

Ghost roads and the destruction of Asia-Pacific tropical forests

<https://doi.org/10.1038/s41586-024-07303-5>

Received: 1 August 2022

Accepted: 13 March 2024

Published online: 10 April 2024

Open access

 Check for updates

Jayden E. Engert^{1,6✉}, Mason J. Campbell^{1,6}, Joshua E. Cinner^{2,4}, Yoko Ishida¹, Sean Sloan^{1,5}, Jatna Supriatna³, Mohammed Alamgir¹, Jaime Cislowski¹ & William F. Laurance^{1,5✉}

Roads are expanding at the fastest pace in human history. This is the case especially in biodiversity-rich tropical nations, where roads can result in forest loss and fragmentation, wildfires, illicit land invasions and negative societal effects^{1–5}. Many roads are being constructed illegally or informally and do not appear on any existing road map^{6–10}; the toll of such ‘ghost roads’ on ecosystems is poorly understood. Here we use around 7,000 h of effort by trained volunteers to map ghost roads across the tropical Asia-Pacific region, sampling 1.42 million plots, each 1 km² in area. Our intensive sampling revealed a total of 1.37 million km of roads in our plots—from 3.0 to 6.6 times more roads than were found in leading datasets of roads globally. Across our study area, road building almost always preceded local forest loss, and road density was by far the strongest correlate¹¹ of deforestation out of 38 potential biophysical and socioeconomic covariates. The relationship between road density and forest loss was nonlinear, with deforestation peaking soon after roads penetrate a landscape and then declining as roads multiply and remaining accessible forests largely disappear. Notably, after controlling for lower road density inside protected areas, we found that protected areas had only modest additional effects on preventing forest loss, implying that their most vital conservation function is limiting roads and road-related environmental disruption. Collectively, our findings suggest that burgeoning, poorly studied ghost roads are among the gravest of all direct threats to tropical forests.

By the middle of this century, Earth is expected to have some 25 million km of new paved roads relative to 2010—enough to encircle the planet more than 600 times¹. Roads serve a number of important societal functions, such as promoting trade and increasing access to natural resources and arable land^{7,8,12}. Without effective planning and law enforcement, however, roads can also unleash a Pandora’s box of environmental ills and societal challenges^{2,13–16}. Unfortunately, many new roads are being constructed informally or illegally, especially in lower-income nations where governance is often hindered by corruption and ineffective law enforcement^{7,15}. These ‘ghost roads’, invisible on official road maps, are one of the most vexing direct threats to tropical forests and their wild and human inhabitants^{6,7}.

We define ghost roads operationally as those missing from the two leading global-road datasets: the Global Roads Inventory Project¹⁷ (GRIP) and OpenStreetMap¹⁸ (OSM). Ghost roads include informally or illicitly constructed roads, bulldozed tracks in logged forests, roads in palm-oil plantations and other roads that are missing from existing road datasets for various reasons. Such roads can be either paved or unpaved, although most are unpaved. Ghost roads are being constructed by a range of people, including legal or illegal agriculturalists, miners, loggers, land grabbers, land speculators and drug traffickers, among others^{6–9}.

The accuracy and completeness of existing road maps vary greatly among nations and regions, and are typically poorest in developing nations with large forest estates^{19,20}. To assess the extent of ghost roads, we carried out an intensive sampling effort (1.42 million plots of 1 km² each) across a range of human-altered and native-forested regions of Borneo, Sumatra and New Guinea—three of the world’s largest continental islands. We manually mapped and digitized roads on each island using recent (circa 2019) high-resolution satellite imagery in Google Earth. Mapping was conducted by 210 trained volunteers or researchers whose individual mapping accuracy was quality-checked by one or more co-authors of this study, using test datasets (Supplementary Information and Supplementary Fig. 1). Each mapper was required to attain an accuracy of more than 90% on test datasets (including road omissions and commissions) before commencing road-mapping.

After generating high-accuracy road data, we (1) compared the extent of roads from our data directly with those from the two leading global-road datasets (GRIP¹⁷ and OSM¹⁸); (2) assessed how roads and other key socioeconomic and environmental variables influence forest loss; (3) gauged how protected areas affect the proliferation of roads and associated environmental disruption; and (4) used a temporal analysis to assess whether roads tend to precede, or follow, deforestation across our study area.

¹Centre for Tropical Environmental and Sustainability Science, and College of Science and Engineering, James Cook University, Cairns, Queensland, Australia. ²College of Arts, Society and Education, James Cook University, Townsville, Queensland, Australia. ³Research Center for Climate Change, and Department of Biology, University of Indonesia, Depok, Indonesia. ⁴Present address: Thriving Oceans Research Hub, School of Geosciences, University of Sydney, Camperdown, New South Wales, Australia. ⁵Present address: Department of Geography, Vancouver Island University, Nanaimo, British Columbia, Canada. ⁶These authors contributed equally: Jayden E. Engert, Mason J. Campbell. [✉]e-mail: engert_ecospacial@outlook.com; bill.laurance@jcu.edu.au

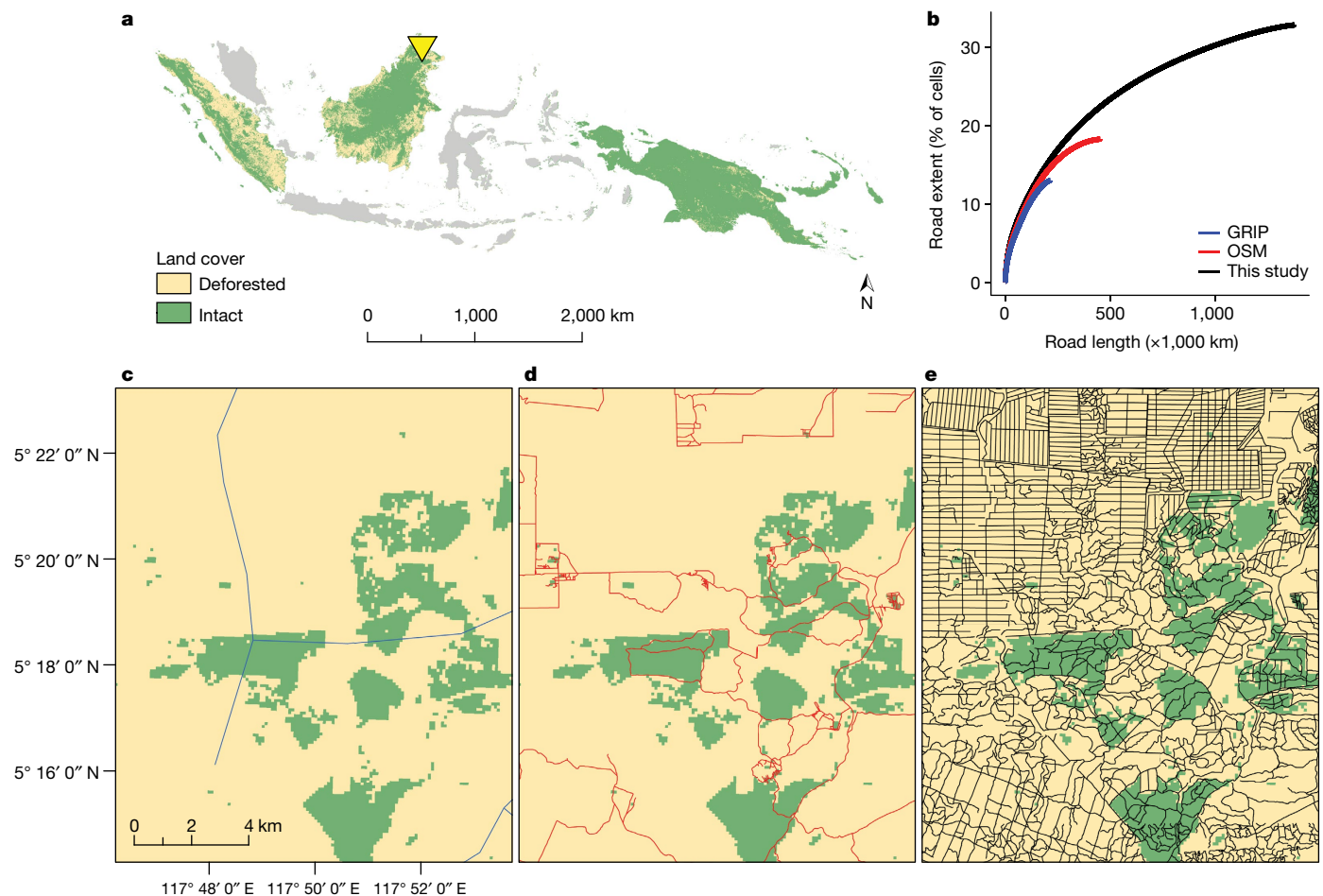


Fig. 1 | Road density in the tropical Asia-Pacific region is much higher than indicated by available global datasets. **a**, The study region, comprising part or all of Indonesia, Malaysia and Papua New Guinea (the yellow triangle shows the location of inset panels **c–e**). **b**, Cumulative plots comparing the total length of roads and proportion of land potentially affected by roads (road extent,

percentage of 1-km² cells containing roads) in this study versus data from OSM and GRIP. Sites are ordered from highest to lowest road length. **c–e**, Mapped roads in a landscape in Sabah, Malaysian Borneo, as shown by GRIP (**c**; blue lines, imagery circa 2018), OSM (**d**; red lines, circa 2020) and this study (**e**; black lines, circa 2019), respectively.

Road extent and density

We compared our road data with those from the two global-road databases, GRIP and OSM, using the same 1.42 million plots for all datasets. Road extent (the percentage of mapped 1-km² cells containing at least one road) was 13.2% using GRIP and 18.3% using OSM, but a much higher 32.9% when using our road data (Fig. 1b). In addition, the total length of mapped roads was 3.0–6.6 times greater when using our dataset (1.37 million km) than when using the GRIP (0.21 million km) and OSM (0.45 million km) datasets. Compared with GRIP and OSM, our data revealed that 35–45% of unmapped roads were in oil-palm or other plantations (23–33% in large plantations; 11–12% in small plantations), 31–39% were in intact forests and 17–28% were in non-plantation agriculture (see Supplementary Fig. 2 and Supplementary Information for land-use definitions). Unmapped roads were less prevalent in urban areas, degraded forest and other land-use types (Supplementary Fig. 2).

Our findings show that the extent and length of roads, at least in our study area, are severely underestimated in leading road databases and official government statistics (Fig. 1 and Supplementary Table 2). Moreover, these badly deficient road data partly underlie popular conservation metrics, such as the ‘human footprint’ index^{21,22} and ‘roadless’ or ‘wilderness’ areas^{5,23}, that are widely used in conservation research and management (see below).

Modelling forest loss

Next, we tested the relative importance of roads and other potential spatial predictors in driving forest loss in our 1.42 million plots. To do this we first created a comprehensive land-cover map for our study region and then quantified the percentage of land cleared per plot (hereafter termed ‘forest loss’) as our response variable. Our map was explicitly designed to accurately detect forest loss while not misclassifying current land covers, such as oil-palm or wood-pulp plantations, as forested land, or open vegetation, such as wetlands, as deforested land (Supplementary Information). We then identified 38 key environmental, demographic or socioeconomic variables potentially related to deforestation (Supplementary Table 4). Included among these were neighbourhood road density (total length of roads within a 5-km radius of each plot) and road proximity (linear distance of the plot to the nearest road). Much road building in the tropics is linked to agriculture—the largest ultimate driver of deforestation in the Asia-Pacific region^{24–26}—which itself is influenced by underlying socioeconomic and demographic factors^{27,28}. Roads also promote deforestation by markedly reducing the costs of transporting timber, bulk minerals, fossil fuels and poached wildlife to domestic or international markets^{27,29}.

To model forest loss on the basis of our 38 potential predictor variables, we developed a generalized linear model with LASSO regularization¹¹ (a technique that encourages simple, sparse models, with

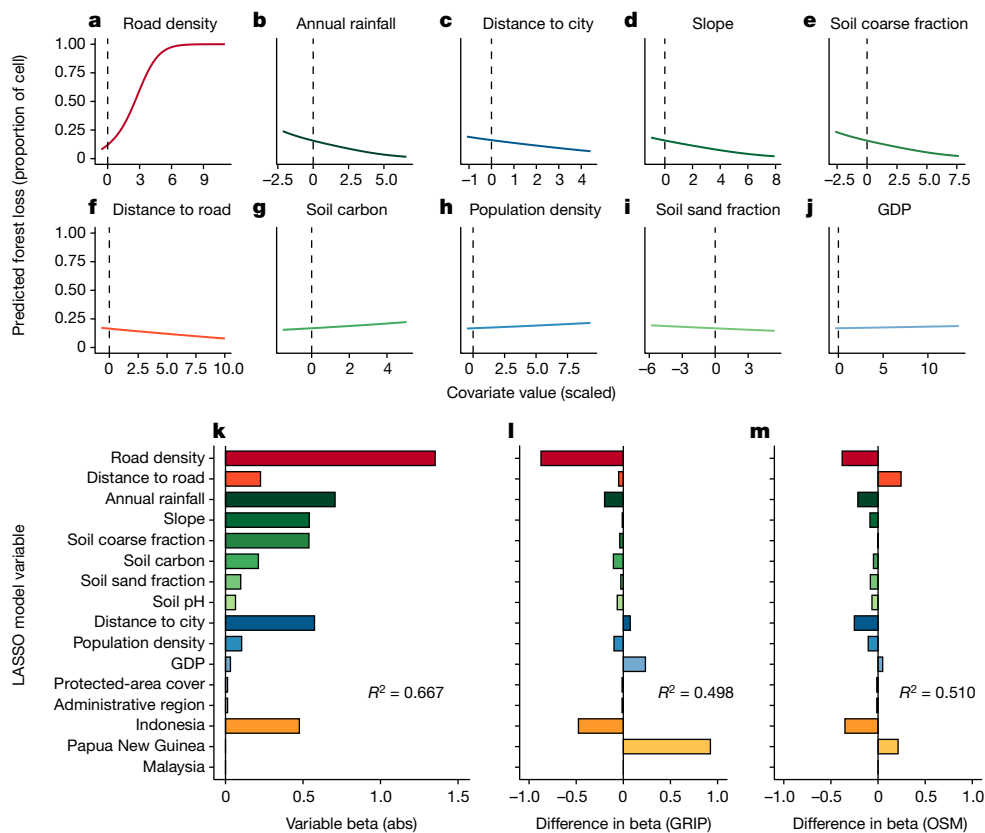


Fig. 2 | Environmental and socioeconomic features that influence forest loss across the tropical Asia-Pacific region. **a–j**, Partial differential plots showing relationships between the 10 most influential features and forest cover (road density (**a**), annual rainfall (**b**), distance to city (**c**), slope (**d**), soil coarse fraction (**e**), distance to road (**f**), soil carbon (**g**), population density (**h**),

soil sand fraction (**i**) and gross domestic product (GDP) (**j**). The *x*-axis values indicate the number of standard deviations from the mean; see Supplementary Table 5. **k–m**, Spatial predictors of deforestation, showing slope (beta) values for model using our road data (**k**) and the difference in slope values when using alternative road data from GRIP (**l**) and OSM (**m**). Abs, absolute values.

fewer parameters and less model variance and bias). Out of these 38 potential predictors, 14 had a discernible relationship with forest loss (Supplementary Information), and their effects were then contrasted using road datasets from this study, GRIP and OSM (Fig. 2). Notably, the marginal relationship between road density and forest loss was distinctly nonlinear (Fig. 2a). This sigmoidal curve suggests a general threshold effect of roads, with deforestation rates being highest when new roads are first constructed in landscape, and then gradually decreasing as road density increases. Forests are expected to decline most sharply when roads initially encroach, up to a road density of around 4 km km⁻², with accessible forests becoming largely depleted if road density exceeds around 7.5 km km⁻². Broadly similar dynamics have been observed in rural communities experiencing ‘boom-and-bust’ development in the Brazilian Amazon³⁰, where initial road building triggers rapid forest loss followed by declines in environmental and human welfare as forest resources are increasingly exhausted.

In our final LASSO regression model, several other variables—annual rainfall, distance to nearest city, topographic slope, soil coarse fraction, distance to nearest road, and country—had modest explanatory power, with each having beta (slope) values significantly smaller than that of road density (Fig. 2b–j). Marginal relationships of these variables with forest loss largely followed expected trends (that is, forest loss was highest near townships or cities, in flatter areas and in less-rainy locales where forest burning is easier) (Fig. 2b–j). Notably, protected-area coverage (with a beta value of just -0.01) had little influence on model performance. We did not evaluate various other potential drivers of deforestation, especially ultimate factors (for example, poverty, access to global markets and social norms)²⁷ for which we lacked adequately spatially resolved data. Thus, although road density was the strongest

spatial predictor of forest loss in our study, we were unable to consider every conceivable driver of deforestation in our model.

We also ran separate LASSO models for each country and then compared their performance with that of our region-wide LASSO model, which indicated that Indonesia had a higher marginal rate of forest loss than did either Malaysia or Papua New Guinea. Notably, the region-wide model performed better (pseudo $R^2 = 0.667$) than the three country-level models (pseudo $R^2 = 0.540$, generated by using area-weighted averages for each nation) (Supplementary Fig. 3). In addition, we reran our LASSO regression while excluding large-scale oil-palm and pulpwood plantations (Supplementary Fig. 4), which are associated with considerable deforestation in the Asia-Pacific region²⁴. This produced only negligible changes in model slope parameters and overall outcome (Supplementary Information), underscoring the robustness of our region-wide model.

The LASSO model based on our road data, which included ghost roads, differed in three important ways from those based on the GRIP and OSM datasets (Fig. 2k–m). First, the model with our improved road data was considerably stronger, explaining more of the total deviance in the response variable (66.7%) than did either the GRIP- or the OSM-based models (49.8% and 51.0%, respectively). As a result, our model was better at predicting spatial patterns of forest loss across our study area (Fig. 2). Second, when using our road data, road density was a much stronger correlate of forest loss (with a beta value of 1.35, which is around 1.4–2.8 times greater than OSM- and GRIP-based values, respectively). Third, the effect of country on forest-conversion rates differed substantially (particularly for Indonesia and Papua New Guinea) when using GRIP or OSM data, compared with our comprehensive road dataset. Hence, the widely used GRIP and OSM datasets are

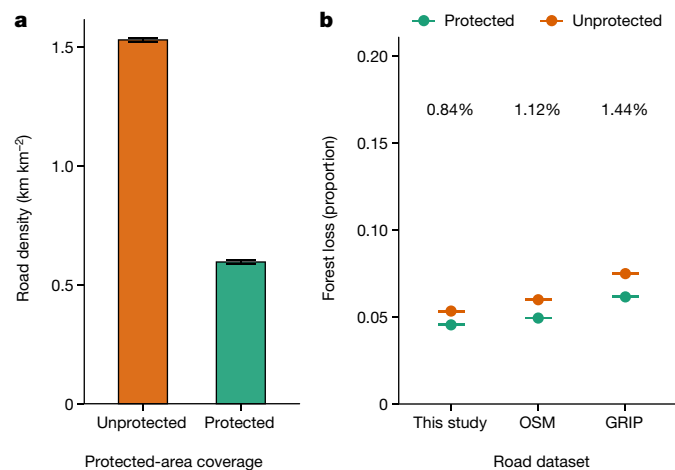


Fig. 3 | Effects of protected areas in limiting road construction and forest loss. **a**, Differences in road density between protected and unprotected areas after site-matching analysis. **b**, Marginal difference in forest loss between protected and unprotected cells after site matching with the full dataset from this study. For both panels, error bars show 95% confidence limits (in **b**, the error is too small for the gap between bars to be visible).

not just seriously incomplete but also markedly inconsistent among nations or geographic regions (Supplementary Table 2)—with developing nations generally having much poorer road data than do wealthier nations^{5,18}.

Roads and protected areas

Next, we assessed the degree to which areas that are designated as protected by the International Union for Conservation of Nature (categories I–VI) limit road incursions and forest loss, relative to non-protected areas, using the three road datasets. We first used propensity-score matching³¹ to account for non-random locations of protected areas, such as biases toward steeper or less productive lands (Supplementary Information). We then used separate propensity-score analyses to assess the capacity of protected areas to reduce both road incursions and forest loss.

When comparing matched sites, we found that average road density was more than twice as high (256.7%) outside protected areas than inside them (Fig. 3a). However, after accounting for lower road densities inside protected areas, the marginal effects of protected-area coverage on forest loss were surprisingly modest: less than 1% in magnitude when based on the road datasets from this study, and less than 1.5% in magnitude when based on data from OSM or GRIP (Fig. 3b). This suggests that, on a per-kilometre basis, roads inside protected areas lead to nearly as much forest loss as do roads outside protected areas. We assert that the most crucial conservation function of terrestrial protected areas, at least in the Asia-Pacific region, is limiting road incursions and their many associated impacts on forests.

Roads precede forest loss

Finally, to test whether roads tend to precede deforestation, or rather, follow it, we evaluated the temporal sequence of land-use change in 12 large land parcels (each around 400 km² in area) arrayed across Sumatra, Borneo and New Guinea (Fig. 4a). We created 35 annual road maps using annual Landsat imagery from 1985 to 2020 and then identified the spatio-temporal relationship between road construction and deforestation using published annual deforestation data³² (Supplementary Information). We summarized this relationship by classifying areas in each parcel that were deforested before, during or after road

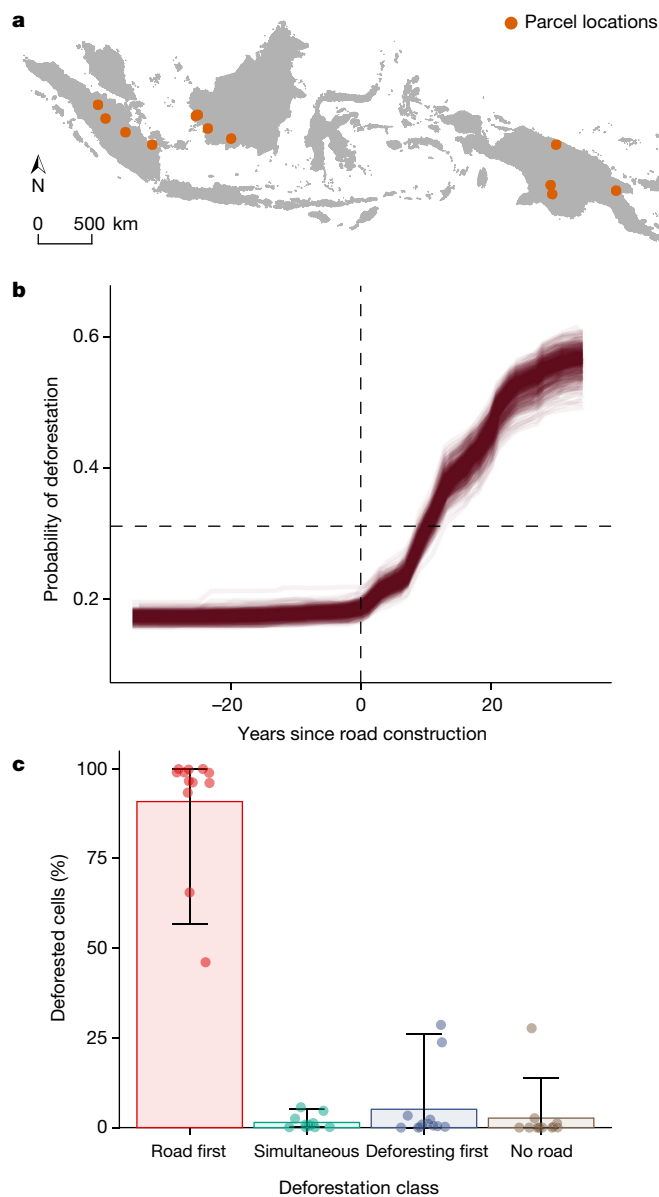


Fig. 4 | Roads usually precede deforestation. **a–c**, Temporal relationships between road construction and deforestation for 12 study sites arrayed across the continental islands of Sumatra, Borneo and New Guinea. **a**, Locations of the 12 study sites. **b**, Partial plots for random forest models testing the temporal relationship between nearby road construction and deforestation (negative values for ‘Years since road construction’ indicate the number of years before road building, whereas the horizontal dotted line shows where true positive and true negative rates are maximized—where cells are more likely to be deforested than not). Each individual line shows the partial plot from a single model iteration. **c**, Median deforestation rates associated with different road-proximity categories. Error bars show the 5–95% interpercentile range for each category.

construction, as well as areas that were deforested independently of roads (more than 2 km from the nearest road).

In our 12 study locations, the probability of deforestation was low before road construction, but spiked immediately after nearby roads were created (Fig. 4b). Our assessment showed that the large majority of deforestation—92.2%, on average—occurred after, or concurrently with, the construction of nearby roads (Fig. 4c). Forest loss preceded road construction in just 5.1% of the total area sampled. These trends indicate that forest loss in our study region is overwhelmingly triggered by ongoing road expansion, rather than vice versa. The 12 study locations include some large-scale oil-palm and pulpwood

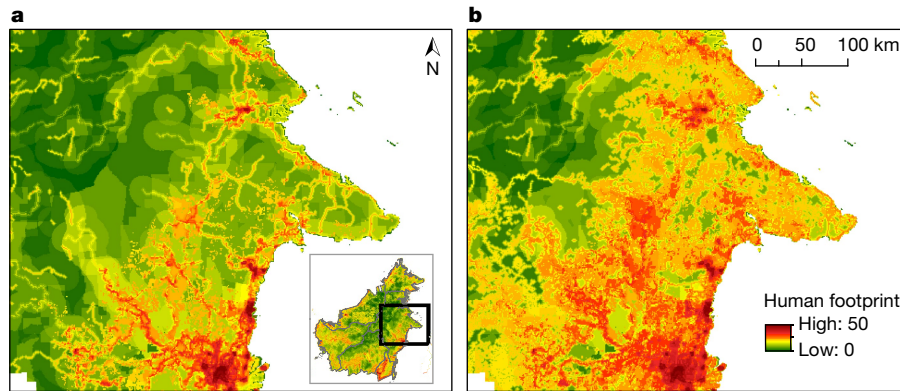


Fig. 5 | Two versions of the human footprint for eastern and central Borneo, using data from 2020. a, b, These maps are based on incomplete road data from OpenStreetMap²² (a) and more-complete road data from this study (b).

plantations, in which forest loss also typically followed road construction (Supplementary Fig. 3).

Summary and conclusions

Using road data generated by trained volunteers, we recorded 3.0–6.6 times more roads in the Asia-Pacific region than were found in leading global-road datasets, while revealing vast numbers of unmapped ‘ghost roads’. These findings have key implications for forest conservation. As a consequence of rapidly proliferating ghost roads, government datasets on roads often have large blind spots and inconsistencies, inhibiting spatial planning, law enforcement and the collection of government rents and royalties on exploited natural resources (Supplementary Table 2).

Striking gaps in road maps are not at all unusual, especially for developing nations^{2–10}. For instance, studies in the Brazilian Amazon^{6,33,34}, Cameroon³⁵ and the Solomon Islands^{10,36} also detected many unmapped or illegal roads, ranging from 2.8 to 9.9 times those recorded in OSM or government sources—values that broadly overlap with and even exceed those observed in our Asia-Pacific study area. Protected areas in this region provided considerable protection against road incursions, containing just a third as many roads as did comparable unprotected areas (Fig. 3a). On a per-kilometre basis, however, roads inside protected areas caused nearly as much forest loss as did those in unprotected areas (Fig. 3b). This underscores, in our view, an urgent need to limit unregulated road expansion in protected areas as a general conservation strategy^{37–40}.

Although global-road databases are gradually improving in quality⁴¹, their many gaps and inconsistencies greatly limit their value for comparing different nations, regions and ecosystem types. Furthermore, popular conservation metrics, such as the human footprint^{21,22} and roadless or wilderness areas²³, are being based in part on seriously incomplete road data. For example, the estimated human footprint in the environmentally critical region of east–central Borneo differs markedly when it is based on a recent OSM road map (Fig. 5a), compared with when it is based on our road data (Fig. 5b). Among these differences, the mapped region in Borneo had twice as much land area with ‘very high’ disturbance (28.4% versus 14.5%), and only half as much land with ‘low’ disturbance (6.6% versus 13.6%), when based on our updated road map and forest-disturbance classifications from the human-footprint study²².

The road-mapping element of this study required around 7,000 h of effort by more than 200 trained volunteers or study authors. Such an intensive undertaking is justified only because human eyes still outperform artificial intelligence (AI)-based methods for identifying and mapping roads (especially when more-accurate, higher-resolution images are used, as in this study). At larger spatial scales, the required

effort is even more daunting. For example, a global-scale analysis using our methods would require around 640,000 h of effort simply to map all of Earth’s current roads just once. For this reason, a viable AI-based road-mapping system is urgently needed⁴². Such schemes are under development^{33,43–45} and could potentially be trained using major datasets such as ours, aiming to provide accurate, global-scale road coverage in near real time. In practical terms, such an automated system is one of the most urgent conservation needs for tropical forests today. Nothing else will keep pace with the contemporary avalanche of proliferating roads.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-024-07303-5>.

- Dulac, J. *Global Land Transport Infrastructure Requirements: Estimating Road and Railway Infrastructure Capacity and Costs to 2050* (International Energy Agency, Paris, 2013).
- Laurance, W. F. et al. A global strategy for road building. *Nature* **513**, 229–232 (2014).
- Ascensão, F. et al. Environmental challenges for the Belt and Road Initiative. *Nat. Sustain.* **1**, 206–209 (2018).
- Kleinschroth, F. et al. Road expansion and persistence in forests of the Congo Basin. *Nat. Sustain.* **2**, 628–634 (2019).
- Ibisch, P. L. et al. A global map of roadless areas and their conservation status. *Science* **354**, 1423–1427 (2016).
- Barber, C. P., Cochrane, M. A., Souza, C. M. Jr & Laurance, W. F. Roads, deforestation, and the mitigating effect of protected areas in the Amazon. *Biol. Conserv.* **177**, 203–209 (2014).
- Laurance, W. F., Goosem, M. & Laurance, S. G. Impacts of roads and linear clearings on tropical forests. *Trends Ecol. Evol.* **24**, 659–669 (2009).
- Alamgir, M. et al. Economic, socio-political and environmental risks of road development in the tropics. *Curr. Biol.* **27**, 1130–1140 (2017).
- Fearnside, P. M. in *Handbook of Road Ecology* (eds van der Ree, R. et al.) 414–424 (John Wiley & Sons, 2015).
- Hughes, A. C. Have Indo-Malaysian forests reached the end of the road? *Biol. Conserv.* **223**, 129–137 (2018).
- Schrieber-Gregory, D. *Regulation Techniques for Multicollinearity: Lasso, Ridge, and Elastic Nets* https://www.lexjansen.com/wuss/2018/131_Final_Paper_PDF.pdf (2018).
- Hettiger, H. *When Do Rural Roads Benefit the Poor and How? An In-Depth Analysis* (Asian Development Bank, 2006).
- Laurance, W. F. et al. The future of the Brazilian Amazon. *Science* **291**, 438–439 (2001).
- Vilela, T. et al. A better Amazon road network for people and the environment. *Proc. Natl Acad. Sci. USA* **117**, 7095–7102 (2020).
- van der Ree, R., Smith, D. & Grilo, C. (eds) *Handbook of Road Ecology* (John Wiley & Sons, 2015).
- Bebbington, A. et al. Priorities for governing large-scale infrastructure in the tropics. *Proc. Natl Acad. Sci. USA* **117**, 21829–21833 (2020).
- Meijer, J. R., Huijbregts, M. A. J., Schotten, K. C. G. J. & Schipper, A. M. Global patterns of current and future road infrastructure. *Environ. Res. Lett.* **13**, 064006 (2018).

18. Ramm, F., Topf, J. & Chilton, S. *OpenStreetMap: Using and Enhancing the Free Map of the World* (UIT Cambridge, 2010).
19. Laurance, W. F. et al. Reducing the global environmental impacts of rapid infrastructure expansion. *Curr. Biol.* **25**, 259–262 (2015).
20. Sloan, S. et al. Infrastructure development and contested forest governance threaten the Leuser Ecosystem, Indonesia. *Land Use Policy* **77**, 298–309 (2018).
21. Venter, O. et al. Sixteen years of change in the global terrestrial human footprint and implications for biodiversity conservation. *Nat. Commun.* **7**, 12558 (2016).
22. Williams, B. A. et al. Change in terrestrial human footprint drives continued loss of intact ecosystems. *One Earth* **3**, 371–383 (2020).
23. Watson, J. E. M. et al. Catastrophic declines in wilderness areas undermine global environment targets. *Curr. Biol.* **26**, 2929–2934 (2016).
24. Curtis, P. G., Slay, C., Harris, N., Tyukavina, A. & Hansen, M. C. Classifying drivers of global forest loss. *Science* **361**, 1108–1111 (2018).
25. Pendrill, F. et al. Agricultural and forestry trade drives large share of tropical deforestation emissions. *Glob. Environ. Change* **56**, 1–10 (2019).
26. Pendrill, F. et al. Disentangling the numbers behind agriculture-driven tropical deforestation. *Science* **377**, eabm9267 (2022).
27. Geist, H. J. & Lambin, E. F. Proximate causes and underlying driving forces of tropical deforestation. *BioScience* **52**, 143–150 (2002).
28. Thünen, J. H. V. *The Isolated State in its Relation to Agriculture and National Economy* (Pergamon Press, 1966).
29. Meyfroidt, P. et al. Middle-range theories of land system change. *Glob. Environ. Change* **53**, 52–67 (2018).
30. Rodrigues, A., Ewers, R., Souza, C., Verissimo, A. & Balmford, A. Boom-and-bust development patterns across the Amazon deforestation frontier. *Science* **324**, 1435–1437 (2009).
31. Geldmann, J., Manica, A., Burgess, N. D., Coad, L. & Balmford, A. A global-level assessment of the effectiveness of protected areas at resisting anthropogenic pressures. *Proc. Natl Acad. Sci. USA* **116**, 23209–23215 (2019).
32. Vancutsem, C. et al. Long-term (1990–2019) monitoring of forest cover changes in the humid tropics. *Sci. Adv.* **7**, eabe1603 (2021).
33. Botelho, J., Costa, S., Ribeiro, J. & Souza, C. M. Mapping roads in the Brazilian Amazon with artificial intelligence and Sentinel-2. *Remote Sens.* **14**, 3625 (2022).
34. das Neves, P. et al. Amazon rainforest deforestation influenced by clandestine and regular roadway network. *Land Use Policy* **108**, 105510 (2021).
35. *Cameroon Road Network* (World Food Programme, Logistics Cluster, accessed 1 July 2022); <https://dlca.logcluster.org/23-cameroon-road-network>
36. Katovai, E., Sirikolo, M., Srinivasan, U., Edwards, W. & Laurance, W. F. Post-logging recovery dynamics and their effects on tree diversity in tropical forests of the Solomon Islands. *For. Ecol. Manag.* **373**, 53–63 (2016).
37. Qin, S. et al. Protected area downgrading, downsizing, and degazettement as a threat to iconic protected areas. *Conserv. Biol.* **33**, 1275–1285 (2019).
38. Laurance, W. F. et al. Averting biodiversity collapse in tropical forest protected areas. *Nature* **489**, 290–294 (2012).
39. Sloan, S., Bertzky, B. & Laurance, W. F. African development corridors intersect key protected areas. *Afr. J. Ecol.* **55**, 731–737 (2017).
40. Haddad, N. M. et al. Habitat fragmentation and its lasting impact on Earth's ecosystems. *Sci. Adv.* **1**, e1500052 (2015).
41. Herfort, B., Lautenbach, S., Albuquerque, J., Anderson, J. & Zipf, A. The evolution of humanitarian mapping within the OpenStreetMap community. *Sci. Rep.* **11**, 3037 (2021).
42. Laurance, W. F. Wanted: AI experts to map road-building boom. *Nature* **558**, 30 (2018).
43. Sloan, S., Talkhani, R. R., Huang, T., Engert, J. & Laurance, W. F. Mapping remote roads using artificial intelligence and satellite imagery. *Remote Sens.* **16**, 839 (2024).
44. Wang, W. et al. A review of road extraction from remote sensing images. *J. Traffic Transp. Eng.* **3**, 271–282 (2016).
45. De Clerk, E. Facebook uses AI technology to map roads of Thailand. *Intelligent Transport* <https://www.intelligenttransport.com/transport-news/85277/facebook-ai-map-roads-thailand/> (2019).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The datasets used for this study (including comprehensive road maps as a raster of road density at 1-km² resolution) are available in the Supplementary Information, on request to J.E.E. and W.F.L. or as follows. OpenStreetMap data are available from <https://download.geofabrik.de/>, and GRIP road data from <https://www.globio.info/download-grip-dataset>. National and subnational administrative-region data were obtained from GADM (<https://gadm.org/>). Population-density data were from WorldPop (<https://www.worldpop.org/>). GDP data were accessed at <https://datadryad.org/stash/dataset/doi:10.5061/dryad.dk1j0>. Protected-area data were from Protected Planet (<https://www.protectedplanet.net/en>). Waterways locations were obtained from the Global River Widths from Landsat Database (<https://zenodo.org/records/1297434>). Elevation data were accessed at <http://srtm.csi.cgiar.org>,

and rainfall data at <https://doi.org/10.16904/envidat.211>. Data for all soil variables were obtained from Soil Grids (<https://soilgrids.org/>).

Acknowledgements We thank J. Barlow, R. Chazdon, M. Cochrane, S. Das, M. Goosem, J. Jaeger, P. Negret, C. Souza Jr, R. Talkhani and H. Tao for many helpful comments, and our volunteer road-mappers, including Wild Green Memes for Ecological Fiends, for their efforts. The Australian Research Council, James Cook University and an anonymous philanthropic donor provided support.

Author contributions W.F.L., J.E.E., M.J.C., M.A. and S.S. conceived the study; J.E.E., J.E.C. and W.F.L. coordinated data analysis. W.F.L., J.E.E., J.E.C. and M.J.C. wrote the manuscript. J.E.E., Y.I., J.C. and S.S. generated key road datasets. M.J.C., J.E.C., M.A., S.S. and J.S. provided ideas and critical feedback.

Competing interests The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-024-07303-5>.

Correspondence and requests for materials should be addressed to Jayden E. Engert or William F. Laurance.

Peer review information *Nature* thanks Robin Chazdon, Jochen Jaeger, Pablo Negret and Carlos Souza Jr for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection Roads were digitized in Google Earth (version dependent on user). Model variables and other geo-spatial data were processed using Arcmap 10.8. Landcover patch sizes were calculated using the Region Group tool. Model variables measured at different focal window sizes (e.g. population density) were created using the Focal Statistics tool. Distance to road and distances to populated area variables were created using the Euclidean Distance tool. Values for model variables were extracted at sample locations using the Extract Multi Values to Points tool.

Data analysis All analyses were conducted in the R open source statistical platform v4.1.2. LASSO regressions were conducted using package 'glmnet' and assessed using package 'plotmo'. Spatial autoregressive terms were created using package 'raster' and Moran's I calculated using package 'moranfast'. Site-matching analysis and balance checks were conducted using package 'MatchIt' and marginal effects of protected areas on road density were calculated using packaged 'lmtest'. Random forest classification analysis was conducted using package 'ranger'. Partial plots were created using package 'pdp' and AUC values were calculated using package 'dismo'.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Correspondence and requests for materials should be addressed to J.E.E (jayden.engert@my.jcu.edu.au) and W.F.L. (bill.laurance@jcu.edu.au). The datasets used for this study (including comprehensive roadmaps as a raster of road density at 1-km² resolution) are available in the Supplementary Information, on request to J.E.E and W.F.L., or as follows. OpenStreetMap data are available from <https://download.geofabrik.de/>, and GRIP road data from <https://www.globio.info/download-grip-dataset>. National and subnational administrative-region data were obtained from GADM (<https://gadm.org/>). Population-density data were from WorldPop (<https://www.worldpop.org/>). Gross Domestic Product data were accessed at <https://datadryad.org/stash/dataset/doi:10.5061/dryad.dk1j0>. Protected area data were from Protected Planet (<https://www.protectedplanet.net/en>). Waterways locations were obtained from the Global River Widths from Landsat Database (<https://zenodo.org/records/1297434>). Elevation data was accessed at <http://srtm.csi.cgiar.org>. Rainfall data was accessed at <https://doi.org/10.16904/envividat.211>. Data for all soil variables were obtained from Soil Grids (<https://soilgrids.org/>).

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	NA
Reporting on race, ethnicity, or other socially relevant groupings	NA
Population characteristics	NA
Recruitment	NA
Ethics oversight	NA

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	<i>Describe how sample size was determined, detailing any statistical methods used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.</i>
Data exclusions	<i>Describe any data exclusions. If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.</i>
Replication	<i>Describe the measures taken to verify the reproducibility of the experimental findings. If all attempts at replication were successful, confirm this OR if there are any findings that were not replicated or cannot be reproduced, note this and describe why.</i>
Randomization	<i>Describe how samples/organisms/participants were allocated into experimental groups. If allocation was not random, describe how covariates were controlled OR if this is not relevant to your study, explain why.</i>
Blinding	<i>Describe whether the investigators were blinded to group allocation during data collection and/or analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.</i>

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	<i>Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).</i>
Research sample	<i>State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.</i>
Sampling strategy	<i>Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.</i>
Data collection	<i>Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.</i>
Timing	<i>Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.</i>
Data exclusions	<i>If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.</i>
Non-participation	<i>State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.</i>
Randomization	<i>If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.</i>

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We identified unmapped 'ghost roads' across the islands of Sumatra, Borneo, and New Guinea, and assessed their effect on forest loss on the region and the ability of protected areas to mitigate the effects of mapped and unmapped roads. Specifically, we (1) digitized unmapped roads across the region and calculated the length of unmapped road and the extent of the road networks; (2) conducted LASSO regression to identify the drivers of deforestation and compare model performance when using existing road datasets (OpenStreetMap and Global Roads Inventory Project) or when using our new road dataset; (3) used propensity-scoring matching to assess the ability of protected areas to limit road-related deforestation; and (4) conducted spatio-temporal analysis to assess if roads preceded or followed deforestation.
Research sample	For LASSO regression, the research sample was mapped areas of Sumatra, Borneo, and New Guinea (1,418,755 1 km ² grid cells). This study region was chosen as it contains a substantial area of high-biodiversity and high-carbon tropical forest that is under imminent threat from the expansion of road networks, agriculture, and extractive industry. 'Mapped areas' refers to the portion of this study region in which we digitised roads that were absent from existing road maps. For the spatio-temporal analysis, the research sample was 12 plots of approximately 400 km ² , arrayed across the aforementioned study region.
Sampling strategy	For LASSO regression, the sampling strategy was to include all mapped areas that also had available data for other model variables. For the spatio-temporal analysis, the sampling strategy involved choosing plots that (1) were arrayed evenly (four plots each) across the three large islands in this study; (2) occurred in sites that had little if any road construction or forest loss prior to 1990 (because most deforestation in the region has occurred since that date); (3) broadly sampled prevailing landcover types (small- and large-scale plantations and various agricultural lands) in the region; and (4) each parcel had >50% forest loss by 2020 to ensure it included sizable areas of roads, forests, and deforested lands to evaluate in our analysis.
Data collection	We mapped and digitized all visible roads that were not recorded in existing road datasets, using Google Earth (circa 2019). Roads were added to new kml files using the New Path tool and by tracing the location of all visible roads in the satellite image, and then exported to shapefiles in Arcmap 10.8. Road-mapping was conducted by 5 coauthors of this study and by 205 trained volunteers, whose individual mapping accuracy was independently quality-checked by one or more study coauthors. Using a standardized protocol that we developed, the newly created road files were required to achieve a minimum accuracy of ≥90%, including road omissions and commissions, verified using standardized test datasets. For the spatio-temporal analysis we digitized all detectable roads for each year between 1985 and 2020, and then used these roadmaps to calculate the linear distance to the nearest road for each year, at 1-ha raster resolution.
Timing and spatial scale	For the spatio-temporal analysis, the temporal scale was 1985-2020 inclusive, with roads digitized at yearly intervals. The LASSO regression had no temporal dimension. The spatial scale for all analyses was the islands of Sumatra, Borneo, and New Guinea.

Data exclusions	Data samples were only excluded when one of the model variables did not have available information for that location, or when the most common land-cover type was water.
Reproducibility	The study was not an experiment and hence we did not replicate the findings. However we did conduct six different sensitivity analyses to ensure results were not influenced by certain factors (such as high road densities in industrial plantations, or differences in road densities and forest loss between countries). All data used in the study (apart from our new road dataset) are available from open data sources and referenced in the manuscript.
Randomization	Randomized controlled trials were not appropriate for the analyses conducted. Data were not randomized but represent a large area of natural and anthropogenic landscapes across the islands of Sumatra, Borneo, and New Guinea. Rather than randomize the regions in which we digitized unmapped roads, we (1) initially prioritised minimally modified landscapes to identify unmapped roads in natural systems, and (2) subsequently prioritised areas representing the variety of different land-cover types present in the study region. To create a parsimonious model of the drivers of forest loss, we used spatial-fold cross validation in our LASSO model, which is more appropriate for spatially autocorrelated data than is random cross-fold validation. For the assessment of protected area effectiveness, we used propensity-score matching to improve balance between samples inside and outside protected areas.
Blinding	Blinding was not necessary and was not used as no experimental procedures that could be affected by observer bias were conducted.
Did the study involve field work?	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No

Field work, collection and transport

Field conditions	NA
Location	NA
Access & import/export	NA
Disturbance	NA

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	<i>Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.</i>
Validation	<i>Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.</i>

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)	<i>State the source of each cell line used and the sex of all primary cell lines and cells derived from human participants or vertebrate models.</i>
Authentication	<i>Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.</i>

Mycoplasma contamination

Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.

Commonly misidentified lines
(See [ICLAC](#) register)

Name any commonly misidentified cell lines used in the study and provide a rationale for their use.

Palaeontology and Archaeology

Specimen provenance

Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information). Permits should encompass collection and, where applicable, export.

Specimen deposition

Indicate where the specimens have been deposited to permit free access by other researchers.

Dating methods

If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.

Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight

Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals

For laboratory animals, report species, strain and age OR state that the study did not involve laboratory animals.

Wild animals

Provide details on animals observed in or captured in the field; report species and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.

Reporting on sex

Indicate if findings apply to only one sex; describe whether sex was considered in study design, methods used for assigning sex. Provide data disaggregated for sex where this information has been collected in the source data as appropriate; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex-based analyses where performed, justify reasons for lack of sex-based analysis.

Field-collected samples

For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.

Ethics oversight

Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration

Provide the trial registration number from [ClinicalTrials.gov](#) or an equivalent agency.

Study protocol

Note where the full trial protocol can be accessed OR if not available, explain why.

Data collection

Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.

Outcomes

Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.

Dual use research of concern

Policy information about [dual use research of concern](#)

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

- | | | |
|--------------------------|--------------------------|----------------------------|
| No | Yes | |
| <input type="checkbox"/> | <input type="checkbox"/> | Public health |
| <input type="checkbox"/> | <input type="checkbox"/> | National security |
| <input type="checkbox"/> | <input type="checkbox"/> | Crops and/or livestock |
| <input type="checkbox"/> | <input type="checkbox"/> | Ecosystems |
| <input type="checkbox"/> | <input type="checkbox"/> | Any other significant area |

Experiments of concern

Does the work involve any of these experiments of concern:

- | | | |
|--------------------------|--------------------------|---|
| No | Yes | |
| <input type="checkbox"/> | <input type="checkbox"/> | Demonstrate how to render a vaccine ineffective |
| <input type="checkbox"/> | <input type="checkbox"/> | Confer resistance to therapeutically useful antibiotics or antiviral agents |
| <input type="checkbox"/> | <input type="checkbox"/> | Enhance the virulence of a pathogen or render a nonpathogen virulent |
| <input type="checkbox"/> | <input type="checkbox"/> | Increase transmissibility of a pathogen |
| <input type="checkbox"/> | <input type="checkbox"/> | Alter the host range of a pathogen |
| <input type="checkbox"/> | <input type="checkbox"/> | Enable evasion of diagnostic/detection modalities |
| <input type="checkbox"/> | <input type="checkbox"/> | Enable the weaponization of a biological agent or toxin |
| <input type="checkbox"/> | <input type="checkbox"/> | Any other potentially harmful combination of experiments and agents |

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.

ChIP-seq

Data deposition

- Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).
- Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links

May remain private before publication.

For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.

Files in database submission

Provide a list of all files available in the database submission.

Genome browser session

(e.g. [UCSC](#))

Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.

Methodology

Replicates

Describe the experimental replicates, specifying number, type and replicate agreement.

Sequencing depth

Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.

Antibodies

Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.

Peak calling parameters	<i>Specify the command line program and parameters used for read mapping and peak calling, including the CHIP, control and index files used.</i>
Data quality	<i>Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.</i>
Software	<i>Describe the software used to collect and analyze the CHIP-seq data. For custom code that has been deposited into a community repository, provide accession details.</i>

Flow Cytometry

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	<i>Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.</i>
Instrument	<i>Identify the instrument used for data collection, specifying make and model number.</i>
Software	<i>Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.</i>
Cell population abundance	<i>Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.</i>
Gating strategy	<i>Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.</i>

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	<i>Indicate task or resting state; event-related or block design.</i>
Design specifications	<i>Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.</i>
Behavioral performance measures	<i>State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).</i>

Acquisition

Imaging type(s)	<i>Specify: functional, structural, diffusion, perfusion.</i>
Field strength	<i>Specify in Tesla</i>
Sequence & imaging parameters	<i>Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.</i>
Area of acquisition	<i>State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.</i>
Diffusion MRI	<input type="checkbox"/> Used <input type="checkbox"/> Not used

Preprocessing

Preprocessing software	<i>Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).</i>
------------------------	--

Normalization	<i>If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.</i>
Normalization template	<i>Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.</i>
Noise and artifact removal	<i>Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).</i>
Volume censoring	<i>Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.</i>

Statistical modeling & inference

Model type and settings	<i>Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).</i>
Effect(s) tested	<i>Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.</i>
Specify type of analysis:	<input type="checkbox"/> Whole brain <input type="checkbox"/> ROI-based <input type="checkbox"/> Both
Statistic type for inference	<i>Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.</i>
(See Eklund et al. 2016)	
Correction	<i>Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).</i>

Models & analysis

n/a | Involved in the study

- Functional and/or effective connectivity
- Graph analysis
- Multivariate modeling or predictive analysis

Functional and/or effective connectivity	<i>Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).</i>
Graph analysis	<i>Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).</i>
Multivariate modeling and predictive analysis	<i>Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.</i>