

**The following resources related to this article are available online at
www.sciencemag.org (this information is current as of November 5, 2009):**

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/326/5954/865>

Supporting Online Material can be found at:

<http://www.sciencemag.org/cgi/content/full/326/5954/865/DC1>

This article **cites 17 articles**, 6 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/326/5954/865#otherarticles>

This article appears in the following **subject collections**:

Genetics

<http://www.sciencemag.org/cgi/collection/genetics>

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:

<http://www.sciencemag.org/about/permissions.dtl>

processes in the oceans. We posit that future studies will elucidate the relative importance of biotic (viral infection as in this study) and abiotic [phosphorus limitation (17)] stress conditions, in shaping community structure of marine microbes, through the detection of different classes of stress-specific lipids.

Although the origin of PCD in unicellular organisms is still unclear, its functional conservation among phylogenetically diverse phytoplankton lineages suggests key evolutionary and ecological drivers in aquatic environments (29). The retention and expression of a nearly complete, virus-based sphingolipid biosynthetic pathway, along with its requirement for viral replication and regulation of the host cell fate, now underscores the pivotal role of a chemical-based, coevolutionary “arms race” in mediating host-virus interactions.

References and Notes

- M. J. Behrenfeld *et al.*, *Science* **291**, 2594 (2001).
- G. Bratbak, J. K. Egge, M. Haldal, *Mar. Ecol. Prog. Ser.* **93**, 39 (1993).
- D. C. Schroeder, J. Oke, G. Malin, W. H. Wilson, *Arch. Virol.* **147**, 1685 (2002).
- J. L. Van Etten, M. V. Graves, D. G. Müller, W. Boland, N. Delaroque, *Arch. Virol.* **147**, 1479 (2002).
- J. A. Fuhrman, *Nature* **399**, 541 (1999).
- C. A. Suttle, *Nat. Rev. Microbiol.* **5**, 801 (2007).
- D. Lindell, J. D. Jaffe, Z. I. Johnson, G. M. Church, S. W. Chisholm, *Nature* **438**, 86 (2005).
- K. D. Bidle, L. Haramaty, J. B. E. Ramos, P. Falkowski, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 6049 (2007).
- E. huxleyi* CCMP1516 Main Genome Assembly, <http://genome.jgi-psf.org/Emihu1/Emihu1.home.html>.
- W. H. Wilson *et al.*, *Science* **309**, 1090 (2005).
- A. Monier *et al.*, *Genome Biol.* **10**.1101/gr.091686.109 (2009).
- K. Hanada, *Biochim. Biophys. Acta* **1632**, 16 (2003).
- Y. A. Hannun, L. M. Obeid, *Trends Biochem. Sci.* **20**, 73 (1995).
- H. Liang *et al.*, *Genes Dev.* **17**, 2636 (2003).
- M. J. Allen *et al.*, *J. Virol.* **80**, 7699 (2006).
- G. Han *et al.*, *J. Biol. Chem.* **281**, 39935 (2006).
- B. A. S. Van Mooy *et al.*, *Nature* **458**, 69 (2009).
- Materials and methods are available as supporting material on Science Online.
- D. Lynch, T. Dunn, *New Phytol.* **161**, 677 (2004).
- B. Brügger *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 2641 (2006).
- M. J. Allen, J. A. Howard, K. S. Lilley, W. H. Wilson, *Proteome Sci.* **6**, 11 (2008).
- H. Sakamoto *et al.*, *Nat. Chem. Biol.* **1**, 333 (2005).
- A. Vardi *et al.*, *Curr. Biol.* **9**, 1061 (1999).
- M. Frada, I. Probert, M. J. Allen, W. H. Wilson, C. de Vargas, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 15944 (2008).
- M. Turkoglu, *J. Mar. Biol. Assoc. U. K.* **88**, 433 (2008).
- A. Vardi *et al.*, *Curr. Biol.* **18**, 895 (2008).
- A. Vardi *et al.*, *PLoS Biol.* **4**, e60 (2006).
- A. Pagarete, M. J. Allen, W. H. Wilson, S. A. Kimmance, C. de Vargas, *Environ. Microbiol.* **10**.1111/j.1462 (2009).
- K. D. Bidle, P. G. Falkowski, *Nat. Rev. Microbiol.* **2**, 643 (2004).
- The authors acknowledge M. Soule and E. Kujawinski and the funding sources of the Woods Hole Oceanographic Institution Fourier-Transform Mass Spectrometry Facility (NSF's Major Research Instrumentation Program OCE-0619608 and the Gordon and Betty Moore Foundation) for assistance with MS data acquisition. We thank D. Repeta for assistance and advice on pigment analysis, V. Starovoytov for TEM analysis, and C. Brown and M. Frada for helpful discussions and valuable feedback. This work was supported by NSF grant IOS-0717494 to K.D.B. and A.V., as well as NSF grant OCE-0646944 to B.V.M.

Supporting Online Material

www.sciencemag.org/cgi/content/full/326/5954/861/DC1

Materials and Methods

SOM Text

Figs. S1 to S4

References

5 June 2009; accepted 2 September 2009

10.1126/science.1177322

Genome Sequence, Comparative Analysis, and Population Genetics of the Domestic Horse

C. M. Wade,^{1,2,3*} E. Giulotto,⁴ S. Sigurdsson,^{1,5} M. Zoli,⁶ S. Gnerre,¹ F. Imsland,⁵ T. L. Lear,⁷ D. L. Adelson,⁸ E. Bailey,⁷ R. R. Bellone,⁹ H. Blöcker,¹⁰ O. Distl,¹¹ R. C. Edgar,¹² M. Garber,¹ T. Leeb,^{11,13} E. Mauceli,¹ J. N. MacLeod,⁷ M. C. T. Penedo,¹⁴ J. M. Raison,⁸ T. Sharpe,¹ J. Vogel,¹⁵ L. Andersson,⁵ D. F. Antczak,¹⁶ T. Biagi,¹ M. M. Binns,¹⁷ B. P. Chowdhary,⁸ S. J. Coleman,⁷ G. Della Valle,⁶ S. Fryc,¹ G. Guérin,¹⁹ T. Hasegawa,²⁰ E. W. Hill,²¹ J. Jurka,²² A. Kiialainen,²³ G. Lindgren,²⁴ J. Liu,²⁵ E. Magnani,⁴ J. R. Mickelson,²⁶ J. Murray,²⁷ S. G. Nergadze,⁴ R. Onofrio,¹ S. Pedroni,¹⁴ M. F. Piras,⁴ T. Raudsepp,⁸ M. Rocchi,²⁸ K. H. Røed,⁹ O. A. Ryder,³⁰ S. Searle,¹⁵ L. Skow,¹⁸ J. E. Swinburne,³¹ A. C. Syvänen,²³ T. Tozaki,³² S. J. Valberg,²⁶ M. Vaudin,³¹ J. R. White,¹ M. C. Zody,^{1,5} Broad Institute Genome Sequencing Platform,¹ Broad Institute Whole Genome Assembly Team,¹ E. S. Lander,^{1,33} K. Lindblad-Toh^{1,5*}

We report a high-quality draft sequence of the genome of the horse (*Equus caballus*). The genome is relatively repetitive but has little segmental duplication. Chromosomes appear to have undergone few historical rearrangements: 53% of equine chromosomes show conserved synteny to a single human chromosome. Equine chromosome 11 is shown to have an evolutionary new centromere devoid of centromeric satellite DNA, suggesting that centromeric function may arise before satellite repeat accumulation. Linkage disequilibrium, showing the influences of early domestication of large herds of female horses, is intermediate in length between dog and human, and there is long-range haplotype sharing among breeds.

As one of the earliest domesticated species, the horse, *Equus caballus*, has played an important role in human exploration of novel territories. Belonging to the order perissodactyla (i.e., odd-toed animals with hooves), the genus *Equus* radiated into 8 or 9 species around three million years ago (1). Members of the family equidae exhibit diverged karyotypes (2) and variable centromeric positioning (1). With over 90 hereditary conditions, which may serve as models for human disorders (3, 4) (such as infertility,

inflammatory diseases, and muscle disorders), the horse has much to offer as a model species.

DNA from a single mare of the Thoroughbred breed was sequenced to 6.8× coverage [supporting online material (SOM) text], resulting in a high-quality draft assembly (designated EquCab2.0) with a 112-kb N50 contig size (SOM) and a 46-Mb N50 scaffold size (tables S1 and S2), and >95% of the sequence anchored to the 64 (2N) equine chromosomes. The 2.5- to 2.7-Gb genome size is somewhat larger than the dog genome (2.5 Gb)

and smaller than the human and bovine genomes (2.9 Gb) (5–7). Segmental duplications (8) make up <1% of the equine genome, and most are intrachromosomal duplications such as are seen in many other mammalian genomes (SOM). Repetitive sequences, many equine-specific, make up 46% of the genome assembly (SOM). The predominant repeat classes include long interspersed nuclear elements, dominated by L1 and L2 types (tables S3 and S4) (19% of bases), and short interspersed nuclear elements, including the recent ERE1 and ERE2 and the ancestral main immunogenic regions (7% of bases). Comparison of horse and human chromosomes reveals strong conserved synteny between these species (fig. S1). Indeed, 17 horse chromosomes (53%) comprise material from a single human chromosome (in the dog, it is 29%).

One unexpected feature of the horse genome landscape was the identification of an evolutionary new centromere (ENC) on chromosome 11 (ECA11), captured in an immature state. Several ENCs have been generated in the genus *Equus* by centromere repositioning (a shift of centromeric position without chromosome rearrangement) (1). Mammalian centromeres are typically complex structures characterized by the presence of satellite tandem repeats. ENCs are believed to form initially by unknown mechanisms in repeat-free regions and then progressively acquire extended arrays of satellite tandem repeats that may contribute to functional stability (9). The centromere of ECA11 resides in a large region of conserved synteny in many mammals, where the horse is the only species with a centromere present, strongly suggesting that this centromere is evolutionarily new. The ECA11 centromere is the only horse centromere lacking any hybridization signal in fluorescence in situ hybridization

¹Broad Institute, 7 Cambridge Center, Cambridge, MA 02142, USA. ²Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA 02114, USA. ³Faculty of Veterinary Sciences, University of Sydney, Sydney, New South Wales 2006, Australia. ⁴Dipartimento di Genetica e Microbiologia, Università di Pavia, Via Ferrata 1, 27100 Pavia, Italy. ⁵Department of Medical Biochemistry and Microbiology, Uppsala University, Box 582, SE-751 24 Uppsala, Sweden. ⁶Dipartimento di Biologia, Università di Bologna, Via Selmi 3, 40126 Bologna, Italy. ⁷Maxwell H. Gluck Equine Research Center, Department of Veterinary Science, University of Kentucky, Lexington, KY 40546, USA. ⁸University of Adelaide, Adelaide, South Australia 5005, Australia. ⁹University of Tampa, 401 West Kennedy Boulevard, Box 3F, Tampa, FL, USA. ¹⁰Helmholtz Centre for Infection Research, Braunschweig, Germany. ¹¹Institute of Animal Breeding and Genetics, University of Veterinary Medicine Hannover, Bünteweg 17p, 30559 Hannover, Germany. ¹²45 Monterey Drive, Tiburon, CA 94920, USA. ¹³Institute of Genetics, Vetsuisse Faculty, University of Berne, Bremgartenstrasse 109a, 3001 Berne, Switzerland. ¹⁴Veterinary Genetics Laboratory, University of California, Davis, CA, USA. ¹⁵Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK. ¹⁶Baker Institute for Animal Health, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA. ¹⁷The Royal Veterinary College, Royal College Street, London NW1 0TU, UK. ¹⁸College of Veterinary Medicine, Texas A&M University, College Station, TX 77843, USA. ¹⁹INRA, UMR 1313, Génétique Animale et Biologie Intégrative, Biologie Intégrative et

Génétique Equine, Bâtiment 440.78350, Jouy-en-Josas, France. ²⁰Equine Research Institute, Japan Racing Association, 321-4 Tokami-cho, Utsunomiya, Tochigi 320-0856, Japan. ²¹Animal Genomics Laboratory, School of Agriculture, Food Science and Veterinary Medicine, University College Dublin, Belfield, Dublin 4, Ireland. ²²Genetic Information Research Institute, 1925 Landings Drive, Mountain View, CA 94043, USA. ²³Department of Medical Sciences, Uppsala University, 75185 Uppsala, Sweden. ²⁴Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Box 597, SE-751 24 Uppsala, Sweden. ²⁵Department of Computer Science, University of Kentucky, Lexington, KY 40506, USA. ²⁶College of Veterinary Medicine, University of Minnesota, St. Paul, MN 55108, USA. ²⁷Population Health and Reproduction, School of Veterinary Medicine, University of California, Davis CA, USA. ²⁸Department of Genetics and Microbiology, University of Bari, Via Amendola 165, 70126 Bari, Italy. ²⁹Department of Basic Sciences and Aquatic Medicine, Norwegian School of Veterinary Science, N-0033 Oslo, Norway. ³⁰San Diego Zoo's Institute for Conservation Research, Escondido, CA 92029, USA. ³¹Animal Health Trust, Suffolk, CB8 7UU, UK. ³²Department of Molecular Genetics, Laboratory of Racing Chemistry, 1731-2 Tsurutamachi Utsunomiya, Tochigi 320-0851, Japan. ³³Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA.

*To whom correspondence should be addressed. E-mail: c.wade@usyd.edu.au (C.M.W.); kersli@broadinstitute.org (K.L.-T.)

experiments probing with the two major horse satellite sequences (fig. S2A, table S8, and SOM text), as if it had not had enough time to acquire satellite DNA. We cytogenetically localized the primary constriction (fig. S2B), then precisely mapped, at the sequence level, the centromeric function using chromatin immunoprecipitation (ChIP)-on-chip experiments (fig. S5). In this region, we found only five sequence gaps [none >200 base pairs (bp)], no protein coding sequences, normal levels of noncoding conserved elements, and typical levels of interspersed repetitive sequences, but no satellite tandem repeated sequences (Fig. 1A). We also found no evidence of accumulation of L1 transposons (*I0*) or KERV-1 elements (*I1*), which were previously hypothesized to influence ENC formation. We propose that the ECA11 centromere was formed very recently during the evolution of the horse lineage, and, in spite of being functional and stable in all horses, has not yet acquired the marks typical of mammalian centromeres.

The equine gene set is similar to those of other eutherian mammals and has a predicted 20,322

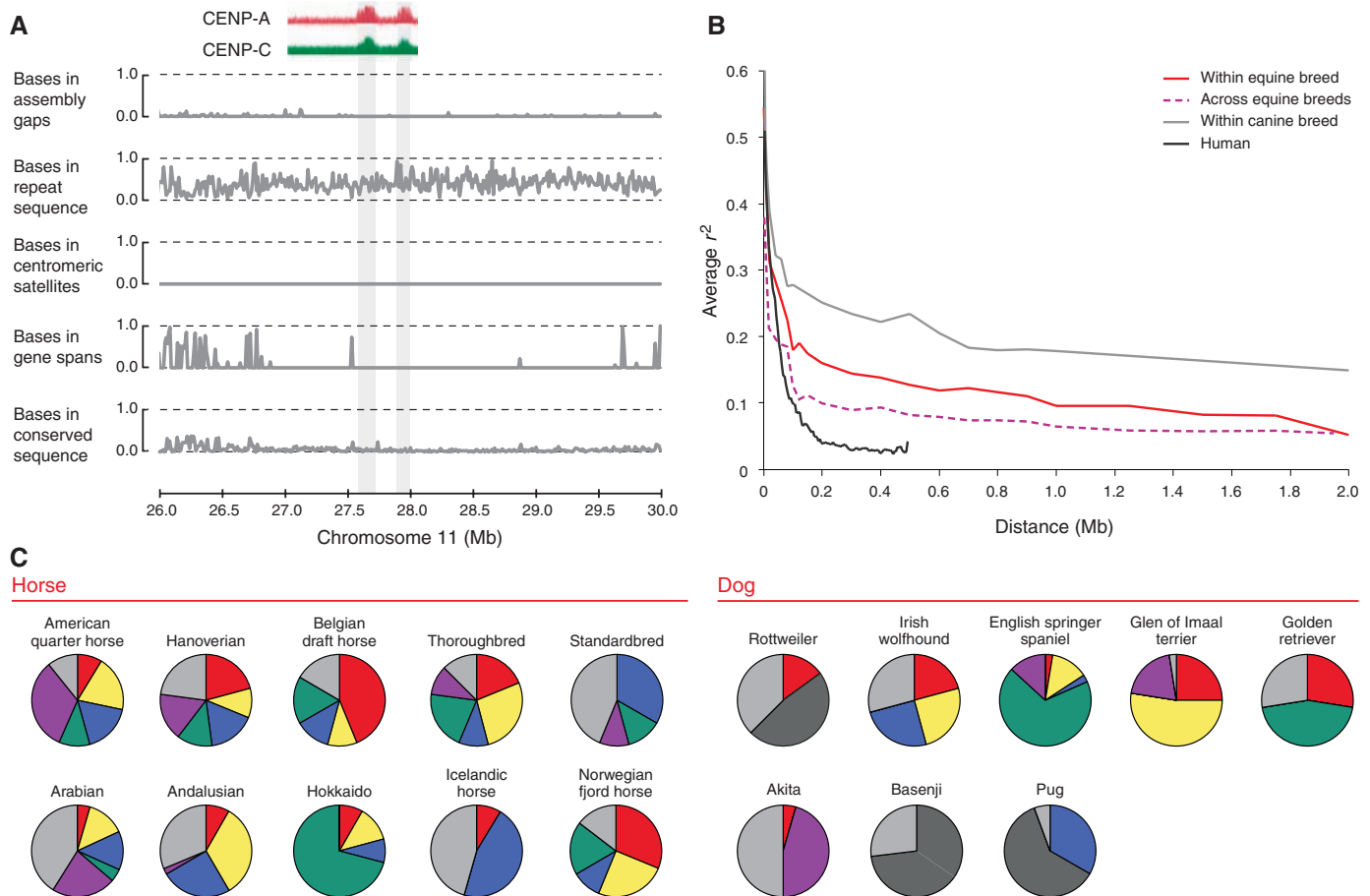


Fig. 1. Major findings of the genome analysis. **(A)** Analysis of the primary centromeric constriction of ECA11: 26,000,000 to 30,000,000 bases. ChIP-on-chip analysis with antibodies against centromeric proteins (CENP-A and CENP-C) shows two regions (136 and 99 kb) bound by kinetochore proteins. There are no uncaptured and few captured gaps, a normal fraction of bases in repeat sequences, no satellite tandem repeats, no protein-coding sequences present nearby, and normal levels of noncoding conserved elements (29

eutherians). **(B)** Horse LD is intermediate between human and dog. **(C)** Horses exhibit more long-range across-breed haplotype sharing than dogs. Haplotypes have the same color across breeds. Haplotypes in <5% of all individuals are light gray, and haplotypes in >5% of all individuals but a single breed are dark gray. Data show LD regions on ECA18 (first 100 kb) and dog chromosome 12 (first 100 kb), which are representative. Full data are in table S11.

protein-coding genes (ENSEMBL build 52.2b), of which 16,617, 17,106, and 17,106 have evidenced orthology to human, mouse, and dog, respectively. The remainder is composed of projected protein-coding genes, novel protein-coding genes, and pseudogenes. One-to-one orthologs with the human account for 15,027 horse gene predictions (SOM). Transcriptome analysis of eight equine samples confirmed the expression of 87% of the 18,039 nonoverlapping genes predicted by ENSEMBL and 88% of the 169,073 predicted exons. Gene family analysis shows paralogous expansion in horses as compared to both human and bovine (SOM) for several interesting families, such as keratin genes related to the condition of pachyonychia (nail bed thickening) in humans (12), perhaps affecting hoof formation; and opsin genes for photoreception, possibly advantageous for visual perception of predators (table S9).

The history of horse domestication, which has important implications for trait mapping strategies, differs in important ways from that of the domestic dog but is perhaps similar to that of the cow. Horses do not appear to have undergone a tight domestication bottleneck, and the presence of many matrilineal lines in domestic horse history has been postulated (13). Screening the horse Y chromosome revealed a limited number of patrilineal lines, consistent with a strong sex bias in the domestication process (14).

We first generated a single-nucleotide polymorphism (SNP) map of more than one million markers at an average density of one SNP per 2 kb by lightly sequencing seven horses from different breeds and by mining the assembly for SNPs (table S10).

We characterized the haplotype structure within and across breeds by genotyping 1,007 SNPs from 10 regions of the genome (SOM) in 12 populations, including 11 breed sets (each with 24 representatives), and 1 set of individual representatives from 24 other breeds and equids. 98% of SNPs were validated, with an average of 69% being polymorphic in alternate breeds (SOM). Like the bovine (15), within-breed linkage disequilibrium (LD) is moderate, dropping to twice the background levels (r^2) at 100 to 150 kb (Fig. 1B). The majority of breeds showed similar LD (SOM and fig. S7), and major haplotypes were frequently shared among diverse populations (Fig. 1C). Based on the length of LD in the horse, the number of haplotypes within haplotype blocks, and the polymorphism rate, power calculations suggest that ~100,000 SNPs are sufficient for association mapping within all breeds as well as across breeds (SOM and fig. S8).

Phylogenetic relationships among breeds were inconsistent across resequenced regions (fig. S9), which is most likely a consequence of the close relationships of horse breeds worldwide. We were unable to phylogenetically separate *E. przewalskii* from the domesticated horses, despite its different karyotype (2N = 66 versus 2N = 64 for the domesticated horse), which is in agreement with recent findings (16), whereas the donkey (*E. africanus*) is clearly a distinct taxon (fig. S9, table S14, and SOM text). This suggests that

either intermixing of *E. przewalskii* and *E. caballus* occurred after subspecies separation or that *E. przewalskii* is recently derived from *E. caballus*.

We demonstrated the utility of the equine genome sequence and a SNP map by applying these resources to mutation detection for the Leopard Complex (LP) spotting locus (SOM). LP (Appaloosa spotting) is defined by patterns of white occurring with or without pigmented spots (fig. S10). Homozygosity confers a phenotype associated with congenital stationary night blindness in the Appaloosa breed (17). Fine mapping of a 2-Mb region followed by regional sequence capture and sequencing (300 kb) found no indications of associated copy number variants or insertions or deletions but found 42 associated SNPs. Of these, 21 reside within an associated haplotype near a candidate gene *melastatin 1* (*TRPM1*), which is expressed in the eye and melanocytes (18). Two conserved SNPs may be good candidates for the causal mutation.

Our analysis of the first high-quality draft sequence of a horse (*E. caballus*) distinguishes *E. caballus* from earlier eutherian genomes by its large synteny with humans and the identification of a centromere repositioning event that may provide an effective model to study epigenetic factors responsible for centromere function. Our results demonstrate that horse population history has led to across-breed haplotype sharing, increasing the feasibility of across-breed mapping. Mapping projects in the horse are likely to accelerate in the coming years and will identify mutations in genes related to morphology, immunology, and metabolism, which may benefit human health.

References and Notes

1. L. Carbone *et al.*, *Genomics* **87**, 777 (2006).
2. V. A. Trifonov *et al.*, *Chromosome Res.* **16**, 89 (2008).

3. Online Mendelian Inheritance in Animals (University of Sydney, 2009), <http://omia.angis.org.au/>.
4. B. P. Chowdhary, N. Paria, T. Raudsepp, *Anim. Reprod. Sci.* **107**, 208 (2008).
5. K. Lindblad-Toh *et al.*, *Nature* **438**, 803 (2005).
6. E. S. Lander *et al.*, *Nature* **409**, 860 (2001).
7. Bovine Genome Sequencing and Analysis Consortium, *Science* **324**, 522 (2009).
8. T. S. Mikkelsen *et al.*, *Nature* **447**, 167 (2007).
9. M. Ventura *et al.*, *Science* **316**, 243 (2007).
10. A. C. Chueh, E. L. Northrop, K. H. Brettingham-Moore, K. H. Choo, L. H. Wong, *PLoS Genet.* **5**, e1000354 (2009).
11. D. M. Carone *et al.*, *Chromosoma* **118**, 113 (2009).
12. J. W. Wu *et al.*, *J. Eur. Acad. Dermatol. Venereol.* **23**, 174 (2009).
13. C. Vila *et al.*, *Science* **291**, 474 (2001).
14. G. Lindgren *et al.*, *Nat. Genet.* **36**, 335 (2004).
15. Bovine HapMap Consortium, *Science* **324**, 528 (2009).
16. A. N. Lau *et al.*, *Mol. Biol. Evol.* **26**, 199 (2009).
17. L. S. Sandmeyer, C. B. Breaux, S. Archer, B. H. Grahn, *Vet. Ophthalmol.* **10**, 368 (2007).
18. R. R. Bellone *et al.*, *Genetics* **179**, 1861 (2008).
19. We thank the Kentucky Horse Park and L. Chemnick for samples, L. Gaffney for graphics, and M. Daly for useful discussions. Supported by the National Human Genome Research Institute, the Dorothy Russell Havemeyer Foundation, the Volkswagen Foundation, the Morris Animal Foundation, the Centro di Eccellenza di Genomica in Campo Biomedico e Agrario, and the Programmi di Ricerca Scientifica di Rilevante Interesse Nazionale (PRIN-2006). K.L.T. is the recipient of a European Young Investigator award funded by the European Science Foundation. Sequences have GenBank accession numbers AAWR02000001 to AAWR02055316. 1,163,466 discovery SNPs have accession numbers rs6844103 to rs69617090 in dbSNP 130.

Supporting Online Material

www.sciencemag.org/cgi/content/full/326/5954/865/DC1
Materials and Methods

SOM Text

Figs. S1 to S11

Tables S1 to S14

References

23 June 2009; accepted 10 September 2009

10.1126/science.1178158

MafB/c-Maf Deficiency Enables Self-Renewal of Differentiated Functional Macrophages

Athar Aziz,^{1,2,3*} Erinn Soucie,^{1,2,3*} Sandrine Sarrazin,^{1,2,3} Michael H. Sieweke^{1,2,3,†}

In metazoan organisms, terminal differentiation is generally tightly linked to cell cycle exit, whereas the undifferentiated state of pluripotent stem cells is associated with unlimited self-renewal. Here, we report that combined deficiency for the transcription factors MafB and c-Maf enables extended expansion of mature monocytes and macrophages in culture without loss of differentiated phenotype and function. Upon transplantation, the expanded cells are nontumorigenic and contribute to functional macrophage populations in vivo. Small hairpin RNA inactivation shows that continuous proliferation of MafB/c-Maf deficient macrophages requires concomitant up-regulation of two pluripotent stem cell-inducing factors, KLF4 and c-Myc. Our results indicate that MafB/c-MafB deficiency renders self-renewal compatible with terminal differentiation. It thus appears possible to amplify functional differentiated cells without malignant transformation or stem cell intermediates.

The nonproliferative state of terminally differentiated cells is assured by robust, often redundant mechanisms (1, 2), and in rare exceptions where fully mature cells can re-enter the cycle, proliferation remains transient and/or

involves de-differentiation (3). It remains unknown what renders differentiated cells refractory to the same mitogen signals that stimulate the proliferation of their direct precursors. For example, the proliferative response of myelomonocytic