a gene of unknown function called *MRS-A*[4,7]. Recently a few *Arabidopsis* and rice MULEs harbouring fragments of host genes were reported[5,6]. Such MULEs are here referred to as Pack-MULEs.

To assess the impact of Pack-MULEs on gene and genome evolution it is necessary first to evaluate the abundance of these elements. To determine the prevalence of Pack-MULEs on a genome-wide basis, we turned to rice with its nearly complete genome sequence[8–10]. Using a computational approach (see Methods and Supplementary Methods for details), 266 families of MULE-related sequences were identified and used to search 440 megabases (Mb) of Nipponbare sequence, yielding about 67,000 non-redundant hits. These hits were used to identify Pack-MULEs by employing the following criteria: first, TIRs separated by less than 5 kilobases (kb) and internal sequences should not contain another transposable element (TE); second, TIRs should belong to the same MULE family (as identified by RECON, software developed for *de novo* repeat identification[11]); third, TIRs should be in inverted orientation with terminal sequences outwards (as in previously described MULEs); last, the sequence between TIRs must be highly similar (BLASTX $E < 10^{-9}$) to non-transposase and non-hypothetical proteins in GenBank and in the *indica* rice genome (see Methods for details)[10].

With these criteria, 1,380 Pack-MULEs were detected. To test whether the search identified genuine Pack-MULEs or whether most of the 1,380 sequences were fortuitous associations of independent MULE TIRs reflecting the high local density of MULEs (150 hits per Mb) in the genome, two control experiments were performed. First, if the TIRs were generated by independent events, their orientation should be uncorrelated. Thus, the same search but with TIRs pointed in the 'wrong' direction (inwards instead of outwards) should produce a similar number of Pack-MULEs. However, this search yielded only 40 hits, compared with the original 1,380. Second, a random sample of 100 elements and their flanking sequences was screened manually for target site duplications (TSDs; see Methods) that are generated on insertion. Because MULEs have no significant target sequence preference, most insertions should have a unique TSD[5,6,12–14]. Thus, a recognizable TSD flanking the two TIRs indicates that the TIRs belong to a single element, rather than independent MULE ends that fortuitously flank a gene. Seventeen of the 100 elements were not flanked by identifiable TSDs, indicating that about 17% of the 1,380 elements might be false positives (see Methods and Supplementary Tables 1 and 2 for results and criteria used to identify TSDs). This value is probably an overestimate because the TSDs of old Pack-MULEs might have sustained mutations making them unrecognizable, or TSDs might have been deleted owing to 'abortive transposition'[15,16]. Taken together, the results indicate that the vast majority of the original search output is Pack-MULEs.

Structures of representative Pack-MULEs are shown in Fig. 1, in which all elements shown have similar TIRs (all belong to the same RECON family) but sequences between the TIRs are largely different and include putative genes (sequences highly similar to known genes; Fig. 1a, b) and 'unknown' proteins (sequences matching cDNAs but not known genes; Fig. 1c). Collectively, the protein hits of the 1,380 Pack-MULEs involve diverse cellular functions including metabolism, transcription, cell defence and signal transduction, indicating that a wide variety of sequences have been captured.

Given the stringent criteria used, it is likely that genuine Pack-MULEs were missed and that there are significantly more than 1,380 Pack-MULEs in the genome. To estimate how many elements were missed by the genome-wide analysis, a second search was conducted of chromosomes 1 and 10, for which the sequences are nearly complete and systematic annotation is available[17,18]. In this search, TIRs separated by more than 5 kb and MULEs with other repeats inserted in their internal sequence were not excluded. Instead, insertions were eliminated manually, and the remaining MULE sequences were subjected to the protein database search as described above, followed by examination of flanking sequences for the presence of a TSD (see Supplementary Tables 3 and 4 for TSDs). Using these criteria, 475 Pack-MULEs were found on the 65.7 Mb of chromosomes 1 and 10 (Fig. 2; Supplementary Tables 3 and 4). Of these, only 172 were identified in the genome-wide search, indicating that more than 60% ($(475 - 172)/475 = 64\%$) were missed.

Correcting for false positives (17%) and missed elements (64%) results in an estimate of 3,200 Pack-MULEs ($1,380 \times (100 - 17) \div (100 - 64)$) in 440 Mb of genome sequence. This assumes that the distribution of Pack-MULEs on chromosomes 1 and 10 is representative of those in the rest of genome. The distributions of the 319 Pack-MULEs on chromosome 1 and the 156 Pack-MULEs on chromosome 10 are shown in Fig. 2. Like the

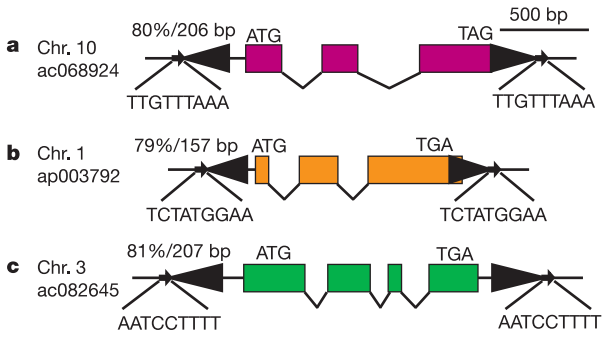

Figure 1 Pack-MULEs in Nipponbare. **a**, Putative peptide transporter; **b**, mitogen activated protein kinase-like protein; **c**, unknown protein. Pack-MULE TIRs are shown as black arrowheads, and black horizontal arrows indicate TSDs, with their sequences shown underneath. Exons are depicted as coloured boxes and introns as the lines connecting exons; other sequences are shown as horizontal lines. The GenBank accession number of the PAC or BAC sequence where the MULE is found and the chromosome where it is located are indicated. All elements depicted here have TIRs with high sequence identity. The gene prediction was furnished by the corresponding sequence provider. The similarity (percentage) and the length of the TIR of each element are also shown.

rice genes annotated on these chromosomes[17,18], Pack-MULEs are less abundant in the heterochromatic regions around centromeres and on the short arm of chromosome 10. Despite this uneven distribution, the average density of Pack-MULEs on these chromosomes is comparable (7.4 per Mb for chromosome 1 versus 7.0 per Mb for chromosome 10; see Supplementary Tables 3 and 4). If on average there are 7.2 Pack-MULEs per Mb, there would be about 3,100 Pack-MULEs in the genome ($7.2 \times 430 = 3{,}100$).

Among the 475 Pack-MULEs identified on chromosomes 1 and 10, 94 (20%) have two or more copies on the two chromosomes. These paralogous Pack-MULEs were probably generated by transposition rather than by large-scale genome duplication because in all cases distinct TSDs flanked duplicate elements (Supplementary Table 5). The genomic copy number of individual Pack-MULEs was estimated by randomly selecting 100 of the 475 Pack-MULEs on chromosomes 1 and 10 and using them to search rice genomic sequence for the presence of elements with both similar TIRs and internal regions (see Methods). Of the 100 Pack-MULEs, 27 are single-copy elements, 63 have two to five copies, and 10 have more than five copies (ranging from 6 to 14). With an average of three copies in the genome, most Pack-MULEs are low-copy-number elements.

To investigate the origin of the sequences captured by Pack-MULEs, the 100 randomly chosen Pack-MULEs from chromosomes 1 and 10 were used to query the rice genome for similar sequences that are not flanked by MULE TIRs. Among the 100 Pack-MULEs, 80 have one or more significant homologues (BLASTN $E < 10^{-10}$) (Supplementary Table 6) with a DNA sequence identity of 78–100% (mean 91%). Assuming that the captured sequence and the genomic homologue were initially identical, sequence acquisition must have occurred over a long time frame, perhaps millions of years, and might still be occurring. Of the 80 Pack-MULEs with identifiable genomic homologues, 73 have one or more homologues corresponding to the coding region of a predicted gene, and 46 have captured genomic sequences that are represented in cDNA collections (Supplementary Table 6). Most of these captured sequences are probably gene fragments, not complete genes, because the average length of captured fragments is only 325 bp (from 47 to 986 bp; Fig. 3, Supplementary Table 6). In addition, sequence acquisition by Pack-MULEs seems to occur at the DNA level as indicated by the conservation of introns between Pack-MULEs and their genomic homologues (where the corresponding cDNA sequences are available; see below and Fig. 3a, c).

There are several possible fates for the sequences in the remaining Pack-MULEs (20 of 100) that display significant similarity to proteins in the database but not to rice genomic DNA. Acquired sequences might be from unsequenced regions (for example, heterochromatic regions), from rapidly evolving sequences or from sequences lost from the genome after capture.

From an evolutionary standpoint, of potentially greater interest is the discovery that Pack-MULEs can acquire genomic fragments from two or more chromosomal loci (Fig. 3). Of the 80 (out of 100) Pack-MULEs with identified genomic homologues, 18 (23%) contain sequences from two or more loci. The significant number of elements with sequences acquired from multiple loci led to a concern that such elements might be an artefact of sequence assembly errors. However, assembly errors would not generate elements that are flanked by a TSD. In this search all of the 18 chimaeric elements have a TSD. In addition, the chimaeric structure of several Pack-MULEs could be verified by the identification of the corresponding cDNA (see below, Fig. 3a, b and Supplementary Fig. 1D, F, G, H). Assembly error was also ruled out by PCR amplification of fragments of the predicted size from Nipponbare genomic DNA by using primers designed to amplify two overlapping fragments of each of 13 Pack-MULEs and flanking sequences (Supplementary Methods, Supplementary Table 7 and Supplementary Fig. 2).

Sequence divergence between captured gene fragments and genomic homologues permitted a determination of whether captured fragments are expressed (Supplementary Table 8). A search of the 475 elements on chromosomes 1 and 10 revealed that 25 (5%) are transcribed, on the basis of exact matches with rice full-length cDNAs[19]. Among the 34 cDNAs that correspond to the 25 Pack-MULEs (seven elements having two or more cDNA matches; Supplementary Table 8), the 5′ ends of 23 (68%) are located within or adjacent to (within 40 bp) the TIR, and 8 (23%) were mapped to other parts of the element. Only three (9%) transcripts seem to be due to transcriptional readthrough, because their 5′ ends were mapped to sequences flanking the Pack-MULE. Thus, the vast majority of Pack-MULE transcription is initiated from promoters in element sequences; either by promoters in the TIR or in acquired genomic sequence.

As another measure of the expression of captured DNA, the 475 Pack-MULEs from chromosomes 1 and 10, and the 1,380 Pack-MULEs from the genome-wide search, were used to query a database of 2,528 unique rice peptides[20]. Although this database might represent less than 5% of rice proteins, six perfect matches were detected, with no other identical matches in the genome (see Methods and Supplementary Table 9).

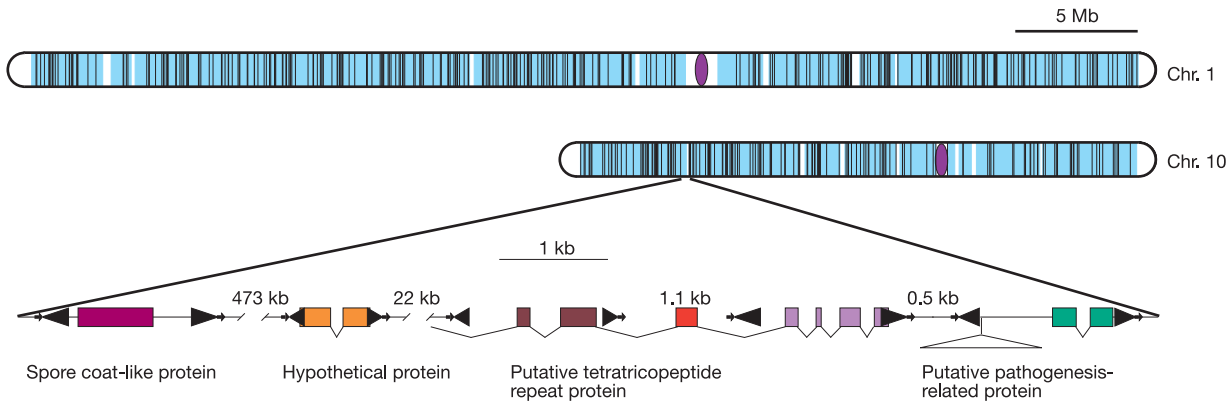

Figure 2 Distribution of Pack-MULEs on chromosomes 1 and 10. Blue blocks represent sequenced regions, whereas white spaces indicate gaps. Individual Pack-MULEs are represented by a vertical black line, and the purple ovals locate the centromeres. Below chromosome 10 a region has been expanded, and individual Pack-MULEs are depicted as described in Fig. 1. Distances between adjacent Pack-MULEs are noted. A TE insertion in a Pack-MULE is noted with an open triangle.

Recognizing that rice expression libraries are incomplete, we sought an independent measure of gene function. To this end, ratios of the non-synonymous (K_a) to synonymous (K_s) substitution rates were calculated for the internal sequences of the randomly selected Pack-MULEs on chromosomes 1 and 10 (see above) and their genomic homologues. This analysis was restricted to the 54 sequence pairs (MULE versus genomic homologue; DNA sequence similarity 80–98%) sharing more than 50 amino acids. A K_a/K_s ratio significantly less than 1 suggests that the function of the acquired sequence either is being maintained or has been maintained for a significant period since acquisition. Of the 54 pairs, 18 have a ratio significantly smaller than 1 ($P < 0.05$; see Methods, Supplementary Table 10). Theoretically, a low K_a/K_s might be due to the erroneous alignment of an element with a gene that is a paralogue of the gene acquired by the Pack-MULE. However, the results of a control experiment designed to test for such artefacts indicated that this was unlikely (see Supplementary Methods for details) and that the low K_a/K_s might reflect a functional constraint. Because the 54 sequence pairs were chosen from 100 Pack-MULEs and 5% of them might have a low K_a/K_s simply by chance ($P < 0.05$), the data suggest that more than 10% (18% − 5%) of the Pack-MULE sequences might have been functionally constrained despite the observation that the vast majority of captured DNA are gene fragments that are often rearranged (Fig. 3B, Supplementary Fig. 1B, 1C, 1E, 1H).

In summary, we report that there are over 3,000 Pack-MULEs in the rice genome with an average copy number of three. Because some elements contain two or more fragments, we estimate that more than 1,000 different gene fragments have been captured by MULEs. Although the mechanism of sequence capture is not yet understood, it is likely to involve the acquisition of genomic DNA rather than the cDNA copies of cellular transcripts as reported previously for human L1 retrotransposons[21,22]. Furthermore, about one-fifth of the identified Pack-MULEs contain fragments acquired from multiple genomic loci, thus demonstrating their potential to create novel genes through the duplication, rearrangement and fusion of diverse genomic sequences. □

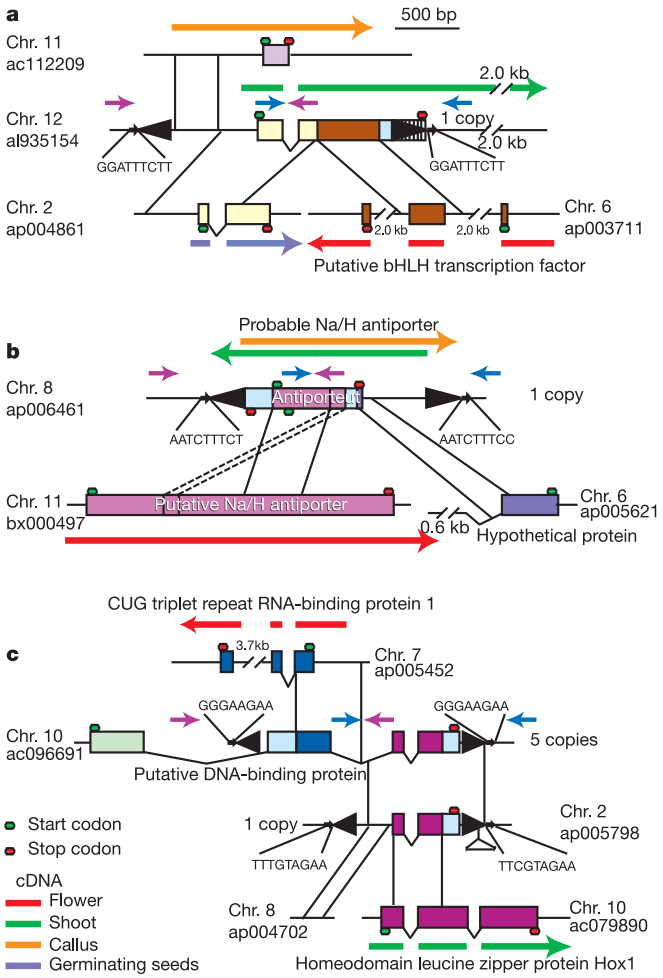

Figure 3 Structure and genomic origin of chimaeric Pack-MULEs. **a**, A Pack-MULE containing gene fragments from three genomic loci including one intron. **b**, A Pack-MULE with a new ORF derived from a rearranged Na/H antiporter gene. **c**, Possible step-wise formation of a chimaeric Pack-MULE: a Pack-MULE on chromosome 10 with sequences acquired from three loci (on chromosomes 7, 8 and 10) and an apparent intermediate element (on chromosome 2) with the gene fragments from chromosomes 8 and 10. Homologous regions are associated with solid or dashed lines. Light-blue boxes represent exons (or part of an exon) where the origin of the sequence is not clear. The striped box in **a** indicates that the TIR overlaps with the putative exon. Long coloured arrows indicate sequences matching cDNAs from the designated tissues. Small arrows (purple and blue) over the Pack-MULEs indicate the location of the primers used in PCR amplification of Pack-MULE fragments (see the text, Supplementary Methods, Supplementary Table 7 and Supplementary Fig. 2 for details). The gene name is given for putative genes and hypothetical proteins; all other genes encode 'unknown proteins'.

Methods

Computer-assisted identification of repeats in rice

Sequences of repeat families in rice were identified with RECON (version 1.03) as described[11,23], with 400 Mb of Nipponbare genomic sequence (downloaded from http://rgp.dna.affrc.go.jp on 23 August 2002) for the initial all-versus-all comparison. The resulting 3,300 repeat families (within each family, more than 90% of the sequence can be aligned between any two members on the basis of the shorter sequence) were examined individually and those derived from TEs were analysed further. MULEs were identified by features of known MULEs, including the presence of TIRs and TSDs (see Supplementary Methods for details). Curated TE sequences, including MULE TIRs, were used to mask the 440 Mb of Nipponbare sequence (downloaded from http://rgp.dna.affrc.go.jp on 7 January 2003) with RepeatMasker (version 07/07/2001, using default parameters; http:/ftp.genome.washington.edu/RM/webrepeatmaskerhelp.html). RepeatMasker output files contain annotations of all sequences that matched TEs as well as their position in the input genomic sequence and are the basis for both the genome-wide search and the search of chromosomes 1 and 10. Specifically, a putative MULE must have both ends hit by the same RECON TIR family using RepeatMasker (see Supplementary Methods and Supplementary Table 1 for how a MULE TIR is defined from RECON output. The size of Pack-MULE TIR consensus ranges from 80 to 496 bp, and the similarity of MULE TIRs range from 68% to 98%). Before the search, the RepeatMasker output was further processed so that only the best hit remained when the same locus was hit by several TE consensus sequences. The relative position of MULE and other TEs (for example, whether there are additional TEs between TIRs) was also determined using the output.

The genome-wide search of Pack-MULEs

In the genome-wide search, all possible pairs of annotated MULE TIRs located within 5 kb of each other (with no additional TEs between the TIR) were examined by following the criteria listed in the text. MULEs recovered in this way included both Pack-MULEs and other MULEs that do not contain cellular genes (or gene fragments). To distinguish the two, BLASTX was used (with the option wordmask = seg) to identify MULEs containing sequences that are similar (with $E < 10^{-9}$) to proteins in GenBank or to the annotated proteins of the 93-11 (*indica*) genome (downloaded from ftp://ftp.gramene.org/pub/gramene/protein/sequence/Os_indica_protein_BGI.txt on 17 April 2003). For each MULE, the best protein hit was excluded because it might have been a self-hit. Those MULEs with hits to non-hypothetical, non-transposase proteins were defined as Pack-MULEs.

TSDs and copy number estimates for individual Pack-MULEs

The length of Pack-MULE TSDs ranges from 7 to 11 bp. The most common TSD is 9 bp (417 out of 475, or 88%; Supplementary Tables 2, 3 and 4). For TSDs that are 9 bp or longer, a maximum of two mismatches (or one mismatch plus one single-base indel) was allowed. For TSDs that are 8 bp and 7 bp, a maximum of one mismatch or one single-base indel was allowed (see Supplementary Tables 3 and 4 for all TSD sequences). For individual Pack-MULEs, if the TIRs of the two elements belonged to the same family identified by RECON, and more than 50% of the sequence between the TIRs could be aligned (BLASTN, $E < 10^{-10}$), the two elements were defined as copies that arose from the same element.

Gene prediction

If there was a cDNA match, gene predictions (except in Fig. 3a; see below) were based on the annotation provided by the rice full-length cDNA consortium[19]. If not, the annotation used was from the corresponding sequence provider or the annotation of rice contigs (version 07232003) from the Institute of Genomic Research (TIGR) (http://www.tigr.org). In Fig. 3a, the open reading frame (ORF) inside the Pack-MULE was defined by ORFfinder (http://www.ncbi.nlm.nih.gov) on the basis of the corresponding cDNA sequence[19].

Expression analysis

The full-length cDNA data set was downloaded from http://cdna01.dna.affrc.go.jp/cDNA/ on 15 November 2003. A Pack-MULE is considered to have a cDNA match if, first, sequence similarity between the Pack-MULE and the cDNA is higher than 99.5% over the entire length of the cDNA and, second, the chromosomal position of the Pack-MULE is consistent with the genomic position of the particular cDNA provided by the rice full-length cDNA consortium[19]. Peptide sequences representing 2,528 unique proteins were downloaded from ref. 20 (http://www.pnas.org) and used as queries to search against all Pack-MULE sequences recovered in both the genome-wide search and the search of chromosomes 1 and 10 with TBLASTN[20]. Peptide sequences that generated perfect hits with Pack-MULE sequences were then used to search the rice genomic database. A particular Pack-MULE sequence was considered to have a peptide match if it was the only perfect hit in the genome.

K_a/K_s analysis

The sequences of each Pack-MULE and its corresponding genomic homologue were aligned using the 'pileup' program from the University of Wisconsin GCG program suite (version 10.1) accessed through Research Computing Resources at the University of Georgia. The alignment was based on the ORF of the genomic homologue (see Supplementary Tables 6 and 10 for details). If the genomic homologue was not predicted as a coding sequence, the ORF of the MULE was used. K_a and K_s were calculated with MEGA using the Pamilo–Bianchi–Li method[24]. The significance of purifying selection (P value) was evaluated with the z-test in MEGA.

Received 9 June; accepted 13 August 2004; doi:10.1038/nature02953.

1. Lisch, D. *Mutator* transposons. *Trends Plant Sci.* **7,** 498–504 (2002).
2. Walbot, V. & Rudenko, G. N. in *Mobile DNA II* (eds Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M.) 533–564 (American Society of Microbiology Press, Washington, D.C., 2002).
3. Chalvet, F., Grimaldi, C., Kaper, F., Langin, T. & Daboussi, M. J. *Hop*, an active *Mutator*-like element in the genome of the fungus *Fusarium oxysporum*. *Mol. Biol. Evol.* **20,** 1362–1375 (2003).
4. Talbert, L. E. & Chandler, V. L. Characterization of a highly conserved sequence related to *mutator* transposable elements in maize. *Mol. Biol. Evol.* **5,** 519–529 (1988).
5. Yu, Z., Wright, S. I. & Bureau, T. E. *Mutator*-like elements in *Arabidopsis thaliana*. Structure, diversity and evolution. *Genetics* **156,** 2019–2031 (2000).
6. Turcotte, K., Srinivasan, S. & Bureau, T. Survey of transposable elements from rice genomic sequences. *Plant J.* **25,** 169–179 (2001).
7. Bennetzen, J. L. & Springer, P. S. The generation of *mutator* transposable element subfamilies in maize. *Theor. Appl. Genet.* **87,** 657–667 (1994).
8. Goff, S. A. *et al.* A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* **296,** 92–100 (2002).
9. Normile, D. & Pennisi, E. The rice genome. Rice: boiled down to bare essentials. *Science* **296,** 32–36 (2002).
10. Yu, J. *et al.* A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* **296,** 79–92 (2002).
11. Bao, Z. & Eddy, S. R. Automated *de novo* identification of repeat sequence families in sequenced genomes. *Genome Res.* **12,** 1269–1276 (2002).
12. Cresse, A. D., Hulbert, S. H., Brown, W. E., Lucas, J. R. & Bennetzen, J. L. *Mu1*-related transposable elements of maize preferentially insert into low copy number DNA. *Genetics* **140,** 315–324 (1995).
13. Hanley, S. *et al.* Identification of transposon-tagged genes by the random sequencing of *Mutator*-tagged DNA fragments from *Zea mays*. *Plant J.* **23,** 557–566 (2000).
14. Le, Q. H., Wright, S., Yu, Z. & Bureau, T. Transposon diversity in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **97,** 7376–7381 (2000).
15. Taylor, L. P. & Walbot, V. A deletion adjacent to the maize transposable element *Mu-1* accompanies loss of *Adh 1* expression. *EMBO J.* **4,** 369–376 (1985).
16. Levy, A. A. & Walbot, V. Molecular analysis of the loss of somatic instability in the *bz2-mu1* allele of maize. *Mol. Gen. Genet.* **229,** 147–151 (1991).
17. Sasaki, T. *et al.* The genome sequence and structure of rice chromosome 1. *Nature* **420,** 312–316 (2002).
18. The Rice Chromosome 10 Sequencing Consortium. In-depth view of structure, activity, and evolution of rice chromosome 10. *Science* **300,** 1566–1599 (2003).
19. Kikuchi, S. *et al.* Collection, mapping, and annotation of over 28,000 cDNA clones from *japonica* rice. *Science* **301,** 376–379 (2003).
20. Koller, A. *et al.* Proteomic survey of metabolic pathways in rice. *Proc. Natl Acad. Sci. USA* **99,** 11969–11974 (2002).
21. Moran, J. V., DeBerardinis, R. J. & Kazazian, H. H. Jr. Exon shuffling by L1 retrotransposition. *Science* **283,** 1530–1534 (1999).
22. Emerson, J. J., Kaessmann, H., Betran, E. & Long, M. Extensive gene traffic on the mammalian X chromosome. *Science* **303,** 537–540 (2004).
23. Jiang, N. *et al.* An active DNA transposon in rice. *Nature* **421,** 163–167 (2003).
24. Kumar, S., Tamura, K., Jakobsen, I. B. & Nei, M. *MEGA2: Molecular Evolutionary Genetics Analysis software* (Arizona State Univ., Tempe, Arizona, 2001).

Supplementary Information accompanies the paper on **www.nature.com/nature**.

Acknowledgements We thank I. K. Jordan and E. K. Kentner for assistance with K_a/K_s analysis, R. Liu for verification of TSD sequences, C. Feschotte for advice, and J. Bennetzen and E. Pritham for critical reading of the manuscript. This study was supported by a grant from the NSF Plant Genome Program to S.R.W. and S.R.E. N.J., X.Z. and S.R.W. were also supported in part by grants from the NIH, and Z.B. and S.R.E. were also supported in part by grants from the NIH, the Howard Hughes Medical Institute, and Alvin Goldfarb.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.R.W. (suew@plantbio.uga.edu).

Small modulation of ongoing cortical dynamics by sensory input during natural vision

József Fiser[1,3], Chiayu Chiu[2] & Michael Weliky[1,3]

[1]*Center for Visual Science,* [2]*Interdepartmental Program in Neuroscience, and* [3]*Department of Brain and Cognitive Sciences, University of Rochester, Rochester, New York 14627, USA*

During vision, it is believed that neural activity in the primary visual cortex is predominantly driven by sensory input from the environment. However, visual cortical neurons respond to repeated presentations of the same stimulus with a high degree of variability[1–4]. Although this variability has been considered to be noise owing to random spontaneous activity within the cortex[5–7], recent studies show that spontaneous activity has a highly coherent spatio-temporal structure[8–13]. This raises the possibility that the pattern of this spontaneous activity may shape neural responses during natural viewing conditions to a larger extent than previously thought. Here, we examine the relationship between spontaneous activity and the response of primary visual cortical neurons to dynamic natural-scene and random-noise film images in awake, freely viewing ferrets from the time of eye opening to maturity. The correspondence between evoked neural activity and the structure of the input signal was weak in young animals, but systematically improved with age. This improvement was linked to a shift in the dynamics of spontaneous activity. At all ages including the mature animal, correlations in spontaneous neural firing were only slightly modified by visual stimulation, irrespective of the sensory input. These results suggest that in both the developing and mature visual cortex, sensory evoked neural activity represents the modulation and triggering of ongoing circuit dynamics by input signals, rather than directly reflecting the structure of the input signal itself.

Using an implanted multi-electrode array, we measured population activity within the visual cortex of awake, freely viewing ferrets at three different developmental stages: immediately after eye opening at postnatal day (P) 30–32 ($n = 3$), immediately after the maturation of orientation tuning and long-range horizontal connections at P44–45 ($n = 3$), and in the mature animal at P83–90 ($n = 4$). A linear array of 16 microwire electrodes, spanning 9.0 mm, was placed at 300–500-μm depth in layer 2/3 of the striate cortex, which allowed recordings to be obtained from cortical sites with varying degrees of receptive field overlap (Fig. 1a). Fifteen 100-s recording trials were acquired in head-restrained ferrets under each of three interleaved conditions: (1) presentation of dynamic natural scenes from a film; (2) presentation of dynamic random-noise stimuli; and (3) complete darkness (see Methods). At all ages,