

Selection for Human Immunodeficiency Virus Type 1 Envelope Glycosylation Variants with Shorter V1-V2 Loop Sequences Occurs during Transmission of Certain Genetic Subtypes and May Impact Viral RNA Levels

Bhavna Chohan,^{1,2†} Dorothy Lang,^{3†} Manish Sagar,^{1,4†} Bette Korber,³ Ludo Lavreys,⁵ Barbra Richardson,^{6,7} and Julie Overbaugh^{1,6*}

Divisions of Human Biology¹ and Public Health Sciences,⁶ Fred Hutchinson Cancer Research Center, Seattle, Washington 98109; Departments of Pathobiology,² Medicine,⁴ Epidemiology,⁵ and Biostatistics,⁷ University of Washington, Seattle, Washington 98195; and Los Alamos National Laboratory, Los Alamos, New Mexico 87544³

Received 5 November 2004/Accepted 23 December 2004

Designing an effective human immunodeficiency virus type 1 (HIV-1) vaccine will rely on understanding which variants, from among the myriad of circulating HIV-1 strains, are most commonly transmitted and determining whether such variants have an Achilles heel. Here we show that heterosexually acquired subtype A HIV-1 envelopes have signature sequences that include shorter V1-V2 loop sequences and fewer predicted N-linked glycosylation sites relative to the overall population of circulating variants. In contrast, recently transmitted subtype B variants did not, and this was true for cases where the major risk factor was homosexual contact, as well as for cases where it was heterosexual contact. This suggests that selection during HIV-1 transmission may vary depending on the infecting subtype. There was evidence from 23 subtype A-infected women for whom there was longitudinal data that those who were infected with viruses with fewer potential N-linked glycosylation sites in V1-V2 had lower viral set point levels. Thus, our study also suggests that the extent of glycosylation in the infecting virus could impact disease progression.

Studies in the simian immunodeficiency virus (SIV)/macaque model of AIDS have shown that the viruses that evolve over the course of disease are selected in part because they increase the number and/or vary the position of the carbohydrates to shield them from the host antibody response (1, 15). Subsequent studies indicate that a similar evolutionary process may occur in the human immunodeficiency virus type 1 (HIV-1) envelope during both simian/human immunodeficiency virus infection in macaques (2) and HIV-1 infection in humans (17). A recent study of eight individuals infected with subtype C HIV-1 suggested that there may be counterselection at transmission against variants with long hypervariable loops and relatively large numbers of potential N-linked glycosylation sites, which are predicted to have a more recessed receptor-binding domain (4). The transmitted subtype C HIV-1s had signature sequence characteristics, which included shorter envelope variable loop domains and fewer potential N-linked glycosylation sites (PNGS) (4). Because the study was limited to a small number of cases, all of one subtype, it is unclear whether transmission of viruses with these characteristics is typical, a point that is of importance for designing globally effective vaccines and other interventions to block transmitted viruses.

We examined HIV-1 sequences within a median of 70 days

(interquartile range [IQR], 49 to 161) post negative serology (PNS) from 27 women and eight men from Kenya who acquired subtype A HIV-1 through heterosexual transmission (8, 9, 13; unpublished data). The days PNS was defined as the time from the last HIV-1-negative serological test to when the sample used to obtain sequences was taken. The sequences were compared to subtype A sequences in the Los Alamos database to determine if they differed in V1-V2 length or number of PNGS. The Los Alamos database includes sequences from subjects at all stages of infection; the sequences known to be from within 1 year of infection were excluded from the data set used for this analysis. The Wilcoxon rank-sum test was used for these comparisons. The subtype A sequences from early in infection ($n = 35$) had significantly shorter V1-V2 loop sequences ($P = 0.008$) and fewer PNGS ($P = 0.017$) than subtype A sequences from the database ($n = 51$; Fig. 1A). A comparison that included two sequences from subjects with a heterogeneous virus population early in infection (8, 9, 13) was also performed to further ensure that random sampling of sequences from subjects with a heterogeneous virus did not influence the outcome. This analysis yielded nearly identical results as the analysis using a single, randomly selected early sequence; there was a significant difference in V1-V2 length ($P = 0.008$) and the number of PNGS ($P = 0.012$) when the data set of early V1-V2 sequences ($n = 49$) that included two sequences from subjects with a heterogeneous virus was compared to the Los Alamos database sequences (data not shown). Finally, a comparison of V1-V2 sequences that was restricted to only the 25 Kenyan A subtype sequences from the database was performed to ensure that the observed effect with the

* Corresponding author. Mailing address: Division of Human Biology, Fred Hutchinson Cancer Center, 1100 Fairview Ave. N., C3-168 Seattle, WA 98109-1024. Phone: (206) 667-3524. Fax: (206) 667-1535. E-mail: joverbau@fhcrc.org.

† B.C., D.L., and M.S. contributed equally to this work.

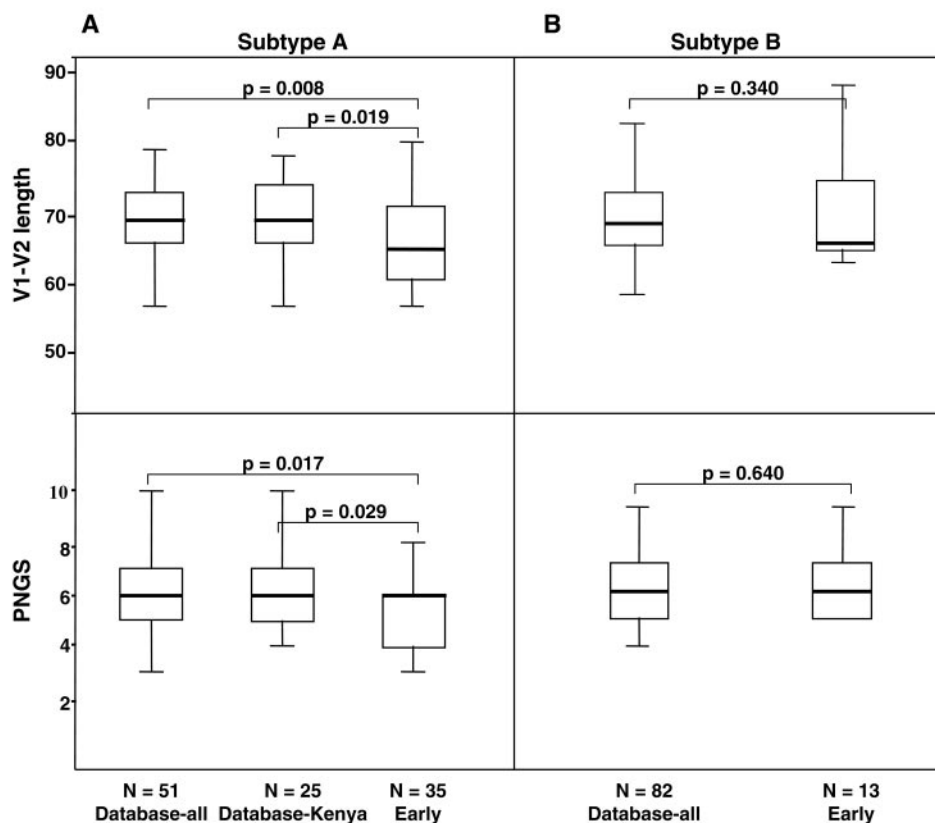


FIG. 1. Box plots of V1-V2 length and PNGS for early sequences versus database sequences. Panel A shows the results of subtype A sequence comparisons, and panel B shows the results of subtype B sequence comparisons. The boxes indicated the IQR of the data, and the line within the box represents the median value for each group. The number of sequences in each group is indicated at the bottom. Early sequences are from subjects within 1 year PNS, which was defined as the time from the last HIV-1-negative serological test to when the sample used to obtain sequence was taken. This interval was estimated in the same manner for both subtype A- and subtype B-infected subjects. Thirty-five and 13 subjects contributed to the early sequence subtype A and B data sets, respectively (8, 9, 13, 16). Database-all refers to sequences from the Los Alamos HIV-1 database and includes one randomly selected sequence per subject; sequences known to be within 1 year of infection were excluded from this data set. Database-Kenya refers to sequences from the Los Alamos database that were isolated from Kenyan subjects during chronic infection. The *P* values for the comparisons between early and database sequences are shown (Wilcoxon rank-sum test; one-sided test). Sequence analyses were confined to the V1-V2 region.

larger subtype A data set was not due to a regional lineage circulating in Kenya. The early subtype A Kenyan V1-V2 sequences were found to be significantly shorter and have fewer PNGS than the database subtype A Kenyan sequences ($P = 0.019$ and $P = 0.029$, respectively; Fig. 1A). These findings suggest that transmitted subtype A HIV-1s, like subtype C HIVs, may have condensed V1-V2 loops with fewer glycosylation sites.

We also examined viral sequences from 13 individuals (10 women and three men) infected with subtype B HIV-1 a median of 142 days PNS (IQR, 114 to 234) (16). As with the subtype A analyses, these sequences were compared to the Los Alamos database V1-V2 sequences. Sequences reported to be from within 1 year postinfection were excluded from the Los Alamos sequences, leaving 82 remaining V1-V2 sequences from different individuals. Interestingly, the early sequences from the 13 individuals did not show evidence of having shorter V1-V2 sequences than the subtype B HIV-1s in the database ($n = 82$), nor did they differ in the number of PNGS ($P = 0.340$ and $P = 0.640$, respectively; Fig. 1B). There was also no significant difference in V1-V2 length or PNGS when all the early

subtype B sequences from the Los Alamos database ($n = 71$ including the 13 discussed above) were compared to the remaining 82 subtype B sequences ($P = 0.132$ and $P = 0.323$, respectively). Interestingly, the early A sequences ($n = 35$) were found to be significantly shorter and have fewer PNGS than these early B sequences ($n = 71$; $P = 0.001$ and $P = 0.05$, respectively). Finally, 27 of the 71 early database subtype B infected individuals were known to have been infected via heterosexual transmission; many of these were samples from a transmission study conducted in Trinidad and Tobago (3). There was also no difference observed in terms of V1-V2 length or number of PNGS in these B clade heterosexual early samples relative to the 82 database samples ($n = 27$, $P = 0.708$ or $P = 0.6215$, respectively). One potential caveat of this analysis is that detailed information was not available on when these sequences, described as from early after seroconversion, were collected in relation to a negative HIV-1 test; therefore, we could not exclude sequences from later than 1 year PNS, as we did for the other analyses of early sequences.

The interval PNS was estimated in the same manner for the 13 subtype B-infected subjects as it was for the 35 subtype

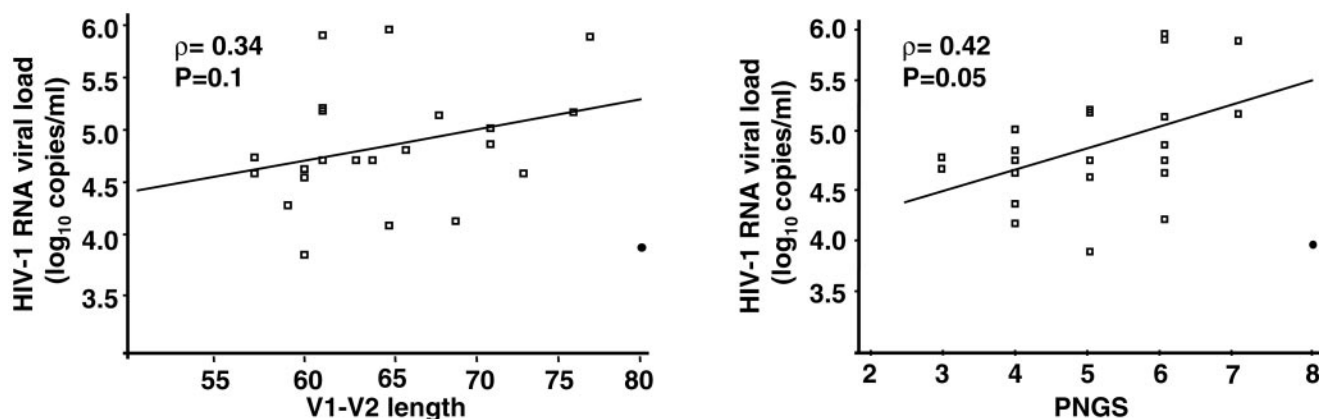


FIG. 2. Comparison of plasma viral RNA levels at set point to V1-V2 length and PNGS. This comparison was limited to 23 women for whom we had longitudinal viral RNA levels that included a plasma RNA measurement at 4 to 24 months PNS (7). In each of the 23 cases, the viral load measurement was also within a 4- to 24-month period postinfection as defined previously, using both RNA and serology data to more precisely define the time of infection (7). Each point represents the data for one individual. The line representing the linear regression model and the Spearman's correlation coefficient were generated from a data set that excludes one influential point. That influential point is shown with a closed circle.

A-infected subjects. The early subtype B sequences analyzed here were from a similar time after infection as the subtype C sequences in the study of Derdeyn et al., which were taken at ~3 to 4 months after a negative serological test (4), but slightly later in infection than the early subtype A sequences discussed above. However, we found that in the 35 subtype A-infected subjects, there was no association between V1-V2 sequence length or PNGS and the time interval from infection ($P = 0.468$ and $P = 0.682$, respectively; Spearman's rank correlation), which ranged from 32 days to 335 days PNS. The subtype B-infected cases were all within this time frame (49 to 250 days PNS), indicating that significant de novo length variation was not likely to be a confounding factor.

In SIVMne infection of macaques, an increased number of PNGS in V1-V2 of the infecting virus is associated with a blunted neutralizing antibody response, higher set point HIV-1 RNA levels, and a more rapid disease progression (6). The converse has also been demonstrated in the SIVmac model—namely, elimination of glycosylation sites in a highly pathogenic strain of SIV reduces virulence (14). Together, these data suggest that the number and/or position of the PNGS of the infecting virus influence disease outcome, presumably because the sugars mask viral epitopes. To determine whether there was a similar effect for V1-V2 glycosylation and/or length on HIV-1 pathogenesis, we examined these two parameters in relation to the first HIV-1 plasma RNA measurement available at 4 to 24 months PNS in women for whom we had longitudinal HIV-1 RNA data (7). We found that there was no association between viral set point levels and V1-V2 length ($P = 0.2$) or PNGS ($P = 0.4$). In this analysis, one woman's data was deemed an influential point through visual examination of a scatter plot and by several formal tests (Cook's distance, Mahalanobis' distance, Studentized deleted residual, and standardized difference beta). Examination of the data from this subject showed that her first viral RNA measurement (at ~10 months PNS) was 1.3 to 1.9 \log_{10} units lower than at visits 20 to 45 months later (data not shown). This was unexpected, because the average increase in viral RNA level in this cohort

is 0.0094 \log_{10} copies/ml/month (95% confidence interval, 0.0057 to 0.0130) (7). When this influential point was excluded from the analyses, there was a significant association between the number of PNGS and viral set point level (Spearman correlation coefficient, 0.42; $P = 0.05$) and a trend for an association between V1-V2 length and viral set point level (Spearman correlation coefficient, 0.34; $P = 0.1$) (Fig. 2). Both nonparametric and parametric estimates of the correlation coefficients were nearly identical (data not shown). Because HIV-1 set point levels are a strong predictor of AIDS progression (10–12), these data suggest that individuals with viruses that are more heavily glycosylated in V1-V2 early in infection may have a more rapid disease course.

The findings presented here, together with the study of eight subtype C transmission cases (4), indicate that there is selection for viruses with fewer N-linked glycosylation sites and condensed V1-V2 loop sequences during heterosexual transmission of subtypes A and C. These findings are of particular significance given that heterosexual transmission represents the major route of HIV-1 spread and that subtypes A and C are the main circulating subtypes of HIV-1 in the world. The apparent selection for viruses with less glycosylated, condensed V1-V2 loops may reflect their fitness advantage during productive infection of early target cells in the new host or during the rapid expansion phase during acute infection prior to the virus-specific antibody response. This phenotype may be associated fortuitously with neutralization sensitivity (3), but it seems unlikely that neutralization sensitivity itself provides a selective advantage to the virus during transmission. Indeed, in some settings, such as in mother-to-child transmission where HIV-1-specific passive antibody may be present, there is likely to be immune pressure against infection by neutralization-sensitive viruses.

Our findings on subtype A-infected women suggest that the characteristics of infecting viruses, including their glycosylation profile, may influence viral replication levels. These results closely parallel findings in the SIV/macaque model, where the extent of glycosylation of the infecting virus is a key determi-

nant of set point viral RNA levels (6). Thus, the findings here reinforce the notion that the properties of the infecting strain are important determinants for HIV-1 disease outcome. However, because these results are with a small cohort, it will be important to examine this association more fully in a larger natural history study, ideally including other clinical measures of disease in addition to set point viral RNA levels.

Interestingly, there was no evidence for selection of viruses with fewer PNGS or smaller V1-V2 loop sequences in individuals recently infected with subtype B HIV-1. Consistent with our findings, Frost et al. have found no distinction in the sequence characteristics or neutralization sensitivity of viruses found in eight subtype B male-to-male transmission pairs near the time of HIV-1 infection (5). It is perhaps noteworthy that the subtype B infection cases in our study and the study of Frost et al. were primarily the result of homosexual transmission or injection drug use (5), whereas the subtype A infections examined here, and the subtype C cases examined previously (4), were all cases of heterosexual transmission. This raised the possibility that selection for viruses that bear envelope proteins with condensed loops and fewer PNGS may take place in some modes of transmission but not in others. Alternatively, the differences in the bottleneck to V1-V2 length between cohorts and/or subtypes observed here may indicate that there are different structural solutions to gaining this selective advantage that depend on the backbone of the virus. In support of the latter model, analysis of database sequences from subtype B early infection cases from individuals who indicated heterosexual transmission as their major risk factor showed no reduction in V1-V2 length or glycosylation versus other subtype B sequences. Collectively, these data suggest that the distinction between selection at or near transmission of subtypes A and C versus subtype B is a lineage-specific, not transmission mode-driven, effect. Regardless, these results suggest that differences in the characteristic of viruses that are transmitted in these different settings could impact the efficacy of vaccines, especially those that are designed to elicit neutralizing antibody responses.

Nucleotide sequence accession numbers. Some of the early sequences we studied had already been assigned GenBank accession numbers (AF407175, AF407157, AF407160, AF048355, AF048583, AF407151, AF407155, AF407148, AF004888, AF004894, AY525511, AY525509, AY525504, AY525446, AY525454, AY525459, AY525462, AY525467, AY525469, AY525473, AY525475, AY525479, AY525486, AY525491, AY525500, AY525502). The remaining early sequences included in this publication were also submitted to GenBank, and the accession numbers are AY850734, AY850735, and AY849808 to AY849827.

This work was supported by NIH grant AI38518.

We gratefully acknowledge the contributions of Jared Baeten, Kishor Mandalia, and other members of our Mombasa/Seattle research team and HIVNET, who provided samples for the study of subtype B infection.

REFERENCES

1. Chackerian, B., L. Rudensey, and J. Overbaugh. 1997. Specific N-linked and O-linked glycosylation additions in the envelope V1 domain of simian immunodeficiency virus variants that evolve in the host alter neutralizing antibody recognition. *J. Virol.* **71**:7719–7727.
2. Cheng-Mayer, C., A. Brown, J. Harouse, P. A. Luciw, and A. J. Mayer. 1999. Selection for neutralization resistance of the simian/human immunodeficiency virus SHIVSF33A variant in vivo by virtue of sequence changes in the extracellular envelope glycoprotein that modify N-linked glycosylation. *J. Virol.* **73**:5294–5300.
3. Cleghorn, F. R., N. Jack, J. K. Carr, J. Edwards, B. Mahabir, A. Sill, C. B. McDanal, S. M. Connolly, D. Goodman, R. Q. Bennetts, T. R. O'Brien, K. J. Weinhold, C. Bartholomew, W. A. Blattner, and M. L. Greenberg. 2000. A distinctive clade B HIV type 1 is heterosexually transmitted in Trinidad and Tobago. *Proc. Natl. Acad. Sci. USA* **97**:10532–10537.
4. Derdeyn, C. A., J. M. Decker, F. Bibollet-Ruche, J. L. Mokili, M. Muldoon, S. A. Denham, M. L. Heil, F. Kasolo, R. Musonda, B. H. Hahn, G. M. Shaw, B. T. Korber, S. Allen, and E. Hunter. 2004. Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. *Science* **303**:2019–2022.
5. Frost, S. D. W., Y. Liu, S. L. Kosakovsky Pond, C. Chappey, S. T. Wrin, C. J. Petropoulos, S. J. Little, and D. D. Richman. 2005. Characterization of human immunodeficiency virus type 1 (HIV-1) envelope variation and neutralizing antibody responses during transmission of HIV-1 subtype B. *J. Virol.* **79**:6523–6527.
6. Kimata, J. T., L. Kuller, D. B. Anderson, P. Dailey, and J. Overbaugh. 1999. Emerging cytopathic and antigenic simian immunodeficiency virus variants influence AIDS progression. *Nat. Med.* **5**:535–541.
7. Lavreys, L., J. M. Baeten, J. K. Kreiss, B. A. Richardson, B. H. Chohan, W. Hassan, D. D. Panteleeff, K. Mandaliya, J. O. Ndinya-Achola, and J. Overbaugh. 2004. Injectable contraceptive use and genital ulcer disease during the early phase of HIV-1 infection increase plasma virus load in women. *J. Infect. Dis.* **189**:303–311.
8. Long, E. M., H. L. Martin, Jr., J. K. Kreiss, S. M. Rainwater, L. Lavreys, D. J. Jackson, J. Rakwar, K. Mandaliya, and J. Overbaugh. 2000. Gender differences in HIV-1 diversity at time of infection. *Nat. Med.* **6**:71–75.
9. Long, E. M., S. M. J. Rainwater, L. Lavreys, K. Mandaliya, and J. Overbaugh. 2002. HIV type 1 variants transmitted to women in Kenya require the CCR5 coreceptor for entry, regardless of the genetic complexity of the infecting virus. *AIDS Res. Hum. Retrovir.* **18**:567–576.
10. Mellors, J. W., C. R. Rinaldo, Jr., P. Gupta, R. M. White, J. A. Todd, and L. A. Kingsley. 1996. Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science* **272**:1167–1170.
11. O'Brien, T. R., W. A. Blattner, D. Waters, E. Eyster, M. W. Hilgartner, A. R. Cohen, N. Luban, A. Hatzakis, L. M. Aledort, P. S. Rosenberg, W. J. Miley, B. L. Kroner, and J. J. Goedert. 1996. Serum HIV-1 RNA levels and time to development of AIDS in the Multicenter Hemophilia Cohort Study. *JAMA* **276**:105–110.
12. O'Brien, W. A., P. M. Hartigan, D. Martin, J. Esinhardt, A. Hill, S. Benoit, M. Rubin, M. S. Simberkoff, J. D. Hamilton, and the Veterans Affairs Cooperative Study Group on AIDS. 1996. Changes in plasma HIV-1 RNA and CD4⁺ lymphocyte counts and the risk of progression to AIDS. *N. Engl. J. Med.* **334**:426–431.
13. Poss, M., H. L. Martin, J. K. Kreiss, L. Granville, B. Chohan, P. Nyange, K. Mandaliya, and J. Overbaugh. 1995. Diversity in virus populations from genital secretions and peripheral blood from women recently infected with human immunodeficiency virus type 1. *J. Virol.* **69**:8118–8122.
14. Reitter, J. N., R. E. Means, and R. C. Desrosiers. 1998. A role for carbohydrates in immune evasion in AIDS. *Nat. Med.* **4**:679–684.
15. Rudensey, L. M., J. T. Kimata, E. M. Long, B. Chackerian, and J. Overbaugh. 1998. Changes in the extracellular envelope glycoprotein of variants that evolve during the course of simian immunodeficiency virus SIVMne infection affect neutralizing antibody recognition, syncytium formation, and macrophage tropism but not replication, cytopathicity, or CCR-5 coreceptor recognition. *J. Virol.* **72**:209–217.
16. Sagar, M., E. Kirkegard, M. E. Long, C. Celum, S. P. Buchbinder, E. S. Daar, and J. Overbaugh. 2004. Human immunodeficiency virus type 1 (HIV-1) diversity at the time of infection is not restricted to certain risk groups or specific HIV-1 subtypes. *J. Virol.* **78**:7279–7283.
17. Wei, X., J. M. Decker, S. Wang, H. Hui, J. C. Kappes, X. Wu, J. F. Salazar-Gonzalez, M. G. Salazar, J. M. Kilby, M. S. Saag, N. L. Komarova, M. A. Nowak, B. H. Hahn, P. D. Kwong, and G. M. Shaw. 2003. Antibody neutralization and escape by HIV-1. *Nature* **422**:307–312.