

Dear Author

Please use this PDF proof to check the layout of your article. If you would like any changes to be made to the layout, you can leave instructions in the online proofing interface. First, return to the online proofing interface by clicking "Edit" at the top page, then insert a Comment in the relevant location. Making your changes directly in the online proofing interface is the quickest, easiest way to correct and submit your proof.

Please note that changes made to the article in the online proofing interface will be added to the article before publication, but are not reflected in this PDF proof.

If you would prefer to submit your corrections by annotating the PDF proof, please download and submit an annotatable PDF proof by clicking the link below.

 [Annotate PDF](#)

Language and ethnobiological skills decline precipitously in Papua New Guinea, the world's most linguistically diverse nation

Alfred Kik^{a,b,c,1}, Martin Adamec^d, Alexandra Y. Aikhenvald^e, Jarmila Bajzekova^{a,c}, Nigel Baro^{f,g}, Claire Bownen^h, Robert K. Colwell^{i,j}, Pavel Drozd^d, Pavel Duda^a, Sentiko Ibalim^{a,c,f}, Leonardo R. Jorge^c, Jane Mogina^k, Ben Rulif^k, Katerina Sam^{a,c}, Hannah Sarvasy^l, Simon Saulei⁹, George D. Weiblen^m, Jan Zrzavy^a, and Vojtech Novotny^{a,c,1}

^aFaculty of Sciences, University of South Bohemia, 37011 Ceske Budejovice, Czech Republic; ^bSchool of Science and Technology, University of Goroka, 441 Goroka, Papua New Guinea; ^cBiology Centre, Czech Academy of Sciences, 37011 Ceske Budejovice, Czech Republic; ^dFaculty of Sciences, University of Ostrava, 70200 Ostrava, Czech Republic; ^eLanguage and Culture Research Centre, James Cook University, QLD 4870 Cairns, Australia; ^fNew Guinea Binatang Research Center, 511 Madang, Papua New Guinea; ^gSchool of Natural and Physical Sciences, University of Papua New Guinea, 121 Port Moresby, Papua New Guinea; ^hDepartment of Linguistics, Yale University, New Haven, CT 06511; ⁱEcology and Evolutionary Biology, University of Connecticut, Storrs, CT 06269; ^jDepartment of Entomology, University of Colorado Museum of Natural History, Boulder, CO 80302; ^kJane Mogina Environment Consultants, 121 Port Moresby, Papua New Guinea; ^lMARCS Institute for Brain, Behaviour and Development, Western Sydney University, NSW 2751 Penrith, Australia; and ^mDepartment of Plant and Microbial Biology, University of Minnesota, St. Paul, MN 55108

Edited by Nicholas Evans, Australian National University, Canberra, Australia, and accepted by Editorial Board Member Elsa M. Redmond April 13, 2021 (received for review January 4, 2021)

Papua New Guinea is home to >10% of the world's languages and rich and varied biocultural knowledge, but the future of this diversity remains unclear. We measured language skills of 6,190 students speaking 392 languages (5.5% of the global total) and modeled their future trends using individual-level variables characterizing family language use, socioeconomic conditions, students' skills, and language traits. This approach showed that only 58% of the students, compared to 91% of their parents, were fluent in indigenous languages, while the trends in key drivers of language skills (language use at home, proportion of mixed-language families, urbanization, students' traditional skills) predicted accelerating decline of fluency to an estimated 26% in the next generation of students. Ethnobiological knowledge declined in close parallel with language skills. Varied medicinal plant uses known to the students speaking indigenous languages are replaced by a few, mostly nonnative species for the students speaking English or Tok Pisin, the national lingua franca. Most (88%) students want to teach indigenous language to their children. While crucial for keeping languages alive, this intention faces powerful external pressures as key factors (education, cash economy, road networks, and urbanization) associated with language attrition are valued in contemporary society.

ethnobiology | language attrition | language endangerment | biocultural diversity | Papua New Guinea

When evaluated against a common set of extinction-risk criteria, the world's ~7,000 extant languages (1) are even more threatened than its biological diversity (2). Orally transmitted cultural knowledge may be threatened by similar forces (3, 4). Language population sizes approximate a log-normal distribution (5), such that the majority of languages have relatively few speakers (1). Nearly half of the world's languages are considered endangered (1, 6). Language extinction is accelerating, with 30% of recorded extinctions having occurred since 1960 (6). Language vulnerability to extinction depends on speakers' attitudes toward their languages as well as on socioeconomic factors (7). However, quantitative evidence on the relative impact of individual drivers of language endangerment is almost nonexistent (8, 9), making it impossible to understand and predict language attrition. Furthermore, language skills and ethnobiological knowledge are rarely examined in relation to socioeconomic variables for individual speakers, as required for mechanistic understanding of language attrition and loss of ethnobiological knowledge (10–12). The present study uses a

modeling approach to assess multiple drivers of language attrition and ethnobiological knowledge loss, based on extensive data for individual speakers, to predict future trends in a global hotspot of linguistic and cultural diversity.

Papua New Guinea (PNG) is the world's most linguistically diverse nation, where ~9 million people speak ~840 languages (5, 13). PNG's languages are highly diverse, classified into at least 33 families (14). Until recently, these languages enjoyed widespread vitality due to the absence of a dominant language in the region, stable small-scale multilingualism (15), and focus on language as a marker of group identity (7, 16). New Guinea is also the world's most floristically diverse island (17), comprising ~5% of the world's biodiversity (18). Throughout PNG, numerous indigenous communities have explored, systematized,

Significance

Around the world, more than 7,000 languages are spoken, most of them by small populations of speakers in the tropics. Globalization puts small languages at a disadvantage, but our understanding of the drivers and rate of language loss remains incomplete. When we tested key factors causing language attrition among Papua New Guinean students speaking 392 different indigenous languages, we found an unexpectedly rapid decline in their language skills compared to their parents and predicted further acceleration of language loss in the next generation. Language attrition was accompanied by decline in the traditional knowledge of nature among the students, pointing to an uncertain future for languages and biocultural knowledge in the most linguistically diverse place on Earth.

Author contributions: A.K. and V.N. designed research; A.K., N.B., S.I., and B.R. performed research; L.R.J. contributed new analytic tools; A.K., M.A., A.Y.A., J.B., N.B., C.B., R.K.C., P. Drozd, P. Duda, S.I., L.R.J., J.M., B.R., K.S., H.S., S.S., G.D.W., J.Z., and V.N. analyzed data; and A.K., M.A., A.Y.A., J.B., N.B., C.B., R.K.C., P. Drozd, P. Duda, S.I., L.R.J., J.M., B.R., K.S., H.S., S.S., G.D.W., J.Z., and V.N. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. N.E. is a guest editor invited by the Editorial Board.

This open access article is distributed under [Creative Commons Attribution License 4.0 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).

¹To whom correspondence may be addressed. Email: akik@prf.jcu.cz or novotny@entu.cas.cz.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2100096118/-DCSupplemental>.

63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124

used, and managed the extraordinary biodiversity in their natural environment, thus generating extensive biocultural knowledge of local ecosystems (19–21). The traditional environmental knowledge of indigenous communities is in decline world-wide in response to the forces of cultural and economic globalization (22). Only 20% of PNG ethnolinguistic groups have any of their traditional plant uses recorded in the literature, and detailed information (>100 plant-use records) exists for only 2.5% of groups (3). Likewise, the contemporary status on this knowledge remains poorly documented.

At present, 32% of indigenous languages in PNG are considered endangered (1) largely due to their replacement by Tok Pisin (an English-based creole and PNG’s major lingua franca) or English (the language of formal education) (23). However, the true status of the country’s languages cannot be assessed in the absence of a national linguistic survey (24). This study presents such a survey and examines the present status and future dynamics of language and biocultural knowledge loss.

Results and Discussion

Language-Skills Drivers. We used questionnaires that compiled information on socioeconomic background and self-reported language fluency for 6,190 secondary-school students followed by tests of their language skills and ethnobiological knowledge. This survey captured 392 languages (46% of languages spoken in PNG and 5.5% worldwide), including 110 languages with ≥ 10 respondents (Fig. 1 and Dataset S1). We have uncovered a dramatic decline in the language skills in a single generation. While 90.8% of students’ parents reportedly speak an indigenous language fluently and only 0.3% of them have no indigenous language skills, just 57.7% of students consider themselves fluent in an indigenous language, whereas 2.0% of students reported a complete lack of indigenous language (Fig. 2A). The 110 languages with ≥ 10 respondents lost, on average, $40 \pm 2.1\%$ (\pm SE) of fluent speakers in the contemporary generation, from parents to the secondary school students we studied (Fig. 2B). The parent–student comparison suggests that language attrition is a recent phenomenon and thus not a direct consequence of the colonial past of PNG (until 1975) but rather a result of economic and social development of a country undergoing globalization.

We tested a set of factors characterizing students’ life skills, family language use, socioeconomic conditions, and language traits that potentially affect language skills (7) (Fig. 3 and SI Appendix, Fig. S1 and Table S1). The language used at home was the most important predictor of language skills. Indigenous

languages, used in 30% of all families, competed with Tok Pisin and English, used in 66% and 4% of families, respectively. More interestingly, home language use was also strongly impacted by mixed-language family background, the second most important predictor of language skills. The effect of mixed-language family remained large even after taking into account its effect on home language use, since only 16% of mixed-language families used an indigenous language at home compared to 38% of same-language families.

The small-scale multilingualism that was historically widespread in PNG and continues in rural parts of the country (25, 26) does not lead to language attrition (27). However, modern urban mixing with communication in Tok Pisin or English is different (28). Presently, 37% of the surveyed students grew up in mixed-language families. The secondary schools we surveyed are a favorable environment for language mixing, attended by students speaking 17 to 124 languages per school (SI Appendix, Table S2). Only 35% of the students speak the same indigenous language as their best friend, which is not very different from the 23% of students expected to do so if friendships were formed randomly with respect to the languages spoken by students. This pattern indicates a potential for further increase in nontraditional mixed-language marriages of these students.

Urbanization, another important factor correlated with language skills, often interrupts contacts between generations crucial for language transfer (7, 10). Urbanization in PNG has been kept low (87% of the population is rural) (29) by customary land ownership (92% of families in our study owned land), since urban dwellers could lose their land rights to relatives who continue to live on their land in villages (30). Urban environment had a strong negative impact on language skills among the 35% of students growing up in towns and cities compared to those growing up in a rural setting, particularly in a remote village.

Parents’ education and employment had only small effects on language skills once the related factors of urbanization and home language use were accounted for (Fig. 3). Students whose parents had salaried employment had lower language skills compared to those with parents growing cash crops or food for subsistence. The statistical importance of parents’ language skills was low, since almost all were fluent in an indigenous language. Indigenous language skills were positively correlated with a student’s reported traditional skills (hunting, fishing, farming, house building, and medicinal plant use) and negatively with contemporary technical skills (mobile phone and computer use). The individual differences in students’ skills thus remain

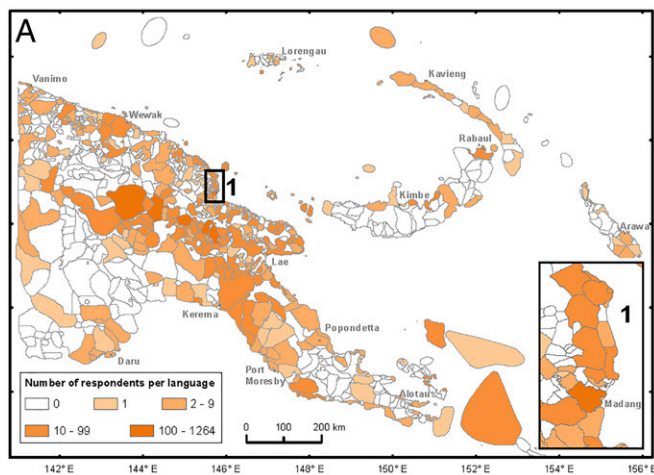


Fig. 1. Languages studied in Papua New Guinea. (A) Language map (1) with the number of students surveyed. (B) Survey of 486 students speaking 37 indigenous languages at the Mt. Hagen Secondary School.

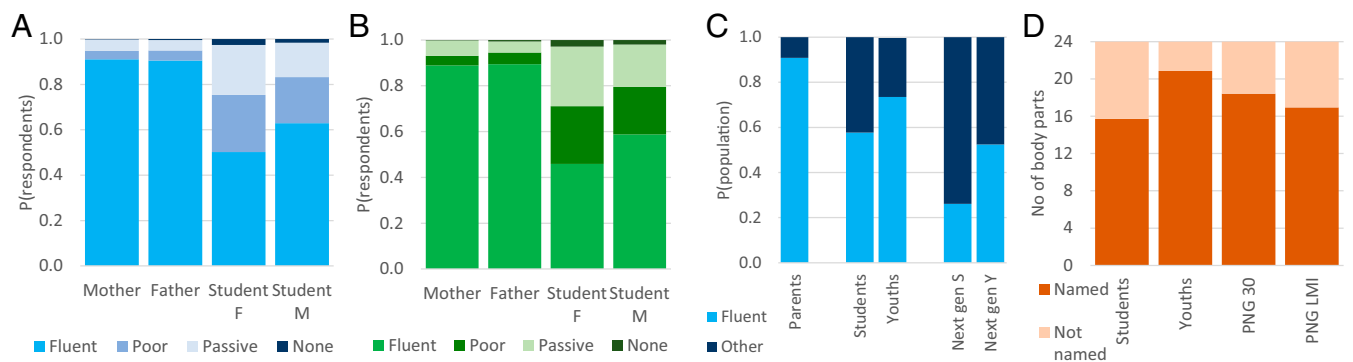


Fig. 2. Indigenous language skills in present and future PNG populations. (A) Language skills (L2) of 6,190 students (female and male) and their parents. (B) Mean language skills (L2) for 110 well-sampled languages ($n \geq 10$ students per language). (C) The proportion of fluent speakers among parents and students extrapolated to the entire 18- to 20-y-old cohort in PNG (Youths) and to the next generation of students (Next gen S) and all 18- to 20-y-olds (Next gen Y). (D) Language skills (L1) of the students, predictions from models characterizing the 18- to 20-y-olds in PNG at present (Youths) and in 30 y (PNG 30), and language skills assuming that PNG will come to match the mean socioeconomic parameters of lower-middle income countries (PNG LMI). Language skills were quantified as the number of body parts (from the total of 24) correctly named from photographs (L1) or by assessment by respondents for themselves and their parents on a four-point scale: no language skills (0), passive understanding (1), speaking but poorly (2), or fluent use (3) (L2).

important within both rural and urban environments, apart from a large decline in traditional skills and improvement in contemporary technical skills associated with transition from rural to urban lifestyle. We did not survey changes in traditional skills between students and their parents, but it is likely that good farming skills, in particular, are almost universal among the parents compared to 68% for the students. Interestingly, the students' English skills and mathematical skills had no effect on language skills, showing the limited direct effect of formal school education compared to lifestyle changes. Finally, language skills did not differ between female and male students.

The EGIDS (Expanded Graded Intergenerational Disruption Scale) (31) language endangerment classification, based on intergenerational transfer of languages and social domains of their use, is a significant predictor of language skills. Our study thus validates this endangerment parameter. Unlike some other measures of endangerment (2, 8), EGIDS does not consider the number of speakers of a language. We tested language size separately and found it had no significant effect on language skills; this finding bodes well for the survival prospects of numerous small languages in PNG (in 2000, the median language had only 1,201 speakers; Dataset S1).

In PNG, 87% of languages have a writing system, but only 15% of those languages have even a limited dictionary (1). Literature is thought to promote language vitality (31), but the existence of a Bible translation, typically the only written text in indigenous languages of PNG, did not improve language skills for the 84% of students who speak indigenous languages with Bible translations. This result could reflect the fact that only one-third of Bible translations are extensively used (32). The students' language skills also differ across geographic regions of PNG, probably reflecting regional differences in environmental or socioeconomic factors not directly captured by the analysis.

While many of the language-attribution drivers we detected have been documented previously (8, 10, 33), our analysis quantified their relative importance and revealed that multiple factors, even when correlated, have significant, statistically independent effects. For instance, urban lifestyle was correlated with better education and salaried employment of parents and with low traditional and high contemporary technical skills of students, but all these variables remained significant, independent predictors of language skills (SI Appendix, Fig. S2).

Future Trends in Language Skills. Only 15% of young people in PNG attend secondary school (34). They tend to come from

towns and cities (35% in our sample versus 13% in the general population), have educated parents (17% with tertiary education versus 5% country wide in the age cohort 45 to 54 y), and rely less on subsistence agriculture (31% versus 57% in the general population) (34) (SI Appendix, Table S1). Rural families can often afford education beyond primary level for only one or a minority of their children.

Considering these selection biases, we estimated indigenous language fluency for the entire 18- to 20-y-old cohort of PNG using country-wide values for urbanization, parents' education, and parents' employment as independent variables. We also used these variables to estimate the country-wide proportion of linguistically mixed families and the proportion of households using any indigenous language. Our model estimated that 73.5% of the 18- to 20-y-olds in PNG are fluent in an indigenous language—a higher proportion than among secondary students but representing a significant decline from their parents, of whom >90% are likely fluent in at least one indigenous language (Fig. 2C). While the share of fluent speakers decreased dramatically from parents to their children, the PNG population almost doubled during the same period from 4.62 million in 1990 to 8.95 million at present (29). It is predicted to grow further to 27 million in 2100 (35). The absolute number of fluent speakers thus probably increased in the past 30 y for most indigenous languages in PNG and may continue to grow in the future while representing a rapidly diminishing share of the total population. Such an increasingly minor position may be detrimental for the survival of indigenous languages irrespective of the number of speakers.

We used extrapolated values of language-skills drivers to model the situation for students and all 18- to 20-y-olds in the next generation. Unlike most other countries, PNG is predicted to remain predominantly (76%) rural in 2050 (36). Higher mobility, including travel for education and employment, will likely lead to an increase in the already high proportion (37%) of linguistically mixed families; a hypothetical random selection of partners would result in 99% of mixed families based on our population size estimates for PNG languages (Dataset S1). We used the proportion of students whose best friend speaks a different first language (65%) as a proxy for the future share of mixed-language families. The share of the population with secondary or tertiary education is expected to increase from 19 to 31% by 2050 (34), but the proportion of the population with salaried employment was modeled as constant (31%), since there has not been a definitive trend over the past 30 y (37).

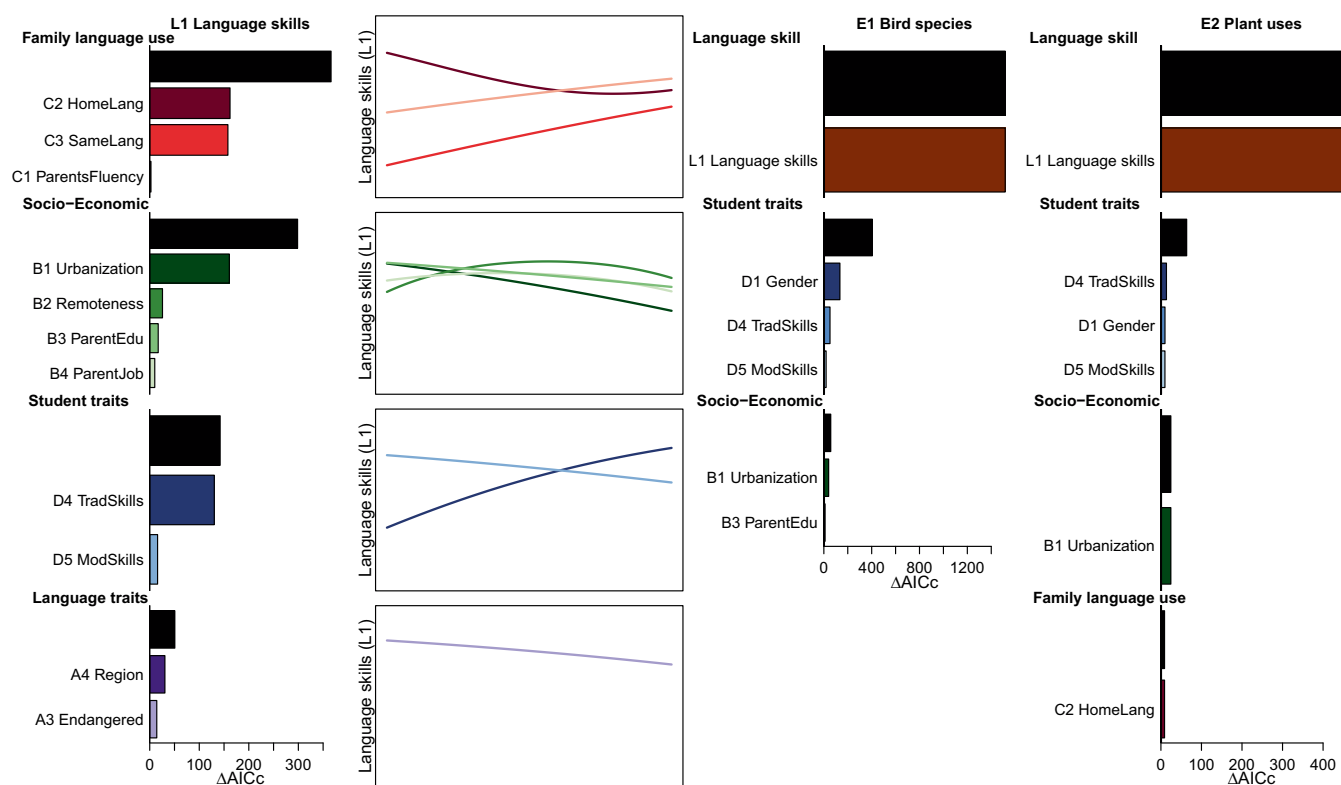


Fig. 3. Effects of language and socioeconomic factors on indigenous language skills and ethnobiological knowledge. GLMMs describe variability in student language skills (L1) and in knowledge of bird species (E1) and traditional plant uses (E2). The language-skills model incorporated 11 fixed variables divided into four classes: (A) language traits, endangerment (A3) and geographic regions (A4); (B) socioeconomic traits, birthplace urbanization (B1) and remoteness (B2) and parents' education (B3) and employment (B4); (C) family language use, parents' language skills (C1), home language use (C2), and whether parents speak the same first language (C3); and (D) student traits: traditional skills (hunting, fishing, growing food, house building, and medicinal plants) (D4) and contemporary technical skills (mobile phone and computer use) (D5). The variables were selected within each class (*SI Appendix, Table S2*) before being included in a global model (*SI Appendix, Table S3*). The bars show the AIC improvement due to the addition of each group (black) and each variable within each group into a model that includes all other variables, quantifying the marginal effect of each class/variable. The line plots show the shape of the effect of each variable across its range (except categorical A4) while keeping the other variables constant. Only significant ($P < 0.05$) variables are shown. The models describing variability in student knowledge of bird species (E1) and traditional plant uses (E2) used language skills (L1) and three classes of explanatory variables (family language use, socioeconomic traits, and student traits, including D1—gender) (*SI Appendix, Tables S5 and S6*). L1 is defined in Fig. 2 and other variables in *Materials and Methods*.

Our model predicted that the current students' 58% fluency in indigenous languages will shrink to 26% for the next generation. Furthermore, we estimated 52% fluent indigenous-language speakers in the entire 18- to 20-y-old cohort of the next generation in PNG (Fig. 2C). We also modeled the scenario of PNG converging to the mean socioeconomic parameters for lower-middle income countries, and this model predicted even greater attrition in language skills in the general population (Fig. 2D).

Ethnobiological Knowledge in Decline. We tested the knowledge of indigenous bird species and traditional uses of plants as two important components of biocultural knowledge (20). The knowledge of both bird species and plant uses was closely predicted by indigenous language skills and, therefore, is in decline (Fig. 3 and *SI Appendix, Fig. S3*). This result was expected, as most indigenous plant and animal names lack established translations into Tok Pisin or English and scientific species identifications (19). The continued maintenance of traditional knowledge in the face of severe language loss is rare, and this knowledge may be lost or restructured even when the indigenous language remains healthy (38, 39). Language shift, together with formal education, transition to a market economy, new technologies, urbanization, interethnic contact, habitat degradation,

modern health care, religious belief, change in values, and modern media have been identified as global drivers of decline in ethnobiological knowledge and its replacement or fusion with new information from external sources (22, 38, 39).

Male students knew birds better than female students, probably because the knowledge of birds was correlated with hunting skills, which were better developed in male students. Several other student and socioeconomic traits were correlated with ethnobiological knowledge, but their importance was low (Fig. 3). The close correlation between language skills and ethnobiological knowledge may result partly from the focus of our ethnobiology tests on naming species. However, the ability to recognize and name species is a prerequisite for acquiring deeper ecological and cultural knowledge of plants and animals, as we have observed when training paraecologists, who use their traditional knowledge of the natural world to build modern research skills (40).

Student traits, including traditional skills and socioeconomic traits, particularly urbanization, were the best predictors of ethnobiological knowledge when language skill itself was not used as an explanatory variable (*SI Appendix, Fig. S4*). The intricate details of biology are often learned during teenage years spent in rainforests (41), an option no longer available to many students growing up in towns or leaving villages for boarding

schools. Even the iconic and culturally important cassowary (*Casuarinus* spp.) (42) could be named in an indigenous language by only 64% of respondents.

The students were asked to list up to 10 plant species with their traditional uses in indigenous languages; when they did not know any, they used Tok Pisin or English names. The majority of the plant uses reported in indigenous languages and in Tok Pisin/English were medicinal, but the proportion was greater in Tok Pisin/English responses, in which 80% were medicinal versus just 53% for plants reported in indigenous languages. Although medicinal use is often one of the most salient across cultures (20), plants reported in indigenous languages had a wide range of reported uses including sorcery, house building, and ceremonies (Fig. 4A) (43). Furthermore, the Tok Pisin/English medicinal uses were dominated by merely 10 plant species, only two of them (*Laportea* sp. and *Morinda citrifolia*) native to PNG (Fig. 4B). *Laportea* is widely distributed and used across PNG, while *M. citrifolia* is a lowland species that has become commercialized throughout the Pacific (44). Students with poor indigenous language skills thus showed severely reduced traditional medicinal knowledge replaced by an impoverished, highly “globalized” knowledge pertaining to a few mostly non-native plant species (e.g., *Carica papaya*, *Citrus* spp., and *Aloe vera*).

Conclusions

We have shown that the drivers of language loss documented for communities around the world (45) are, to variable extents, at play in the world’s most linguistically diverse nation. The traditional multilingualism in indigenous languages in the present oldest generation has given way to bilingualism with the English-based creole Tok Pisin in an intermediate generation and monolingualism in Tok Pisin, with perhaps English from schooling, in a third generation (46). With Greenberg’s language diversity index (the probability that an individual does not share the same language with another randomly selected individual) approaching 0.989 (Dataset S1), the languages of PNG are too localized to be practical for wider communication. Unfortunately, we have shown that ethnobiological knowledge is closely correlated with indigenous language skills and therefore equally at risk.

The factors predicting language and biocultural-knowledge attrition in our models are determined by the factors considered desirable in contemporary PNG society, such as education, cash economy, ease of travel, and skills demanded for employment, or they are a consequence of economic development such as urbanization, which also leads to mixed-language marriages. These powerful forces are making the preservation of traditional knowledge difficult. In 2013, PNG abandoned a decades-long experiment in allowing local communities to deliver early childhood education in local indigenous languages by moving to an English-only plan (47). Furthermore, children often leave their home village to pursue education, which can cause attrition of their indigenous language skills (48). PNG’s extraordinary linguistic diversity and overwhelmingly rural population pose a challenge for state-delivered education but have played an important role in the retention of vast biocultural knowledge that exists outside the education system.

The survival of most indigenous languages and traditional knowledge will be determined by factors other than their practicality. On a positive note, PNG communities prize language as a marker of group identity (24). A majority, 88%, of the students fluent in an indigenous language expressed their intention to teach it to their children, but only 8% were motivated by practicality for communication, while the others valued language as an important part of their culture. It is possible that biocultural knowledge is less consciously prized than language skills and therefore even more in danger of disappearing than indigenous languages (41).

New Guinea’s share of global linguistic diversity is more than twice as high as its share of biological diversity (5, 17). The nation’s linguistic and biological diversity continue to be extensively studied (13, 18) with some sustained efforts at protection (47, 49), but both local and international programs to document and support ethnobiological diversity remain limited (3, 21). A better synergy between traditional biological knowledge and formal biology such as grassroots paraecologist programs could reinvigorate the interest of indigenous communities in their ethnobiological heritage as well as in the preservation of linguistic and biological diversity (17, 21, 40, 41).

Materials and Methods

Language Skills and Ethnobiological Knowledge Variables. We surveyed students attending upper secondary school (grades 11 and 12) at 30 of the 123

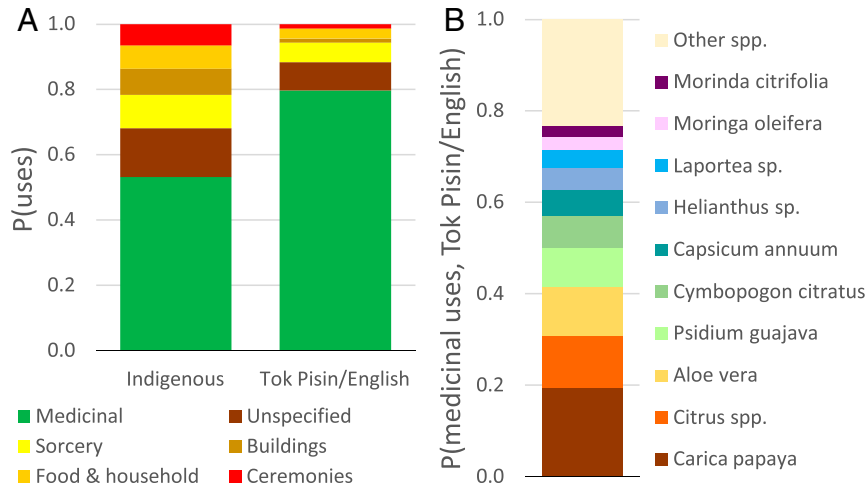


Fig. 4. Language skills and ethnobiological knowledge. (A) Indigenous plant use by categories (Hill’s diversity $^1D = 4.15$ for indigenous and 2.18 for Tok Pisin/English uses). (B) The 10 most common plant species listed in Tok Pisin/English medicinal uses. The respondents were asked to freely list up to 10 plant species with their indigenous names and traditional uses (E2). They provided 21,829 responses in indigenous languages (i.e., 35% of the maximum of 10 uses \times 6,190 respondents) and an additional 5,458 responses in Tok Pisin/English when they could not name any plant in an indigenous language.

621 secondary schools in PNG from April 6, 2015 to November 14, 2018. The
622 schools were selected to represent both rural and urban locations in the
623 lowland and highland regions from several provinces, comprising areas with
624 both low and high language diversity (SI Appendix, Fig. S5 and Table S2). The
625 students completed tests of indigenous language skills and ethnobiological
626 knowledge and a questionnaire on their family background, skills, and
627 lifestyle (Dataset S3). All surveys were voluntary, anonymous, and with in-
628 formed consent given by all participants. The surveys were approved by the
629 PNG Department of Education and approved by the IRB of the New Guinea
630 Binatang Research Center (BRC_03_15.01.2015). The surveys were conducted
631 at schools and attained 100% participation, eliminating the problem of self-
632 selection, whereby poor speakers may be reluctant to volunteer for
633 language tests (Fig. 1B).

634 Indigenous language skills were quantified by two variables: L1, the
635 number of body parts, from a set of 24 that included both frequently and
636 rarely used terms, named by students from photographs (50), and L2, the
637 students' self-assessment on a four-point scale: no language skills (0), passive
638 understanding (1), speaking but poorly (2), and fluent language use (3).
639 Students speaking more than one indigenous language were assessed for
640 the language they knew best. Language-skills measures based on self-
641 assessment (L2) can be biased (51). Individual respondents may have differ-
642 ent ideas about what it means to be "fluent." Younger people may
643 consider themselves less linguistically fluent than elders because they have
644 less cultural knowledge. However, these potential biases are unlikely to be
645 important, since the L1 and L2 variables are closely correlated in our study (SI
646 Appendix, Fig. S6).

647 Ethnobiological knowledge was quantified by two variables: E1, the
648 number of bird species named in an indigenous language from a set of
649 images of 10 species, and E2, the number of plant species freely listed with
650 their indigenous names and traditional uses other than food (10 species
651 maximum). For birds, the students completed two sets including 10 lowland
652 and 10 montane species, respectively, all geographically widespread. Each
653 selection included a range of species from widely known and easily recogniz-
654 able ones (e.g., birds of paradise and cassowary species) to more difficult
655 ones. The set with the higher score was used for each student so as not to
656 penalize students from any geographic location. We combined the image
657 identification for birds with free listing for plants in order to obtain more
658 comprehensive ethnobiological information, as each method of data col-
659 lection has its own strengths and biases (20, 52). The ethnobiological
660 knowledge measures focused on indigenous species names for birds and
661 plants because knowledge of these names is a prerequisite for learning
662 traditional information associated with individual species. Tok Pisin does not
663 have detailed animal or plant taxonomies, and those available in English are
664 not widely used in PNG. For plants, some students listed species by their Tok
665 Pisin or English names only, when they did not know their indigenous
666 names. These data were analyzed separately. Our tests, limited to 10 bird
667 and 10 plant species, did not explore the full scope of ethnobiological
668 knowledge, which often includes several hundred species (20, 43). With their
669 focus on students, they were also not designed to capture improvements of
670 knowledge with age that often take place for people who are immersed in
671 the relevant cultural and natural environment (53).

672 We used 21 independent variables (details in SI Appendix) to explain
673 language skills and ethnobiological knowledge, categorized into four
674 classes.

675 Class A includes language traits. For language population size (A1), we
676 estimated the number of language users by interpolating or extrapolating
677 the number listed in the *Ethnologue* database (1) to the year 2000 (Dataset
678 S1). For language status (A2 and A3), we used either detailed EGIDS cat-
679 egories (A2) as given for each language in *Ethnologue* (1) or the language
680 status (A3) classified as endangered (EGIDS 6b to 10) or not (EGIDS 1 to 6a).
681 For geographic region (A4), the location of the language in one of the four
682 geographic and administrative regions of PNG (Highlands [1], Momase [2],
683 Southern [3], and Islands [4]) is used as a categorical variable to examine
684 geographic differences in language skills. For elevation (A5), each language
685 was characterized by its median elevation (in meters above sea level, log
686 transformed) obtained from the *Ethnologue* (1) language maps. Concerning
687 Bible translation (A6), Bible translations are typically the only written liter-
688 ature in indigenous languages that are used by their speakers.

689 Class B includes family socioeconomic traits. For urbanization (B1), the
690 student's childhood place of residence is given as a village (1), a government
691 outpost (2), or a town or city (3). For remoteness (B2), the student's child-
692 hood place of residence can be accessed by road (1), boat (no road) (2), plane
693 (no road or boat) (3), or only on foot (4). For parents' education (B3), the
694 highest education reached by either of the parents is given as no school (1),
695 lower primary (first to sixth) grade (2), higher primary (seventh to eighth)

685 grade (3), lower secondary (9th to 10th) grade (4), higher secondary (11th to
686 12th) grade (5), or any tertiary education (6). For parents' employment (B4),
687 the highest employment category reached by either of the parents is given
688 as subsistence farming (1), cash crop farming (2), or salaried job or small
689 business (3).

690 Class C includes family language use. For parents' language fluency (C1),
691 the L2 scores were assessed by the respondents for their parents; the higher
692 of the mothers' and fathers' scores was used. Home language use (C2) was
693 given as indigenous language (alone or with other languages, including Tok
694 Pisin and English) (1), Tok Pisin only (2), or English (alone or with Tok Pisin)
695 (3). The parents' languages variable (C3) was given as mother and father
696 speak the same indigenous language (1) or the family is linguistically
697 mixed (0).

698 Class D includes student traits. Gender (D1) is given as female (1) or male
699 (0). Grade 10 test results are given from English (D2) and Mathematics (D3).
700 For traditional and contemporary technical skills in variables D4 to D6, stu-
701 dents self-assessed their skills (as none [0], poor [1], or good [2]) at five
702 traditional tasks (D4), hunting, fishing, growing staple crops, building a
703 house from forest materials, and using plants to treat fever, as well as at two
704 contemporary technical tasks (D5), using a mobile phone and a computer.
705 The difference between traditional and contemporary technical skills was
706 used as an additional explanatory variable (D6).

707 For the best friend's language variable (D7), the participant's best friend
708 speaks the same (1) or a different (0) indigenous language as the informant.
709 This variable was used as a proxy for the proportion of mixed-language
710 families likely to be formed by the surveyed students in the future (C3),
711 viewed as a predictor of language skills for the generation of the students'
712 children. We compared the D7 values for each surveyed school with the
713 expected proportion of best friends speaking the same language as the in-
714 formant, with the assumption that students choose their friends at school
715 and do so irrespective of the indigenous language they speak. The teaching
716 indigenous language variable (D8) was given as the intention to teach one's
717 children an indigenous language (yes [1] or no [0]), from those who have the
718 language skills to do so, with a predefined list of five motivations to justify
719 this choice: no, because 1) the indigenous language belongs to an old cul-
720 ture or 2) it is not a useful skill for my child or yes, because 1) everyone in my
721 village/town does it, 2) it is a useful skill for my children, or 3) it is part of my
722 culture. This variable was not used for generalized linear mixed
723 models (GLMMs).

724 **Data Verification.** Identification of the indigenous language used by each
725 respondent was often difficult (details in SI Appendix). We were able to
726 identify the indigenous language for 6,190 of 8,708 respondents. Both the
727 complete and the verified data sets give similar results for the language skills
728 (L1, L2) and ethnobiological knowledge (E1, E2); only verified data were
729 used in the analysis. We also verified the body part test results, as detailed in
730 the SI Appendix.

731 **Language Skills and Ethnobiological Knowledge Analysis.** The data used for
732 analysis are provided in Dataset S2. We used GLMMs to assess the effect of
733 the four classes of variables on the language skills of the students. The re-
734 sponse variable was the number of correct/incorrect body parts identified by
735 students in their indigenous language (L1). The probability of getting correct
736 responses was modeled as a binomial variable, with students and individual
737 languages treated as random variables in all models. Except for A4, all other
738 potential predictors are either binary or ordinal variables, allowing us to
739 model these variables as numeric (and A1 as natively numeric), with or-
740 thogonal polynomials of order $N - 1$ representing the number of levels in
741 each variable. This approach is equivalent to representing contrasts in a
742 categorical variable but allows for numeric extrapolation of noninteger
743 values.

744 We employed a hierarchical model selection approach using the Akaike
745 Information Criterion (AIC) (54) to compare the fit among candidate models.
746 First, we used model selection separately for each class (A through D) of
747 predictor variables. For the variables that had more than one level (except
748 categorical A4), we built models with polynomials of different order, from
749 $N - 1$ levels to a simple linear relationship. For each class, we considered all
750 the variable combinations within that class, using different polynomial or-
751 ders when applicable. When more than one variable represented alternative
752 expressions of the same factor (A2 versus A3, D6 versus D4 and D5), we ex-
753 cluded models that included these variables together. In order to make in-
754 terpretations easier, avoid inflation in the number of candidate models, and
755 limit degrees of freedom in the models, we did not consider interactions
756 among the variables. In the end, we obtained, for each class, one best-
757 performing model with the optimal set of variables belonging to that

807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826

834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868

class (*SI Appendix, Table S2*). Subsequently, we combined the variables from each of these class-specific, best-performing models to test whether different classes of variables acted jointly on language skills. We built these models by combining all the variables from the best-performing model in each class into models with two, three, and all four classes in all possible combinations, again with no interactions (*SI Appendix, Table S3*).

Once we obtained the best-performing, overall model, we investigated the relative role of each class and individual variable by calculating how much the AIC value was increased by removing the focal variable or class from the full model. In case any variable came out with a nonsignificant marginal effect (using a threshold of 2 points of AIC) at this stage, we removed it from the final model, as such loss of effect would be due to a higher predictive power in other correlated variables from another class. In addition, we assessed the direction and shape of the effects for individual predictor variables (Fig. 3). We predicted the response variable while varying each of the predictor variables in the best-performing model across its range, while keeping the other predictors at their original mean values across the whole population of test scores. This procedure was not possible for geographic regions (A4), which is categorical. We kept the A4 values fixed on the most abundant region when predicting the effect of the other variables because an average does not apply to this categorical variable.

We also used the GLMMs to analyze the ethnobiological knowledge of students (E1 and E2). We used the same variables and model-building strategy as for L1, except that we omitted the Language traits (A) class of variables and added language skills (L1) as a new independent variable.

Language-Skills Extrapolation. The empirical relationships among the predictor variables included in the best model were used to extrapolate language skills (L1) for hypothetical populations characterized by values for the model parameters different from the observed parameters: for the 18- to 20-y-olds in PNG at present (Model L1A), characterized by parameters extrapolated 30 y into the future (Model L1B) and assuming PNG reaches the current mean socioeconomic parameters for the lower-middle income countries (55) (Model L1C). We used the estimated parameters for the effects of each variable from the best model described above to predict the response variable for different values of the predictor variables in these populations. Because variable A4 is categorical, we made separate predictions for each geographic region and then made an average prediction by weighting the predicted value for each region by its proportion in the overall population.

1. D. M. Eberhard, G. F. Simons, C. D. Fennig, Eds., *Ethnologue: Languages of the World* (SIL International, ed. 23, 2020).
2. W. J. Sutherland, Parallel extinction risk and global distribution of languages and species. *Nature* **423**, 276–279 (2003).
3. R. Cámara-Leret, Z. Dennehy, Indigenous knowledge of new Guinea's useful plants: A review. *Econ. Bot.* **73**, 405–415 (2019).
4. L. Maffi, Linguistic, cultural, and biological diversity. *Annu. Rev. Anthropol.* **29**, 599–617 (2005).
5. V. Novotny, P. Droz, The size distribution of conspecific populations: The peoples of New Guinea. *Proc. Biol. Sci.* **267**, 947–952 (2000).
6. L. Campbell, E. Okura, "New knowledge produced by the *Catalogue of Endangered Languages*" in *Cataloguing the World's Endangered Languages*, L. Campbell, A. Belew, Eds. (Routledge, 2018), pp. 79–84.
7. M. L. Landweer, Methods of language endangerment research: A perspective from Melanesia. *Int. J. Soc. Lang.* **214**, 153–178 (2012).
8. T. Amano et al., Global distribution and drivers of language extinction risk. *Proc. Biol. Sci.* **281**, 20141574 (2014).
9. K. Prochazka, G. Vogl, Quantifying the driving factors for language shift in a bilingual region. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 4365–4369 (2017).
10. L. Bromham, X. Hua, C. Algy, F. Meakins, Language endangerment: A multidimensional analysis of risk factors. *J. Lang. Evol.* **5**, 75–91 (2020).
11. S. Aswani, A. Lemahieu, W. H. H. Sauer, Global trends of local ecological knowledge and future implications. *PLoS One* **13**, e0195440 (2018).
12. T. Furusava, Changing ethnobotanical knowledge of the Roviana people, Solomon Islands: Quantitative approaches to its correlation with modernization. *Hum. Ecol.* **37**, 147–159 (2009).
13. B. Palmer, *The Languages and Linguistics of the New Guinea Area: A Comprehensive Guide* (De Gruyter Mouton, 2017).
14. H. Hammarström, R. Forkel, M. Haspelmath, S. Bank, Glottolog (Version 4.2.1, Max Planck Institute for the Science of Human History, Jena, Germany, 2020). glottolog.org.
15. J. Vaughan, R. Singer, Indigenous multilingualisms past and present. *Lang. Commun.* **62**, 83–90 (2018).
16. S. A. Wurm, "Threatened languages in the Western Pacific area from Taiwan to, and including, Papua New Guinea" in *Language Diversity Endangered*, M. Brenzinger, Ed. (Mouton de Gruyter, 2007), pp. 374–390.

In Model L1A, we used parameter values characterizing 18- to 20-y-olds in PNG for urbanization (B1), parents' education (B3), parents' employment (B4), geographic region (A4), and language status (A3) (1, 29, 34, 37) (*SI Appendix, Table S4*). These variables were also used to adjust the remaining variables for which PNG-wide data were not available. For instance, there are no country-wide data for remoteness (B2), but the distribution of remoteness values differs between two levels of urbanization—village and town/city. The remoteness variable was therefore adjusted as a weighted mean between village and town/city values, reflecting the change in the share of village residents from 58% in the original data to 87% in Model L1A. More complex adjustments included several explanatory variables, using distributions of the adjusted variable for all possible combinations of their values. In Model L1A, in addition to remoteness adjusted by urbanization, we adjusted four variables (C2, C3, D4, and D5) by a combination of urbanization and parental education (*SI Appendix, Table S4*). The parameters for the L1B and L1C models extracted from literature (1, 34, 36, 37, 56) and adjusted using other variables appear in *SI Appendix, Table S4*. Finally, we adjusted the proportion of respondents fluent in an indigenous language (L2 = 3) from the student respondents in our study to the entire 18- to 20-y-old cohort in PNG (Model L2A) and used this approach to estimate the proportion of fluent speakers in the next generation, both for students (Model L2B) and the entire 18- to 20-y-old cohort in PNG (Model L2C) (*SI Appendix, Table S4*).

Data Availability. All data used for the analysis are included in the article, the *SI Appendix*, and *Datasets S1* and *S2*.

ACKNOWLEDGMENTS. We thank students, teachers, and heads of the 30 secondary schools that participated in the study; the PNG Department of Education for support; the PNG National Research Institute, particularly G. Kaipu, for permitting the research; and the New Guinea Binatang Research Center for logistical support. G. Luke, M. Sisol, M. Kigl, and E. Messimato provided information on individual languages; K. Kama assisted data collection; and J. Leps advised on statistical analysis. C. Volker commented on the manuscript. The New Guinea Binatang Research Center has provided the ethics oversight of the study, including data collection. The study was supported by the Czech Science Foundation (18-23889S), The European Research Council (669609), The Christensen Fund (2017-10237), and The Darwin Initiative for the Survival of the Species (DIR2551100123).

17. V. Novotny, K. Molem, An inventory of plants for the land of the unexpected. *Nature* **584**, 531–533 (2020).
18. V. Novotny, P. Toko, "Ecological research in Papua New Guinean rainforests: Insects, plants and people" in *The State of the Forests of Papua New Guinea 2015. Measuring change over period 2002-2014*, J. E. Bryan, P. L. Shearman, Eds. (University of Papua New Guinea, 2015), pp. 71–85.
19. P. Mühlhäusler, "The development of the life form lexicon in Tok Pisin" in *Degrees of Restructuring in Creole Languages*, I. Neumann-Holzschuh, E. W. Schneider, Eds. (J. Benjamins, 2000), pp. 337–360.
20. B. Berlin, *Ethnobiological classification. Principles of Classification of Plants and Animals in Traditional Societies* (Princeton University, 1992).
21. L. Maffi, E. Woodley, *Biocultural Diversity Conservation. A Global Sourcebook* (Earthscan, 2010).
22. S. Zent, "Processual perspectives on traditional environmental knowledge. Continuity, erosion, transformation, innovation" in *Understanding Cultural Transmission in Anthropology: A Critical Synthesis*, R. Ellen, S. J. Lycett, S. E. Johns, Eds. (Berghahn Books, 2013), pp. 213–265.
23. D. A. Kulick, *Death in the Rainforest. How a Language and a Way of Life Came to an End on Papua New Guinea* (Algonquin Books, 2019).
24. K. M. Sumbuk, "Papua New Guinea's languages: Will they survive?" in *Language Diversity in the Pacific: Endangerment and Survival*, D. D. Cunningham, P. D. E. Ingram, P. K. Sumbuk, Eds. (Multilingual Matters, 2006), pp. 85–96.
25. A. Y. Aikhenvald, Language contact along the Sepik River. *Anthropol. Linguist.* **50**, 1–66 (2008).
26. F. Lüpke, Uncovering small-scale multilingualism. *Crit. Multilingualism Stud.* **4**, 35–74 (2016).
27. A. Y. Aikhenvald, "Language endangerment in the Sepik area of Papua New Guinea" in *Lectures on Endangered Languages: 5. Endangered Languages of the Pacific Rim*, O. Sakiyama, F. Endo, Eds. (ELPR, 2004), pp. 97–142.
28. S. Romaine, *Language, Education, and Development: Urban and Rural Tok Pisin in Papua New Guinea* (Clarendon Press, 1992).
29. Worldometers, Papua New Guinea Population. <https://www.worldometers.info/world-population/papua-new-guinea-population/>. Accessed 5 March 2020.
30. G. Koczberski, G. N. Curry, B. Imbun, Property rights for social inclusion: Migrant strategies for securing land and livelihoods in Papua New Guinea. *Asia Pac. Viewp.* **50**, 29–42 (2009).

869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930

31. M. P. Lewis, G. F. Simons, Assessing endangerment: Expanding Fishman's GIDS. *Rev. Roum. Linguist.* **55**, 103–120 (2010).
32. R. van den Berg, *Scripture Use Research and Ministry* (SIL International, 2020).
33. P. K. Austin, J. Sallabank, "Introduction" in *Cambridge Handbook of Endangered Languages*, P. K. Austin, J. Sallabank, Eds. (Cambridge University, 2011), pp. 3–24.
34. National Statistical Office (NSO), *Papua New Guinea demographic and health survey 2016-18* (NSO, Port Moresby, Papua New Guinea, 2019).
35. S. E. Vollset et al., Fertility, mortality, migration, and population scenarios for 195 countries and territories from 2017 to 2100: A forecasting analysis for the Global Burden of Disease study. *Lancet* **396**, 1285–1306 (2020).
36. United Nations, *World Urbanization Prospects: The 2018 Revision* (United Nations, 2019).
37. National Statistical Office (NSO), *2009-2010 Papua New Guinea Household Income and Expenditure Survey* (NSO, Port Moresby, Papua New Guinea, 2011).
38. A. Si, Patterns in the transmission of traditional ecological knowledge: A case study from Arnhem Land, Australia. *J. Ethnobiol. Ethnomed.* **16**, 52 (2020).
39. E. S. Hunn, *A Zapotec Natural History: Trees, Herbs, and Flowers, Birds, Beasts, and Bugs in the Life of San Juan Gbëë*; (Arizona University Press, 2008).
40. V. Novotny, G. D. Weiblen, S. E. Miller, Y. Basset, "The role of paraecologists in 21st century tropical forest research" in *Methods in Forest Canopy Research*, M. D. Lowman, T. D. Schowalter, J. F. Franklin, Eds. (University of California, 2012), pp. 154–157.
41. I. S. Majnep, R. Bulmer, *Birds of My Kalam Country* (Auckland University & Oxford University, 1977).
42. R. Bulmer, Why is the cassowary not a bird? A problem of zoological taxonomy among the Karam of the New Guinea highlands. *Man (Lond.)* **2**, 5–25 (1967).
43. R. O. Gardner, Plant names of the Kalam (Upper Kaironk Valley, Schrader Range, Papua New Guinea). *Rec. Auckland Mus.* **47**, 5–50 (2010).
44. O. Potterat, M. Hamburger, *Morinda citrifolia* (Noni) fruit—Phytochemistry, pharmacology, safety. *Planta Med.* **73**, 191–199 (2007).
45. P. K. Austin, J. Sallabank, Eds., *The Cambridge Handbook of Endangered Languages* (Cambridge University, 2011).
46. A. Y. Aikhenvald, "Traditional multilingualism and language endangerment" in *Language Endangerment and Language Maintenance*, D. Bradley, M. Bradley, Eds. (Routledge, 2002), pp. 24–33.
47. P. Paraide, Challenges with the implementation of vernacular and bilingual education in Papua New Guinea. *Contemp. PNG Stud.: DWU Res. J.* **21**, 44–57 (2014).
48. B. Köpke, Neurolinguistic aspects of attrition. *J. Neurol.* **17**, 3–30 (2004).
49. V. M. Adams et al., Scheduling incremental actions to build a comprehensive national protected area network for Papua New Guinea. *Conserv. Sci. Pract.* **3**, e354 (2021).
50. W. O. O'Grady, A. J. Schafer, J. Perla, O.-S. Lee, J. Wieting, A psycholinguistic tool for the assessment of language loss: The HALA Project. *Lang. Doc. Conserv.* **3**, 100–112 (2009).
51. C. Grinevald, M. Bert, "Speakers and communities" in *Cambridge Handbook of Endangered Languages*, P. K. Austin, J. Sallabank, Eds. (Cambridge Univ., 2011), pp. 45–65.
52. V. Reyes-García, N. Martí, T. McDade, S. Tanner, V. Vadez, Concepts and methods in studies measuring individual ethnobotanical knowledge. *J. Ethnobiol.* **27**, 182–203 (2007).
53. J. Koster et al., The life history of human foraging: Cross-cultural and individual variation. *Sci. Adv.* **6**, eaax9070 (2020).
54. K. P. Burnham, D. R. Anderson, *Model Selection and Multimodel Inference. A Practical Information-Theoretic Approach* (Springer, ed. 2, 2002).
55. World Bank, Country Classification. <https://datahelpdesk.worldbank.org/knowledgebase/topics/19280-country-classification>. Accessed 25 March 2020.
56. World Bank, Data Catalog: Educational Statistics. <https://datacatalog.worldbank.org/dataset/education-statistics>. Accessed 25 March 2020.

931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992