

1 **Running Title:** SARS-CoV-2 rural seroprevalence, Peru

2 **Title:** High SARS-CoV-2 seroprevalence in rural Peru, 2021; a cross-sectional population-  
3 based study

4 **Authors:**

5 Andres Moreira-Soto<sup>1,2</sup>, Johanna Maribel Pachamora Diaz<sup>3</sup>, Lilian González-Auza<sup>1</sup>, Xiomara  
6 Jeanleny Merino Merino<sup>4,5</sup>, Alvaro Schwalb<sup>4,5</sup>, Christian Drosten<sup>1</sup>, Eduardo Gotuzzo<sup>4,5</sup>, Michael  
7 Talledo<sup>4,5</sup>, Heriberto Arévalo Ramirez<sup>3</sup>, Roxana Peralta Delgado<sup>3</sup>, Spassky Bocanegra Vargas<sup>6</sup>,  
8 Jan Felix Drexler<sup>#1,7</sup>

9 **Affiliations:**

10 <sup>1</sup>Charité-Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-  
11 Universität zu Berlin, and Berlin Institute of Health, Institute of Virology, Berlin, Germany.

12 <sup>2</sup>Virology-CIET, Faculty of Microbiology, University of Costa Rica, San José, Costa Rica

13 <sup>3</sup>Dirección Regional de Salud de San Martín, Perú

14 <sup>4</sup>Instituto de Medicina Tropical Alexander von Humboldt, Universidad Peruana Cayetano Heredia,  
15 Lima, Peru

16 <sup>5</sup>Departamento de Medicina, Facultad de Medicina, Universidad Peruana Cayetano Heredia, Lima,  
17 Peru

18 <sup>6</sup>Moyobamba Hospital, Perú

19 <sup>7</sup>German Centre for Infection Research (DZIF), associated partner site Charité, Berlin, Germany

20 **\*Correspondence:**

21 Jan Felix Drexler, Helmut-Ruska-Haus, Institute of Virology, Campus Charité Mitte, Charitéplatz

22 1, 10098 Berlin, Germany. Email: [felix.drexler@charite.de](mailto:felix.drexler@charite.de)

23

24 **Keywords:** SARS-CoV-2, San Martin, Peru, COVID-19, serology, seroprevalence

25

26 **Abstract word count:** 247

27 **Main text word count:** 1541

28 **Abstract**

29 Latin America has been severely affected by the COVID-19 pandemic. The COVID-19 burden in  
30 rural settings in Latin America is unclear. We performed a cross-sectional, population-based,  
31 random-selection SARS-CoV-2 serological study during March 2021 in the rural population of  
32 San Martin region, northern Peru. The study enrolled 563 persons from 288 houses across 10  
33 provinces, reaching 0.19% of the total rural population of San Martin. Screening for SARS-CoV-  
34 2 IgG antibodies was done using a chemiluminescence immunoassay (CLIA) and reactive sera  
35 were confirmed using a SARS-CoV-2 surrogate virus neutralization test (sVNT). Validation using  
36 pre-pandemic sera from two regions of Peru showed false-positive results in the CLIA (23/84 sera;  
37 27%), but not in the sVNT, highlighting the pitfalls of SARS-CoV-2 antibody testing in tropical  
38 regions and the high specificity of the two-step testing algorithm. An overall 59.0% seroprevalence  
39 (95% CI: 55-63%) corroborated intense SARS-CoV-2 spread in San Martin. Seroprevalence rates  
40 between the 10 provinces varied from 41.3-74.0% (95% CI: 30-84). Higher seroprevalence was  
41 neither associated with population size, population density, surface area, mean altitude or poverty

42 index in spearman correlations. Seroprevalence and reported incidence diverged substantially  
43 between provinces, suggesting regional biases of COVID-19 surveillance data. Potentially, limited  
44 healthcare access due to environmental, geographic, economic, and cultural factors, might lead to  
45 undetected infections in rural populations. Additionally, test avoidance to evade mandatory  
46 quarantine might affect rural regions more than urban regions. Serologic diagnostics should be  
47 pursued in resource-limited settings to inform country-level surveillance, vaccination strategies  
48 and support control measures for COVID-19.

49

## 50 **Importance**

51 Latin America is a global hotspot of the COVID-19 pandemic. Serological studies in Latin  
52 America have been mostly performed in urban settings. Rural populations comprise 20% of the  
53 total Latin American population. Nevertheless, information of COVID-19 spread and transmission  
54 in rural settings is scarce. Using a representative population-based seroprevalence study, we  
55 detected a high seroprevalence in rural populations in San Martin, northern Peru in 2021, reaching  
56 41 to 74 %. However, seroprevalence and reported incidence diverged substantially between  
57 regions, suggesting either limited healthcare access or test avoidance due to mandatory quarantine.  
58 Our results suggest that rural populations are highly affected by SARS-CoV-2 even though they  
59 are socio-demographically distinct from urban populations, and that highly specific serological  
60 diagnostics should be performed in resource-limited settings to support public-health strategies of  
61 COVID-19 surveillance and control.

62

## 63 **Background**

64 Peru has been severely affected by the COVID-19 pandemic, with the highest mortality per capita  
65 reported worldwide since the start of the pandemic, reaching 6,132 deaths per million as of October  
66 2021 (1, 2). The determinants of SARS-CoV-2 spread in Latin America are poorly defined. An  
67 epidemiological study from European and Asian urban centers correlated population density with  
68 increased SARS-CoV-2 infection rates (3). In contrast, epidemiological studies from Latin  
69 American urban centers yielded diverse prevalence estimates that seemed uncorrelated to  
70 population density, exemplified by three available seroprevalence studies from Peru, conducted  
71 during mid-late 2020. The first study from the Peruvian capital Lima (population: 9.5 million;  
72 density: 12,000/km<sup>2</sup>) reported a 21% seroprevalence rate (4). The second study from the  
73 Lambayeque department in northern Peru (population: 1.2 million; density: 84/km<sup>2</sup>) reported a  
74 29% seroprevalence rate (5). The third study from Iquitos city, Peruvian Amazon (population:  
75 470,000; density: 417/km<sup>2</sup>) reported a 70% seroprevalence rate, implying local herd immunity had  
76 been reached (6). Despite the high seroprevalence already during 2020, a second SARS-CoV-2  
77 wave occurred in Iquitos in January 2021. Potentially, mobile susceptible populations in the  
78 Amazon region might explain occurrence of a second wave (6). Hypothetically, susceptible  
79 populations might correspond to rural and indigenous populations, comprising  $\leq 19\%$  of the total  
80 Latin American population based on World Bank estimates  
81 (<https://data.worldbank.org/indicator/SP.RUR.TOTL.ZS?locations=ZJ>). The high geographical  
82 dispersion and low population density of these communities, in addition to regular travelling to  
83 urban centers for commerce, might contribute to a differential dispersion of SARS-CoV-2  
84 compared to urban settings.

## 85 **Seroprevalence study in San Martin**

86 Our study focused on the rural region of San Martin, Peru (total population: 899,648 density:  
87 18/km<sup>2</sup>). The San Martin department (51,000 km<sup>2</sup>) is located in the north of Peru (**Figure 1A**).  
88 The department encompasses diverse ecosystems such as valleys, Andes mountains and Amazon  
89 rainforests, reaching altitudes from 190 to 3,080 m. (**Figure 1A**) and comprises  $\leq 5\%$  of Peruvian  
90 forest cover. The San Martin population is divided into Mestizo (83.6%) followed by Quechua  
91 (5.3%) and Afro-Peruvian (4.9%) ethnicities as per census data of the National Institute of  
92 Statistics and Census of Peru (INEI; [www.inei.gob.pe/](http://www.inei.gob.pe/)).

93 We performed a cross-sectional population-based random-selection study during March 2021  
94 (institutional ethics committee Via Libre; number 6532-2021a) in the ten provinces of San Martin  
95 (**Figure 1B**). Sample size calculation considered the rural population of San Martin (286 988), a  
96 95% confidence level and an estimated SARS-CoV-2 prevalence of 50%, reaching 576  
97 individuals. As the average household in Peru consists of 4 inhabitants ([www.inei.gob.pe/](http://www.inei.gob.pe/)), the  
98 study aimed at sampling two individuals from 288 individual houses. First, a conglomerate of 100  
99 rural houses was selected based on estimates of the census data of 2017 and on geographic  
100 information systems of the INEI. Next, a random household selection in the conglomerate was  
101 performed. Exclusion criteria encompassed population residing in collective dwellings such as  
102 barracks, police stations, convents, boarding schools and hotels, used as temporary housing of  
103 unrelated people with unstable population dynamics, age  $\leq 5$  years, skin lesions in the venous  
104 puncture site, usage of alcohol or psychoactive drugs, not being permanent resident and not signing  
105 the informed consent. Participation was voluntary and participant selection in households was  
106 performed using a kish grid (7). A 10 mL blood sample was taken after the informed consent was  
107 signed by the person or the caretaker if the person had  $\leq 18$  years of age. We obtained samples  
108 from 563 persons visiting 288 houses in the 10 provinces, comprising 0.19% of the total rural

109 population of San Martin. The sampled cohort included persons aged 6-89 years (mean: 35.8,  
110 standard deviation (SD): 21.15) consistent with San Martin's age distribution data from the INEI  
111 2017 census (**Annex Figure 1**; [www.inei.gov.pe/](http://www.inei.gov.pe/)). From the cohort, 37.7% (212/563) were male  
112 and 62.3% (351/563) female, contrasting with the sex distribution in San Martin (51.3% male,  
113 48.7% female). The difference between the distribution can be explained by the sampling strategy,  
114 limited at visiting the homes of the participants during the day due to safety and operational  
115 reasons, while males were mostly out working.

## 116 **Two-step serological testing and validation**

117 Using a chemiluminescence immunoassay (CLIA) (SARS-CoV-2 S-RBD IgG kit, Snibe  
118 Diagnostic, China), we detected SARS-CoV-2-specific IgG antibodies in 63.6% of the samples  
119 (358/563) (**Figure 1C** and **Annex Figure 2**). To validate the CLIA results, we used 84 pre-  
120 COVID-19 samples from our previous study in arbovirus serology in Peru (8) (**Figure 1D** and  
121 **Annex Figure 2**). The samples were collected in the municipalities of Cusco (altitude 3,400 m.)  
122 and Loreto (altitude 150 m.) during 2013 and 2018, respectively. A total of 27% (23/84; 11 from  
123 Cuzco and 12 from Loreto) of pre-COVID-19 samples yielded positive results in the CLIA  
124 screening test irrespective of geographical location (**Figure 1D** and **Annex Figure 2**). The  
125 apparently positive pre-COVID-19 samples showed statistically significantly lower signal  
126 concentrations in the CLIA than those samples from 2021 (median pre-COVID-19= 1.6 vs  
127 COVID-19= 32.81; t-test:  $p < 0.001$ ) (**Annex Figure 2**), suggesting unspecific reactivity.  
128 Unspecific reactivity may be elicited by common-cold coronaviruses (9) or endemic tropical  
129 diseases such as Malaria, Dengue and Zika (10, 11). Therefore, a confirmatory SARS-CoV-2  
130 surrogate virus neutralization test (sVNT, GenScript, USA) was performed in all CLIA-reactive  
131 samples. A total of 92.7% (332/358) of the 2021 CLIA-positive samples were confirmed using the

132 sVNT which can be attributed to differential sensitivity of the tests. (**Figure 1C** and **Annex Figure**  
133 **2**). None of the CLIA-positive pre-pandemic samples yielded positive results in the sVNT. These  
134 results corroborate that unspecific reactivity of serological tests in tropical areas must be carefully  
135 evaluated (10) and highlight robustness of our serological testing algorithm. Therefore, only  
136 samples yielding positive results in both serological assays were considered for further analyses in  
137 our study to guarantee specificity.

### 138 **Seroprevalence and statistical analyses**

139 Overall, the seroprevalence for San Martin was 59.0% (332/563; 95% Confidence Interval (CI):  
140 55-63%). No statistically significant difference of seroprevalence per sex was observed, using a  
141 chi-square test ( $\chi^2=0.05$ ;  $p=0.83$ ). However, a statistically significant difference was observed  
142 between the age of SARS-CoV-2-seropositive versus SARS-CoV-2-seronegative persons using a  
143 t-test (33 years (range: 6-89); versus 38 years (range: 7-57;  $p=0.01$ )). This observation was in  
144 concordance to the Iquitos study, likely associated with higher contact rates in younger age groups  
145 (6).

146  
147 Seroprevalence between the 10 provinces varied from 41.3-74.0% (CI: 30-84%) (**Supplementary**  
148 **Table 1, Figure 2A** and **2B**). Higher seroprevalences were neither associated with population size  
149 (Spearman correlation test:  $r_s=-0.26$ ;  $p=0.46$ ), nor population density ( $r_s=-0.14$ ;  $p=0.68$ ), nor  
150 surface area ( $r_s=0.33$ ;  $p=0.34$ ), nor mean altitude of the province ( $r_s=-0.59$ ;  $p=0.06$ ), nor poverty  
151 index ( $r_s=0.10$ ;  $p=0.76$ ) (**Figure 2C** and **Annex Figure 3**). Interestingly, the cumulative  
152 epidemiological surveillance data from the Regional Health Directorate of Peru (DIRESA) relying  
153 on notified cases of acute SARS-CoV-2 infection confirmed either by RT-PCR or an antigen test

154 (Figure 2B), showed no correlation with our serological data although cumulatively measuring  
155 similar time spans, encompassing early 2020-March 2021 ( $r_s=-0.14$ ;  $p=0.68$ ) (Figure 1B and  
156 Figure 2D). On the one hand, overall higher seroprevalence than reported cases is consistent with  
157 high numbers of undetected asymptomatic COVID-19 cases (12). On the other hand, we detected  
158 two provinces with a relatively higher divergence of seroprevalence compared to incidence data  
159 (Mariscal Caceres and Bellavista, Figure 2A, 2B and 2D). Hypothetically, individuals in those  
160 provinces may lack healthcare access due to environmental, geographic, economic, and cultural  
161 factors (13), that might lead to more undetected infections in rural populations. First, test refusal  
162 to avoid mandatory quarantine or stigmatization has been reported for SARS-CoV-2 from different  
163 countries (14, 15). Test avoidance might be greater in rural regions from resource-limited areas,  
164 were a high share of the population work in jobs that do not allow quarantine for long periods of  
165 time and were lower incomes may force them to work and not visit or afford the hospital. Second,  
166 local health centers might have difficult access, or not be sufficiently equipped for the high  
167 COVID-19 burden (16), limiting access to diagnostics in rural areas. High infection rates in rural  
168 populations might be explained by economic, or cultural factors. Lack of access to basic housing  
169 elements such as sewage disposal system, as observed in rural populations in Ecuador may  
170 contribute to the high seroprevalence in rural Peru (17). A survey analyzing rural residents in the  
171 USA found that they were less likely to participate in COVID-19 preventive measures such as the  
172 use of masks in public or working from home, increasing the possibility of infection (18). Another  
173 province (Huallaga, Figure 2A, 2B and 2D) showed higher cumulative incidence than other  
174 provinces for unknown reasons, substantiating inconsistencies between serological and incidence  
175 data.  
176

177 Irrespective of the underlying reasons leading to the overall high seroprevalence detected, our data  
178 suggest a large number of undiagnosed COVID-19 cases potentially challenging test-trace-isolate  
179 interventions in the region (19). Previous studies in the USA have stressed that rural regions are  
180 particularly vulnerable to COVID-19, leading to higher mortality rates in rural than in urban  
181 regions, and with higher mortality rates associated with black and Hispanic populations (20, 21),  
182 suggesting further studies are needed in Latin American vulnerable populations. Limitations of  
183 our study include lack of metadata and potential sampling biases. However, thorough study design,  
184 exhaustive serological testing and high seroprevalence consistent with data from other Latin  
185 American settings suggest robustness of our results (6, 17). Serological diagnostics should be  
186 pursued in resource-limited settings to inform country-level surveillance, vaccination strategies  
187 and support control measures for COVID-19.

188

189

## 190 **Funding**

191 This work was supported by the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ)  
192 GmbH contract number 88114104.

193

## 194 **Acknowledgments**

195 We thank the DIRESA workers that aided in the sample collection and all study participants. We  
196 thank Victor Carvalho for technical support.

197

198 **Conflicts of interest**

199 The authors report no conflicts of interest.

200

201 **References**

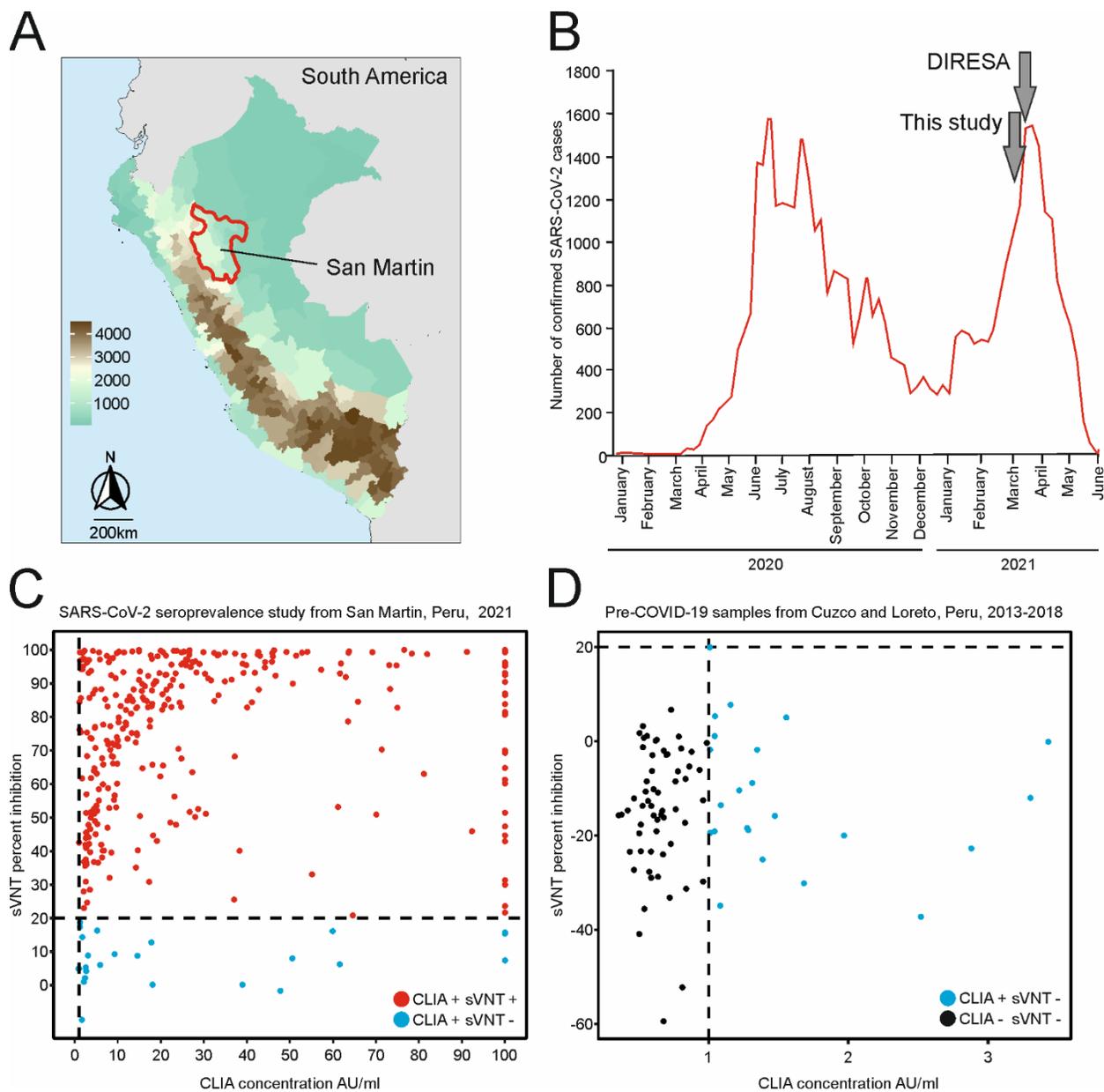
- 202 1. Munayco C, Chowell G, Tariq A, Undurraga EA, Mizumoto K. 2020. Risk of death by age and gender  
203 from CoVID-19 in Peru, March-May, 2020. *Aging (Albany NY)* 12:13869-13881.
- 204 2. Roser M. 2020. Coronavirus Pandemic (COVID-19). *Our World in Data*.
- 205 3. Diao Y, Koderá S, Anzai D, Gomez-Tames J, Rashed EA, Hirata A. 2021. Influence of population  
206 density, temperature, and absolute humidity on spread and decay durations of COVID-19: A  
207 comparative study of scenarios in China, England, Germany, and Japan. *One Health* 12:100203.
- 208 4. Reyes-Vega MF, Soto-Cabezas MG, Cárdenas F, Martel KS, Valle A, Valverde J, Vidal-Anzardo M,  
209 Falcón ME, Munayco CV. 2021. SARS-CoV-2 prevalence associated to low socioeconomic status  
210 and overcrowding in an LMIC megacity: A population-based seroepidemiological survey in Lima,  
211 Peru. *EClinicalMedicine* 34:30.
- 212 5. Díaz-Vélez C, Failoc-Rojas VE, Valladares-Garrido MJ, Colchado J, Carrera-Acosta L, Becerra M,  
213 Moreno Paico D, Ocampo-Salazar ET. 2021. SARS-CoV-2 seroprevalence study in Lambayeque,  
214 Peru. June-July 2020. *PeerJ* 9:e11210-e11210.
- 215 6. Alvarez-Antonio C, Meza-Sanchez G, Calampa C, Casanova W, Carey C, Alava F, Rodriguez-Ferrucci  
216 H, Quispe AM. 2021. Seroprevalence of anti-SARS-CoV-2 antibodies in Iquitos, Peru in July and  
217 August, 2020: a population-based study. *Lancet Glob Health* 9:e925-e931.
- 218 7. Kish L. 1949. A Procedure for Objective Respondent Selection Within the Household. *Journal of*  
219 *the American Statistical Association* 44:380–87.
- 220 8. Fischer C, Bozza F, Merino Merino XJ, Pedroso C, De Oliveira Filho EF, Moreira-Soto A, Schwalb A,  
221 De Lamballerie X, Netto EM, Bozza PT, Sarno M, Brites C, Gotuzzo E, Talledo M, Drexler JF. 2020.  
222 Robustness of Serologic Investigations for Chikungunya and Mayaro Viruses following  
223 Coemergence. *mSphere* 5.
- 224 9. Hicks J, Klumpp-Thomas C, Kalish H, Shunmugavel A, Mehalko J, Denson J-P, Snead KR, Drew M,  
225 Corbett KS, Graham BS, Hall MD, Memoli MJ, Esposito D, Sadtler K. 2021. Serologic Cross-  
226 Reactivity of SARS-CoV-2 with Endemic and Seasonal Betacoronaviruses. *Journal of Clinical*  
227 *Immunology* 41:906-913.
- 228 10. Yadouleton A, Sander AL, Moreira-Soto A, Tchiboza C, Hounkanrin G, Badou Y, Fischer C, Krause  
229 N, Akogbeto P, de Oliveira Filho EF, Dossou A, Brunink S, Aissi MAJ, Djingarey MH, Hounkpatin B,  
230 Nagel M, Drexler JF. 2021. Limited Specificity of Serologic Tests for SARS-CoV-2 Antibody  
231 Detection, Benin. *Emerg Infect Dis* 27.
- 232 11. Lustig Y, Keler S, Kolodny R, Ben-Tal N, Atias-Varon D, Shlush E, Gerlic M, Munitz A, Doolman R,  
233 Asraf K, Shlush LI, Vivante A. 2020. Potential Antigenic Cross-reactivity Between Severe Acute  
234 Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) and Dengue Viruses. *Clinical Infectious*  
235 *Diseases* doi:10.1093/cid/ciaa1207.
- 236 12. McArthur L, Sakthivel D, Ataíde R, Chan F, Richards JS, Narh CA. 2020. Review of Burden, Clinical  
237 Definitions, and Management of COVID-19 Cases. *Am J Trop Med Hyg* 103:625-638.

- 238 13. Peters DJ. 2020. Community Susceptibility and Resiliency to COVID-19 Across the Rural-Urban  
239 Continuum in the United States. *J Rural Health* 36:446-456.
- 240 14. Rubin R. 2020. First It Was Masks; Now Some Refuse Testing for SARS-CoV-2. *JAMA* 324:2015-  
241 2016.
- 242 15. McDermott JH, Newman WG. 2020. Refusal of viral testing during the SARS-CoV-2 pandemic. *Clin*  
243 *Med (Lond)* 20:e163-e164.
- 244 16. Sander AL, Yadouleton A, Moreira-Soto A, Tchiboza C, Hounkanrin G, Badou Y, Fischer C, Krause  
245 N, Akogbeto P, E FdOF, Dossou A, Brunink S, Drosten C, Aissi MAJ, Harouna Djingarey M,  
246 Hounkpatin B, Nagel M, Drexler JF. 2021. An Observational Laboratory-Based Assessment of SARS-  
247 CoV-2 Molecular Diagnostics in Benin, Western Africa. *mSphere* 6.
- 248 17. Del Brutto OH, Costa AF, Mera RM, Recalde BY, Bustos JA, Garcia HH. 2020. SARS-CoV-2 in rural  
249 Latin America. A population-based study in coastal Ecuador. *Clin Infect Dis*  
250 doi:10.1093/cid/ciaa1055.
- 251 18. Callaghan T, Lueck JA, Trujillo KL, Ferdinand AO. 2021. Rural and Urban Differences in COVID-19  
252 Prevention Behaviors. *The Journal of Rural Health* 37:287-295.
- 253 19. Contreras S, Dehning J, Loidolt M, Zierenberg J, Spitzner FP, Urrea-Quintero JH, Mohr SB, Wilczek  
254 M, Wibral M, Priesemann V. 2021. The challenges of containing SARS-CoV-2 via test-trace-and-  
255 isolate. *Nat Commun* 12:378.
- 256 20. Cheng KJG, Sun Y, Monnat SM. 2020. COVID-19 Death Rates Are Higher in Rural Counties With  
257 Larger Shares of Blacks and Hispanics. *The Journal of Rural Health* 36:602-608.
- 258 21. Huang Q, Jackson S, Derakhshan S, Lee L, Pham E, Jackson A, Cutter SL. 2021. Urban-rural  
259 differences in COVID-19 exposures and outcomes in the South: A preliminary analysis of South  
260 Carolina. *PLOS ONE* 16:e0246548.

261

262

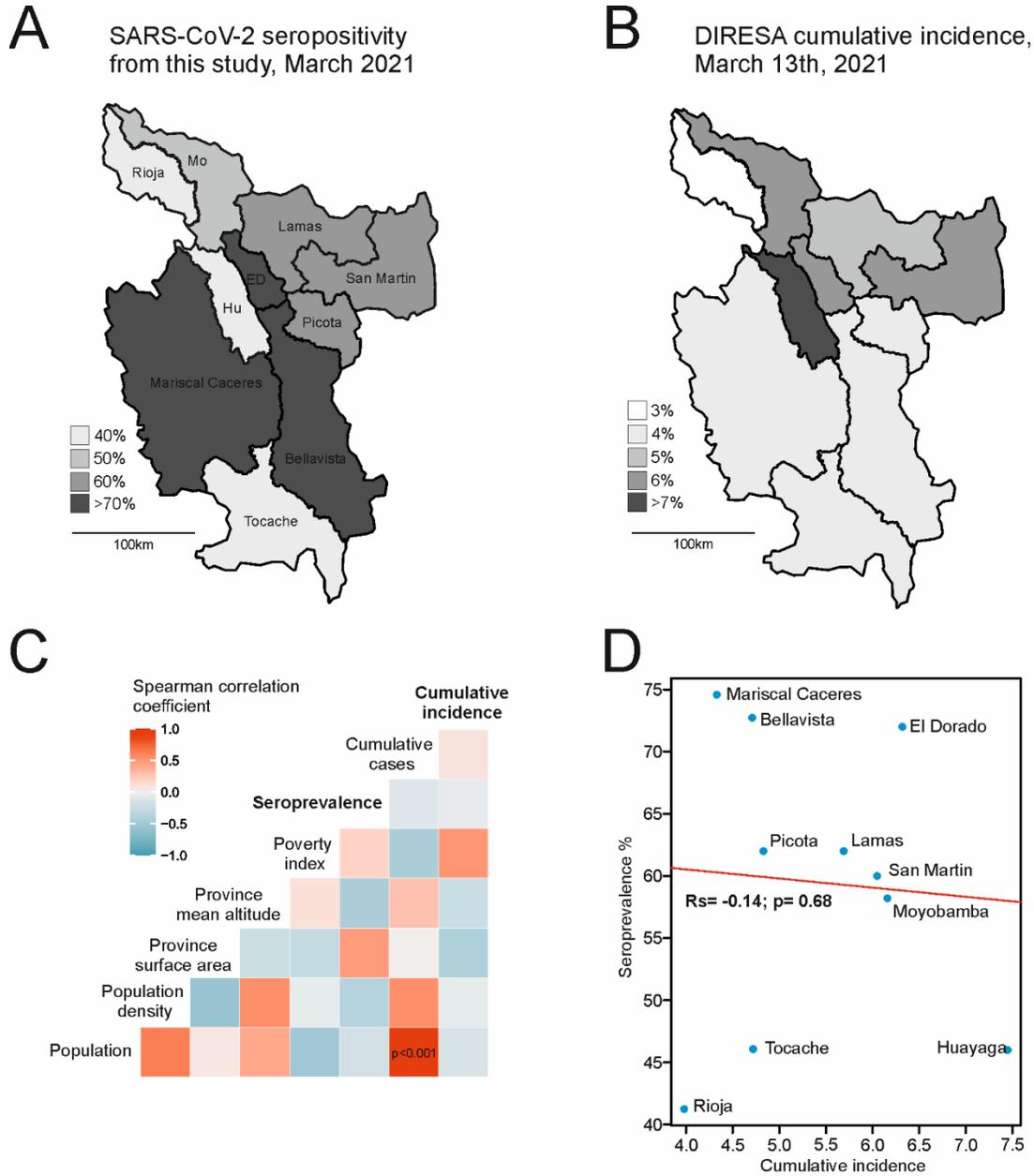
263



264

265 **Figure 1.** Epidemiological surveillance and serological diagnostics of SARS-CoV-2 from San  
266 Martin, Peru. (A) Mean altitude map of Peru. In red the department of San Martin. (B) Number of  
267 confirmed cases in San Martin as of June 2021. The time where incidence data and samples for  
268 the serological study were taken is marked with arrows. Surveillance data were gathered from  
269 <https://diressanmartin.gob.pe/> and <https://www.gob.pe/minsa/>. Reactivity of serum samples from  
270 (C) SARS-CoV-2 seroprevalence study in San Martin in 2021 and (D) pre-COVID-19 cohort from

271 2013-2018 from Peru in a chemiluminescence immunoassay (CLIA) and a SARS-CoV-2 surrogate  
 272 virus neutralization test (sVNT, GenScript, USA).



273

274 **Figure 2.** Correlation of seroprevalence and incidence data with different indicators. Comparison  
 275 of serological (A) and incidence (B) data from San Martin. Seropositivity by provinces of San  
 276 Martin. Hu= Huallaga, Mo= Moyobamba, ED= El Dorado. Right: confirmed cases by province  
 277 from the Regional Health Directorate of Peru (DIRESA) as of 13/03/2021. Scale denotes  
 278 percentage of cumulatively confirmed cases per total of the province population. (C) Heatmap of  
 279 Spearman's rank correlation test against different social, economic, and geographical indicators.  
 280 Significant correlations are depicted inside the square. (D) Spearman's rank correlation test of  
 281 seroprevalence and cumulative incidence per province. Rs=Spearman's correlation coefficient.

282

### 283 **Supplementary Material**

284 **Annex Table 1.** Demographics per province, seropositivity and incidence data from San Martin,  
 285 Peru, 2021

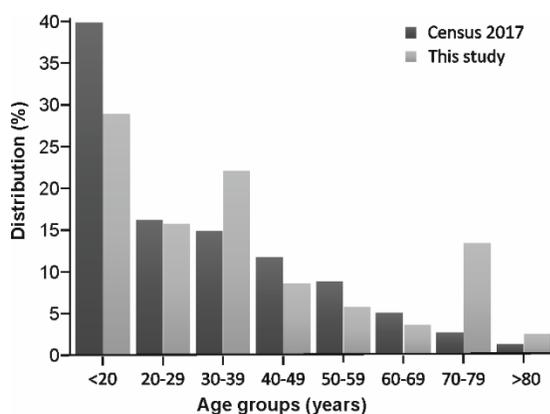
Province	Population (Projection 2020)*	Popula tion density (/ km <sup>2</sup> )*	Surface area ( km <sup>2</sup> )	Mean altitu de (m)	Povert y index (%) <sup>s</sup>	N° positive s / Total analyze d	Seroprevalen ce % (Confidence Interval)	DIRESA** cumulativ e cases	DIRESA cumulativ e incidence (%) #
<b>Rioja</b>	131 651	51.93	2 535.04	841	42.50	33/80	41.25 (31-52)	5 240	3.98
<b>Huallaga</b>	30 097	12.64	2 380.85	303	57.10	23/50	46.00 (33-59)	2 242	7.45
<b>Tocache</b>	76 450	13.03	5 865.44	502	26.80	14/30	46.70 (30-63)	3 610	4.72
<b>Moyobam ba</b>	133 631	35.42	3 772.31	878	43.20	46/79	58.20 (47-68)	6 903	5.16

<b>San</b>	218 074	38.67	5 639.82	356	28.00	30/50	60.00	13 188	6.05
<b>Martin</b>							(46-72)		
<b>Picota</b>	44 039	20.28	2 171.41	223	37.10	31/50	62.00	2 128	4.83
							(48-74)		
<b>Lamas</b>	88 637	17.58	5 082.54	791	56.80	50/80	62.00	5 050	5,69
							(52-72)		
<b>El Dorado</b>	42 983	33.11	1 298.14	346	59.50	36/50	72.00	2 719	6.32
							(58-83)		
<b>Bellavista</b>	60 893	7.56	8 033.19	285	41.30	32/44	72.73	2 868	4.71
							(58-83)		
<b>Mariscal</b>	73 193	5.05	14 498.73	282	42.60	37/50	74.00	3 168	4.33
<b>Caceres</b>							(60-84)		
<b>Total</b>	899 648	17.55	51 253.31	1609	43.49	332/563	58.97	47 116	12.67
							(55-63)		

286 \*2020 projection data gathered from the Peruvian National Institute of Statistics and Informatics (inei.gob.pe). \*\* DIRESA: Regional  
 287 Health Directorate of Peru. #Cumulative incidence was calculated with the cumulative cases/ total population per province. \$Poverty  
 288 index is calculated as the percent of the population that have salaries below the regional poverty line corresponding to 328  
 289 Soles/monthly/per capita, corresponding to the minimum wage for their nutritional and non-nutritional necessities. Taken from  
 290 INEI.gob.pe and from the Regional government of San Martin.

291

292



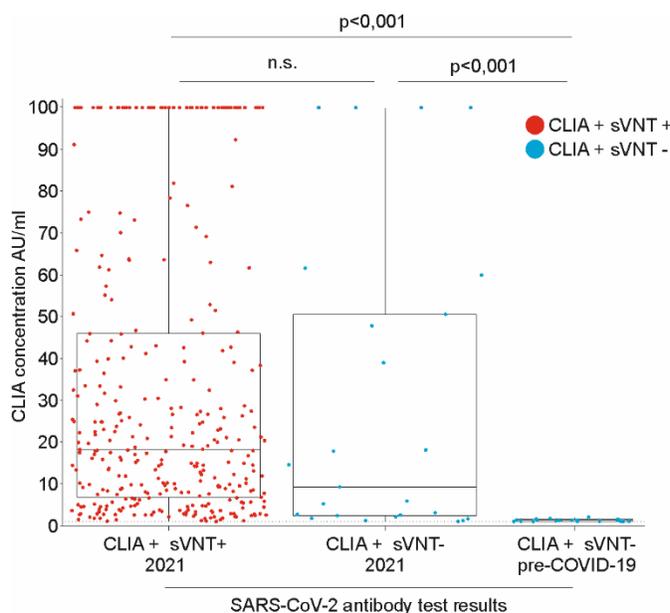
293

294

295 **Annex Figure 1.** Age distribution from Census data in San Martin in comparison to this study.

296 Data from the National Institute of Statistics and Informatics of Peru (<https://www.inei.gob.pe/>).

297



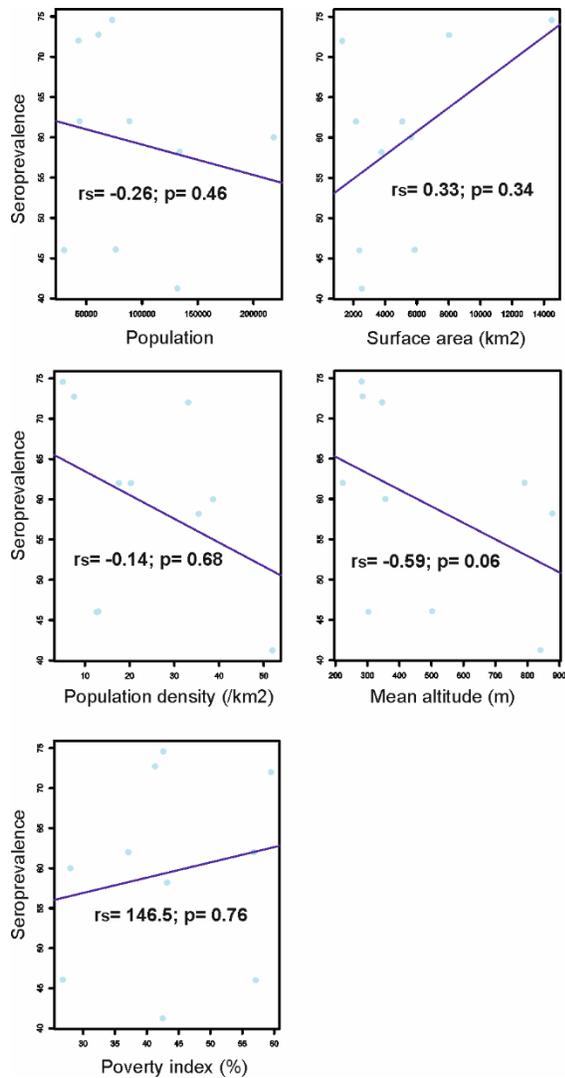
298

299 **Annex Figure 2.** Validation of the testing algorithm using pre-COVID-19 samples from Peru.

300 CLIA concentration comparison between SARS-CoV-2 positive samples from the 2021 cohort

301 from this study and 84 pre-COVID-19 samples from Peru. Line inside the boxplots denotes the

302 median. N.s. not significant.



303

304 **Annex Figure 3.** Spearman's rank correlation test of seroprevalence and population, population

305 density, surface area, poverty index and altitude.  $R_s$ = Spearman's correlation coefficient.

306 Epidemiological data taken from: <http://sial.minam.gob.pe/sanmartin/indicador/850> ,

307 <https://www.citypopulation.de/en/peru/sanmartin/admin/>