and LED 8 (541 observed, 456 expected, $P <$ 0.0006)—providing evidence in plants for a link between genome organization and gene regulation.

Together these data provide an organ expression map, revealing putative localized hormone-response domains and a complex pattern of regulatory genes that could mediate primary developmental cues. These data should help identify candidate genes involved in pattern formation and cell specificity in the root, which is a model for organogenesis. The expression map will also facilitate both computational and experimental methods aimed at decoding regulatory mechanisms in the root. Thus, these results can now be used to explore how the hundreds of different expression patterns they reveal are established and interpreted at the cellular level to generate a complex organ.

References and Notes
1. N. M. Kerk, T. Ceserani, S. L. Tausta, I. M. Sussex, T. M. Nelson, *Plant Physiol.* **132**, 27 (2003).
2. T. Asano *et al.*, *Plant J.* **32**, 401 (2002).
3. D. Milioni, P. E. Sado, N. J. Stacey, K. Roberts, M. C. McCann, *Plant Cell* **14**, 2813 (2002).
4. P. J. Roy, J. M. Stuart, J. Lund, S. K. Kim, *Nature* **418**, 975 (2002).
5. H. Jasper *et al.*, *Dev. Cell* **3**, 511 (2002).
6. P. N. Benfey, J. W. Schiefelbein, *Trends Genet.* **10**, 84 (1994).
7. Materials and methods are available as supporting material on *Science* Online.
8. J. Sheen, *Plant Physiol.* **127**, 1466 (2001).
9. J. Quackenbush, *Nature Rev. Genet.* **2**, 418 (2001).
10. The program Cluster was used in the analysis and downloaded from http://rana.lbl.gov/EisenSoftware.htm.
11. M. B. Eisen, P. T. Spellman, P. O. Brown, D. Botstein, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14863 (1998).
12. K. Birnbaum *et al.*, unpublished data.
13. T. Berleth, J. Mattsson, *Curr. Opin. Plant Biol.* **3**, 406 (2000).
14. U. Wittstock, B. A. Halkier, *Trends Plant Sci.* **7**, 263 (2002).
15. L. L. Murdock, R. E. Shade, *J. Agric. Food Chem.* **50**, 6605 (2002).
16. B. A. Cohen, R. D. Mitra, J. D. Hughes, G. M. Church, *Nature Genet.* **26**, 183 (2000).
17. H. Caron *et al.*, *Science* **291**, 1289 (2001).
18. A. P. Mahonen *et al.*, *Genes Dev.* **14**, 2938 (2000).
19. M. Bonke, S. Thitamadee, A. P. Mahonen, M. T. Hauser, Y. Helariutta, *Nature*, in press.
20. J. W. Wysocka-Diller, Y. Helariutta, H. Fukaki, J. E. Malamy, P. N. Benfey, *Development* **127**, 595 (2000).
21. The plant line was generated by the Haseloff laboratory (www.plantsci.cam.ac.uk/Haseloff/Home.html). The lines were obtained through the *Arabidopsis* Information Resource (www.arabidopsis.org/).
22. Y. Lin, J. Schiefelbein, *Development* **128**, 3697 (2001).
23. M. M. Lee, J. Schiefelbein, *Cell* **99**, 473 (1999).
24. E. Truernit, N. Sauer, *Planta* **196**, 564 (1995).
25. We thank J. Malamy for valuable ideas on the protoplasting technique; H. Petri, K. Gordon, and J. Hirst for assistance in cell sorting; H. Dressman and the Duke Microarray Core Facility for assistance with microarrays; A. Pekka Mähönen and Y. Helariutta for use of the pWOL::GFP line and M. Cilia and D. Jackson for the pSUC2::GFP line, both before publication; M. Levesque for valuable discussions; and G. Sena and T. Nawy for photos. This work was supported by NSF grants MCB-020975 (P.N.B. and D.E.S.), DBI-9813360 (D.W.G.), DBI-0211857 (D.W.G.), and a Small Grant for Exploratory Research (P.N.B. and D.E.S.). The NIH supported K.B. with a postdoctoral fellowship grant (5 F32 GM20716-03).

Inferring Nonneutral Evolution from Human-Chimp-Mouse Orthologous Gene Trios

Andrew G. Clark,[1] Stephen Glanowski,[3] Rasmus Nielsen,[2] Paul D. Thomas,[4] Anish Kejariwal,[4] Melissa A. Todd,[2] David M. Tanenbaum,[5] Daniel Civello,[6] Fu Lu,[5] Brian Murphy,[3] Steve Ferriera,[3] Gary Wang,[3] Xianqgun Zheng,[5] Thomas J. White,[6] John J. Sninsky,[6] Mark D. Adams,[5]* Michele Cargill[6]†

Even though human and chimpanzee gene sequences are nearly 99% identical, sequence comparisons can nevertheless be highly informative in identifying biologically important changes that have occurred since our ancestral lineages diverged. We analyzed alignments of 7645 chimpanzee gene sequences to their human and mouse orthologs. These three-species sequence alignments allowed us to identify genes undergoing natural selection along the human and chimp lineage by fitting models that include parameters specifying rates of synonymous and nonsynonymous nucleotide substitution. This evolutionary approach revealed an informative set of genes with significantly different patterns of substitution on the human lineage compared with the chimpanzee and mouse lineages. Partitions of genes into inferred biological classes identified accelerated evolution in several functional classes, including olfaction and nuclear transport. In addition to suggesting adaptive physiological differences between chimps and humans, human-accelerated genes are significantly more likely to underlie major known Mendelian disorders.

Although the human genome project will allow us to compare our genome to that of other primates and discover features that are uniquely human, there is no guarantee that such features are responsible for any of our unique biological attributes. To identify genes and biological processes that have been most altered by our recent evolutionary divergence from other primates, we need to fit the data to models of sequence divergence that allow us to distinguish between divergence caused by random drift and divergence driven by natural selection. Early observations of unexpectedly low levels of protein divergence between humans and chimpanzees led to the hypothesis that most of the evolutionary changes must have occurred at the level of gene regulation (*1*). Recently, much more extensive efforts at DNA sequencing in nonhuman primates has confirmed the very close evolutionary relationship between humans and chimps (*2*), with an average nucleotide divergence of just 1.2% (*3–5*). The role of protein divergence in causing morphological, physiological, and behavioral differences between these two species, however, remains unknown.

Here we apply evolutionary tests to identify genes and pathways from a new collection of more than 200,000 chimpanzee exonic sequences that show patterns of divergence consistent with natural selection along the human and chimpanzee lineages.

To construct the human-chimp-mouse alignments, we sequenced PCR amplifications using primers designed to essentially all human exons from one male chimpanzee, resulting in more than 20,000 human-chimp gene alignments spanning 18.5 Mb (*6–8*). To identify changes that are specific to the divergence in the human lineage, we compared the human-chimp aligned genes to their mouse ortholog. Inference of orthology involved a combination of reciprocal best matches and syntenic evidence between human and mouse gene annotations (*9, 10*). This genome-wide set of orthologs underwent a series of filtering steps to remove ambiguities, orthologs with little sequence data, and genes with suspect annotation (*6*). The filtered ortholog set was compared to

[1]Molecular Biology and Genetics, [2]Biological Statistics and Computational Biology, Cornell University, Ithaca, NY 14853, USA. [3]Applied Biosystems, 45 West Gude Drive, Rockville, MD 20850, USA. [4]Protein Informatics, Celera Genomics, 850 Lincoln Centre Drive, Foster City, CA 94404, USA. [5]Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA. [6]Celera Diagnostics, 1401 Harbor Bay Parkway, Alameda, CA 94502, USA.

*Present address: Department of Genetics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, USA.
†To whom correspondence should be addressed. E-mail: michele_cargill@celeradiagnostics.com

other public sets and found to be highly consistent (*11*) (table S1). We used the most conservative set of 7645 genes for which we had the highest confidence in orthology and sequence annotation (*12*) (Database S1).

To identify genes that have undergone adaptive protein evolution, we applied two formal statistical tests that fit models of molecular evolution at the codon level. Both tests fit models of the nucleotide-substitution process by maximum likelihood (ML) (*13*), and both include parameters specifying rates of synonymous and nonsynonymous substitution (*14–16*). In the first (Model 1), we performed a classic test of the null hypothesis of $d_N/d_S = 1$ in the human lineage (*17, 18*). The second model is a modification of the method described by Yang and Nielsen (*16*), which allows variation in the d_N/d_S ratio among lineages and among sites at the same time. In this method (Model 2), a likelihood ratio test of the hypothesis of no positive selection is performed by comparing the likelihood values for two hypotheses. Under the null hypothesis, it is assumed that all sites are either neutral ($d_N/d_S =1$) or evolve under negative selection ($d_N/d_S < 1$). Under the alternative hypothesis, some of the sites are allowed to evolve with $d_N > d_S$ in the human lineage only (Fig. 1). We refer to this as Model 2, and to the *P*-value of neutrality as P_2 (*6*). The test based on Model 2 is not as conservative as the test based on Model 1 and may tend to detect genes with accelerated amino acid substitution rates in humans even if the average d_N/d_S rate is not larger than 1.

There were 1547 human genes and 1534 chimp genes, which met the criteria for positive selection (with $d_N/d_S >1$). The neutral null hypothesis of Model 1 was rejected for 72 genes (0.94% of the tests) at $P < 0.001$, 414 genes (5.4%) at $P < 0.01$, and 1216 genes (15.9%) at $P < 0.05$ (*12*). There were six human genes for which the neutral null hypothesis of Model 1 was rejected at $P < 0.05$ and d_N/d_S was greater than 1 (*12*). The neutral null hypothesis of Model 2 was rejected for 28 genes (0.38%) at $P_2 < 0.001$, 178 genes (2.3%) at $P_2 < 0.01$, and 667 genes (8.7%) at $P_2 < 0.05$. The relatively low overlap of these sets reflects the different nature of the tests. Of the 1547 human genes that exhibited $d_N/d_S > 1$, only 125 also fell into the class of 178 human genes with a $P_2 < 0.01$. Similarly, Model 2 can detect cases where a protein has a domain undergoing positive selection, but the overall d_N/d_S may not be elevated, and thus would be missed by Model 1. For this reason, the remainder of the analysis considers only the Model 2 test results.

Before attempting any biological inference from the results of the statistical tests, it is important to consider whether attributes like GC content, repeat density, local recombination rate, and segmental duplications might affect the rates and patterns of substitution (*19, 20*). In principle,

the ML estimation procedure corrects for variation in base composition; however, if the true substitution rate differs across the genome in a manner that is correlated with GC content, then we should be able to detect this by simple correlation (*6, 12*) (Database S2). The synonymous substitution rate was significantly correlated with the following attributes: GC content (0.164, $P < 0.0001$), local recombination rate in cM/Mb (*21*) (0.100, $P < 0.001$), and LINE (long interspersed nuclear element) density (−0.091, $P < 0.0001$). None of these factors was significantly correlated with either nonsynonymous substitution rate or P_2-value; however, genes associated with some biological processes, such as olfaction, do show nonrandom associations with genomic location [$P < 10^{-4}$, Kolmogorov-Smirnov (K-S) test] and GC content ($P < 10^{-9}$, K-S test). We also verified that segmental duplications were not responsible for distortions in the patterns of substitution seen in our tests, mostly because genes with close duplicates were underrepresented in our set because of the requirement for strict human-mouse orthology. Interestingly, the genes with P_2-values <0.05 are overrepresented in the Online Mendelian Inheritance in Man (OMIM) catalog of genes associated with genetic disease ($P = 0.009$), demonstrating the relevance of interspecific comparisons (ftp.ncbi.nih.gov/repository/OMIM/morbidmap).

Many of the 7645 genes have been classified into inferred functional categories based on the

Panther classification system (*6, 22*). We asked, for the subset of genes in each functional category, whether the distribution of P_2 values for those genes differed significantly from the P_2 distribution for the full set of 7645 genes (*6*) (tables S2 and S3). In this way, we can gain insight into higher-order biological processes and molecular functions that may be under selective pressure in a given lineage (Tables 1 and 2). The statistical tests of significance are valid as formal inferences, and these lead immediately to tentative biological hypotheses, only some of which we describe here.

In the human lineage, genes involved in olfaction show a significant tendency to be under positive selection ($P_{MW} < 0.005$) (Table 1 and Fig. 2). Nearly all the genes classified to olfaction are olfactory receptors (ORs). It seems likely that the different lifestyles of chimps and humans might have led to divergent selection pressure on these receptors. There has been a rapid acceleration of pseudogene formation in human ORs (*23*), and the acceleration of apparent amino acid substitution in pseudogenes could potentially lead to a spurious inference of selection. However, we verified that most of the OR genes in our set are bona fide genes (http://bioinformatics.weizmann.ac.il/HORDE/), indicating that these genes are either undergoing positive selection or are in the process of pseudogenization (*24*).

Fig. 1. Graphical representation of the test of positive selection (Model 2). The null hypothesis (H₀) assumes all three branches have two classes of amino acid residues: those that are neutrally evolving (p_1: $d_N = d_S$) and those that are under constraint (p_0: $d_N/d_S<$ 1). The alternative hypothesis (Hₐ) allows the human lineage to have a subset of sites (p_s) with accelerated amino acid substitution ($d_N > d_S$).

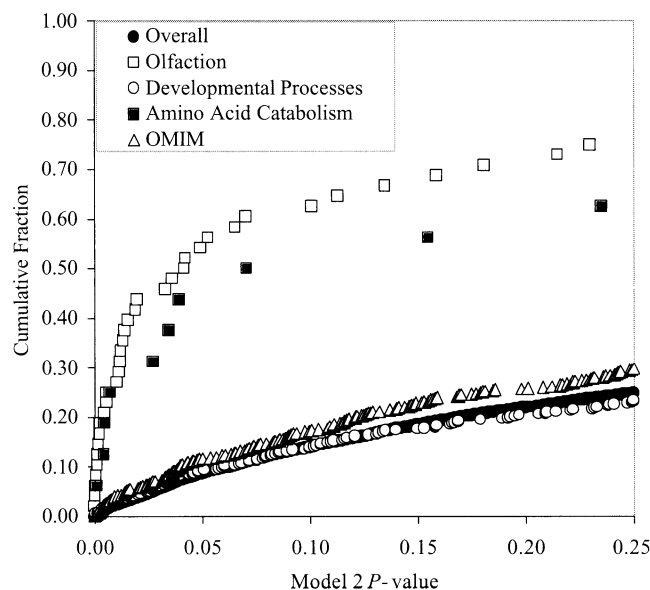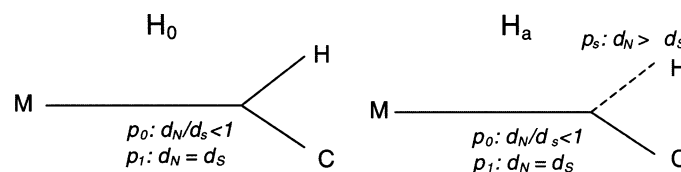**Fig. 2.** P_2-value distributions of selected groups of genes. The plot shows the cumulative fraction of selected biological processes showing the excess of cases of significant positive selection in genes for olfaction, amino acid catabolism, and Mendelian disease genes (OMIM) relative to the overall distribution of genes. The distribution of developmental genes that do not show a significant excess is shown for comparison.

Several other classes of genes (amino acid catabolism, developmental processes, reproduction, neurogenesis, and hearing) show many genes with low P_2 values, although these classes do not show significant P_{MW} values or contain fewer than 20 genes (table S1 and Fig. 2). It is possible that individual genes within these categories account disproportionately for specific phenotypic effects. For example, 7 (GSTZ1, HGD, PAH, ALDH6A1, BCKDHA, PCCB, and HAL) of the 16 genes in the amino acid catabolism category have P_2 values less than 0.05. A speculative suggestion is that this signal of positive selection may arise from different dietary habits or pressures in the two lineages. For example, branched-chain amino acid catabolism, which involves the ALDH6A1, BCKDHA, and PCCB genes, is the primary pathway for energy production from muscle protein under starvation conditions (25). For all seven genes, mutations have been found that result in human metabolic disorders, consistent with the idea that natural selection shifted these genes in a manner that is relevant to reproductive fitness.

Most of the human developmental genes with low P_2 values fall into two main categories: skeletal development (TLL2, ALPL, BMP4, SDC2, MMP20, and MGP) and neurogenesis (NLGN3, SEMA3B, PLXNC1, NTF3, WNT2, WIF1, EPHB6, NEUROG1, and SIM2). In addition, several of the genes with low P_2 values are homeotic transcription factor genes (CDX4, HOXA5, HOXD4, MEOX2, POU2F3, MIXL1, and PHTF), which play key roles in early development. Several genes associated with pregnancy, such as the progesterone receptor (PGR), GNRHR, MTNR1A, and PAPPA, appear to exhibit nonneutral divergence between humans and chimps. PGR is involved not only in maintenance of the uterus, but is also expressed on the cell membrane of sperm, where it may play a role in the acrosome reaction (26), so the physiological basis for the adaptive evolution remains unclear.

Speech is considered to be a defining characteristic of humans. The forkhead-box P2 transcription factor ($P_2 = 0.0027$) has been implicated in speech development, and has previously been identified as undergoing an unusual human-specific pattern of substitution (27). Several genes involved in the development of hearing also appear to have undergone adaptive evolution in the human lineage, and we speculate that understanding spoken language may have required tuning of hearing acuity. The gene with the most significant pattern of human-specific positive selection is alpha tectorin, whose protein product plays a vital role in the tectorial membrane of the inner ear. Single–amino acid polymorphisms are associated with familial high-frequency hearing loss (28), and knockout mice are deaf. These results strongly motivate a detailed assessment of the nature of hearing differences between humans and chimpanzees. Other genes involved in hearing that appear to be under human-specific selection include DIAPH1, FOXI1, EYA4, EYA1, and OTOR.

The inference of lineage-specific evolutionary acceleration requires a phylogenetic tree. By simply adding mouse to our alignments, we went from a directionless pairwise comparison of human and chimp to having reasonable ability to infer common ancestral state, and lineage-specific changes. These approaches will gain in both statistical and biological power as additional primate or other mammalian genomes are sequenced, enabling identification of genes that exhibited accelerated amino acid substitution since our most recent common ancestor. Although it is tempting to conclude that this will

Table 1. Biological processes showing the strongest evidence for positive selection. The top panel includes the categories showing the greatest acceleration in human lineage, and the bottom panel includes categories with the greatest acceleration in the chimp lineage.

Biological process	Number of genes*	P_{MW} (human/ Model 2)*	P_{MW} (chimp/ Model 2)*
Categories showing the greatest acceleration in human lineage			
Olfaction	48	0	0.9184
Sensory perception	146 (98)	0 (0.026)	0.9691 (0.9079)
Cell surface receptor–mediated signal transduction	505 (464)	0 (0.0386)	0.199 (0.0864)
Chemosensory perception	54 (6)	0 (0.1157)	0.9365 (0.7289)
Nuclear transport	26	0.0003	0.2001
G-protein–mediated signaling	252 (211)	0.0003 (0.1205)	0.2526 (0.0773)
Signal transduction	1030 (989)	0.0004 (0.0255)	0.0276 (0.0092)
Cell adhesion	132	0.0136	0.3718
Ion transport	237	0.0247	0.8025
Intracellular protein traffic	278	0.0257	0.8099
Transport	391	0.0326	0.7199
Metabolism of cyclic nucleotides	20	0.0408	0.1324
Amino acid metabolism	78	0.0454	0.0075
Cation transport	179	0.0458	0.8486
Developmental processes	542	0.0493	0.2322
Hearing	21	0.0494	0.9634
Categories with the greatest acceleration in the chimp lineage			
Signal transduction	1030 (989)	0.0004 (0.0255)	0.0276 (0.0092)
Amino acid metabolism	78	0.0454	0.0075
Amino acid transport	23	0.1015	0.0102
Cell proliferation and differentiation	82	0.3116	0.0182
Cell structure	174	0.2633	0.0233
Oncogenesis	201	0.3132	0.0267
Cell structure and motility	239	0.2208	0.0299
Purine metabolism	35	0.9127	0.0423
Skeletal development	44	0.2876	0.0438
Mesoderm development	168	0.5813	0.0439
Other oncogenesis	39	0.2777	0.0469
DNA repair	49	0.9363	0.0477

*The number of genes and the P_{MW} values excluding olfactory receptor genes are shown in parentheses.

Table 2. Molecular functions showing the strongest evidence for positive selection. The table includes only human-accelerated categories, because the only categories accelerated in the chimp lineage are chaperones ($P = 0.0124$), cell adhesion molecules ($P = 0.0220$), and extracellular matrix ($P = 0.0333$).

Molecular function	Number of genes*	P_{MW} (human/ Model 2)*	P_{MW} (chimp/ Model 2)*
G protein coupled receptor	199 (153)	0 (0.2533)	0.8689 (0.6776)
G protein modulator	62	0.0008	0.3776
Receptor	448	0.0030	0.9798
Ion channel	134	0.0043	0.8993
Extracellular matrix	97 (95)	0.0120 (0.0178)	0.1482 (0.1593)
Other G protein modulator	32	0.0149	0.4441
Extracellular matrix glycoprotein	44 (42)	0.0178 (0.0269)	0.1579 (0.1765)
Voltage-gated ion channel	62	0.0219	0.6692
Other hydrolase	95	0.0260	0.4823
Oxygenase	46	0.0303	0.4792
Protein kinase receptor	37	0.0314	0.6911
Transporter	214	0.0338	0.1836
Ligand-gated ion channel	45	0.0405	0.9503
Microtubule binding motor protein	22	0.0421	0.6385
Microtubule family cytoskeletal protein	54	0.0467	0.2815

*The number of genes and the P_{MW} values excluding olfactory receptor genes are shown in parentheses.

constitute a list of genes that "make us human," one has to take a step back to see the gulf that exists between understanding at this narrowly focused molecular level and at the organismal level. A large number of human genes, when transformed into mutant yeast or *Drosophila*, can rescue the mutant phenotype, but this does not make these genetically modified organisms any more human. This study has focused only on protein-coding genes, and it will require examination of regulatory sequences to determine the contribution of regulation of gene expression to the evolutionary divergence between humans and chimps.

Perhaps the best way to understand the relation between DNA sequence divergence and the differences between human and chimpanzee physiology and morphology is to compare these differences to the variability among humans. Human-chimp DNA sequence divergence is roughly 10 times the divergence between random pairs of humans. Contrasts that are under way to place human polymorphism in the context of human-specific divergence further empower these models to identify molecular targets of natural selection. Evolutionary analysis will be extended to include comparison of the X chromosome and autosomes, the impact of local recombination rates and GC content, codon-usage patterns, and divergence in regulatory sequences. Additional insight will be gained by examining sequence divergence in the context of gene-expression differences. The informativeness of all these approaches will increase by inclusion of additional mammalian genome sequences, and realization of the goal to ascribe functional significance to the complex landscape of our own genome will most effectively be made in the context of our close relatives.

References and Notes
1. M. C. King, A. C. Wilson, *Science* **188**, 107 (1975).
2. Y. Satta, J. Klein, B. Takahata, *Mol. Phylogenet. Evol.* **14**, 259 (2000).
3. F. C. Chen, W.-H. Li, *Am. J. Hum. Genet.* **68**, 444 (2001).
4. I. Ebersberger, D. Metzler, C. Schwarz, S. Pääbo, *Am. J. Hum. Genet.* **70**, 1490 (2002).
5. R. Sakate et al., *Genome Res.* **13**, 1022 (2003).
6. Detailed materials and methods are available as supporting material on *Science* Online.
7. A total of 201,805 primer pairs were successfully designed to 23,363 human coding sequences (27.6 Mb).
8. Primer pairs were amplified in 39 female human individuals (19 African-Americans, 20 Caucasians) and 1 male chimpanzee (4X0033, Southwest National Primate Research Center) by a standard PCR and sequencing protocols. Trimmed chimp sequences were BLASTed against human exon sequence (9) to create virtual transcripts.
9. J. C. Venter et al., *Science* **291**, 1304 (2001).
10. R. J. Mural et al., *Science* **296**, 1661 (2002).
11. Mouse-human orthologs were downloaded from National Center for Biotechnology Information (NCBI) HomoloGene; NCBI Homol_seq_pairs; NCBI Homology Map; and Mouse Genome Database, Mouse Genome Informatics Web Site, The Jackson Laboratory (Bar Harbor, ME).
12. All 7645 alignments in Phylip format (13) and a flatfile of genes and their associated statistics are available at http://panther.celera.com/appleraHCM_alignments/index.jsp. Sequences have been deposited in GenBank under accession codes AY398769-AY421703.
13. J. Felsenstein, *J. Mol. Evol.* **17**, 368 (1981).
14. N. Goldman, Z. Yang, *Mol. Biol. Evol.* **11**, 725 (1994).
15. S. V. Muse, B. S. Gaut, *Mol. Biol. Evol.* **11**, 715 (1994).
16. Z. Yang, R. Nielsen, *Mol. Biol. Evol.* **19**, 908 (2002).
17. Z. Yang, R. Nielsen, *J. Mol. Evol.* **46**, 409 (1998).
18. Z. Yang, R. Nielsen, *Mol. Biol. Evol.* **17**, 32 (2000).
19. I. Hellmann et al., *Genome Res.* **13**, 831 (2003).
20. J. A. Bailey et al., *Science* **297**, 1003 (2002).
21. A. Kong et al., *Nature Genet.* **31**, 241 (2003).
22. P. D. Thomas et al., *Nucleic Acids Res.* **31**, 334 (2003).
23. Y. Gilad, O. Man, S. Pääbo, D. Lancet, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3324 (2003).
24. Y. Gilad, C. D. Bustamante, D. Lancet, S. Paabo, *Am. J. Hum. Genet.* **73**, 489 (2003).
25. H. R. Freund, M. Hanani, *Nutrition* **18**, 287 (2002).
26. S. Gadkar et al., *Biol. Reprod.* **67**, 1327 (2002).
27. W. Enard et al., *Nature* **418**, 869 (2002).
28. S. Naz et al., *J. Med. Genet.* **40**, 360 (2003).
29. The data in this paper were obtained from more than 18 million sequencing reads obtained from the Celera Genomics sequencing center in Rockville, MD. We thank J. Duff, C. Gire, M. A. Rydland, C. Forbes, and B. Small for development and maintenance of software systems, laboratory information management systems, and analysis programs. S. Hannenhalli and S. Levy provided particularly helpful discussions. C. Aquadro, B. Lazzaro, K. Montooth, T. Schlenke, and P. Wittkopp provided helpful comments on the manuscript.

Supporting Online Material
www.sciencemag.org/cgi/content/full/302/5652/1960/DC1
Materials and Methods
Tables S1 to S3
Databases S1 and S2

7 July 2003; accepted 24 October 2003

The Proteasome of *Mycobacterium tuberculosis* Is Required for Resistance to Nitric Oxide

K. Heran Darwin,[1] Sabine Ehrt,[1] José-Carlos Gutierrez-Ramos,[2] Nadine Weich,[2] Carl F. Nathan[1,3]*

The production of nitric oxide and other reactive nitrogen intermediates (RNI) by macrophages helps to control infection by *Mycobacterium tuberculosis* (*Mtb*). However, the protection is imperfect and infection persists. To identify genes that *Mtb* requires to resist RNI, we screened 10,100 *Mtb* transposon mutants for hypersusceptibility to acidified nitrite. We found 12 mutants with insertions in seven genes representing six pathways, including the repair of DNA (*uvrB*) and the synthesis of a flavin cofactor (*fbiC*). Five mutants had insertions in proteasome-associated genes. An *Mtb* mutant deficient in a presumptive proteasomal adenosine triphosphatase was attenuated in mice, and exposure to proteasomal protease inhibitors markedly sensitized wild-type *Mtb* to RNI. Thus, the mycobacterial proteasome serves as a defense against oxidative or nitrosative stress.

Mtb persistently infects about two billion people. The identification of pathways used by the microbe to resist elimination by the host immune response may suggest new targets for prevention or treatment of tuberculosis. During latent infection, the primary residence of *Mtb* is the macrophage. The antimicrobial arsenal of the activated macrophage includes inducible nitric oxide synthase (iNOS or NOS2) (*1*). At the acidic pH ($\leq$5.5) prevalent in the phagosome of activated macrophages (*2*), nitrite, a major oxidation product of NO, is partially protonated to nitrous acid, which dismutates to form NO and another radical, $\cdot NO_2$ (*3*).

[1]Department of Microbiology and Immunology, Weill Medical College of Cornell University, New York, NY 10021, USA. [2]Millennium Pharmaceuticals, 75 Sidney Street, Cambridge, MA 02139, USA. [3]Programs in Immunology and Molecular Biology, Weill Graduate School of Medical Sciences of Cornell University, New York, NY 10021, USA.

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

Thus, mildly acidified nitrite is a physiologic antimicrobial system. RNI may inflict not only nitrosative but also oxidative injury, such as when NO combines with superoxide from bacterial metabolism to generate peroxynitrite (*4*). Reagent NO kills *Mtb* with a molar potency exceeding that of most antituberculosis drugs (*5*, *6*). In humans and mice with tuberculosis, macrophages in infected tissues and airways express enzymatically active iNOS (*7–9*), and mice lacking iNOS cannot control *Mtb* infection (*10*). Despite the protective effects of RNI, a small number of viable mycobacteria usually persist for the lifetime of the infected host (*11*) and sometimes resume growth.

To identify *Mtb* genes required for resistance against RNI, we screened 10,100 transposon mutants individually for increased sensitivity to nitrite at pH 5.5 [supporting online material (SOM text)]. Twelve mutants were hypersensitive. To quantify their phenotype, bacteria were exposed to pH 5.5 with or without 3 mM

1963