Intra- and Interspecific Variation in Primate Gene Expression Patterns

Wolfgang Enard,[1]* Philipp Khaitovich,[1]* Joachim Klose,[2] Sebastian Zöllner,[1] Florian Heissig,[1] Patrick Giavalisco,[3] Kay Nieselt-Struwe,[4] Elaine Muchmore,[5,6] Ajit Varki,[5] Rivka Ravid,[7] Gaby M. Doxiadis,[8] Ronald E. Bontrop,[8] Svante Pääbo[1]†

Although humans and their closest evolutionary relatives, the chimpanzees, are 98.7% identical in their genomic DNA sequences, they differ in many morphological, behavioral, and cognitive aspects. The underlying genetic basis of many of these differences may be altered gene expression. We have compared the transcriptome in blood leukocytes, liver, and brain of humans, chimpanzees, orangutans, and macaques using microarrays, as well as protein expression patterns of humans and chimpanzees using two-dimensional gel electrophoresis. We also studied three mouse species that are approximately as related to each other as are humans, chimpanzees, and orangutans. We identified species-specific gene expression patterns indicating that changes in protein and gene expression have been particularly pronounced in the human brain.

Striking differences in morphology and cognitive abilities exist between humans and their closest evolutionary relatives, the chimpanzees. At least some of these differences can be assumed to form the basis for the complex and rapid cultural evolution and demographic explosions that have characterized recent human evolution (*1*). In addition, humans and chimpanzees differ in several other traits that are of medical interest, such as susceptibility to AIDS, epithelial neoplasms, malaria, and Alzheimer's disease (*2*, *3*). Although it was pointed out 25 years ago (*4*) that many of these differences may be due to quantitative differences in gene expression rather than structural changes in gene products, nothing is known about how gene expression profiles differ between humans and chimpanzees. In order to take a first step toward understanding the evolution of the mammalian transcriptome and proteome, we studied mRNA expression levels, as well as protein expression patterns, in different tissues of humans, chimpanzees (*Pan troglodytes*), orangutans (*Pongo pygmaeus*), and rhesus macaques (*Macaca mulatta*). For comparative purposes, we performed similar studies in rodent species that have diverged from each other approximately as much as humans and the great apes.

First, we compared mRNA levels in brain and liver of humans, chimpanzees, and a orangutan using Affymetrix U95A arrays (*5*), which contain oligonucleotides that examine approximately 12,000 human genes. From the brain, gray matter from the left prefrontal lobe (Brodmann area 9) was removed at autopsies from three adult male humans, three adult male chimpanzees, and one adult male orangutan. For brain and liver, two independent isolations of RNA from adjacent tissue samples were performed for each individual and analyzed independently (*5*).

All possible pairwise comparisons among the six human, six chimpanzee, and two orangutan samples were made for each tissue, and the differences in apparent expression levels were used to calculate an overall distance summarized over all genes (*6*). For the brain samples, the distances measured among the duplicates from the same individual constituted less than 14% of the distances between individuals. For the liver samples, the corresponding value was less than 12%. Because experimental variation between the tissue samples from the same individual was small, the averages of the pairwise distances measured between the duplicates for each tissue sample were used to estimate a tree depicting the overall differences in gene expression measured between individuals. The results (Fig. 1A) show that the variation in gene expression between individuals within the species is substantial, relative to the differences between humans and chimpanzee. For example, one human brain sample differs more from the other human samples than the latter differ from the chimpanzee samples. However, for both the brain and liver samples, the humans, as well as the chimpanzees, fall into two mutually exclusive groups when their gene expression patterns are related to that seen in the orangutan, which is evolutionarily further removed from humans and chimpanzees than these are from each other. When statistically tested by a bootstrap approach, this observation is supported in both liver and brain (*7*). Thus, a number of gene expression differences between humans and chimpanzees are shared among all individuals analyzed from each species. The amount of gene expression differences shared among all humans is larger than those shared among all chimpanzees. One likely factor contributing to this is that oligonucleotides complementary to human cDNAs are used to assay RNA levels not only in humans but also in chimpanzees and orangutans. Thus, nucleotide sequence differences between the last-named species and humans can be expected to reduce the apparent expression levels measured in the apes. Such differences will be assigned to the human lineage. However, the apparent acceleration on the human lineage is larger in the brain (3.8-fold) than in the liver (1.7-fold), raising the possibility that gene expression patterns may have changed more in the brain than in the liver during recent human evolution.

To investigate the latter possibility, we performed a second set of experiments using membrane-based cDNA arrays carrying

[1]Max-Planck-Institute for Evolutionary Anthropology, Inselstrasse 22, D-04103 Leipzig, Germany. [2]Institut fuer Humangenetik Charité, D-13353 Berlin, Germany. [3]Max-Planck-Institute for Molecular Genetics, D-14195 Berlin, Germany. [4]Max-Planck-Institute of Biophysical Chemistry, D-37077 Göttingen, Germany. [5]University of California San Diego, La Jolla, CA 92093, USA. [6]VA Medical Center, San Diego, La Jolla, CA 92093, USA. [7]The Netherlands Brain Bank, Amsterdam, the Netherlands. [8]Biomedical Primate Research Centre, Rijswijk, the Netherlands.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: paabo@eva.mpg.de

Table 1. Brain protein pattern differences between humans and chimpanzees as analyzed by 2D gel electrophoresis (*16*). Differences between humans and chimpanzees were scored if confirmed in three individual human-chimpanzee pairs and were analyzed in the same way as in a larger mouse study comparing *M. musculus* and *M. spretus* (*23*). Qualitative differences represent changes in electrophoretic mobility of spots, which likely result from amino acid substitutions, whereas quantitative differences reflect changes in the amount of protein.

Comparison	Analyzed spots	Differences	
		Qualitative	Quantitative
Human–chimpanzee	538	41 (7.6%)	169 (31.4%)
M. musculus–M. spretus	8767	668 (7.6%)	656 (7.5%)

21,504 DNA sequences of an average length of ~1,000 bp, amplified from 17,997 human genes of the Unigene set (*8*). When such long target sequences are used, the average nucleotide sequence difference of around 0.8% between human and chimpanzee cDNAs (*9*) is not expected to influence the results significantly. For these experiments, brain neocortex samples from the autopsies of seven humans, four chimpanzees and two macaques were used, as well as liver samples from six humans, five chimpanzees, and four macaques. In addition, blood samples were collected from 10 humans, 10 chimpanzees, and 10 rhesus macaques. To allow the same filter arrays to be used throughout the experiments, equal amounts of RNA from a given species and tissue were pooled, labeled, and hybridized to the cDNA arrays (*10*).

The relative rates of evolutionary change in the transcriptomes of the three tissues were estimated (*11*), using the macaque as an outgroup (Fig. 2). For both blood leukocytes and liver, the human expression patterns are more similar to those of the chimpanzees than to those of the macaques, reflecting the evolutionary relationships of the species. Furthermore, the extent of change on the lineages leading to the chimpanzees and the humans are equal in leukocytes and 1.3-fold different in liver. In stark contrast, the expression pattern in the chimpanzee brain cortex is more similar to that of the macaques than to that of humans. This is due to a 5.5-fold acceleration of the rate of change in gene expression levels on the lineage leading to humans. Thus, the results show that the rate of evolutionary change of gene expression levels in the brain is accelerated in the human evolutionary lineage relative to the chimpanzee, whereas no such acceleration is evident in liver or blood. It should be noted, however, that the extent of the acceleration is highly dependent on the metric used.

To gauge whether the observations made among the primate species are typical of mammals, we investigated the three mouse species, *Mus spretus*, *M. caroli*, and *M. musculus*, among which the former two species differ from *M. musculus* at silent sites, i.e., at sites that do not change the encoded amino acids, by approximately 2.5% and 4.5%, respectively (*12*). Thus, their extent of divergence from *M. musculus* is in the same order of magnitude as that of chimpanzees (1.08%) and orangutans (2.98%), respectively, from humans (*13*, *14*). Affymetrix arrays carrying oligonucleotides specific for 12,000 *M. musculus* genes (*5*) were used to analyze samples from the frontal part of the brains and livers from three individuals of *M. musculus*, three individuals of *M. spretus*, and one individual of *M. caroli*. To make the experiments as comparable as possible to the analysis of the humans and apes, outbred mice were used, and only gray matter was sampled from the frontal cortex. As in the primates, the gene expression patterns within species show great variation (Fig. 1B), as recently reported even for inbred mice (*15*). However, when the more distantly related *M. caroli* is taken into account, it is clear that all *M. musculus* and *M. spretus* individuals share gene expression patterns that separate them from the other species, as is the case for humans and chimpanzees. When these species-specific differences are compared, it is found that the change on the line to *M. musculus* is 2.1-fold and 2.3-fold that in brain and liver, respectively. Thus, as in the case of the primate analyses, the species for which the oligonucleotide array was designed shows an apparent acceleration, which is likely to be due to nucleotide sequence differences between the species analyzed. However, in the rodents, this acceleration is of similar magnitude in brain and liver, and as expected from the slightly higher genomic divergence, it is slightly higher than that seen in primate liver. Thus, these results show that gene expression differences are substantial between closely related mammalian species and supports the notion that changes in gene expression levels in the brain may have been especially pro-

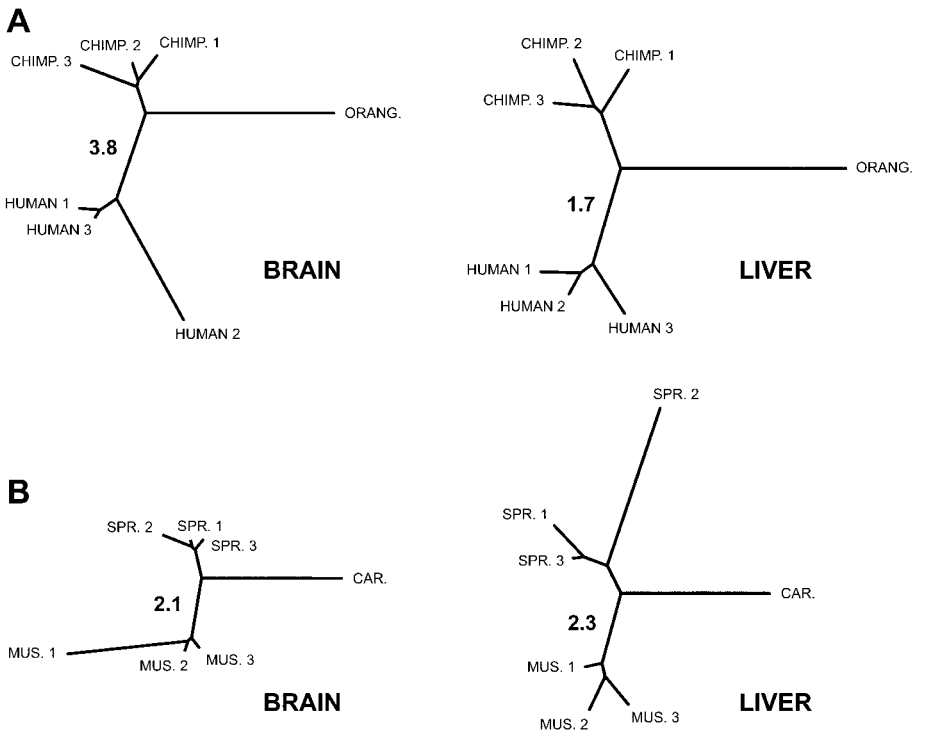

Fig. 1. Distance trees representing the relative extent of expression changes in brain and liver among (**A**) three primate and (**B**) three mouse species: MUS., *M. musculus*; SPR., *M. spretus*; and CAR, *M. caroli* (*6*). Numbers refer to the ratio between the changes common to humans and chimpanzees, and *M. musculus* and *M. spretus*, respectively.

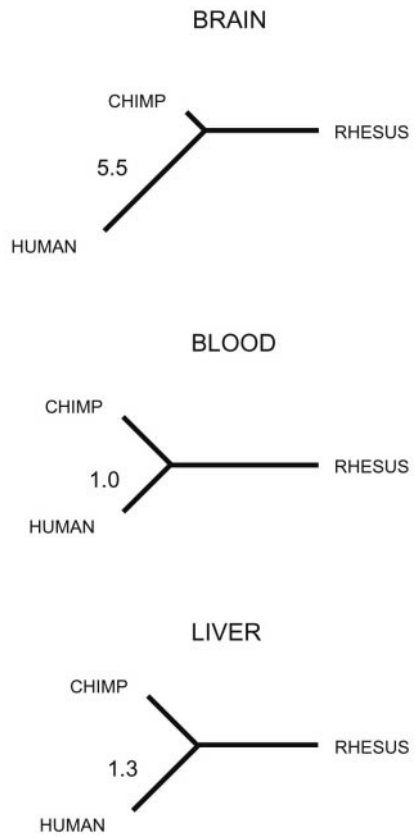

Fig. 2. Distance trees representing the relative extent of expression changes among three primate species and three tissues as assayed by the cDNA arrays (*11*). Numbers refer to the ratio between the changes common to humans and chimpanzees.

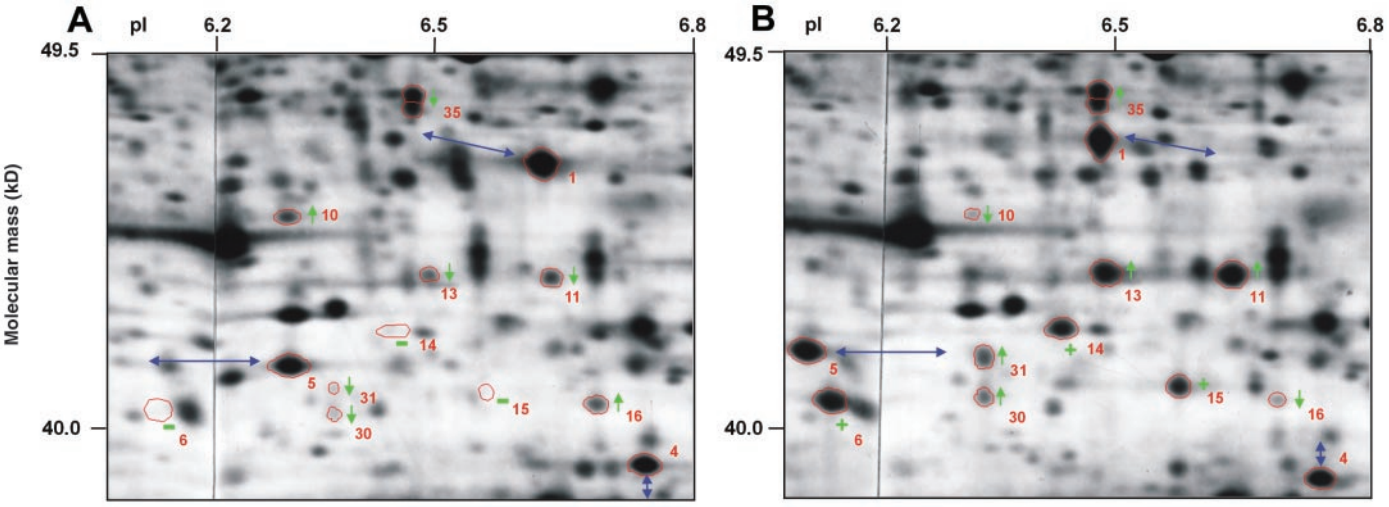

Fig. 3. Two-dimensional gel electrophoresis of proteins from the cytosolic fraction of human (**A**) and chimpanzee (**B**) *cortex frontalis*. From the total pattern, comprising about 8500 protein spots, a representative section consisting of about 200 spots is shown. Protein patterns from human and chimpanzee were compared, and changes between homologous spots found in all three human-chimpanzee pairs were scored. Three different types of variations were registered: (i) variations in electrophoretic mobility of spots (↔), most likely due to mutations affecting the structure of proteins (e.g., amino acid substitutions); (ii) variations in spot intensity (↑ or ↓) reflecting alterations in protein amount, possibly due to mutations in regulatory genes; and (iii) presence or absence variations (+ or −), which may also result from quantitative changes. For each type of variation, a few examples are indicated. From identification by mass spectrometry in both humans and chimpanzees, these proteins spots are 1, aldose reductase (gi|576365); 4, carbonic dehydratase (gi|4502517); 5, electron transfer flavoprotein (gi|2781202); 6, hypothetical protein DKFZp564D1378 (gi|14149777); 10, δ-aminolevulinate dehydratase (gi|2118316); 11, CGI-105 protein (gi|11431155); 13, hypothetical protein XP_047816 (gi|14743583); 14, malate dehydrogenase 2 (gi|5174541); 15, MAWD-binding protein (gi|16307296); 16, uncharacterized hypothalamus protein HCDASE (gi|8923864); 30, purine nucleoside phosphorylase (gi|4557801); 31, purine nucleoside phosphorylase (gi|4557801); and 35, aldehyde reductase (gi|1633300).

nounced during recent human evolution.

Differences in mRNA levels do not necessarily translate into differences in protein levels. Therefore, we investigated whether quantitative changes not only in RNA levels but also in protein levels are especially pronounced in the brain during recent human evolution. We studied protein patterns in the brains of humans and chimpanzees, as well as in *M. musculus* and *M. spretus* to put the primate differences into perspective (*16*). In each case, the tissue samples were removed from sites adjacent to the ones used in the first set of mRNA analyses from the same individuals. Soluble proteins were isolated by differential centrifugation, separated on two-dimensional (2D) polyacrylamide gels, and visualized by silver staining (Fig. 3). Two types of differences were scored: (i) shifts in the migration positions of proteins, which represent a shift in size or charge of the protein, i.e., covalent difference that in most cases are changes in amino acid sequence; (ii) differences in quantity of proteins without a shift in migration position which represent differences in amounts of protein expressed in the tissue. The relative amounts of qualitative protein differences observed between humans and chimpanzees and between *M. musculus* and *M. spretus*, respectively, are similar (Table 1), as expected from the similar extent of genomic DNA sequence differences between the two species pairs. For the two rodents, the relative amounts of quantitative protein differences are similar to the qualitative differences. In contrast, quantitative differences are approximately 6 times as common as qualitative differences when chimpanzee and human brains are compared. Thus, the human brain has probably experienced more evolutionary changes in gene expression both at the mRNA and protein levels than the two mouse species. In this regard, a recent comparison of human and great ape blood plasma proteins (*17*) found only one human-specific difference. This is in contrast to the many differences found here for soluble brain proteins and supports a more rapid rate of evolution of protein expression levels in the brain.

Our results show that that large numbers of quantitative changes in gene expression can be detected between closely related mammals. They furthermore suggest that such changes have been particularly pronounced during recent evolution of the human brain. The underlying reasons for such expression differences are likely to be manifold, for example, duplications and deletions of genes, promotor changes, changes in levels of transcription factors, and changes in cellular composition of tissues. A challenge for the future is to investigate the relative contributions of these factors to the expression differences observed. A further challenge is to clarify how many of the differences have functional consequences.

References and Notes

1. G. Klein, *The Human Career: Human Biological and Cultural Origins* (The Univ. of Chicago Press, Chicago, 1989).
2. P. Gagneux, A. Varki, *Mol. Phylogenet. Evol.* **18**, 2 (2001).
3. A. Varki, *Genome Res.* **10**, 1065 (2000).
4. M. C. King, A. C. Wilson, *Science* **188**, 107 (1975).
5. All apes used in this study died of natural causes. In all cases, postmortem times were shorter than 6 hours, and only minimal RNA degradation was seen by agarose electrophoresis. Preparation of the samples for the Affymetrix arrays, hybridization, and scanning were performed as described (*18*). Nine of 10 genes that differed at least twofold between human and chimpanzee brains could be verified by a Northern analysis. Details of experimental procedures are available on *Science* Online at www.sciencemag.org/cgi/content/full/296/5566/340/DC1 and on http://email.eva.mpg.de/~khaitovi/supplement1.html.
6. Affymetrix array results were carried out with Microarray Suite, version 4.0 (Affymetrix) by using default settings. All arrays were normalized to the same target intensity using all probe sets. The difference in scaling factor was less than threefold among all arrays. In order to build distance trees, pairwise distances between samples were calculated as the sum of the base-two logarithms of the absolute values of the "fold change" for all 12,000 genes represented on a chip. When "absent calls" were assigned to both samples in a comparison, and when the difference call for the gene was "no change," the fold change value was set to zero. The resulting distance matrix was used to build neighbor joining trees (*19*) as implemented in the PHYLIP package (*20*). The full data set is available at http://email.eva.mpg.de/~khaitovi/supplement1.html
7. The reliability of the distance trees branching pattern was estimated by 1000 bootstrap samples of the 12,000 genes. The bootstrap values for the species were >99.9% in all cases except for the chimpanzee brain branch, where in the remaining 16% of cases, the orangutan fell among the chimpanzees. There

was 100% bootstrap support for the separation of the mouse species in both tissues.
8. M. S. Boguski, G. D. Schuler, *Nature Genet.* **10**, 369 (1995).
9. I. Hellmann, unpublished data.
10. All RNA pools were hybridized 4 to 6 times to the same set of filters in order to estimate interexperimental errors and to minimize their effects through the combined analysis of several experiments. Signals that were at least 5 times above background and not influenced to more than 25% by neighboring spots were further analyzed. A gene was regarded as differently expressed if it fulfilled two criteria: (i) The difference in signal between two species was at least two-fold; and (ii) the signal between the two species was significantly different as determined by a paired *t* test. Sixteen differently expressed genes were analyzed by Northern blots, and 1 out of 12 that were detected by the Northern analyses yielded results contradictory to the arrays, whereas the remaining 11 showed expression patterns that were both qualitatively and quantitatively similar in all three species to that detected by the arrays. Details of experimental procedures are available on *Science* Online at www.sciencemag.org/cgi/content/full/296/5566/340/DC1and on http://email.eva.mpg.de/~khaitovi/supplement1.html.
11. The distance between two expression profiles of two species in a given tissue was determined by summing up the absolute ratios of the included genes given by the formula: $\sum_{i=1}^{n} \left| \log_2 \sum \frac{x_i^j}{x_i^k} \right|$, where n is the number of included genes, and is the normalized intensity of gene i as measured in species j. In order to avoid the contribution of genomic differences, only those differently expressed genes were considered that did not show the same expression difference in two or more tissues. The resulting distance matrix was used to build neighbor joining trees (*19*) as implemented in the PHYLIP package (*20*). The data are available at http://email.eva.mpg.de/~khaitovi/supplement1.html
12. We retrieved nonmitochondrial nucleotide sequences from *M. spretus* (10 sequences) and *M. caroli* (11 sequences) from GenBank and compared them with the corresponding *M. musculus* sequence. The average number of substitutions at silent sites was estimated to be 0.025 (± 0.006) for *M. spretus* and 0.045 (± 0.008) for *M. caroli*.
13. F. C. Chen, W. H. Li, *Am. J. Hum. Genet.* **68**, 444 (2001).
14. A. Fujiyama *et al.*, *Science* **295**, 131 (2002).
15. C. C. Pritchard, L. Hsu, J. Delrow, P. S. Nelson, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13266 (2001).
16. Fractionation, 2D gel electrophoresis, and matrix-assisted laser desorption time-of-flight (MALDI-TOF) mass spectrometry were done as described in (*21–23*).
17. P. Gagneux *et al.*, *Am. J. Phys. Anthropol.* **115**, 99 (2001).
18. L. Wodicka, H. Dong, M. Mittmann, M. H. Ho, D. J. Lockhart, *Nature Biotechnol.* **15**, 1359 (1997).
19. N. Saitou, M. Nei, *Mol. Biol. Evol.* **4**, 406 (1987).
20. J. Felsenstein, PHYLIP (Phylogeny Inference Package, version 3.5c; Department of Genetics, Univ. of Washington, Seattle (1993).
21. J. Klose, *Methods Mol. Biol.* **112**, 67 (1999).
22. _____, *Methods Mol. Biol.* **112**, 147 (1999).
23. _____ *et al.*, *Nature Genet.* **30**, 385 (2002); published online 25 March 2000 (10.1038/ng 861).
24. We thank C. Allen and T. Insel of the Yerkes Primate Center, Atlanta, for chimpanzee tissues; T. Arendt and U. Überham of the Paul Flechsig Institute, Leipzig, for human and mouse tissue dissections; E. Kuhn at the Biomedical Primate Research Centre, Rijswijk, for autopsies; H. Zischler and K. Mätz-Rensing from the German Primate Center, Göttingen, for macaque samples; C. Wittekind and U. Gütz from the University of Leipzig for liver resections; E. Edel from the University of Leipzig for human blood samples; F. Bonhomme and A. Orth of the Université Montpellier II for the mouse tissues; J. Retief at Affymetrix for help and support; K. Chowdhury from the Max-Planck-Institute of Biophysical Chemistry, Göttingen, and K. Krohn and Petra Süptitz from the Interdisziplinäres Zentrum für Klinische Forschung (IZKF), Leipzig, for the help with microarray experiments; T. Kitano for analyzing the mouse divergence; D. Kuhl for initial advice in the project; and many people at Max Planck Institute for Evolutionary Anthropology for stimulating discussions. We also thank the Bundesministerium für Bildung und Forschung, the Saxonium Ministry for Science, and the Max Planck Gesellschaft for financial support.

14 December 2001; accepted 27 February 2002

A MADS-Box Gene Necessary for Fruit Ripening at the Tomato *Ripening-Inhibitor* (*Rin*) Locus

Julia Vrebalov,[1,2] Diane Ruezinsky,[2] Veeraragavan Padmanabhan,[2] Ruth White,[1,2] Diana Medrano,[1,2] Rachel Drake,[3] Wolfgang Schuch,[3] Jim Giovannoni[1]*

Tomato plants harboring the *ripening-inhibitor* (*rin*) mutation yield fruits that fail to ripen. Additionally, *rin* plants display enlarged sepals and loss of inflorescence determinacy. Positional cloning of the *rin* locus revealed two tandem MADS-box genes (*LeMADS-RIN* and *LeMADS-MC*), whose expression patterns suggested roles in fruit ripening and sepal development, respectively. The *rin* mutation alters expression of both genes. Gene repression and mutant complementation demonstrate that *LeMADS-RIN* regulates ripening, whereas *LeMADS-MC* affects sepal development and inflorescence determinacy. *LeMADS-RIN* demonstrates an agriculturally important function of plant MADS-box genes and provides molecular insight into nonhormonal (developmental) regulation of ripening.

The maturation and ripening of fleshy fruits is a developmental process unique to plants and affects the quality and nutritional content of a significant portion of the human diet. Although specific fruit-ripening characteristics vary among species, ripening can be generally described as the coordinated manifestation of changes in color, texture, flavor, aroma, and nutritional characteristics that render fruit attractive to organisms receiving sustenance in exchange for assisting in seed dispersal (*1*, *2*).

Fruit species are classically defined as one of two ripening types, climacteric and nonclimacteric, where the former display a burst in respiration at the onset of ripening, in contrast to the latter. Climacteric fruit typically increase biosynthesis of the gaseous hormone ethylene, which is required for the ripening of fruit such as tomatos, bananas, apples, pears, and most stone fruit. Nonclimacteric fruit, including strawberries, grapes, and citrus fruits, do not require climacteric respiration or increased ethylene for maturation. Molecular ripening research has focused primarily on ethylene, but little is known of control before ethylene induction, nor of common regulatory mechanisms shared by climacteric and nonclimacteric species (*3*).

The tomato is a model for analysis of ripening due originally to its significance as a food source and diverse germplasm, and more recently, the availability of molecular tools (*4*) and efficient transformation (*5*). A number of tomato-ripening mutants have been useful for research and breeding (*3*). Especially interesting is the recessive *ripening-inhibitor* (*rin*) mutation that inhibits all measured ripening phenomena, including the respiratory climacteric and associated ethylene evolution, pro-vitamin A carotenoid accumulation, softening, and production of flavor compounds (*6*). The *rin* mutant exhibits ethylene sensitivity, including the seedling triple response (*7*), floral abscission, and petal and leaf senescence. Nevertheless, *rin* fruit do not ripen in response to exogenous ethylene, yet they display induction of at least some ethylene-responsive genes, indicating retention of fruit ethylene sensitivity (*8*). We interpret these results to mean that the *RIN* gene encodes a genetic regulatory component necessary to trigger climacteric respiration and ripening-related ethylene biosynthesis in addition to requisite factors whose regulation is

[1]U.S. Department of Agriculture–Agricultural Research Service (USDA-ARS) Plant, Soil and Nutrition Lab and Boyce Thompson Institute for Plant Research, Cornell University, Ithaca, NY 14853, USA. [2]Texas A&M University, Department of Horticultural Sciences, College Station, TX 77843, USA. [3]Zeneca Plant Sciences (Syngenta), Jeallots Hill Research Station, Bracknell, Berkshire RG12 6EY, UK.

*To whom correspondence should be addressed. E-mail: jjg33@cornell.edu

Intra- and Interspecific Variation in Primate Gene Expression Patterns

Wolfgang Enard *et al.*

Science **296**, 340 (2002);
DOI: 10.1126/science.1068996