Traces of Human Migrations in *Helicobacter pylori* Populations

Daniel Falush,[1] Thierry Wirth,[1] Bodo Linz,[1]
Jonathan K. Pritchard,[2] Matthew Stephens,[3] Mark Kidd,[4]
Martin J. Blaser,[5] David Y. Graham,[6] Sylvie Vacher,[7]
Guillermo I. Perez-Perez,[5] Yoshio Yamaoka,[6] Francis Mégraud,[7]
Kristina Otto,[8] Ulrike Reichard,[1] Elena Katzowitsch,[8]
Xiaoyan Wang,[1] Mark Achtman,[1]* Sebastian Suerbaum[8]

Helicobacter pylori, a chronic gastric pathogen of human beings, can be divided into seven populations and subpopulations with distinct geographical distributions. These modern populations derive their gene pools from ancestral populations that arose in Africa, Central Asia, and East Asia. Subsequent spread can be attributed to human migratory fluxes such as the prehistoric colonization of Polynesia and the Americas, the neolithic introduction of farming to Europe, the Bantu expansion within Africa, and the slave trade.

Geographic subdivisions exist for a variety of human pathogens and commensals, including JC virus (*1*), *Mycobacterium tuberculosis* (*2*), *Haemophilus influenzae* (*3*), and *Helicobacter pylori* (*4–8*). *H. pylori,* a Gram-negative bacterium that colonizes the human gastric mucosa for decades and does not spread epidemically (*9*), has the potential to be informative about human migrations (*10*). Sequence diversity within *H. pylori* is greater than that of most other bacteria (*4*) and about 50-fold greater than that of human beings (*11*). Furthermore, frequent recombination between different *H. pylori* strains (*12–14*) implies that only partial linkage disequilibrium exists between polymorphic nucleotides within genes (*15*), which increases the information content for population genetic analysis. In this report, we use a population genetic tool that we have developed (*16*) on a large, global sample of *H. pylori* isolates to define modern populations and reconstruct their ancestral sources.

Previous data with 20 *H. pylori* isolates from East Asia, Europe, and Africa show that the sequences of fragments of seven housekeeping genes and one virulence-associated gene (*vacA*) differ according to the continent of origin (*4*). We sequenced the same fragments from 370 strains isolated from 27 geographical, ethnic, and/or linguistic human groupings (Table 1). Of the 3850 nucleotides sequenced for each isolate, 1418 were polymorphic and were used to define bacterial populations (*15*).

The program STRUCTURE (*16*, *17*) implements a Bayesian approach for deducing population structure from multilocus data by a variety of models, including the no-admixture model, which assumes that each individual has derived all of its ancestry from only one population. We used this model to identify four modern populations (*15*), designated hpAfrica1, hpAfrica2, hpEastAsia, and hpEurope on the basis of their current distributions (Table 1 and Fig. 1A). Further analyses split hpEastAsia into the hspAmerind, hspEAsia, and hspMaori subpopulations, and hpAfrica1 into hspWAfrica and hspSAfrica (Fig. 1B). These results confirm and extend previous data showing geographical subdivisions (*4*, *7*, *8*).

Almost all *H. pylori* strains isolated from various countries in East Asia were assigned to the hspEAsia subpopulation. The hspMaori subpopulation was isolated exclusively from Maoris and other Polynesians in New Zealand, whereas the hspAmerind strains were isolated from Inuits and from Amerinds in North and South America.

The hspSAfrica and hpAfrica2 populations were found only in South Africa, where they made up a majority of the strains isolated. The hspWAfrica strains were found at
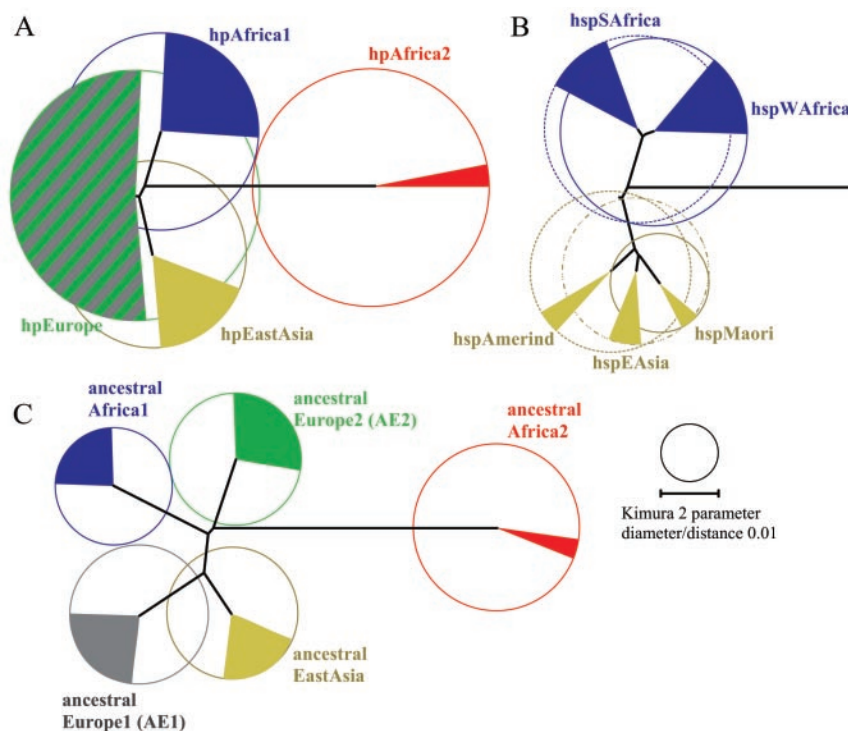

Fig. 1. Relationships between modern populations (**A**), modern subpopulations (**B**), and ancestral populations (**C**) of *H. pylori*. The black lines show neighbor-joining population trees as measured by $\hat{\delta}$, the net nucleotide distance between populations (*15*). The circle diameters indicate their genetic diversity, measured as the average genetic distance between random pairs of individuals. The larger circles in (A) versus (C) reflect the effects of admixture between ancestral populations. Filled arcs reflect the number of isolates (A and B) or nucleotides (C) in each population. Color coding is consistent in different parts of the figure, except for modern hpEurope, which is an admixture between the ancestral AE1 and AE2 populations. Scales are at lower right.

[1]Department of Molecular Biology, Max-Planck Institut für Infektionsbiologie, 10117 Berlin, Germany. [2]Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA. [3]Department of Statistics, University of Washington, Seattle, WA 98195–4322, USA. [4]Department of Surgery, Yale University School of Medicine, New Haven, CT 06520–8062, USA. [5]Department of Medicine, New York University School of Medicine, New York, NY 10016–9196, USA. [6]VA Medical Center, Houston, TX 77030, USA. [7]Université Victor Segalen Bordeaux 2, 33076 Bordeaux, France. [8]Institut für Hygiene und Mikrobiologie, Universität Würzburg, Josef-Schneider Straße 2, 97080 Würzburg, Germany.

*To whom correspondence should be addressed. E-mail: achtman@mpiib-berlin.mpg.de

low frequency in South Africa but at high frequency in West Africa and also in the Americas, particularly among African Americans in Louisiana and Tennessee. The hpEurope population contained almost all *H. pylori* from Europeans as well as from Turks, Israelis, Bangladeshis, Ladakhis, and Sudanese. These bacteria were also isolated from the Americas and Australia, and from whites, blacks, and Cape Coloured in South Africa, where they were predominantly associated with whites.

The current global sample is still incomplete, and additional isolates from large parts of Asia and Africa and from aboriginal groups around the world will be needed to determine whether additional populations exist. However, our definition of seven modern populations and subpopulations provides a solid basis for deducing the global patterns of spread of *H. pylori* with their human hosts.

Our attempts to define subpopulations by the same method among the 200 hpEurope isolates were not successful because of inconsistent clustering (*15*). We hypothesized that this inconsistency reflected the complex history of Europe, which was populated in several independent waves of migration (*18*) of unknown genetic composition (*19*). We have therefore developed an approach, the linkage model in STRUCTURE, that can reconstruct ancestral populations even after substantial genetic hybridization (*16*). This approach uses the mosaic ancestry of genomes within breeding species, assigning individual nucleotides to ancestral populations on the basis of their linkage to neighboring nucleotides.

Analysis of the global *H. pylori* sample with the linkage model defined five ancestral populations (*15*), which we named ancestral Africa1, Africa2, EastAsia, Europe1 (AE1), and Europe2 (AE2) (Fig. 1C). *H. pylori* strains within modern hpEurope are recombinants between AE1 and AE2 bacteria. No single isolate possesses more than 80% estimated ancestry from either of these populations (fig. S1); instead, each genome is a mosaic of multiple small chromosomal chunks (Fig. 2, F and G; fig. S2). In contrast, the other populations are more homogeneous.

Despite clear evidence for occasional import (Fig. 2, C and D), many isolates have derived 85 to 98% of their nucleotides from the ancestral population (Fig. 2, A and B; fig. S2).

Recombination between populations alters their genetic distances and blurs the branching order of trees (*20*). The ability to infer nucleotide pools in ancestral populations now allows more accurate estimates of ancestral relationships and evolutionary history. The ancestral population tree (Fig. 1C) suggests that Africa2 evolved before the other populations split and that AE1 and ancestral East Asia diverged from each other most recently. Additional detailed analyses (*15*) support these inferences.

Knowledge of ancestral gene pools also allows inferences about gene flow between populations. The high diversity in hpEurope (Fig. 1A) is due to fusion between AE1 and AE2. Within our sample, the proportion of AE1 nucleotides is highest in Finland, Estonia, and Ladakh (Fig. 3A). However, all European isolates also possess AE2 nucleotides, but only 3 of 17 isolates from Ladakh do so (fig. S1).

Table 1. Assignments of 370 *H. pylori* isolates from diverse continents to seven populations and subpopulations by STRUCTURE (no admixture). Language classifications were based on the Ethnologue online database (www. ethnologue.com). The two strains of *H. pylori* whose genomes have been sequenced belong to hpEurope (26695, isolated in the UK) and hpAfrica1 (J99, isolated from a white American in Tennessee).

Code	Source				No. of isolates assigned to						
	Region	Country	Ethnic	Linguistic	hpEurope	hpAfrica2	hpAfrica1		hpEastAsia		
							hspWAfrica	hspSAfrica	hspMaori	hspAmerind	hspEAsia
1	East Asia	Korea		Korean							11
2	East Asia	Singapore		Sino-Tibetan	2						9
3	India	Ladakh	North Indian	Sino-Tibetan	17						
4	India	Bangladeshi*		Indo-European	9						
5	Africa	South Africa	Black	Niger-Congo	2	7		17			
6	Africa	South Africa	White	Indo-European	10	3		9			
7	Africa	South Africa	Cape Coloured†	Indo-European	4		6	25			
8	Africa	Burkina Faso		Niger-Congo			12				
9	Africa	Senegal		Niger-Congo			5				
10	Africa	Sudan		Semitic	2						
11	N. America	USA	African American	Indo-European	3		10				
12	N. America	USA	White	Indo-European	2		3				
13	N. America	Canada/USA	Inuit	Eskimo-Aleut	4					8	
14	N. America	Canada	Athabaskan‡	Na-Dene						4	
15	S. America	Colombia	Mestizo	Indo-European	11		1				
16	S. America	Colombia	Huitoto‡	Witotoan	12					4	
17	S. America	Venezuela	Piaroa‡	Salivan	2		1			1	
18	Australasia	New Zealand	Polynesian§	Austronesian	3		2		23		
19	Australasia	Australia	White	Indo-European	3						
20	Europe	UK		Indo-European	19		1				1
21	Europe	Estonia		Uralic	11						
22	Europe	Finland		Uralic	9						
23	Europe	Germany		Indo-European	12						
24	Europe	Italy		Indo-European	6						
25	Europe	Spain‖		Indo-European	37						
26	Europe	Germany	Turkish	Altaic	10						
27	Near East	Israel		Semitic	5						
	Other¶				5		2				5
Total					200	10	43	51	23	17	26

*Isolates from Bangladeshis resident in the UK are listed here as being from India.   †Speak English but with elements of Khoisan.   ‡Collectively referred to as Amerinds in the text.   §Polynesian isolates were from 18 Maoris, 8 Samoans, and 2 Tongans in New Zealand.   ‖Includes two Basque speakers.   ¶ "Other" summarizes unique isolates from the following sources: hspEastAsia: Japan, China, Hong Kong, Thailand, and a Japanese from Peru; hpEurope: France, Lithuania, Holland, Thailand, and an Asian in Cape Town, South Africa; hspWAfrica: Gambia and Guatemala.

Similarly, AE2 nucleotides are most frequent in Spain, Sudan, and Israel, but the isolates from Sudan and Israel possess lower levels of AE1 than do European isolates. Thus, AE1 and AE2 probably reached Europe from different sources, AE1 primarily from the direction of central Asia and AE2 primarily from the Near East and North Africa.

Further reconstruction of the history of *H. pylori* is best done in the context of current knowledge about human migration. As with a human population tree (*21*), hpEurope derives from a short central branch between hpEastAsia and hpAfrica1 (Fig. 1A), hinting at a parallel history of intercontinental gene flow to Europe for humans and bacteria. Furthermore, the relative contribution of AE2 versus AE1 correlates significantly with the first principle component of European human variation (table S1), which is thought to reflect the entry of neolithic farmers into Europe from the Near East (*20*). The second principle component has been tentatively attributed to the migratory fluxes that brought Uralic languages to Europe, and indeed correlated weakly with AE1 versus AE2 ($r = 0.6$, $P = .13$) (table S1). It seems that neither AE1 nor AE2 was harbored by the original Paleolithic hunter-gatherers in Europe, because considerable AE1 or AE2 ancestry is found outside Europe, whereas paleolithic Y-chromosome haplotypes are largely restricted to Europe (*18*).

Known human migrations can also explain the spread of hpEastAsia and hpAfrica1 populations (Fig. 3B). Current models (*22, 23*) agree that speakers of Austronesian languages (Maoris and other Polynesians) arrived in New Zealand after sequential island-hopping that is likely to have resulted in repeated human popula-

tion bottlenecks. Indeed, consistent with population bottlenecks, the genetic diversity within the hspMaori sample is extremely low (Fig. 1), and the pattern of nucleotide polymorphisms within subpopulations implies that there has been strong drift in the evolution of the hspMaori population (*15*) (fig. S3). The isolation of hpEastAsia from Native Americans (*7, 8*) can be similarly explained by hpEastAsia's being carried during the colonization of the Americas that began at least 12,000 years ago. Unlike hspMaori, hspAmerind did not show signs of strong drift, implying that *H. pylori* accompanied the ancestors of modern Amerinds and Inuits in large numbers of individuals and/or was introduced on multiple occasions.

The high degree of similarity between hspWAfrica and hspSAfrica (Fig. 1B, fig. S3) is concordant with the low genetic distances (*20*) observed between speakers of the Niger-Congo family of languages and is consistent with hspSAfrica's being carried to Southern Africa during the rapid expansion of Bantu farmers from central West Africa (*24*). Given this scenario, one possibility to account for the extremely distinct hpAfrica2 population is that they colonized the Khoisan hunter-gatherer inhabitants of Southern Africa, who fall on one of the deepest branches of an African human population tree (*20*) and are very distinct from Bantu.

Modern migrations of slaves from West Africa to the Americas and of Europeans to South Africa, the Americas, and Australasia are probably responsible for the current existence of hspWAfrica and hpEurope in these and other locations (Table 1). According to this interpretation, the past few centuries since modern human migrations were too short for the distinctions between multiple bacterial populations to become blurred.

The assignments of particular human migrations to migrations of *H. pylori* populations can allow dating of the bacterial population tree by archaeological events. The five ancestral populations existed before the separation of hspAmerind from the other hpEastAsia populations (Fig. 1, B and C), which is estimated to have occurred at least 12,000 years ago. Accordingly, *H. pylori* has probably accompanied anatomically modern humans since their origins.
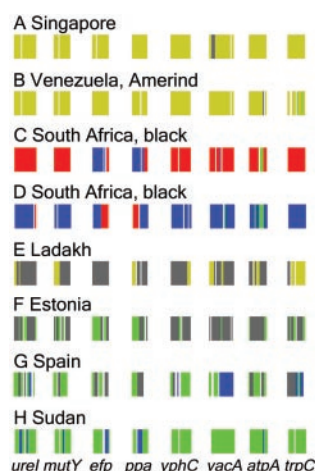
Fig. 2. Ancestral sources of individual nucleotides in eight selected isolates. The origin of each polymorphic nucleotide (colors as in Fig. 1C) is shown for each of the eight gene fragments. The geographical sources of each isolate are shown above each graph.
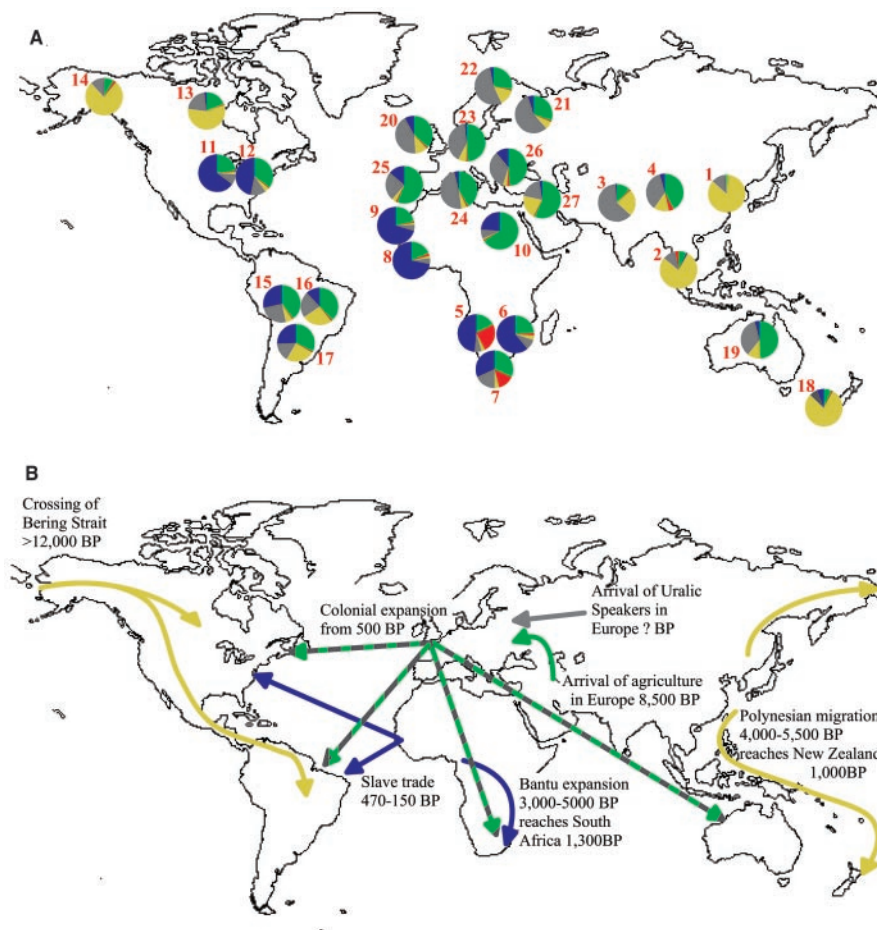
Fig. 3. Putative modern and ancient migrations of *H. pylori*. (**A**) Average proportion of ancestral nucleotides by source. Numbers correspond to the codes in Table 1 and colors are as in Fig. 1C. (**B**) Interpretation. Arrows indicate specific migrations of humans and *H. pylori* populations. BP, years before present.

The high sequence diversity in *H. pylori* allows the recognition of distinct populations after centuries of coexistence in individual geographic locations, as demonstrated in the Americas and South Africa. Even after thousands of years of contact in Europe between bacteria introduced by distinct waves of migration, residual short-range linkage disequilibrium has allowed us to identify ancestral chunks of chromosome. Thus, analysis of *H. pylori* from human populations could also help resolve details of human migrations.

Elucidation of the pattern of population subdivision is also of medical relevance (*25*). Geographically variable results regarding the association of putative virulence factors with disease (*26*) might well reflect differences in the local prevalence of the individual *H. pylori* populations. Similarly, the development of diagnostic tests, antibiotics, and vaccines needs to account for global diversity and will be aided by the availability of representative isolates.

References and Notes

1. H. T. Agostini, R. Yanagihara, V. Davis, C. F. Ryschkewitsch, G. L. Stoner, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 14542 (1997).
2. K. Kremer *et al.*, *J. Clin. Microbiol.* **37**, 2607 (1999).
3. J. M. Musser *et al.*, *Rev. Infect. Dis.* **12**, 75 (1990).
4. M. Achtman *et al.*, *Mol. Microbiol.* **32**, 459 (1999).
5. D. Kersulyte *et al.*, *J. Bacteriol.* **182**, 3210 (2000).
6. A. K. Mukhopadhyay *et al.*, *J. Bacteriol.* **182**, 3219 (2000).
7. Y. Yamaoka *et al.*, *FEBS Lett.* **517**, 180 (2002).
8. C. Ghose *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 15107 (2002).
9. R. A. Feldman, in Helicobacter pylori: *Molecular and Cellular Biology*, M. Achtman, S. Suerbaum, Eds. (Horizon Scientific, Wymondham, UK, 2001), pp. 29–51.
10. A. Covacci, J. L. Telford, G. Del Giudice, J. Parsonnet, R. Rappuoli, *Science* **284**, 1328 (1999).
11. W. H. Li, L. A. Sadler, *Genetics* **129**, 513 (1991).
12. D. Kersulyte, H. Chalkauskas, D. E. Berg, *Mol. Microbiol.* **31**, 31 (1999).
13. D. Falush *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 15056 (2001).
14. S. Suerbaum *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 12619 (1998).
15. Materials and methods, details of the STRUCTURE analysis, and analysis of the pattern of divergence between populations are available on *Science* Online.
16. D. Falush, M. Stephens, J. K. Pritchard, in preparation; available at www.mpiib-berlin.mpg.de/str2.pdf.
17. J. K. Pritchard, M. Stephens, P. Donnelly, *Genetics* **155**, 945 (2000).
18. O. Semino *et al.*, *Science* **290**, 1155 (2000).
19. L. Chikhi, R. A. Nichols, G. Barbujani, M. A. Beaumont, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11008 (2002).
20. L. L. Cavalli-Sforza, P. Menozzi, A. Piazza, *The History and Geography of Human Genes* (Princeton Univ. Press, Princeton, NJ, 1994).
21. E. S. Poloni, L. Excoffier, J. L. Mountain, A. Langaney, L. L. Cavalli-Sforza, *Ann. Hum. Genet.* **59**, 43 (1995).
22. J. M. Diamond, *Nature* **403**, 709 (2000).
23. S. Oppenheimer, M. Richards, *Sci. Prog.* **84**, 157 (2001).
24. C. Ehret, *Int. J. Afr. Hist. Stud.* **34**, 5 (2001).
25. J. F. Wilson *et al.*, *Nature Genet.* **29**, 265 (2001).
26. D. Y. Graham, Y. Yamaoka, *Helicobacter* **5** (suppl. 1), S3 (2000).
27. We thank all the colleagues listed in the supporting online text who have supplied bacterial isolates, DNA, and information and C. Josenhans for critical reading of the manuscript. Expert technical assistance was provided by S. Friedrich, A. Wirsing, and E. Bernard. Supported by grants from the Deutsche Forschungsgemeinschaft (Ac 39/10-3, SFB479/TP A5), the Bundesministerium für Bildung und Forschung Pathogenomics Network, and NIH (RO2GM63270).

Experience Strengthening Transmission by Driving AMPA Receptors into Synapses

Takuya Takahashi,[1] Karel Svoboda,[2] Roberto Malinow[1]*

The mechanisms underlying experience-dependent plasticity in the brain may depend on the AMPA subclass of glutamate receptors (AMPA-Rs). We examined the trafficking of AMPA-Rs into synapses in the developing rat barrel cortex. In vivo gene delivery was combined with in vitro recordings to show that experience drives recombinant GluR1, an AMPA-R subunit, into synapses formed between layer 4 and layer 2/3 neurons. Moreover, expression of the GluR1 cytoplasmic tail, a construct that inhibits synaptic delivery of endogenous AMPA-Rs during long-term potentiation, blocked experience-driven synaptic potentiation. In general, synaptic incorporation of AMPA-Rs in vivo conforms to rules identified in vitro and contributes to plasticity driven by natural stimuli in the mammalian brain.

The modifications that occur in the vertebrate brain as a consequence of experience are poorly understood, although changes at excitatory synapses may encode learning (*1–4*). Fast excitatory transmission in the central nervous system of vertebrates is largely mediated by the actions of glutamate on AMPA-Rs (*5*). The number of AMPA-Rs at a synapse can control the coupling strength between pre- and postsynaptic neurons, and thus serves as a key control site for neural function (*6–10*).

AMPA-Rs are multimeric complexes composed of subunits GluR1 to GluR4 (*11*). In vitro, subunit-specific rules govern the incorporation of AMPA-Rs into synapses (*12–14*). For example, plasticity-inducing protocols and *N*-methyl-D-aspartate receptor (NMDA-R) activation are required before AMPA-Rs with subunits containing long cytoplasmic tails (such as GluR1) can be driven into synapses, thus enhancing transmission. In contrast, AMPA-Rs comprising only subunits containing short cytoplasmic tails (GluR2 and GluR3) continuously replace synaptic receptors in a manner that maintains transmission (fig. S1C).

To examine the effects of experience on AMPA-R trafficking in vivo, we delivered AMPA-R subunits to a small number of neurons (~100) in rat barrel cortex by in vivo microinjection of an expressing Sindbis virus (Fig. 1A) (fig. S1, A and B) (*15–17*) at postnatal day (PND) 12, an age characterized by rapid experience-dependent development of barrel cortex circuitry (*18, 19*). Sensory experience was controlled by either preserving or trimming whiskers (*19–21*), and 2 days later, coronal brain slices were prepared (fig. S1B).

To determine whether AMPA-Rs are driven into synapses by experience, we examined brain slices from animals infected with a virus producing GluR1 (*22*). Infected neurons in layer 2/3 were identified by the green fluorescent protein (GFP) tag on GluR1 (fig. S1B). Whole-cell recordings were obtained from nearby infected and uninfected neurons, allowing direct comparison of synaptic responses evoked with stimulating electrodes placed in layer 4. AMPA-R–mediated responses were isolated pharmacologically (*22*). Transmission mediated by activation of AMPA-Rs on neurons expressing GluR1-GFP showed significantly increased rectification (ratio of response at -60 mV to response at $+40$ mV) compared to transmission onto nearby control neurons (Fig. 1). Neurons expressing only GFP in animals exposed to the same experimental protocol showed no change in rectification (Fig. 1D). The increased rectification is a signature of recombinant homomeric GluR1 receptors and indicates their delivery into synapses (*12, 23, 24*).

In vitro, GluR1-GFP cannot be driven into synapses unless long-term potentiation (LTP)

---

[1]Jones Laboratory, [2]Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA.

*To whom correspondence should be addressed. E-mail: malinow@cshl.org