Global urban environmental change drives adaptation in white clover

James Santangelo, Rob Ness, Beata Cohan, Connor Fitzpatrick, Simon Innes, Sophie Koch, Lindsay Miles, Samreen Munim, Pedro Peres-Neto, Cindy Prashad, et al.

▶ **To cite this version:**

James Santangelo, Rob Ness, Beata Cohan, Connor Fitzpatrick, Simon Innes, et al.. Global urban environmental change drives adaptation in white clover. Science, 2022, 375 (6586), pp.1275-1281. 10.1126/science.abk0989 . hal-03656967

HAL Id: hal-03656967

https://univ-rennes.hal.science/hal-03656967

Submitted on 25 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Title: Global urban environmental change drives adaptation in white clover

Authors: James S. Santangelo[1,2]†, Rob W. Ness[1,2]†, Beata Cohan[1]†, Connor R. Fitzpatrick[3]†, Simon G. Innes[4,1]†, Sophie Koch[1]†, Lindsay S. Miles[1,2]†, Samreen Munim[5,1]†, Pedro R. Peres-Neto[6]†, Cindy Prashad[1]†, Alex T. Tong[1]†, Windsor E. Aguirre[7], Philips O. Akinwole[8], Marina Alberti[9], Jackie Álvarez[10], Jill T. Anderson[11], Joseph J. Anderson[12], Yoshino Ando[13], Nigel R. Andrew[14], Fabio Angeoletto[15], Daniel N. Anstett[16], Julia Anstett[17,18], Felipe Aoki-Gonçalves[19], A.Z. Andis Arietta[20], Mary T.K. Arroyo[21,22], Emily J. Austen[23], Fernanda Baena-Díaz[24], Cory A. Barker[25], Howard A. Baylis[26], Julia M. Beliz[27,28], Alfonso Benitez-Mora[29], David Bickford[30], Gabriela Biedebach[30], Gwylim S. Blackburn[31], Mannfred M. A. Boehm[16], Stephen P. Bonser[32], Dries Bonte[33], Jesse R. Bragger[34], Cristina Branquinho[35], Kristien I. Brans[36], Jorge C. Bresciano[37], Peta D. Brom[38], Anna Bucharova[39], Briana Burt[40], James F. Cahill[41], Katelyn D. Campbell[25], Elizabeth J. Carlen[42], Diego Carmona[43], Maria Clara Castellanos[44], Giada Centenaro[45], Izan Chalen[10,46], Jaime A. Chaves[10,47], Mariana Chávez-Pesqueira[48], Xiao-Yong Chen[49,50], Angela M. Chilton[51], Kristina M. Chomiak[40], Diego F. Cisneros-Heredia[10,46], Ibrahim K. Cisse[40], Aimée T. Classen[52], Mattheau S. Comerford[53], Camila Cordoba Fradinger[54], Hannah Corney[55], Andrew J. Crawford[56], Kerri M. Crawford[57], Maxime Dahirel[58], Santiago David[59], Robert De Haan[60], Nicholas J. Deacon[61], Clare Dean[62], Ek del-Val[63], Eleftherios K. Deligiannis[64], Derek Denney[11], Margarete A. Dettlaff[41], Michelle F. DiLeo[65], Yuan-Yuan Ding[49], Moisés E. Domínguez-López[66,67], Davide M. Dominoni[68], Savannah L. Draud[69], Karen Dyson[9], Jacintha Ellers[70], Carlos I. Espinosa[71], Liliana Essi[72], Mohsen Falahati-Anbaran[73,74], Jéssica C. F. Falcão[75], Hayden T. Fargo[1], Mark D. E. Fellowes[76], Raina M. Fitzpatrick[77], Leah E. Flaherty[78], Pádraic J. Flood[79], María F. Flores[22], Juan Fornoni[80], Amy G. Foster[81], Christopher J. Frost[82], Tracy L. Fuentes[9], Justin R. Fulkerson[83], Edeline Gagnon[84,85], Frauke Garbsch[81], Colin J. Garroway[86], Aleeza C. Gerstein[87], Mischa M. Giasson[88], E. Binney Girdler[89], Spyros Gkelis[64], William Godsoe[90], Anneke M. Golemiec[5], Mireille Golemiec[1], César González-Lagos[29,91], Amanda J. Gorton[92], Kiyoko M. Gotanda[93,26], Gustaf Granath[12], Stephan Greiner[81], Joanna S. Griffiths[94], Filipa Grilo[35], Pedro E. Gundel[95,54], Benjamin Hamilton[40], Joyce M. Hardin[69], Tianhua He[96,97], Stephen B. Heard[88], André F. Henriques[35], Melissa Hernández-Poveda[56], Molly C. Hetherington-Rauth[1], Sarah J. Hill[14], Dieter F. Hochuli[98], Kathryn A. Hodgins[99], Glen R. Hood[100], Gareth R. Hopkins[101], Katherine A. Hovanes[102], Ava R. Howard[101], Sierra C. Hubbard[69], Carlos N. Ibarra-Cerdeña[103], Carlos Iñiguez-Armijos[71], Paola Jara-Arancio[104,105], Benjamin J. M. Jarrett[106,26], Manon Jeannot[107], Vania Jiménez-Lobato[108], Mae Johnson[109], Oscar Johnson[110], Philip P. Johnson[111], Reagan Johnson[112], Matthew P. Josephson[113], Meen Chel Jung[9], Michael G. Just[114], Aapo Kahilainen[65], Otto S. Kailing[115], Eunice Kariñho-Betancourt[116], Regina Karousou[64], Lauren A. Kirn[99], Anna Kirschbaum[117], Anna-Liisa Laine[118,65], Jalene M. LaMontagne[7,119], Christian Lampei[39], Carlos Lara[120], Erica L. Larson[121], Adrián Lázaro-Lobo[122], Jennifer H. Le[123], Deleon S. Leandro[124], Christopher Lee[99], Yunting Lei[125], Carolina A. León[29], Manuel E. Lequerica Tamara[98], Danica C. Levesque[126], Wan-Jin Liao[127], Megan Ljubotina[41], Hannah Locke[57], Martin T. Lockett[128], Tiffany C. Longo[34], Jeremy T. Lundholm[55], Thomas MacGillavry[68], Christopher R. Mackin[44], Alex R. Mahmoud[27], Isaac A. Manju[101], Janine Mariën[70], D. Nayeli Martínez[63,129], Marina Martínez-Bartolomé[130,122], Emily K. Meineke[131], Wendy Mendoza-Arroyo[116], Thomas J. S. Merritt[126], Lila Elizabeth L. Merritt[126], Giuditta Migiani[68], Emily S. Minor[111], Nora Mitchell[132,133], Mitra Mohammadi Bazargani[134], Angela T.

ACCEPTED MANUSCRIPT - CLEAN COPY

Moles[32], Julia D. Monk[20], Christopher M. Moore[135], Paula A. Morales-Morales[136], Brook T. Moyers[137,138], Miriam Muñoz-Rojas[51,139], Jason Munshi-South[42], Shannon M. Murphy[121], Maureen M. Murúa[140], Melisa Neila[29], Ourania Nikolaidis[123], Iva Njunjić[141], Peter Nosko[142], Juan Núñez-Farfán[80], Takayuki Ohgushi[143], Kenneth M. Olsen[27], Øystein H. Opedal[106], Cristina Ornelas[144], Amy L. Parachnowitsch[88,12], Aaron S. Paratore[40], Angela M. Parody-Merino[37], Juraj Paule[145], Octávio S. Paulo[35], João Carlos Pena[146], Vera W. Pfeiffer[147], Pedro Pinho[35], Anthony Piot[31], Ilga M. Porth[31], Nicholas Poulos[148], Adriana Puentes[149], Jiao Qu[33], Estela Quintero-Vallejo[150], Steve M. Raciti[151], Joost A. M. Raeymaekers[152], Krista M. Raveala[65], Diana J. Rennison[153], Milton C. Ribeiro[146], Jonathan L. Richardson[154], Gonzalo Rivas-Torres[10,155], Benjamin J. Rivera[89], Adam B. Roddy[156], Erika Rodriguez-Muñoz[56], José Raúl Román[157], Laura S. Rossi[142], Jennifer K. Rowntree[62], Travis J. Ryan[158], Santiago Salinas[89], Nathan J. Sanders[52], Luis Y. Santiago-Rosario[159], Amy M. Savage[123], J.F. Scheepens[160,117], Menno Schilthuizen[161], Adam C. Schneider[69,1], Tiffany Scholier[149,162], Jared L. Scott[163], Summer A. Shaheed[34], Richard P. Shefferson[164], Caralee A. Shepard[69], Jacqui A. Shykoff[165], Georgianna Silveira[166], Alexis D. Smith[111], Lizet Solis-Gabriel[63], Antonella Soro[167], Katie V. Spellman[168,144], Kaitlin Stack Whitney[169], Indra Starke-Ottich[145], Jörg G. Stephan[170,149], Jessica D. Stephens[171], Justyna Szulc[172], Marta Szulkin[172], Ayco J. M. Tack[45], Ítalo Tamburrino[22], Tayler D. Tate[101], Emmanuel Tergemina[79], Panagiotis Theodorou[167], Ken A. Thompson[59,173], Caragh G. Threlfall[98], Robin M. Tinghitella[121], Lilibeth Toledo-Chelala[63], Xin Tong[49], Léa Uroy[58,174], Shunsuke Utsumi[13], Martijn L. Vandegehuchte[107,33], Acer VanWallendael[175], Paula M. Vidal[22], Susana M. Wadgymar[176], Ai-Ying Wang[127], Nian Wang[177], Montana L. Warbrick[142], Kenneth D. Whitney[132], Miriam Wiesmeier[178], J. Tristian Wiles[69], Jianqiang Wu[125], Zoe A. Xirocostas[32], Zhaogui Yan[177], Jiahe Yao[179], Jeremy B. Yoder[148], Owen Yoshida[55], Jingxiong Zhang[125], Zhigang Zhao[179], Carly D. Ziter[6], Matthew P. Zuellig[180], Rebecca A. Zufall[57], Juan E. Zurita[10], Sharon E. Zytynska[178,181], Marc T.J. Johnson[1,2]*†

[1]Department of Biology, University of Toronto Mississauga; Mississauga, ON, Canada.
[2]Centre for Urban Environments, University of Toronto Mississauga; Mississauga, ON, Canada.
[3]Department of Biology, University of North Carolina; Chapel Hill, NC, USA.
[4]Department of Biology, University of Louisiana; Lafayette, LA, USA.
[5]Department of Biology, Queen's University; Kingston, ON, Canada.
[6]Department of Biology, Concordia University; Montreal, QC, Canada.
[7]Department of Biological Sciences, DePaul University; Chicago, IL, USA.
[8]Department of Biology, DePauw University; Greencastle, IN, USA.
[9]Department of Urban Design and Planning, University of Washington; Seattle, WA, USA.
[10]Colegio de Ciencias Biológicas y Ambientales, Universidad San Francisco de Quito USFQ; Quito, Ecuador.
[11]Department of Genetics, University of Georgia; Athens, GA, USA.
[12]Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University; Uppsala, Sweden.
[13]Field Science Center for Northern Biosphere, Hokkaido University; Sapporo, Hokkaido, Japan.
[14]Natural History Museum, Zoology, University of New England; Armidale, NSW, Australia.
[15]Programa de Pós-Graduação em Geografia da UFMT, campus de Rondonópolis; Cuiabá, Brazil.
[16]Department of Botany and Biodiversity Research Centre, University of British Columbia; Vancouver, BC, Canada.

ACCEPTED MANUSCRIPT - CLEAN COPY

[17]Graduate Program in Genome Sciences and Technology, Genome Sciences Centre, University of British Columbia, Vancouver, British Columbia, Canada.
[18]Department of Microbiology and Immunology, University of British Columbia, Vancouver, British Columbia, Canada.
[19]Red de Biología Evolutiva, Instituto de Ecología, A. C.; Xalapa, Mexico.
[20]School of the Environment, Yale University; New Haven, CT, USA.
[21]Departamento de Ciencias Ecológicas, Facultad de Ciencias, Universidad de Chile, Santiago, Chile.
[22]Instituto de Ecología y Biodiversidad, Universidad de Chile; Santiago, Chile.
[23]Department of Biology, Mount Allison University; Sackville, NB, Canada.
[24]Red de Ecoetología, Instituto de Ecología A. C.; Xalapa, Mexico.
[25]Department of Biology, University of Ottawa; Ottawa, ON, Canada.
[26]Department of Zoology, University of Cambridge; Cambridge, UK.
[27]Department of Biology, Washington University in St. Louis; St. Louis, MO, USA.
[28]Department of Biology, University of Miami; Miami, FL, USA.
[29]Centro de Investigación en Recursos Naturales y Sustentabilidad (CIRENYS), Universidad Bernardo O'Higgins; Santiago, Chile.
[30]Department of Biology, University of La Verne; La Verne, CA, USA.
[31]Département des sciences du bois et de la forêt, Université Laval; Quebec, QC, Canada.
[32]Evolution & Ecology Research Centre, School of Biological, Earth and Environmental Sciences, UNSW Sydney; Sydney, NSW, Australia.
[33]Department of Biology, Ghent University; Ghent, Belgium.
[34]Department of Biology, Monmouth University; West Long Branch, NJ, USA.
[35]Centre for Ecology, Evolution and Environmental Changes, Faculdade de Ciências, Universidade de Lisboa; Campo Grande, Lisboa, Portugal.
[36]Department of Biology, KU Leuven; Leuven, Belgium.
[37]School of Agriculture and Environment, Wildlife and Ecology group, Massey University; Palmerston North, Manawatu, New Zealand.
[38]Department of Biological Sciences, University of Cape Town; Cape Town, South Africa.
[39]Institute of Landscape Ecology, University of Münster; Münster, Germany.
[40]Gosnell School of Life Sciences, Rochester Institute of Technology; Rochester, NY, USA.
[41]Department of Biological Sciences, University of Alberta; Edmonton, AB, Canada.
[42]Louis Calder Center and Department of Biological Sciences, Fordham University; Armonk, NY, USA.
[43]Departamento de Ecología Tropical, Universidad Autónoma de Yucatán; Mérida, Yucatán, México.
[44]School of Life Sciences, University of Sussex; Brighton, UK.
[45]Department of Ecology, Environment and Plant Sciences, Stockholm University; Stockholm, Sweden.
[46]iBIOTROP Instituto de Biodiversidad Tropical, Universidad San Francisco de Quito; Quito, Ecuador.
[47]Department of Biology, San Francisco State University, San Francisco, CA, USA.
[48]Unidad de Recursos Naturales, Centro de Investigación Científica de Yucatán AC; Mérida, Yucatán, México.
[49]School of Ecological and Environmental Sciences, East China Normal University; Shanghai, China.

ACCEPTED MANUSCRIPT - CLEAN COPY

[50]Shanghai Engineering Research Center of Sustainable Plant Innovation, Shanghai 200231, China.
[51]Centre for Ecosystem Science, School of Biological, Earth and Environmental Sciences, UNSW Sydney; Sydney, NSW, Australia.
[52]Department of Ecology and Evolutionary Biology, University of Michigan; Ann Arbor, MI, USA.
[53]Department of Biosciences, Rice University; Houston, TX, USA.
[54]IFEVA, Universidad de Buenos Aires, CONICET, Facultad de Agronomía; Buenos Aires, Argentina.
[55]Biology Department, Saint Mary's University; Halifax, NS, Canada.
[56]Department of Biological Sciences, Universidad de los Andes; Bogotá, Colombia.
[57]Department of Biology and Biochemistry, University of Houston; Houston, TX, USA.
[58]ECOBIO (Ecosystèmes, biodiversité, évolution), Université de Rennes; Rennes, France.
[59]Department of Zoology and Biodiversity Research Centre, University of British Columbia; Vancouver, BC, Canada.
[60]Department of Environmental Studies, Dordt University; Sioux Center, IA, USA.
[61]Department of Biology, Minneapolis Community and Technical College; Minneapolis, MN, USA.
[62]Department of Natural Sciences, Ecology and Environment Research Centre, Manchester Metropolitan University; Manchester, UK.
[63]Instituto de Investigaciones en Ecosistemas y Sustentabilidad, UNAM; Morelia, Mexico.
[64]Department of Botany, School of Biology, Aristotle University of Thessaloniki; Thessaloniki, Greece.
[65]Faculty of Biological and Environmental Science, Organismal & Evolutionary Biology Research Programme, University of Helsinki; Helsinki, Finland.
[66]Corporación Científica Ingeobosque; Medellín, Antioquia, Colombia.
[67]GTA Colombia S.A.S. Envigado, Antioquia, Colombia.
[68]Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow; Glasgow, Scotland, UK.
[69]Department of Biology, Hendrix College; Conway, AR, USA.
[70]Department of Ecological Science, Vrije Universiteit Amsterdam; Amsterdam, The Netherlands.
[71]Departamento de Ciencias Biológicas y Agropecuarias, Universidad Técnica Particular de Loja; Loja, Ecuador.
[72]Departamento de Biologia, Universidade Federal de Santa Maria (UFSM); Santa Maria, Rio Grande do Sul, Brazil.
[73]Department of Plant Sciences, School of Biology, College of Science, University of Tehran; Tehran, Iran.
[74]NTNU University Museum, Norwegian University of Science and Technology, 7491 Trondheim, Norway.
[75]Red de Estudios Moleculares Avanzados, Instituto de Ecología A. C.; Xalapa, Mexico.
[76]School of Biological Sciences, University of Reading; Whiteknights Park, Reading, Berkshire, UK.
[77]Department of Biology, Northern Arizona University; Flagstaff, AZ, USA.
[78]Department of Biological Sciences, MacEwan University; Edmonton, AB, Canada.
[79]Max Planck Institute for Plant Breeding Research; Cologne, Germany.

ACCEPTED MANUSCRIPT - CLEAN COPY

[80]Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México; Ciudad de México, México.
[81]Max Planck Institute of Molecular Plant Physiology; Potsdam-Golm, Germany.
[82]BIO5 Institute, University of Arizona; Tucson, AZ, USA.
[83]Alaska Center for Conservation Science, University of Alaska Anchorage; Anchorage, AK, USA.
[84]Tropical Diversity, Royal Botanical Garden of Edinburgh; Edinburgh, UK.
[85]Département de biologie, Université de Moncton; Moncton, New Brunswick, Canada.
[86]Department of Biological Sciences, University of Manitoba; Winnipeg, MB, Canada.
[87]Departments of Microbiology & Statistics, University of Manitoba; Winnipeg, MB, Canada.
[88]Department of Biology, University of New Brunswick; Fredericton, NB, Canada.
[89]Department of Biology, Kalamazoo College; Kalamazoo, MI, USA.
[90]BioProtection Research Centre, Lincoln University; Lincoln, Canterbury, New Zealand.
[91]Departamento de Ciencias, Facultad de Artes Liberales, Universidad Adolfo Ibáñez, Santiago, Chile.
[92]Department of Ecology, Evolution, and Behaviour University of Minnesota; Minneapolis, MN, USA.
[93]Department of Biological Sciences, Brock University; St. Catharines, Ontario, Canada.
[94]Department of Environmental Toxicology, University of California; Davis, CA, USA.
[95]ICB - University of Talca, Chile.
[96]School of Molecular and Life Science, Curtin University; Perth, Australia.
[97]College of Science, Health, Engineering and Education, Murdoch University, Murdoch, WA, Australia.
[98]School of Life and Environmental Sciences, The University of Sydney; Sydney, NSW, Australia.
[99]School of Biological Sciences, Monash University; Melbourne, VIC, Australia.
[100]Department of Biological Sciences, Wayne State University; Detroit, MI, USA.
[101]Department of Biology, Western Oregon University; Monmouth, OR, USA.
[102]School of Natural Resources and the Environment, University of Arizona; Tucson, AZ, USA.
[103]Departamento de Ecología Humana, Cinvestav Mérida; Yucatan, México.
[104]Departamento de Ciencias Biológicas y Departamento de Ecología y Biodiversidad, Facultad de Ciencias de la Vida, Universidad Andrés Bello; Santiago, Chile.
[105]Institute of Ecology and Biodiversity (IEB), Chile.
[106]Department of Biology, Lund University; Lund, Sweden.
[107]Department of Biology, Norwegian University of Science and Technology; Trondheim, Norway.
[108]Escuela Superiro de Desarrollo Sustentable, Universidad Autónoma de Guerrero -CONACYT; Las Tunas, Mexico.
[109]Clarkson Secondary School, Peel District School Board; Mississauga, ON, Canada.
[110]Homelands Sr. Public School, Peel District School Board; Mississauga, ON, Canada.
[111]Department of Biological Sciences, University of Illinois at Chicago; Chicago, IL, USA.
[112]St. James Catholic Global Learning Centre, Dufferin-Peel Catholic District School Board; Mississauga ON, Canada.
[113]Department of Biosciences, University of Calgary; Calgary, AB, Canada.
[114]Ecological Processes Branch, U.S. Army ERDC-CERL; Champaign, IL, USA.
[115]Department of Biology, Oberlin College; Oberlin, OH, USA.

ACCEPTED MANUSCRIPT - CLEAN COPY

[116]Escuela Nacional de Estudios Superiores Unidad Morelia, UNAM; Morelia, Mexico.
[117]Institute of Evolution and Ecology, University of Tübingen; Tübingen, Germany.
[118]Department of Evolutionary Biology and Environmental Studies, University of Zurich; Winterthurerstrasse, Zurich, Switzerland.
[119]Urban Wildlife Institute, Department of Conservation and Science, Lincoln Park Zoo, Chicago, IL, USA.
[120]Departamento de Ecología, Universidad Católica de la Santísima Concepción; Concepción, Chile.
[121]Department of Biological Sciences, University of Denver; Denver, CO, USA.
[122]Department of Biological Sciences, Mississippi State University; Starkville, MS, USA.
[123]Department of Biology, Center for Computational & Integrative Biology, Rutgers University-Camden; Camden, NJ, USA.
[124]Programa de Pós-Graduação em Geografia da UFMT, campus de Rondonópolis; Brasil.
[125]Kunming Institute of Botany, Chinese Academy of Sciences; Kunming, Yunnan, China.
[126]Department of Chemistry & Biochemistry, Laurentian University; Sudbury, ON, Canada.
[127]Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, College of Life Sciences, Beijing Normal University; Beijing, China.
[128]School of BioSciences, University of Melbourne; Melbourne, VIC, Australia.
[129]Posgrado en Ciencias Biológicas, Universidad Nacional Autónoma de México, Coyoacán, Mexico City, 04510, Mexico.
[130]Department of Biological Sciences, Auburn University; Auburn, AL, USA.
[131]Department of Entomology and Nematology, University of California; Davis, CA, USA.
[132]Department of Biology, University of New Mexico; Albuquerque, NM, USA.
[133]Department of Biology, University of Wisconsin - Eau Claire; Eau Claire, WI 54701.
[134]Agriculture Institute, Iranian Research Organization for Science and Technology (IROST); Tehran, Iran.
[135]Department of Biology, Colby College; Waterville, ME, USA.
[136]Instituto de Biología, Universidad de Antioquia, Medellín, Colombia.
[137]Department of Biology, University of Massachusetts Boston; Boston, MA, USA.
[138]Agricultural Biology, Colorado State University, Fort Collins, CO, USA.
[139]Departamento de Biología Vegetal y Ecología, Facultad de Biología, Universidad de Sevilla, Av. Reina Mercedes s/n, 41012 Sevilla, Spain.
[140]Facultad de Estudios Interdisciplinarios, Centro GEMA- Genómica, Ecología y Medio Ambiente, Universidad Mayor; Santiago, Chile.
[141]Evolutionary Ecology Group, Naturalis Biodiversity Center; Leiden, The Netherlands.
[142]Department of Biology and Chemistry, Nipissing University; North Bay, ON, Canada.
[143]Center for Ecological Research, Kyoto University; Otsu, Shiga, Japan.
[144]Bonanza Creek Long Term Ecological Research Program, University of Alaska Fairbanks; Fairbanks, AK, USA.
[145]Department of Botany and Molecular Evolution, Senckenberg Research Institute and Natural History Museum Frankfurt; Frankfurt am Main, Germany.
[146]Departamento de Biodiversidade, Instituto de Biociências, Univ Estadual Paulista - UNESP; Rio Claro, São Paulo, Brazil.
[147]Nelson Institute for Environmental Studies, University of Wisconsin-Madison; Madison, WI, USA.
[148]Department of Biology, California State University, Northridge; Los Angeles, CA, USA.

ACCEPTED MANUSCRIPT - CLEAN COPY

[149]Department of Ecology, Swedish University of Agricultural Sciences; Uppsala, Sweden.
[150]Facultad de Ciencias y Biotecnologia, Universidad CES; Medellín, Colombia.
[151]Department of Biology, Hofstra University; Long Island, NY, USA.
[152]Faculty of Biosciences and Aquaculture, Nord University; Bodø, Norway.
[153]Division of Biological Sciences, University of California San Diego; San Diego, CA, USA.
[154]Department of Biology, University of Richmond; Richmond, VA, USA.
[155]Estación de Biodiversidad Tiputini, Colegio de Ciencias Biológicas y Ambientales, Universidad San Francisco de Quito USFQ, Quito, Ecuador.
[156]Department of Biological Sciences, Institute of Environment, Florida International University; Miami, FL, USA.
[157]Agronomy Department, University of Almería; Almería, Spain.
[158]Department of Biological Sciences and Center for Urban Ecology and Sustainability, Butler University; Indianapolis, IN, USA.
[159]Department of Biological Sciences, Louisiana State University; Baton Rouge, LA, USA.
[160]Faculty of Biological Sciences, Goethe University Frankfurt; Frankfurt am Main, Germany.
[161]Institute of Biology Leiden, Leiden University; Leiden, The Netherlands.
[162]Department of Biological and Environmental Science, University of Jyväskylä; Jyväskylä, Finland.
[163]Department of Biology, University of Louisville; Louisville, KY, USA.
[164]Organization for Programs on Environmental Science, University of Tokyo; Tokyo, Japan.
[165]Université Paris-Saclay, CNRS, AgroParisTech, Ecologie Systématique et Evolution, 91405, Orsay, France.
[166]Department of Biology, Providence College; Providence, RI, USA.
[167]General Zoology, Institute for Biology, Martin Luther University Halle-Wittenberg; Halle, Germany.
[168]International Arctic Research Center, University of Alaska Fairbanks; Fairbanks, AK, USA.
[169]Science, Technology and Society Department, Rochester Institute of Technology; Rochester, NY, USA.
[170]SLU Swedish Species Information Centre, Swedish University of Agricultural Sciences; Uppsala, Sweden.
[171]Department of Biology, Westfield State University; Westfield, MA, USA.
[172]Centre of New Technologies, University of Warsaw; Warsaw, Poland.
[173]Department of Biology, Stanford University, Stanford, CA, USA.
[174]UMR 0980 BAGAP, Agrocampus Ouest-ESA-INRA, Rennes, France.
[175]Plant Biology Department, Michigan State University; East Lansing, MI, USA.
[176]Biology Department, Davidson College; Davidson, NC, USA.
[177]College of Horticulture and Forestry Sciences/ Hubei Engineering Technology Research Center for Forestry Information, Huazhong Agricultural University, Wuhan, China; Hubei, China.
[178]School of Life Sciences, Technical University of Munich; Munich, Germany.
[179]School of Life Sciences, Lanzhou University; Lanzhou, China.
[180]Institute of Ecology and Evolution, University of Bern; Bern, Switzerland.
[181]Department of Evolution, Ecology and Behaviour, University of Liverpool; Liverpool, UK.

† These authors were members of the lead team
***Corresponding author. Email: marc.johnson@utoronto.ca**

ACCEPTED MANUSCRIPT - CLEAN COPY

*Corresponding author. Email: **marc.johnson@utoronto.ca**

ACCEPTED MANUSCRIPT - CLEAN COPY

Abstract: Urbanization transforms environments in ways that alter biological evolution. We examined whether urban environmental change drives parallel evolution by sampling 110,019 white clover plants from 6,169 populations in 160 cities globally. Plants were assayed for a Mendelian antiherbivore defence that also affects tolerance to abiotic stressors. Urban-rural gradients were associated with the evolution of clines in defence in 47% of cities throughout the world. Variation in the strength of clines was explained by environmental changes in drought stress and vegetation cover that varied among cities. Sequencing 2,074 genomes from 26 cities revealed that the evolution of urban-rural clines was best explained by adaptive evolution, but the degree of parallel adaptation varied among cities. Our results demonstrate that urbanization leads to adaptation at a global scale.

Once sentence summary: Convergent urban environmental change at a global scale drives adaptation in a cosmopolitan plant

ACCEPTED MANUSCRIPT - CLEAN COPY

Main Text: Urbanization is a driver of both environmental and evolutionary change. Towns and cities are rapidly expanding throughout the world to accommodate human population growth. These urban areas represent novel ecosystems, in which urban development alters multiple environmental factors (*1*). Recent research shows that urban environmental change can influence four evolutionary processes: mutation, genetic drift, gene flow, and adaptation due to natural selection (*2, 3*). Despite numerous examples of how urbanization affects genetic drift and gene flow (*4, 5*), the effects of urbanization on adaptive evolution have received less attention (*6-8*). Adaptation to urban environments can impact species' conservation (*9*), the spread of pests and disease (*2*), eco-evolutionary feedbacks (*10*), as well as urban planning and human society (*11*). However, the few examples of adaptation to urban environments focus on just one or a small number of cities in a single region (*2*). It is therefore unclear whether populations can adapt to urban habitats in similar ways across cities throughout the world.

Parallel adaptive evolution is most likely when populations experience similar environmental selective pressures on the same genes or phenotypes (*12, 13*). For urbanization to drive parallel evolution, urban areas must converge in environmental features that affect an organism's fitness. Urbanization can lead to similar environmental changes across cities (*14*), but whether urban environmental convergence causes parallel evolution has never been examined at a global scale.

Here we test how global urbanization affects environmental change and evolution in a cosmopolitan plant species, white clover (*Trifolium repens* L., Fabaceae). White clover populations are polymorphic for the production of hydrogen cyanide (HCN), an antiherbivore chemical defence controlled by two genes (*15*). At least one functional allele at each of two unlinked loci (*Ac* and *Li*) are required to produce HCN following tissue damage, while plants that are homozygous for gene deletions (*ac* and *li* alleles) at either locus lack HCN (*16, 17*).

ACCEPTED MANUSCRIPT - CLEAN COPY

Notably, these deletions occur throughout the world, resulting in standing genetic variation on which selection can act (*18*). Previous work showed that herbivores select for the production of HCN, and abiotic stressors (e.g., freezing and drought) influence the costs and benefits of the metabolic components underlying the defence (*19, 20*). Variation in these environmental factors is credited with driving the evolution of clines in HCN production at continental and regional scales (*21, 22*), including in response to urban environments (*23-25*). Thus, HCN production could evolve in response to urbanization if there are urban-rural gradients in herbivory, winter temperature, or drought.

We examined global urban environmental and evolutionary change across the diverse climates that white clover inhabits. To this end, we created the Global Urban Evolution Project to test for parallel evolution and urban adaptation in natural populations across white clover's worldwide range. The present study builds on our previous work on white clover (*23-25*) by sampling cities globally across diverse climates in both the native (Europe and western Asia) and introduced ranges, by quantifying many environmental factors from each population, and by integrating evolutionary genomic analyses using whole genome sequence data. This project spanned 160 cities across 26 countries (Fig. 1, *15*) in white clover's native (Europe and western Asia) and introduced ranges (Fig. 1, Fig. S1). From these cities, we phenotyped 110,019 plants from 6,169 sampling sites (hereafter "populations", Table S1). Populations within each city were sampled along an urban-rural transect, with half of each transect in urban and suburban areas (i.e., areas with high building density), and the other half in rural areas (Fig. 2E-G) (*15*).

Across 160 cities, we tested whether urban white clover habitats converged to be more similar among cities and less variable within cities in their environmental characteristics compared to rural habitats (*15*). Urban and rural habitats significantly diverged (H_0: $urban_{mean} = rural_{mean}$,

ACCEPTED MANUSCRIPT - CLEAN COPY

387 $P_{bootstrapped} < 0.01$, Fig. 2A) along two principal component axes that accounted for 65% of the
388 variation in the multivariate environments between the two habitats across cities. Urban locations
389 consistently had more impervious surface, higher summer temperatures and less vegetation than
390 rural populations (Fig. 2B, Fig. S2). The remaining environmental variables changed along
391 urban-rural gradients in many cities, but these changes were less consistent in direction among
392 cities (Fig. S2, Table S2). Although urban and rural environments diverged on average, urban-
393 rural changes in the environment were not always parallel (H_0: parallel urban-rural changes
394 among cities, $P_{bootstrapped} < 0.01$, Fig. 2A). Additionally, environmental variance among urban
395 populations within a city was lower than the environmental variance among rural populations
396 ($F_{9,1570} = 31.76$, $P < 0.001$, Fig. S3). Together these results show that on average urbanization
397 leads to similar and less variable environmental conditions in some factors (e.g., impervious
398 surface, summer temperature, summer vegetation), but not in others (e.g. potential
399 evapotranspiration, snow cover, winter vegetation), which could lead to variation in the degree of
400 parallel evolution.
401 We next tested whether convergent urban environmental change causes parallel evolution in
402 an ecologically important trait of white clover. We examined evolution in response to
403 urbanization by testing for a relationship between HCN production and distance to the urban
404 center (i.e., an "HCN cline"), as well as other metrics of urbanization (*15*). Our model explained
405 28% of the variation in the frequency of HCN production within populations (Table S3). Across
406 160 cities, distance from the city center was positively related to the frequency of HCN-
407 producing plants (distance: $\chi^2_{df=1} = 12.35$, $P < 0.001$). The probability that a plant produced
408 HCN increased by 44% on average from the center of an urban area to the furthest rural
409 population (Fig. 2C, D). However, cities varied in the strength and direction of clines (distance ×

ACCEPTED MANUSCRIPT - CLEAN COPY

410 city interaction: $\chi^2_{df=1}$= 1001, $P < 0.001$, Fig. 2C, D). Overall, 47% of cities exhibited a
411 significant ($P < 0.05$) cline (*15*), with 39% of cities (62 of 160) showing a positive cline in which
412 HCN production was less common in urban than rural populations, and 8% of cities (13 of 160)
413 had negative clines (Fig. 2, Table S4). Positive and negative clines occurred in both the native
414 and introduced ranges, with the former being more prevalent among continents and across
415 diverse climates (Fig. 1).
416 Given the prevalence of HCN clines at a global scale, we sought to identify the evolutionary
417 processes driving variation in the strength and direction of clines. In addition to natural selection,
418 non-adaptive evolution can lead to the evolution of clines (*26*). Importantly, the epistatic genetic
419 architecture of HCN production makes the loss of the trait more likely with increased genetic
420 drift (*26*). Therefore, the prevalence of positive clines could reflect stronger drift in urban
421 populations (*4, 5*). To examine whether urban populations exhibited stronger drift, we estimated
422 pairwise nucleotide diversity (π) of putatively neutral sites using whole genome sequence data
423 from ~80 individuals per city, with samples equally split between urban and rural habitats across
424 26 cities (N = 2,074, *15*). These cities were selected to capture variation in the strength and
425 direction of clines, geography, and climate (Fig. 1) (*15*).
426 Genetic diversity was not consistently different between urban and rural habitats and did not
427 explain variation in the slope of HCN clines along urban-rural gradients. On average, urban and
428 rural habitats did not differ in neutral genetic diversity ($F_{1, 25} = 0.028$, $P = 0.87$; Fig. 3A).
429 Furthermore, the difference in π between urban and rural habitats within a city was not strongly
430 related to the slope of HCN clines ($F_{1, 24} = 0.25$, $P = 0.62$; Fig. 3B, Fig. S4), and urban-rural
431 differences in genetic diversity were similar between cities with and without clines ($F_{1, 24} =$
432 0.017, $P = 0.90$).

ACCEPTED MANUSCRIPT - CLEAN COPY

Variation in the strength of genetic differentiation and gene flow between urban and rural habitats can influence the ability of populations to adapt to urban environments (*27*). To test the association between genetic differentiation and the evolution of HCN clines we estimated population genetic differentiation between urban and rural populations using both F_{ST} and principal components analysis (PCA) (Fig. S5), in addition to urban-rural admixture (Fig. S6) (*15*). Urban-rural F_{ST} was low (mean = 0.012 ± 0.002 [SE]) and did not differ significantly between cities with and without clines ($F_{1, 24}$ = 1.47, P = 0.24; Fig. 3C, Fig. S4). Neither F_{ST} ($F_{1, 24}$ = 1.42, P = 0.25; Fig. 3D) nor urban-rural differentiation measured using PCA ($F_{1, 24}$ = 1.10, P = 0.31, Fig. S5) predicted the strength of clines in HCN production. The absence of strong differentiation was associated with extensive admixture between urban and rural populations (Fig. S6). Since genetic differentiation is consistently low and gene flow appears to be high among urban and rural populations, the repeated evolution of clines suggests strong selection on HCN production along urban-rural gradients. This conclusion is further supported by direct tests of selection on the *Ac* and *Li* loci, as well as HCN production, in which differentiation (using a statistic equivalent to F_{ST}) between urban and rural populations was stronger than expected under neutral evolution in cities with HCN clines compared to cities without clines (Figs. 3E, 3F, *15*).

Multiple environmental stressors are known to influence the evolution of HCN production at continental scales (*20-22, 28*), so we asked: What environmental factors explain variation in the evolution of HCN production along urban-rural gradients? Environmental factors related to drought and vegetation cover were the strongest predictors of variation in HCN clines, accounting for 11.3% of the variation in the strength of clines (Table S5, Table S6). Change in potential evapotranspiration (PET) along urban-rural gradients was one of the most consistent predictors of evolution in HCN production (Table S5); the frequency of HCN production tended

ACCEPTED MANUSCRIPT - CLEAN COPY

456 to be higher in rural than urban populations in cities where PET was also greater in rural habitats
457 (Fig. 4A,C; Fig. S7). Since high PET can lead to plant water stress under low soil moisture, this
458 result is consistent with drought selecting for higher HCN production, a pattern also observed at
459 continental scales (*21*). However, the effect of PET on the evolution of HCN production only
460 occurs when the amount of vegetation in and around cities is low (Fig. 4A). When vegetation
461 cover is relatively high (and impervious surface is low) along the whole urban-rural transect,
462 HCN clines tend to be positive regardless of variation in PET (Fig. 4A-C). Importantly, the
463 amount of vegetation is positively correlated with invertebrate herbivore biomass and diversity
464 (*29*), which can select for increased HCN production (*20*). When combined with the observation
465 that herbivores are often less abundant in urban habitats (*30*), our evidence suggests that
466 herbivores are selecting for greater HCN production in rural than urban areas. The positive
467 association between urban-rural changes in vegetation and the positive slope of HCN clines in
468 some cities further supports this interpretation (Fig. 4D). Put simply, herbivory seems to select
469 for higher HCN production in rural areas, but in the absence of strong herbivore pressure (i.e.,
470 when there is less vegetation across the whole gradient), drought is the main selective agent.
471 Contrary to previous findings, urban-rural changes in temperature and snow cover did not
472 explain changes in HCN production (*24*), suggesting urban-rural changes in these abiotic factors
473 are not a general explanation for the evolution of clines at a global scale.
474 Our results have general implications for understanding how environmental change affects
475 adaptation in widespread species. Parallel evolution is a hallmark of natural selection because it
476 suggests that adaptation proceeds in a repeatable way when populations face similar
477 environments (*12, 13*). However, departures from parallel evolution are common, and a major
478 goal of recent research involves quantifying how ecological and evolutionary factors interact to

ACCEPTED MANUSCRIPT - CLEAN COPY

influence variation in adaptive responses to similar environments (*12*). Our results show that white clover rapidly adapts to urban environments on a global scale, but there is considerable variation in the strength and direction of HCN clines that is driven by variation in particular biotic and abiotic factors that differ in how they change along urban-rural gradients among cities. Variation in additional unmeasured factors (e.g., gene flow from agricultural varieties, pollution, etc.) might further explain variation in the strength of clines, and future work will seek to explore such mechanisms.

Urbanization is increasingly transforming rural and natural environments into unique ecosystems that Earth's biodiversity has never experienced, and these changes are altering the evolution of life. If adaptation to urban environments is common, then this could have cascading effects on populations and ecosystems. This knowledge could help conserve some of Earth's most vulnerable species (*9*), mitigate the impacts of pests (*2*), improve human well-being (*8, 11*), and contribute to understanding fundamental eco-evolutionary processes (*10*).

References and Notes:

1. N. B. Grimm *et al.*, Global change and the ecology of cities. *Science* **319**, 756-760 (2008).
2. M. T. J. Johnson, J. Munshi-South, Evolution of life in urban environments. *Science* **358**, aam8327 (2017).
3. M. Szulkin, J. Munshi-South, A. Charmantier, Eds., *Urban Evolutionary Biology* (Oxford University Press, 2020).
4. L. S. Miles, L. R. Rivkin, M. T. J. Johnson, J. Munshi-South, B. C. Verrelli, Gene flow and genetic drift in urban environments. *Mol. Ecol.* **28**, 4138-4151 (2019).

ACCEPTED MANUSCRIPT - CLEAN COPY

5. C. Schmidt, M. Domaratzki, R. Kinnunen, J. Bowman, C. J. Garroway, Continent-wide effects of urbanization on bird and mammal genetic diversity. *Proc. Biol. Sci.* **287**, 20192497 (2020).
6. E. M. Oziolor *et al.*, Adaptive introgression enables evolutionary rescue from extreme environmental pollution. *Science* **364**, 455-457 (2019).
7. K. M. Winchell, R. G. Reynolds, S. R. Prado-Irwin, A. R. Puente-Rolón, L. J. Revell, Phenotypic shifts in urban areas in the tropical lizard *Anolis cristatellus*. *Evolution* **70**, 1009-1022 (2016).
8. L. R. Rivkin *et al.*, A roadmap for urban evolutionary ecology. *Evol. Appl.* **18**, 384-398 (2019).
9. M. R. Lambert, C. M. Donihue, Urban biodiversity management using evolutionary tools. *Nat. Ecol. Evol.* **4**, 903-910 (2020).
10. M. Alberti, Eco-evolutionary dynamics in an urbanizing planet. *Trends Ecol. Evol.* **30**, 114-126 (2015).
11. C. J. Schell *et al.*, The ecological and evolutionary consequences of systemic racism in urban environments. *Science* **369**, eaay4497 (2020).
12. D. I. Bolnick, R. D. Barrett, K. B. Oke, D. J. Rennison, Y. E. Stuart, (Non)parallel evolution. *Ann. Rev. Ecol. Evol. Syst.* **49**, 303-330 (2018).
13. J. B. Losos, Convergence, adaptation, and constraint. *Evolution* **65**, 1827-1840 (2011).
14. P. M. Groffman *et al.*, Ecological homogenization of urban USA. *Front. Ecol. Environ.* **12**, 74-81 (2014).
15. *See supplementary materials*.

ACCEPTED MANUSCRIPT - CLEAN COPY

16. K. M. Olsen, L. L. Small, Micro-and macroevolutionary adaptation through repeated loss of a complete metabolic pathway. *New Phytol.* **219**, 757-766 (2018).
17. K. M. Olsen, B. L. Sutherland, L. L. Small, Molecular evolution of the *Li/li* chemical defence polymorphism in white clover (*Trifolium repens* L.). *Mol. Ecol.* **16**, 4180-4193 (2007).
18. N. J. Kooyers, K. M. Olsen, Adaptive cyanogenesis clines evolve recurrently through geographical sorting of existing gene deletions. *J. Evol. Biol.* **27**, 2554-2558 (2014).
19. N. J. Kooyers, B. Hartman Bakken, M. Ungerer, K. M. Olsen, Freeze-induced cyanide toxicity does not maintain the cyanogenesis polymorphism in white clover (*Trifolium repens*). *Am. J. Bot.* **105**, 1224-1231 (2018).
20. M. Hughes, The cyanogenic polymorphism in *Trifolium repens* L. (white clover). *Heredity* **66**, 105-115 (1991).
21. N. J. Kooyers, L. R. Gage, A. Al-Lozi, K. M. Olsen, Aridity shapes cyanogenesis cline evolution in white clover (*Trifolium repens* L.). *Mol. Ecol.* **23**, 1053-1070 (2014).
22. H. Daday, Gene frequencies in wild populations of *Trifolium repens* L. *Heredity* **12**, 169-184 (1958).
23. M. T. J. Johnson, C. M. Prashad, M. Lavoignat, H. S. Saini, Contrasting the effects of natural selection, genetic drift and gene flow on urban evolution in white clover (*Trifolium repens*). *Proc. Biol. Sci.* **285**, 20181019 (2018).
24. K. A. Thompson, M. Renaudin, M. T. J. Johnson, Urbanization drives the evolution of parallel clines in plant populations. *Proc. Biol. Sci.* **283**, 20162180 (2016).
25. J. S. Santangelo *et al.*, Predicting the strength of urban-rural clines in a Mendelian polymorphism along a latitudinal gradient. *Evol. Lett.* **4**, 212-225 (2020).

ACCEPTED MANUSCRIPT - CLEAN COPY

26. J. S. Santangelo, M. T. J. Johnson, R. W. Ness, Modern spandrels: the roles of genetic drift, gene flow and natural selection in the evolution of parallel clines. *Proc. Biol. Sci.* **285**, 20180230 (2018).
27. T. Lenormand, Gene flow and the limits to natural selection. *Trends Ecol. Evol.* **17**, 183-189 (2002).
28. N. J. Kooyers, K. Olsen, Searching for the bull's eye: agents and targets of selection vary among geographically disparate cyanogenesis clines in white clover (*Trifolium repens* L.). *Heredity* **111**, 495-504 (2013).
29. M. Fernández-Tizón, T. Emmenegger, J. Perner, S. Hahn, Arthropod biomass increase in spring correlates with NDVI in grassland habitat. *Sci. Nat.* **107**, 42 (2020).
30. L. S. Miles, S. T. Breitbart, H. H. Wagner, M. T. J. Johnson, Urbanization shapes the ecology and evolution of plant-arthropod herbivore interactions. *Front. Ecol. Evol.* **7**, 310 (2019).

Supplementary Materials:

31. A. G. Griffiths *et al.*, Breaking free: the genomics of allopolyploidy-facilitated niche expansion in white clover. *Plant Cell* **31**, 1466-1487 (2019).
32. J. Burdon, *Trifolium repens* L. *J. Ecol.* **71**, 307-330 (1983).
33. T. Kjaergaard, A plant that changed the world: the rise and fall of clover 1000-2000. *Landscape Res.* **28**, 41-49 (2003).
34. J. E. Poulton, Cyanogenesis in plants. *Plant Physiol.* **94**, 401-405 (1990).
35. H. E. Armstrong, E. F. Armstrong, E. Horton, Herbage studies. II. Variation in *Lotus corniculatus* and *Trifolium repens* (cyanophoric plants). *Proc. Biol. Sci.* **86**, 262-269 (1913).

ACCEPTED MANUSCRIPT - CLEAN COPY

36. L. Corkill, Cyanogenesis in white clover (*Trifolium repens* L.) V. The inheritance of cyanogeneiss. *NZ J. Sci. Technol. B* **23**, 178-193 (1942).
37. C. E. Cooper, G. C. Brown, The inhibition of mitochondrial cytochrome oxidase by the gases carbon monoxide, nitric oxide, hydrogen cyanide and hydrogen sulfide: chemical mechanism and physiological significance. *J. Bioenerg. Biomembr* **40**, 533 (2008).
38. E. Antonini, M. Brunori, G. C. Rotilio, C. Greenwood, B. G. Malmström, The interaction of cyanide with cytochrome oxidase. *Eur. J. Biochem* **23**, 396-400 (1971).
39. K. M. Olsen, S. Hsu, L. L. Small, Evidence on the molecular basis of the Ac/ac adaptive cyanogenesis polymorphism in white clover (*Trifolium repens* L.). *Genetics* **179**, 517-526 (2008).
40. K. M. Olsen, L. L. Small, Micro- and macroevolutionary adaptation through repeated loss of a complete metabolic pathway. *New Phytol.* **219**, 757-766 (2018).
41. I. Coop, Cyanogenesis in white clover (*Trifolium repens* L.). III. A Study of linamarase, the enzyme which hydrolyses lotaustralin. *NZ J. Sci. Technol.* **22-23**, 71-83 (1940).
42. T. B. Sackton, D. L. Hartl, Genotypic context and epistasis in individuals and populations. *Cell* **166**, 279-287 (2016).
43. K. M. Olsen, N. J. Kooyers, L. L. Small, Recurrent gene deletions and the evolution of adaptive cyanogenesis polymorphisms in white clover (*Trifolium repens* L.). *Mol. Ecol.* **22**, 724-738 (2013).
44. H. Daday, Gene frequencies in wild populations of *Trifolium repens.* II Distribution by altitude. *Heredity* **8**, 377-384 (1954).
45. H. Daday, Gene frequencies in wild populations of *Trifolium repens.* I. Distribution by latitude. *Heredity* **8**, 61-78 (1954).

ACCEPTED MANUSCRIPT - CLEAN COPY

46. F. Horne, M. T. Bright, Influence of temperature on cyanogenic polymorphisms. *Nature* **265**, 437-438 (1977).
47. F. R. Ganders, Altitudinal clines for cyanogenesis in introduced populations of white clover near Vancouver, Canada. *Heredity* **64**, 387-390 (1990).
48. A. M. de Araújo, The relationship between altitude and cyanogenesis in white clover (*Trifolium repens*, L.). *Heredity* **37**, 291-293 (1976).
49. *Trifolium repens* in GBIF Secretariat, GBIF Backbone Taxonomy. Checklist dataset https://doi.org/10.15468/39omei accessed via GBIF.org on 2021-07-15 (2021).
50. R. J. Hijmans, S. E. Cameron, J. L. Parra, P. G. Jones, A. Jarvis, Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965-1978 (2005).
51. A. Trabucco, R. J. Zomer, Global aridity index (global-aridity) and global potential evapo-transpiration (global-PET) geospatial database. *CGIAR Consortium for Spatial Information* **89**, 1-2 (2009).
52. E. C. Brown de Colstoun *et al.* (NASA Socioeconomic Data and Applications Center (SEDAC), Palisades, NY, 2017).
53. M. Reba, F. Reitsma, K. C. Seto, Spatializing 6,000 years of global urbanization from 3700 BC to AD 2000. *Scientific Data* **3**, 160034 (2016).
54. J. R. Hagler, S. Mueller, L. R. Teuber, S. A. Machtley, A. Van Deynze, Foraging range of honey bees, *Apis mellifera*, in alfalfa seed production fields. *J. Insect Sci.* **11**, 144 (2011).
55. J. S. Santangelo, L. R. Rivkin, C. Advenard, K. A. Thompson, Multivariate phenotypic divergence along an urbanization gradient. *Biol. Lett.* **16**, 20200511 (2020).
56. T. Brinkhoff. (2019), City Population. Available from: http://www.citypopulation.de (2019).

ACCEPTED MANUSCRIPT - CLEAN COPY

57. F. Feigl, V. Anger, Replacement of benzidine by copper ethylacetoacetate and tetra base as spot-test reagent for hydrogen cyanide and cyanogen. *Analyst* **91**, 282-284 (1966).
58. R. Gleadow, N. Bjarnhold, K. Jørgensen, J. Fox, R. Miller, in *Research Methods in Plant Sciences, Volume 1: Soil Allelochemicals* S. Narwal, L. Szajdak, D. Sampietro, Eds. (Studium Press, Houston, TX, 2011).
59. K. M. Olsen, M. C. Ungerer, Freezing tolerance and cyanogenesis in white clover (*Trifolium repens* L. Fabaceae). *Int. J. Plant Sci.* **169**, 1141-1147 (2008).
60. K. A. Thompson, M. T. J. Johnson, Antiherbivore defenses alter natural selection on plant reproductive traits. *Evolution* **70**, 796-810 (2016).
61. A. A. Agrawal, A. P. Hastings, M. T. J. Johnson, J. L. Maron, J.-P. Salminen, Insect herbivores drive real-time ecological and evolutionary change in plant populations. *Science* **338**, 113-116 (2012).
62. S. Fisher *et al.*, A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome Biol.* **12**, R1 (2011).
63. N. Rohland, D. Reich, Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res.* **22**, 939-946 (2012).
64. T. C. Glenn *et al.*, Adapterama I: universal stubs and primers for 384 unique dual-indexed or 147,456 combinatorially-indexed Illumina libraries (iTru & iNext). *PeerJ* **7**, e7755 (2019).
65. M. Meyer, M. Kircher, Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols* **2010**, pdb.prot5448 (2010).
66. G. Kobs, Cloning blunt-end DNA fragments into the pGEM®-T Vector Systems. *Promega Notes* **62**, 15-18 (1997).

ACCEPTED MANUSCRIPT - CLEAN COPY

637 67. H. Li *et al.*, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078-
638 2079 (2009).
639 68. S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ preprocessor.
640 *Bioinformatics* **34**, i884-i890 (2018).
641 69. S. Andrews, *FastQC: A Quality Control Tool for High Throughput Sequence Data*.
642 Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (2010).
643 70. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform.
644 *Bioinformatics* **25**, 1754-1760 (2009).
645 71. K. Okonechnikov, A. Conesa, F. García-Alcalde, Qualimap 2: advanced multi-sample
646 quality control for high-throughput sequencing data. *Bioinformatics* **32**, 292-294 (2016).
647 72. G. Jun, M. K. Wing, G. R. Abecasis, H. M. Kang, An efficient and scalable analysis
648 framework for variant extraction and refinement from population-scale DNA sequence
649 data. *Genome Res.* **25**, 918-925 (2015).
650 73. P. Ewels, M. Magnusson, S. Lundin, M. Käller, MultiQC: summarize analysis results for
651 multiple tools and samples in a single report. *Bioinformatics* **32**, 3047-3048 (2016).
652 74. R. Nielsen, J. S. Paul, A. Albrechtsen, Y. S. Song, Genotype and SNP calling from next-
653 generation sequencing data. *Nat. Rev. Genet.* **12**, 443-451 (2011).
654 75. E. Han, J. S. Sinsheimer, J. Novembre, Characterizing bias in population genetic inferences
655 from low-coverage sequencing data. *Mol. Biol. Evol.* **31**, 723-735 (2014).
656 76. R. Nielsen, T. Korneliussen, A. Albrechtsen, Y. Li, J. Wang, SNP calling, genotype
657 calling, and sample allele frequency estimation from new-generation sequencing data. *PloS*
658 *One* **7**, e37558 (2012).

ACCEPTED MANUSCRIPT - CLEAN COPY

77. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* **15**, 356 (2014).

78. H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987-2993 (2011).

79. H. Li, Improving SNP discovery by base alignment quality. *Bioinformatics* **27**, 1157-1158 (2011).

80. G. Van Rossum, F. Drake, *Python 3 Reference Manual* (CreateSpace, Scotts Valley, CA, 2009).

81. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).

82. R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2021).

83. W. N. Venables, B. D. Ripley, *Modern Applied Statistics with S* (Springer, New York, NY, ed. 4, 2002).

84. M. L. Collyer, D. C. Adams, Analysis of two-state multivariate phenotypic change in ecological studies. *Ecology* **88**, 683-692 (2007).

85. J. H. Zar, *Biostatistical Analysis* (Prentice-Hall, Hoboken, NJ, ed. 4, 1999).

86. D. Bates, M. Maechler, B. Bolker, S. Walker, Fitting linear mixed-effects models using lme4. *J. Stat. Soft.* **67**, 1-48 (2015).

87. J. Fox, S. Weisberg, *An R Companion to Applied Regression* (Sage Publications, Thunder Oaks, CA, ed. 3, 2018).

ACCEPTED MANUSCRIPT - CLEAN COPY

88. A. Kuznetsova, P. B. Brockhoff, R. H. B. Christensen, lmerTest package: tests in linear mixed effects models. *J. Stat. Soft.* **82**, 1-26 (2017).

89. Wildlife Conservation Society - WCS, and Center for International Earth Science Information Network - CIESIN - Columbia University. 2005. Last of the Wild Project, Version 2, 2005 (LWP-2): *Global Human Influence Index (HII) Dataset (Geographic)*. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). Available from: https://doi.org/10.7927/H4BP00QC).

90. G. Bhatia, N. Patterson, S. Sankararaman, A. L. Price, Estimating and interpreting F_{ST}: the impact of rare variants. *Genome Res.* **23**, 1514-1521 (2013).

91. R. R. Hudson, M. Slatkin, W. P. Maddison, Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**, 583-589 (1992).

92. S. Y. Kim *et al.*, Estimation of allele frequency and association mapping using next-generation sequencing data. *BMC Bioinformatics* **12**, 231 (2011).

93. J. Meisner, A. Albrechtsen, Inferring population structure and admixture proportions in low-depth NGS data. *Genetics* **210**, 719-731 (2018).

94. Y. E. Stuart *et al.*, Contrasting effects of environment and genetics generate a continuum of parallel evolution. *Nat. Ecol. Evol.* **1**, 0158 (2017).

95. M. Nei, Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci.* **70**, 3321-3323 (1973).

96. M. Jakobsson, M. D. Edge, N. A. Rosenberg, The relationship between F_{ST} and the frequency of the most frequent allele. *Genetics* **193**, 515-528 (2013).

97. G. James, D. Witten, T. Hastie, R. Tibshirani, *An Introduction to Statistical Learning with Applications in R* (Springer, New York, NY, 2013).

ACCEPTED MANUSCRIPT - CLEAN COPY

98. M. Kuhn, *caret: Classification and Regression Training*. Available from: https://CRAN.R-project.org/package=caret (2020).
99. D. A. Jackson, Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. *Ecology* **74**, 2204-2214 (1993).
100. K. Bartoń, *MuMIn: Multi-Model Inference*. R package version 1.43.17. Available from: https://cran.r-project.org/web/packages/MuMIn/MuMIn.pdf (2016).
101. J. C. Jiménez-Muñoz, J. A. Sobrino, A generalized single-channel method for retrieving land surface temperature from remote sensing data. *J. Geophys. Res.–Atmos.* **108**, D22-4688 (2003).
102. J. C. Jiménez-Muñoz *et al.*, Revision of the single-channel algorithm for land surface temperature retrieval from Landsat thermal-infrared data. *IEEE Trans. Geosci. Remote Sens.* **47**, 339-349 (2009).
103. J. A. Barsi, J. L. Barker, J. R. Schott, in *IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No. 03CH37477)*. (IEEE, 2003), vol. 5, pp. 3014-3016.
104. J. A. Barsi, J. R. Schott, F. D. Palluconi, S. J. Hook, in *Earth Observing Systems X*. (International Society for Optics and Photonics, 2005), vol. 5882, pp. 58820E.
105. Z.-L. Li *et al.*, Satellite-derived land surface temperature: current status and perspectives. *Remote Sens. Environ.* **131**, 14-37 (2013).
106. J. A. Sobrino *et al.*, Land surface emissivity retrieval from different VNIR and TIR sensors. *IEEE Trans. Geosci. Remote Sens.* **46**, 316-327 (2008).

ACCEPTED MANUSCRIPT - CLEAN COPY

107. J. C. Jiménez-Muñoz, J. A. Sobrino, D. Skoković, C. Mattar, J. Cristóbal, Land surface temperature retrieval methods from Landsat-8 thermal infrared sensor data. *IEEE Trans. Geosci. Rem. Sens.* **11**, 1840-1843 (2014).

108. A. A. Lamaro, A. Mariñelarena, S. E. Torrusio, S. E. Sala, Water surface temperature estimation from Landsat 7 ETM+ thermal infrared data using the generalized single-channel method: Case study of Embalse del Río Tercero (Córdoba, Argentina). *Advances in Space Research* **51**, 492-500 (2013).

109. J. Cristóbal *et al.*, An improved single-channel method to retrieve land surface temperature from the Landsat-8 thermal band. *Remote Sens.* **10**, 431 (2018).

110. J. Cristóbal, J. C. Jiménez-Muñoz, J. A. Sobrino, M. Ninyerola, X. Pons, Improvements in land surface temperature retrieval from the Landsat series thermal band using water vapor and air temperature. *J. Geophys. Res.–Atmos.* **114**, D08103 (2009).

111. J. A. Sobrino, J. C. Jiménez-Muñoz, L. Paolini, Land surface temperature retrieval from LANDSAT TM 5. *Remote Sens. Environ.* **90**, 434-440 (2004).

112. J. A. Sobrino *et al.*, Soil emissivity and reflectance spectra measurements. *Appl. Optics* **48**, 3664-3670 (2009).

113. N. Pettorelli *et al.*, Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends Ecol. Evol.* **20**, 503-510 (2005).

114. J. Rouse, R. Haas, J. Schell, D. Deering, J. Harlan, *Monitoring the Vernal Advancement of Retrogradation of Natural Vegetation* (NASA/GSFC, Type III, Final Report , vol. 371, 1974).

ACCEPTED MANUSCRIPT - CLEAN COPY

115. D. K. Hall, G. A. Riggs, V. V. Salomonson, Development of methods for mapping global snow cover using moderate resolution imaging spectroradiometer data. *Remote Sens. Environ.* **54**, 127-140 (1995).
116. H. Kyle, R. Curran, W. Barnes, D. Escoe, in *3rd Conference on Atmospheric Radiation*. (1978), pp. 107-109.
117. F. R. Valovcin, *Snow/Cloud Discrimination* (Air Force Geophysics Laboratories, Air Force Systems Command, U.S., 1976).
118. F. R. Valovcin, *Spectral Radiance of Snow and Clouds in the Near Infrared Spectral Region* (Air Force Geophysics Laboratory, Air Force Systems Command, U.S., vol. 78, 1978).
119. R. J. Hijmans *et al.*, *raster: Geographic Data Analysis and Modeling*. R package version 3.5-2. Available from: https://CRAN.R-project.org/package=raster (2015).
120. R. S. Bivand, E. J. Pebesma, V. Gomez-Rubio, E. J. Pebesma, *Applied Spatial Data Analysis with R* (Springer, New York, NY, 2013).
121. E. Pebesma, R. S. Bivand, S classes and methods for spatial data: the sp package. *R News* **5**, 9-13 (2005).
122. N. Middleton, D. Thomas, *World Atlas of Desertification* (Arnold, Hodder Headline, PLC, London, UK, ed. 2, 1997).
123. J. M. Gillett, N. L. Taylor, *The World of Clovers* (Iowa State University Press, Ames, IA 2001).
124. S. Arruda, D. Pereira, M. Silva-Castro, M. Brito, A. Waldschmidt, An optimized protocol for DNA extraction in plants with a high content of secondary metabolites, based on leaves

ACCEPTED MANUSCRIPT - CLEAN COPY

of *Mimosa tenuiflora* (Willd.) Poir. (Leguminosae). *Genet. Mol. Res.* **16**, gmr16039063 (2017).

125. P. W. Inglis, M. R. Pappas, L. V. Resende, D. Grattapaglia, Fast and inexpensive protocols for consistent extraction of high quality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. *PloS One* **13**, e0206085 (2018).

Acknowledgments: We thank L. Alejandro Giraldo, L. Arboleda-Restrepo, E. Bernal, F. Carrera, M. T. Solano de la Cruz, K. Christensen-Dalsgaard, K. Cuypers, E. Dawaas, A. Giraldo, D. González-Tokman, B. Gravendeel, T. Gregor, J. Hatakoshi, P. Hyttinen, S. Kagiya, H. Kappes, B. Kerr, A. Matsuura, S. Silberhorn, B. Kwan, M. Potter, E. Peñaherrera, J. Rafalski, L. Revell, E. Sparrow, R. Tapia-López, A. Tovar, Y. Wang, J. Wrath, L. Yaneva-Roder, X. Zhu, 2018 MacEwan University BIOL422 students, Minneapolis College's 2018 Plant Biology students, SHAD Mount Allison 2018 students and staff, and University of Wisconsin-Madison's 2018 Field Ecology students for assisting with collecting plants, performing HCN assays or providing equipment and facilities. D. Murray-Stoker, A. Filazzola, L. Albano, S. Breitbart and R. Rivkin provided comments on an earlier draft of the paper. I. Sheoran prepared most genomic libraries. M. Malcolm and X. Xhao assisted with shipping and lab logistics, respectively. High performance computing services were provided by Compute Ontario (www.computeontario.ca/) and Compute Canada (www.computecanada.ca). This work benefited from ideas and collaborations in Future Earth's EvolvES network and the NSF funded RCN Urban Eco-Evo NET.

Funding: The Global Urban Evolution project was primarily funded by an NSERC Discovery Grant, Canada Research Chair and NSERC Steacie Fellowship to M.T.J.J.. J.S.S. received funding from an NSERC CGS and C.R.F. is funded by an NSERC PDF. P.R.P.-N., R.W.N. and J.C.C. were supported by NSERC Discovery grants. M.A. was funded by NSF RCN DEB-1840663. F.A. received funding from CAPES. MTKA was funded by CONICYT PIA APOYO CCTE AFB170008. J.R.B, T.C.L., and S.A.S were supported by Monmouth University Sch. of Sci. SRP. E.G. was funded by D. Biologie, Université de Moncton. C.G.-L. received funding from the Center of Applied Ecology and Sustainability (CAPES), and ANID PIA/BASAL FB0002. S.G. was funded by the Max Planck Society. P.J.-A. was funded by ANID PIA/BASAL FB210006. I.N. and M.S. were supported by Leiden Municipality. K.M.O. was funded by US NSF awards IOS-1557770 and DEB-1601641. J.C.P. thanks FAPESP process 2018/00107-3, and M.C.R. thanks CNPq and FAPESP.

Author contributions: The project's lead team included B.C., C.R.F., S.G.I., M.T.J.J., S.K., L.S.M., S.M., R.W.N., P.R.P.-N., C.P., J.S.S., A.T.. M.T.J.J., R.W.N.. J.S.S. conceived of the project. H.T.F., M.T.J.J, J.S.S., and A.T. collected spatial environmental and city data. M.T.J.J., P.R.P.-N., and J.S.S. performed statistical analyses. R.W.N. and J.S.S. performed bioinformatic and genomic analyses. B.C., C.R.F., S.G.I., M.T.J.J., S.K., L.S.M., S.M., R.W.N., C.P., J.S.S., E.C., and J.M.-S. contributed reagents, materials, technical skills or analysis tools. All remaining authors designed transects, collected samples, and data. M.T.J.J. and J.S.S. wrote the paper with

ACCEPTED MANUSCRIPT - CLEAN COPY

input from the Lead Team; all authors provided comments on drafts of the paper. **Competing interests:** The authors declare no competing financial interests.

Data and materials Availability: All code and environmental and phenotypic data is available on the GitHub page for JSS (https:// https://github.com/James-S-Santangelo/glue_pc) and additionally archived on Zenodo (https://zenodo.org/record/5765252). BAM files have been deposited in the European Nucleotide Archive (ENA BioProject PRJEB48967).

Supplementary Materials

Material and Methods
Figs. S1-S16
Tables S1-S15
Supplementary Text S1-S6
References (31-125)

ACCEPTED MANUSCRIPT - CLEAN COPY

FIGURE LEGENDS

Fig. 1. Cities sampled for urban environmental and evolutionary change. Blue dots indicate cities with positive clines for hydrogen cyanide (HCN) production along urban-rural gradients ($HCN_{urban}<HCN_{rural}$). Red dots show negative clines ($HCN_{urban}>HCN_{rural}$). Grey dots indicate cities without a cline. Plants from the 26 cities outlined in black underwent whole genome sequencing. Inset: White clover and a honeybee.

Fig. 2. Urban environmental and evolutionary change across cities. (A) Principal component analysis showing environmental differences between urban (orange dots) and rural (green dots) habitats; ovals represent 95% CI. Lines connect urban and rural habitats from the same city. (**B**) The eigenvectors for environmental variables, coloured according to their contribution to PC2. The environmental variables included vegetation in winter ($NDVI_{winter}$) and summer ($NDVI_{summer}$), snow accumulation (NDSI), surface temperature in winter (LST_{winter}) and summer (LST_{summer}), aridity index (AI), potential evapotranspiration (PET), impervious surface (GMIS) and elevation (DEM). (**C**) Histogram of the slopes from binomial regressions of the relationship between HCN production within populations and distance from the city center. Distance was standardized to vary between 0 (urban center) and 1 (furthest rural population) in each city, so that cities that varied in size were compared on the same scale. The dashed vertical line corresponds to the mean slope across cities and overlap between bars showing cities with significant (blue and red) and non-significant clines (grey) are shown as muted colours. (**D**) The relationship between HCN production within populations and distance for each city; colours correspond to panel C. The black line shows the positive main effect of distance across cities ($P < 0.001$). (**E-G**) Examples of transects, with the orange lines showing the urban boundary, pie charts (jittered to reduce overlap) showing the proportion of HCN+ plants coloured in yellow.

ACCEPTED MANUSCRIPT - CLEAN COPY

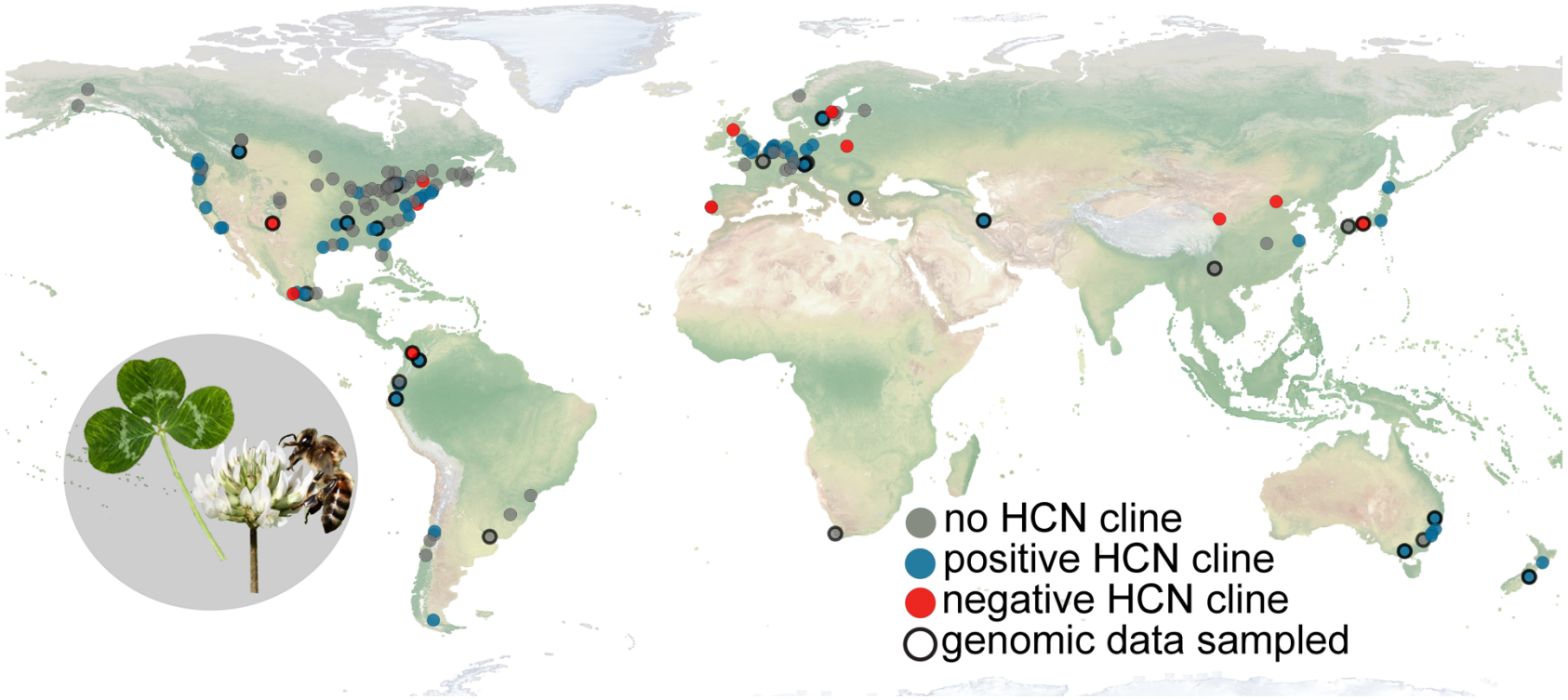
(**H-J**) Frequency of HCN production versus distance for the cities shown in E-G. The line shows the regression line ± 95% CI.

Fig. 3. Genetic diversity and differentiation within and between urban and rural habitats. (**A**) Mean (± SE) pairwise nucleotide diversity (π) for urban (orange) and rural (green) plants across cities. (**B**) The relationship between the slope of HCN clines versus the difference in nucleotide diversity between habitats, where each point is a city. (**C**) Histogram showing the distribution of genetic differentiation (F_{ST}) between urban and rural habitats for each city, coloured with respect to the significance of HCN clines. (**D**) Relationship between the absolute value of the slope of HCN clines versus F_{ST}. (**E**) Percentage of cities in which differentiation between urban and rural habitats at *Ac* or *Li* exceeds neutral expectation in cities with or without significant HCN clines (*15*). (**F**) Percentage of cities with differentiation in HCN production between urban and rural habitats that exceeds neutral expectation in cities with or without significant HCN clines (*15*). *P*-values in E and F correspond to χ^2-test for independence.

Fig. 4. Environmental predictors of urban-rural clines in HCN production. (**A**) Change in potential evapotranspiration along urban-rural gradients (PET_{β}) interacts with the regional amount of summer vegetation (i.e., $NDVI_{summer\_mean}$) to explain variation in the slopes of HCN clines. (**B**) The relationship between the slopes of HCN clines and the regional amount of winter vegetation ($NDVI_{winter\_mean}$). (**C**) PET_{β} interacts with the regional amounts of impervious surface ($GMIS_{mean}$) to predict the slope of HCN clines. (**D**) Change in summer vegetation along urban-rural gradients ($NDVI_{summer\_\beta}$) interacts with regional aridity (AI_{mean}) to explain variation in the slope of HCN clines. Acronyms as in Fig. 2.

ACCEPTED MANUSCRIPT - CLEAN COPY

820 FIGURE 1



821
822

823 **Fig. 1. Cities sampled for urban environmental and evolutionary change.** Blue dots indicate cities with positive clines for

824 hydrogen cyanide (HCN) production along urban-rural gradients ($HCN_{urban}<HCN_{rural}$). Red dots show negative clines

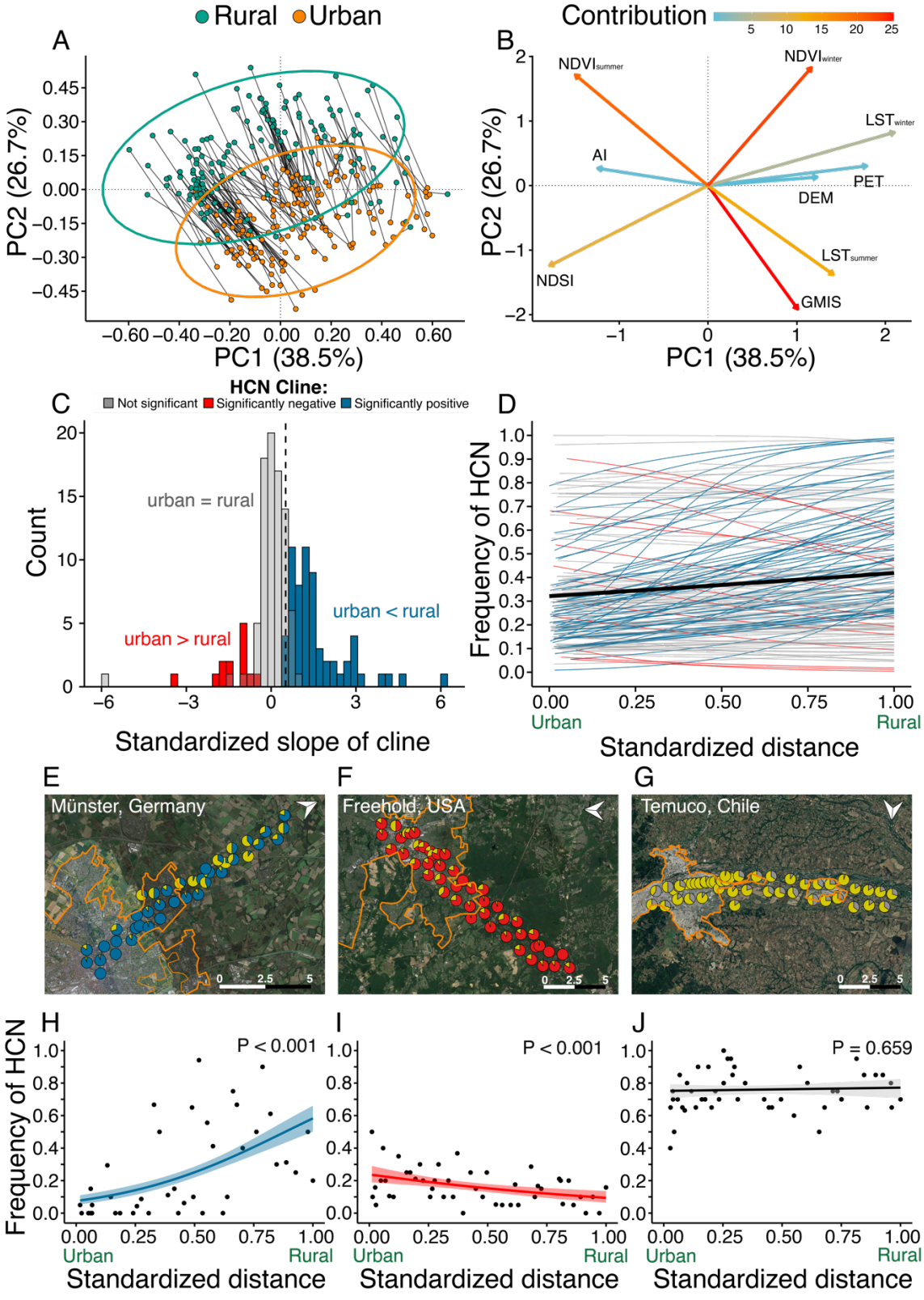
825 ($HCN_{urban}>HCN_{rural}$). Grey dots indicate cities without a cline. Plants from the 26 cities outlined in black underwent whole genome

826 sequencing. Inset: White clover and a honeybee.

827

ACCEPTED MANUSCRIPT - CLEAN COPY

828 FIGURE 2



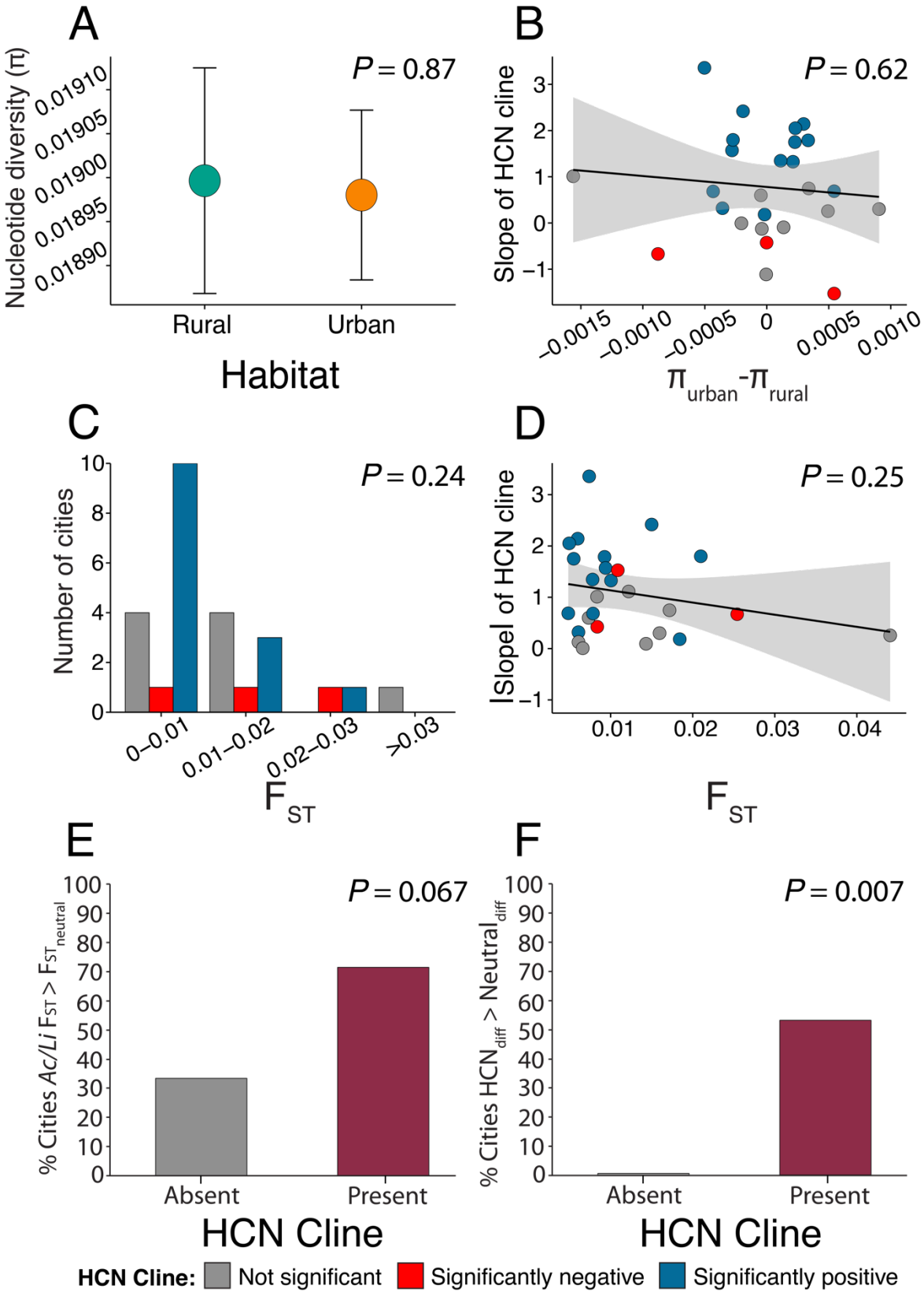
829

ACCEPTED MANUSCRIPT - CLEAN COPY

Fig. 2. Urban environmental and evolutionary change across cities. (A) Principal component analysis showing environmental differences between urban (orange dots) and rural (green dots) habitats; ovals represent 95% CI. Lines connect urban and rural habitats from the same city. (**B**) The eigenvectors for environmental variables, coloured according to their contribution to PC2. The environmental variables included vegetation in winter ($NDVI_{winter}$) and summer ($NDVI_{summer}$), snow accumulation (NDSI), surface temperature in winter (LST_{winter}) and summer (LST_{summer}), aridity index (AI), potential evapotranspiration (PET), impervious surface (GMIS) and elevation (DEM). (**C**) Histogram of the slopes from binomial regressions of the relationship between HCN production within populations and distance from the city center. Distance was standardized to vary between 0 (urban center) and 1 (furthest rural population) in each city. The dashed vertical line corresponds to the mean slope across cities. (**D**) The relationship between HCN production within populations and distance for each city; colours correspond to panel C. The black line shows the positive main effect of distance across cities ($P < 0.001$). (**E-G**) Examples of transects, with the orange lines showing the urban boundary, pie charts (jittered to reduce overlap) showing the proportion of HCN+ plants coloured in yellow. (**H-J**) Frequency of HCN production versus distance for the cities shown in E-G. The line shows the regression line ± 95% CI.

ACCEPTED MANUSCRIPT - CLEAN COPY

848 FIGURE 3



849

ACCEPTED MANUSCRIPT - CLEAN COPY

Fig. 3. Genetic diversity and differentiation within and between urban and rural habitats. (**A**) Mean (± SE) pairwise nucleotide diversity (π) for urban (orange) and rural (green) plants across cities. (**B**) The relationship between the slope of HCN clines versus the difference in nucleotide diversity between habitats, where each point is a city. (**C**) Histogram showing the distribution of genetic differentiation (F_{ST}) between urban and rural habitats for each city, coloured with respect to the significance of HCN clines. (**D**) Relationship between the absolute value of the slope of HCN clines versus F_{ST}. (**E**) Percentage of cities in which differentiation between urban and rural habitats at *Ac* or *Li* exceeds neutral expectation in cities with or without significant HCN clines (*15*). (**F**) Percentage of cities with differentiation in HCN production between urban and rural habitats that exceeds neutral expectation in cities with or without significant HCN clines (*15*). *P*-values in E and F correspond to χ^2-test for independence.

ACCEPTED MANUSCRIPT - CLEAN COPY

FIGURE 4

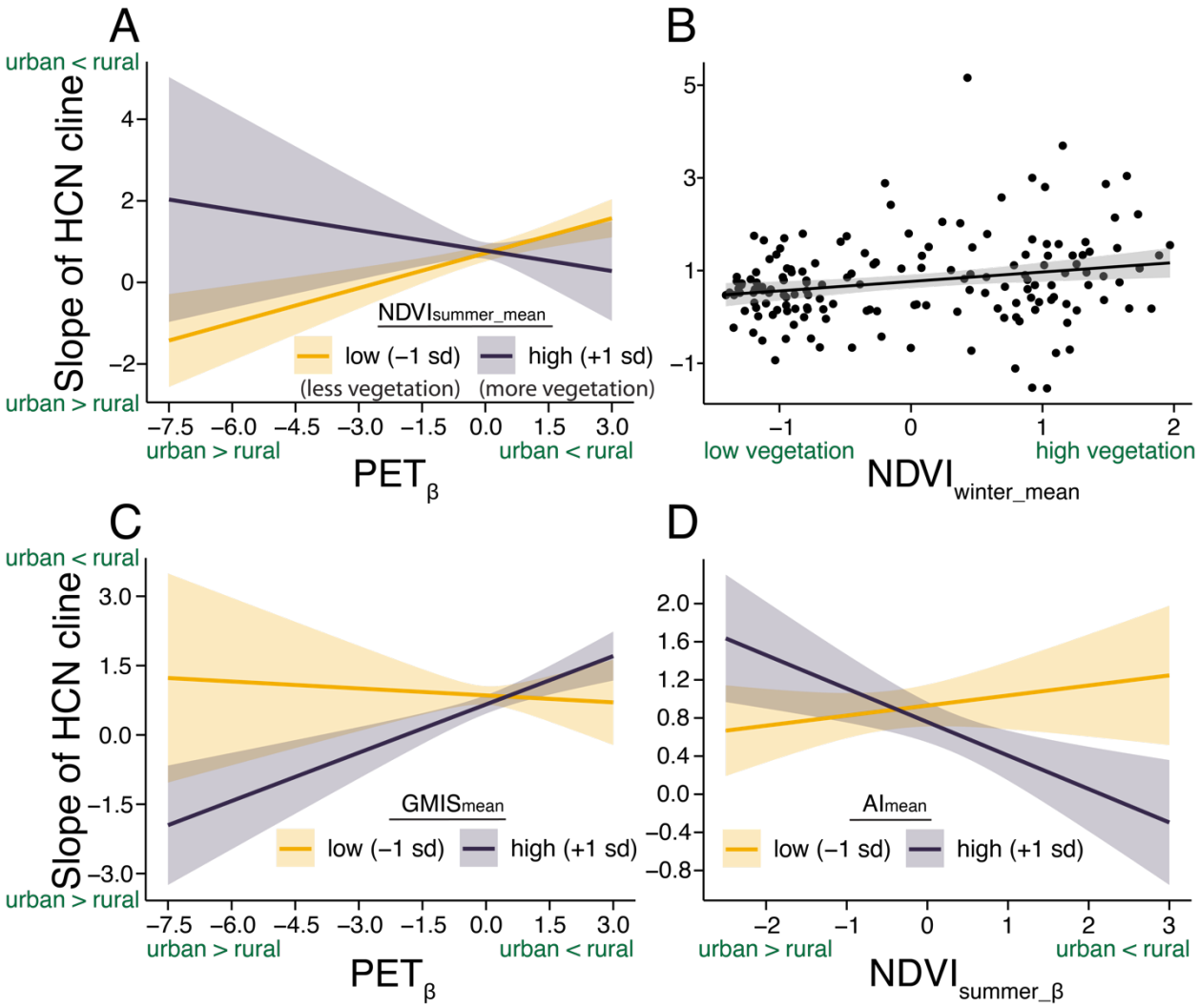


Fig. 4. Environmental predictors of urban-rural clines in HCN production. (**A**) Change in potential evapotranspiration along urban-rural gradients (PET_{β}) interacts with the regional amount of summer vegetation (i.e., $NDVI_{summer\_mean}$) to explain variation in the slopes of HCN clines. (**B**) The relationship between the slopes of HCN clines and the regional amount of winter vegetation ($NDVI_{winter\_mean}$). (**C**) PET_{β} interacts with the regional amounts of impervious surface ($GMIS_{mean}$) to predict the slope of HCN clines. (**D**) Change in summer vegetation along urban-rural gradients ($NDVI_{summer\_\beta}$) interacts with regional aridity (AI_{mean}) to explain variation in the slope of HCN clines. Acronyms as in Fig. 2.

ACCEPTED MANUSCRIPT - CLEAN COPY

Science
AAAS

Supplementary Materials for

Global urban environmental change drives parallel evolution in a cosmopolitan plant

James S. Santangelo, Rob W. Ness, Beata Cohan, Connor R. Fitzpatrick, Simon G. Innes, Sophie Koch, Lindsay S. Miles, Samreen Munim, Pedro Peres-Neto, Cindy Prashad, Alex T. Tong, Windsor E. Aguirre, Philips O. Akinwole, Marina Alberti, Jackie Álvarez, Jill T. Anderson, Joseph J. Anderson, Yoshino Ando, Nigel R. Andrew, Fabio Angeoletto, Daniel N. Anstett, Julia Anstett, Felipe Aoki-Gonçalves, A.Z. Andis Arietta, Mary T.K. Arroyo, Emily J. Austen, Fernanda Baena-Díaz, Cory A. Barker, Howard A. Baylis, Julia M. Beliz, Alfonso Benitez-Mora, David Bickford, Gabriela Biedebach, Gwylim S. Blackburn, Mannfred M. A. Boehm, Stephen P. Bonser, Dries Bonte, Jesse R. Bragger, Cristina Branquinho, Kristien I. Brans, Jorge C. Bresciano, Peta D. Brom, Anna Bucharova, Briana Burt, James F. Cahill, Katelyn D. Campbell, Elizabeth J. Carlen, Diego Carmona, Maria Clara Castellanos, Giada Centenaro, Izan Chalen, Jaime A. Chaves, Mariana Chávez-Pesqueira, Xiao-Yong Chen, Angela M. Chilton, Kristina M. Chomiak, Diego F. Cisneros-Heredia, Ibrahim K. Cisse, Aimee T. Classen, Mattheau S. Comerford, Camila Cordoba Fradinger, Hannah Corney, Andrew J. Crawford, Kerri M. Crawford, Maxime Dahirel, Santiago David, Robert De Haan, Nicholas J. Deacon, Clare Dean, Ek del-Val, Eleftherios K. Deligiannis, Derek Denney, Margarete A. Dettlaff, Michelle F. DiLeo, Yuan-Yuan Ding, Moisés E. Domínguez-López, Davide M. Dominoni, Savannah L. Draud, Karen Dyson, Jacintha Ellers, Carlos I. Espinosa, Liliana Essi, Mohsen Falahati-Anbaran, Jéssica C. F. Falcão, Hayden T. Fargo, Mark D. E. Fellowes, Raina M. Fitzpatrick, Leah E. Flaherty, Pádraic J. Flood, María F. Flores, Juan Fornoni, Amy G. Foster, Christopher J. Frost, Tracy L. Fuentes, Justin R. Fulkerson, Edeline Gagnon, Frauke Garbsch, Colin J. Garroway, Aleeza C. Gerstein, Mischa M. Giasson, E. Binney Girdler, Spyros Gkelis, William Godsoe, Anneke M. Golemiec, Mireille Golemiec, César González-Lagos, Amanda J. Gorton, Kiyoko M. Gotanda, Gustaf Granath, Stephan Greiner, Joanna S. Griffiths, Filipa Grilo, Pedro E. Gundel, Benjamin Hamilton, Joyce M. Hardin, Tianhua He, Stephen B. Heard, André F. Henriques, Melissa Hernández-Poveda, Molly C. Hetherington-Rauth, Sarah J. Hill, Dieter F. Hochuli, Kathryn A. Hodgins, Glen R. Hood, Gareth R. Hopkins, Katherine A. Hovanes, Ava R. Howard, Sierra C. Hubbard, Carlos N. Ibarra-Cerdeña, Carlos Iñiguez-Armijos, Paola Jara-Arancio, Benjamin J. M. Jarrett, Manon Jeannot, Vania Jiménez-Lobato, Mae Johnson, Oscar Johnson, Philip P. Johnson, Reagan Johnson, Matthew P. Josephson, Meen Chel Jung, Michael G. Just, Aapo Kahilainen, Otto S. Kailing, Eunice Kariñho-Betancourt, Regina Karousou, Lauren A. Kirn, Anna Kirschbaum, Anna-Liisa Laine, Jalene M. LaMontagne, Christian Lampei, Carlos Lara, Erica L. Larson, Adrián Lázaro-Lobo, Jennifer H. Le, Deleon S. Leandro, Christopher Lee, Yunting Lei, Carolina A. León, Manuel E. Lequerica Tamara, Danica C. Levesque, Wan-Jin Liao, Megan Ljubotina, Hannah Locke, Martin T. Lockett, Tiffany C. Longo, Jeremy T.

ACCEPTED MANUSCRIPT - CLEAN COPY

Lundholm, Thomas MacGillavry, Christopher R. Mackin, Alex R. Mahmoud, Isaac A. Manju, Janine Mariën, Deysi N. Martínez, Marina Martínez-Bartolomé, Emily K. Meineke, Wendy Mendoza-Arroyo, Thomas J. S. Merritt, Lila Elizabeth L. Merritt, Giuditta Migiani, Emily S. Minor, Nora Mitchell, Mitra Mohammadi Bazargani, Angela T. Moles, Julia D. Monk, Christopher M. Moore, Paula A. Morales-Morales, Brook T. Moyers, Miriam Muñoz-Rojas, Jason Munshi-South, Shannon M. Murphy, Maureen M. Murúa, Melisa Neila, Ourania Nikolaidis, Iva Njunjić, Peter Nosko, Juan Núñez-Farfán, Takayuki Ohgushi, Kenneth M. Olsen, Øystein H. Opedal, Cristina Ornelas, Amy L. Parachnowitsch, Aaron S. Paratore, Angela M. Parody-Merino, Juraj Paule, Octávio S. Paulo, João Carlos Pena, Vera W. Pfeiffer, Pedro Pinho, Anthony Piot, Ilga M. Porth, Nicholas Poulos, Adriana Puentes, Jiao Qu, Estela Quintero-Vallejo, Steve M. Raciti, Joost A. M. Raeymaekers, Krista M. Raveala, Diana J. Rennison, Milton C. Ribeiro, Jonathan L. Richardson, Gonzalo Rivas-Torres, Benjamin J. Rivera, Adam B. Roddy, Erika Rodriguez-Muñoz, José Raúl Román, Laura S. Rossi, Jennifer K. Rowntree, Travis J. Ryan, Santiago Salinas, Nathan J. Sanders, Luis Y. Santiago-Rosario, Amy M. Savage, J.F. Scheepens, Menno Schilthuizen, Adam C. Schneider, Tiffany Scholier, Jared L. Scott, Summer A. Shaheed, Richard P. Shefferson, Caralee A. Shepard, Jacqui A. Shykoff, Georgianna Silveira, Alexis D. Smith, Lizet Solis-Gabriel, Antonella Soro, Katie V. Spellman, Kaitlin Stack Whitney, Indra Starke-Ottich, Jörg G. Stephan, Jessica D. Stephens, Justyna Szulc, Marta Szulkin, Ayco J. M. Tack, Ítalo Tamburrino, Tayler D. Tate, Emmanuel Tergemina, Panagiotis Theodorou, Ken A. Thompson, Caragh G. Threlfall, Robin M. Tinghitella, Lilibeth Toledo-Chelala, Xin Tong, Léa Uroy, Shunsuke Utsumi, Martijn L. Vandegehuchte, Acer VanWallendael, Paula M. Vidal, Susana M. Wadgymar, Ai-Ying Wang, Nian Wang, Montana L. Warbrick, Kenneth D. Whitney, Miriam Wiesmeier, J. Tristian Wiles, Jianqiang Wu, Zoe A. Xirocostas, Zhaogui Yan, Jiahe Yao, Jeremy B. Yoder, Owen Yoshida, Jingxiong Zhang, Zhigang Zhao, Carly D. Ziter, Matthew P. Zuellig, Rebecca A. Zufall, Juan E. Zurita, Sharon E. Zytynska, Marc T.J. Johnson*

*Correspondence to: marc.johnson@utoronto.ca

This PDF file includes:

Material & Methods
Supplementary Text S1-S6
Figs. S1-S16
Table S1-S15 (legends only for Table S1-S9)

Other Supplementary Materials for this manuscript include the following:

Table S1-S9 (in xlsx format)

ACCEPTED MANUSCRIPT - CLEAN COPY

Materials & Methods

Study system

Trifolium repens L. (Fabaceae) is an herbaceous perennial native to Europe and western Asia (*22, 31*). It grows low to the ground, spreading via stolons to form clonal patches up to 1 m across (*32*). Plants reproduce sexually via pollination of self-incompatible flowers that are arranged in dense inflorescences (Fig. 1 inset). *Trifolium repens* originated in the Mediterranean 15 to 28 kya following hybridization between two diploid species, *T. occidentale* and *T. pallescens* (*31*), to form an allotetraploid. It has since expanded from its native range over the past several hundred years through anthropogenic distribution as a cover crop, fodder and for road stabilization. Today, it has a global distribution in temperate, continental, semiarid and tropical climates (*22, 32, 33*). While a recent allotetraploid, *T. repens* exhibits disomic inheritance of its two parental subgenomes (~500 Mb each, *31*). It thrives in anthropogenically modified habitats, particularly mowed grass and grazed pastures, although it requires supplemental watering in drought-prone (e.g., semi-arid) habitats to persist.

Populations of *T. repens* are often polymorphic for the production of hydrogen cyanide (HCN or cyanogenesis), an antiherbivore chemical defense that is present in over 3,000 plant species (*34*). Individual *T. repens* plants either produce hydrogen cyanide (i.e., cyanogenic, HCN+), or completely lack the ability to produce HCN (i.e., acyanogenic, HCN–) (*35, 36*). HCN is a toxic chemical defence produced when tissue is damaged, and it achieves its toxicity by binding to cytochrome oxidase in the mitochondrion where it inhibits cellular respiration (*37, 38*). The cyanogenesis polymorphism is controlled by the epistatic interaction between two independently segregating Mendelian loci (*36*). One encodes a biosynthetic cluster of three tightly-linked genes (hereafter referred to as *Ac*), which includes the cytochrome P450 enzyme *CYP79D15* that

ACCEPTED MANUSCRIPT - CLEAN COPY

catalyzes the initial step in the production of the cyanogenic glucosides linamarin and lotaustralin (*39, 40*). The second locus, *Li*, encodes the hydrolyzing enzyme linamarase (*17, 41*), which cleaves the glucoside's sugar moiety to liberate HCN. The cyanogenic phenotype exhibits complementary epistasis (*42*); plants require at least one functional allele at both loci to produce HCN (i.e., only genotype *Ac– Li–* is HCN+), and the expression of the functional alleles (*Ac* and *Li*) are partially dominant (i.e., heterozygotes have reduced expression) to the respective alternative recessive alleles (*36*). The recessive alleles of both loci are caused by partial or complete gene deletions (denoted *ac* for a deletion in the region containing *CYP79D15*, *li* for a deletion at *Li*). These deletions result in a loss of function for that gene copy and plants are acyanogenic when the recessive allele is homozygous at either locus (i.e., genotypes Ac–*lili, acac Li–*, and *acac lili* are all HCN–) (*17, 18, 39, 43*).

City selection

Our sampling was restricted to regions with suitable climates for *T. repens* (Fig. S1). We focused on cities that had clearly defined urban-rural gradients, and we avoided cities or transects within cities that had major elevational gradients. These criteria were necessary to avoid confounding variation unrelated to urbanization because temperature is an important driver of evolution in HCN production (*19, 44-46*), and HCN clines frequently occur along elevational gradients (*44, 47, 48*). We selected cities that had a minimum distance of 4 km from the urban center to the edge of contiguous suburban development, which allowed for a target of 40 sampling sites (hereafter referred to as "populations" or "sites" interchangeably for convenience) per city with a minimum spacing of 200 m between populations as explained below (see *Sample collection*). In total we sampled *T. repens* from 160 cities (Fig. 1): including 34 cities in Europe, 94 cities in North America, 13 cities in South America, 1 city in Africa, 10 cities in Asia, and 8 cities in

ACCEPTED MANUSCRIPT - CLEAN COPY

Oceania (Australia and New Zealand). Africa had the fewest cities sampled because white clover's distribution on the African continent is largely restricted to South Africa (*49*) (Fig. S1), and because South Africa experienced a prolonged historic drought at the time of sampling, which killed most plants in the region and prevented further sampling.

Sample collection

Trifolium repens samples were collected between 2016-2019 along transects that spanned urban to nonurban (typically rural, although occasionally forest, grassland or shrubland roadcuts in temperate, Mediterranean or semiarid sites) gradients (Fig. 2, Fig. S10). For each city, we aimed to sample between 20-50 *T. repens* populations (mean = 38.6 +/− 3.05 [SE] populations per city). Transect design and sample collection were performed using standardized protocols across all cities (Supplementary Text S2, Supplementary Text S3). Briefly, transects were designed to: 1) sample half urban (city and high-density suburbs) and half rural (periurban and rural) habitats, 2) include urban areas dominated by impervious surfaces, such as roads and buildings, and 3) avoid environmental gradients not directly associated with urbanization, such as changes in elevation or variation in proximity to large bodies of water (i.e., lakes, oceans, seas). The average distance between sampling sites was scaled to the total transect length, with the constraint that each site had to be a minimum of 200 m apart, which prevented sampling sites from overlapping since we could generally sample plants well within ~50 m of a central point. We aimed to sample 15-20 plants from each site (mean = 17.8 +/− 0.22 [SE] plants per site), in which individual plants were represented by 3-4 leaves attached to a single stolon that were stored in a cooler prior to returning to the lab. To avoid sampling from the same clone, each sample was separated by a minimum of 3 m. At the lab, individual samples were transferred to 1.5 mL microcentrifuge

ACCEPTED MANUSCRIPT - CLEAN COPY

tubes and stored at −80°C until HCN phenotyping (see, *HCN Feigl-Anger assays* and Supplementary Text S4). In total, we sampled 110,019 plants from 6,169 populations worldwide.

Environmental data

To determine how urbanization has shaped environmental change, and how urban-rural environmental gradients affect changes in HCN evolution within *T. repens* populations, we extracted environmental data from every sampled population using satellite images. This analysis involved performing data extraction based on pixel-level time series images that were then used for downstream analyses (Fig. S8). To do this, we used custom Python (v. 2.7.4) scripts to automate data extraction, processing and analysis, geoprocessing (e.g., data conversion/projection, spatial operations, and calculations of vector and raster datasets), data scraping, and map production. A detailed description of these methods and workflow is shown in Supplementary Text S1 and Fig. S8. A concise description of these methods follows below.

For each city, a vector dataset was created from population GPS coordinates (latitude and longitude, Table S1) using the World Geodetic System 1984 (WGS 84) coordinate system. We then selected two recent 30 m resolution Landsat images for each city using the USGS-EROS Earth Explorer interface (https://www.usgs.gov/core-science-systems/nli/landsat); one image from summer and one from winter, including snow cover where appropriate (Table S8, Table S9). We took the most recent images possible that met our selection criteria for ensuring high quality images and data (Supplementary Text S1)**,** with all images coming from the last 10 years, except for a single image in winter (Cincinnati: LE07_L1TP_020033_20030115_20160927_01_T1_MTL.txt) that came from 2003 because there was no recent clear image due to atmospheric interference. For each city, summer images were selected for the month with the highest mean temperature based on WorldClim data v.1.4

ACCEPTED MANUSCRIPT - CLEAN COPY

(*50*), and winter images were selected for the month with the lowest mean temperature. When no usable image was available for this time period (e.g., due to cloud/haze cover), we used the most recent image from the previous or following month. We used the projected coordinate datasets for each city to extract the following pixel-level environmental data for every sampled population from Landsat satellite images: summer land surface temperature (LST), winter LST, summer vegetation cover quantified using the normalized vegetation index (NDVI), winter NDVI, snow accumulation according to the normalized difference snow index (NDSI), and elevation above sea-level using a digital elevation model (DEM, Supplementary Text S1). 30 m resolution pixel values were averaged over a 100 m buffer around each collection site to obtain local environmental conditions for each population.

We extracted additional environmental variables from curated raster datasets. Specifically, we estimated population-level annual aridity (AI) and potential evapotranspiration (PET) using 30 arcsecond (~1 km) raster datasets provided by the CGIAR Consortium for Spatial Information (*51*). We also estimated percent impervious surface for each population using the 30 m resolution "Global Man-made Impervious Surface" raster dataset (GMIS) provided by the Socioeconomic Data and Applications Center (SEDAC). As with other variables, GMIS values were averaged over a 100 m buffer to characterize percent impervious surface in and surrounding each population (*52*).

City characteristics

In addition to population-level environmental variables extracted from each sampling location within a city, we examined how a city's relative age, area, proximity to other cities (hereafter "no. cities"), human population size, and human population density were related to the strength of HCN clines (Table S1, Table S7). To estimate a city's relative age, we first selected 30

ACCEPTED MANUSCRIPT - CLEAN COPY

random cities from our 160 sampled cities and used them to estimate the human population size of a hypothetical city with a 4 km radius, the minimum size of cities sampled. Across these 30 cities, we found that on average a city with a 4 km radius contained 150,000 people. The year at which each sampled city reached this population size threshold was determined from historical urban population growth data (*53*). For 11 of the 160 cities, no historical data was available prior to the city reaching this population threshold, therefore, a regression of population size versus year was used to extrapolate the year at which the threshold was reached. The relative age of each city was calculated as the number of years prior to 2020 a city reached a population size of 150,000; 52 cities had a population size below 150,000, so their age was set to 0 for analyses.

To estimate city area (km^2), Google Earth Pro (version 7.3.2, google.com/earth) satellite imagery was used to manually trace the boundaries between urban and rural areas (rural included periurban) using the most recent images available for each city as of December 2019. This manual determination of city area was necessary to ensure that estimates were based on consistent definitions in which urban and suburban habitats had a high level of impervious surface; estimates provided by governments are usually based on political jurisdictions and often include rural areas and may not include contiguous adjacent cities or suburbs. All mapping was performed using the Google Earth Polygon Tool with the viewing plane positioned directly above the city with the entire extent of the urban area in view. We then manually traced the margins of the city based on changes in impervious surface vs natural landcover in nonurban areas, such as vegetation, water, desert and farmland. Calculation of area excluded green spaces (e.g., parks, agricultural land, low-density suburbs, or undeveloped areas) with a diameter of > 3 km within the focal city boundary, as well as green spaces and water bodies on the border of urban areas. Neighboring urban areas were included in a city's area calculation when they were

ACCEPTED MANUSCRIPT - CLEAN COPY

separated by < 3 km of green space. This distance was chosen based on the limits of honeybee dispersal (*54*), one of the most common pollinators of white clover worldwide (*32, 55*, Fig. 1 inset).

The number of cities surrounding the focal city was calculated as the number of cities within a 50 km radius of the sampled city's geographic center. Using Google Earth we counted all cities (i.e., densely populated areas) with a radius of at least 4 km that partially or completely occurred within the 50 km radius circle. The focal city was also included in this count, so the minimum count was equal to 1.

To estimate human population size, we used the City Population Database (*56*). For each city, the specified city population was used from the most recent census data. Some sampled cities were not found on the City Population Database and were supplemented from local census authorities. Finally, human population density was calculated as the human population size divided by the city area.

HCN Feigl-Anger assays

The cyanogenic phenotype (i.e., HCN+ or HCN-), and thus the presence of at least one functional *Ac* and *Li* allele at each locus, was determined using a colorimetric assay (i.e., Feigl-Anger assay) following the protocol in Supplementary Text S4. Briefly, the Feigl-Anger assay uses filter paper soaked in a solution of copper ethylacetoacetate and tetrabase that turns green to blue in the presence of HCN (*57, 58*, Fig. S16). Tests show that the Feigl-Anger assay is >95% congruent with PCR amplification of *Ac* and *Li* alleles (*17, 43, 59, 60*), indicating it is an accurate and rapid method for determining the presence/absence of the functional alleles. We used Feigl-Anger assays to perform high throughput quantification of the frequency of cyanogenic genotypes (*Ac– Li–*) within populations. Whole genome sequencing of 2,074

ACCEPTED MANUSCRIPT - CLEAN COPY

samples confirmed that our phenotyping distinguished plants in which the *Ac*/*ac* alleles and *Li*/*li* alleles were present versus absent (Fig. S9), and was further able to distinguish individuals that were heterozygous vs homozygous (see *Allele frequencies at* Ac *and* Li).

Feigl-Anger assay papers were prepared at the University of Toronto Mississauga using a modified version of the protocol of Gleadow et al (*58*). Solutions of copper ethylacetoacetate and tetrabase were prepared in chloroform and combined at equal volumes. Square filter papers (Whatman #3) were dipped into the solution and air dried, then stored in an opaque container at 4°C to prevent UV degradation before use. Assay papers and 96-well plates were provided to each collaborator before sampling.

When performing the Feigl-Anger assay, a single large leaf (1-2 cm diameter), or two small leaves (< 1 cm) from an individual plant (Figs. S11, S13), was placed into a well of the 96-well plate. We placed leaves from each sampled plant into every other well, grouping individuals by population in plates so that a plate contained a maximum of 48 individuals, ensuring one well separated samples in each direction to prevent bleeding of the blue pigment that would obscure the results from other samples (Fig. S16). Leaves were frozen at −80°C (where possible, −20ºC otherwise) immediately prior to the assay, macerated in 80 µL water (Fig. S14), and incubated with Feigl-Anger assay paper placed evenly over the plate's wells for 3 h at 37°C (Fig. S15). After incubation, samples were immediately scored for the presence/absence of HCN; the presence of HCN was indicated by a pigmented spot (light green to dark blue) on the assay paper above the well, whereas a lack of any pigment indicated the absence of HCN (Fig. S16). The top and bottom of all filter papers were photographed and double-checked in M. Johnson's lab to ensure data consistency.

Tissue preparation & DNA extraction

ACCEPTED MANUSCRIPT - CLEAN COPY

Plant tissue was air-dried, dried using silica, or freeze-dried depending on the equipment available in collaborating labs and then shipped to the University of Toronto Mississauga for genomic DNA extraction. Leaves from each plant used for genomic library preparation (see below) were individually placed in strip tubes arranged in 96-well format, freeze-dried for 48 h in a Martin Christ Epsilon 2-6D LSC plus freeze dryer (Martin Christ, Osterode am Harz, Germany), then homogenized with three 2 mm beads per sample using a FastPrep-96 (MP Biomedicals, Solon, OH, USA) for 120s at 1,800 rpm. Homogenized tissue was stored at -80°C until DNA extraction.

Total genomic DNA was extracted with a modified phenol-chloroform extraction protocol (*61*). The detailed protocol for DNA extraction is outlined in Supplementary Text S5. Briefly, homogenized tissue was incubated at 60°C with a 3% CTAB buffer (2.5M NaCl, 2% polyvinylpyrrolidone (PVP), 1% B-mercaptoethanol) for 30 min. Phase separation washes were performed using phenol-chloroform-isoamyl alcohol (24:24:1; Sigma, P2069) and RNAse A (Fisher Scientific, Mississauga, Canada) incubation was performed for 30 min at 37°C prior to a pure chloroform (Sigma Aldrich, Oakville, ON, Canada) phase separation. DNA was then precipitated overnight in chilled isopropanol at −20°C, eluted in 50 µL TE (Fisher Scientific, Mississauga, Canada), and quantified on a Qubit Fluorometer 3.0 (Life Technologies, Mississauga, Canada) using the dsDNA BR Assay Kit (Fisher Scientific, Mississauga, Canada).

Genomic library preparation

Cities were selected for whole genome sequencing to represent geographic variation among continents, climatic variation within continents, and the presence or absence of urban-rural HCN clines. Our aim was to maximize variation in each of these three categories, with the constraint that a high proportion of the extracted genomic DNA from a city had to be of reasonably high

ACCEPTED MANUSCRIPT - CLEAN COPY

quality (i.e., high molecular weight DNA with low fragmentation when visualized on a 1.5% agarose gel) and all samples needed to have ≥ 10 ng/μL gDNA. Lower quality DNA typically resulted from tissue that was air-dried or incompletely dried before shipping to the GLUE Lead Team. Given these constraints, we prepared 2,074 samples from 26 cities: 7 cities from the native range of Europe and Middle East, and 19 cities from the introduced range, including 6 cities from North America, 5 from South America, 4 from Asia, 1 from Africa, and 4 cities from Australia and New Zealand (Oceania) (Fig. 1, Table S1). In each city, we sampled ~40 individuals from the five populations closest to the urban center and another ~40 individuals from the five furthest rural populations (N ~ 80 individuals per city), ensuring approximately equal representation of individuals from each population (~8 individuals/population). All individuals were diluted to 10 ng/μL in TE (where necessary) to a final volume of 25 μL in 0.2 mL Bioruptor Microtubes (Diagenode Inc., Denville, NJ, USA), and then sheared using a Bioruptor Pico Sonicator (Diagenode Inc., Denville, NJ, USA) for 3 cycles (15s on, 30s off) to achieve a mean fragment size of ~500 bp. When individuals had <10 ng/uL of DNA, we sheared 25 μL.of the undiluted sample.

Dual-indexed genomic DNA libraries were generated using Sera-Mag SpeedBeads (Fisher Scientific, Mississauga, ON, Canada) to clean DNA between each of the following reaction steps: 1) end repair, 2) A-tailing, 3) adapter ligation, and 4) index PCR with i5 and i7 indices (*62-64*). Our detailed protocol is outlined in Supplementary Text S6. Sheared DNA was first cleaned using a 0.8× SPRI bead:DNA ratio to remove fragments <250 bp. End-repair was performed by incubating sample DNA with a mixture of T4 DNA polymerase (Fisher Scientific, Mississauga, ON, Canada) and T4 polynucleotide kinase (Fisher Scientific, Mississauga, ON, Canada) to create 5’ phosphorylated blunt-ended fragments (*65*). Samples were then cleaned

ACCEPTED MANUSCRIPT - CLEAN COPY

with SPRI bead solution at a 2.8× bead:DNA ratio, which allowed us to retain the full distribution of fragment sizes. 3' A-tailing was performed using Taq polymerase (Fisher Scientific, Mississauga, ON, Canada) (*66*) and samples were then cleaned with SPRI bead solution at a 2.2× bead:DNA ratio. Adaptor ligation was performed using T4 DNA Ligase (Thermo Scientific, ON, Canada) to ligate 5 µM annealed iTrusR2-stubRCp and iTrusR1-stub adaptors (*64*). Samples were then cleaned with SPRI bead solution at a 0.9× DNA:bead ratio to remove adaptor dimers and eluted in TE. The indexing PCR was performed using unique 5 µM iTru5 and iTru7 primers (*64*) and Phusion HiFi polymerase (New England Biolabs, Ipswich, MA, USA) under the following conditions: 3 min at 98°C, (30s at 98°C, 30s at 65°C, 60s at 72°C) x 14 cycles, 1 min at 72°C. Finally, the PCR products were cleaned using a 0.8-1× SPRI bead:DNA ratio to remove primer dimers and other small fragments. All libraries were quantified on a Qubit using the high-sensitivity (HS) assay and run on a gel to visualize the size distribution of DNA fragments. We only sequenced genomic libraries with a minimum concentration of ≥ 0.8 ng/µL. Equimolar pools of 0.8 ng/µL were created for each library for sequencing.

Whole genome sequencing and bioinformatics

Sequencing, alignment, and quality control—We sequenced the genomes of 2,074 individual plants from the 26 cities on a Novaseq 6000 S4 platform using 150 bp paired-end reads (Table S1). 1,984 of these plants were sequenced at low coverage (mean = 1.05X) on four separate lanes, and 90 plants from Toronto were sequenced to approximately 10× as part of a separate project and downsampled to ~2.5× using *SAMtools* v1.10 (*67*) for inclusion in the present study. We trimmed raw reads using *fastp* v0.20.1 (*68*) with the *-trim_poly_g* argument to trim polyG tails that are commonly generated by Novaseq platforms. We performed per-sample quality

ACCEPTED MANUSCRIPT - CLEAN COPY

control (QC) of both raw and trimmed reads using *FastQC* v0.11.9 (*69*) and mapped the trimmed reads to the ~1 Gb *T. repens* reference assembly (NCBI BioProject number PRJNA523044, (*31*)) using *BWA MEM* v0.7.17 (*70*). We marked duplicate reads and sorted and indexed resulting bam files using *SAMtools* v1.10 prior to performing QC of mapped reads using *Qualimap* v2.2.2 (*71*), *Bamtools* v2.5.1 (*71*), *BamUtil* v1.0.14 (*72*), and multiQC (*73*). QC of the mapped reads revealed 18 samples with uncharacteristically high alignment error rates as reported by Qualimap. These samples might represent different species that were incorrectly identified as *T. repens* during sampling and were removed from all downstream analyses.

Site frequency spectrum—All population genomic analyses described below (see *Statistical Analyses*) were performed using genotype likelihoods to avoid biases associated with calling genotypes from low or variable coverage data (*74-76*). One- and two-dimensional folded SFS (for pairwise nucleotide diversity and F_{ST}, respectively) were estimated in *ANGSD* v0.933 (*76, 77*). For each city, we first generated separate urban and rural site allele frequency likelihood files (-doSaf 1) using the *SAMtools* genotype likelihood model (-GL 1, *78*) with base alignment quality scores recalculated according to the "extended SAMtools" model (-baq 2) to reduce false-positive SNP calls around INDELS (*79*). To ensure the same major and minor allele calls in both urban and rural habitats for F_{ST} estimates, we polarized alleles by forcing the reference base to be the major allele in each habitat (-doMajorMinor 4 with -ref). We only considered reads with a minimum phred-scaled base quality score of 20 (-minQ 20) and mapping quality of 30 (-minMapQ 30). We performed these analyses using a set of 250K randomly-selected four-fold degenerate sites (i.e., silent sites that result in no change in amino acid; -sites 4-fold random.sites) from across the genome as an estimate of neutral diversity, which we extracted from the *T. repens* genome using a custom script (https://github.com/tvkent/Degeneracy) that

ACCEPTED MANUSCRIPT - CLEAN COPY

requires *Python3* (*80*) and *bedtools* v2.26 (*81*) as dependencies. Finally, for each city, we used the *realSFS* script that is packaged with *ANGSD* to estimate separate folded (-fold 1) one-dimensional urban and rural SFS ($N = 52$), and two-dimensional urban-rural joint SFS (N = 26) for use in estimating urban and rural pairwise nucleotide diversity (π) and urban-rural F_{ST} within cities, respectively.

Statistical Analyses

We used a combination of general linear models and multivariate statistics to address our research questions related to whether global urbanization causes convergent environmental change, parallel evolution of HCN across cities, and to identify the environmental predictors of evolution. All analyses were performed in R version 3.6.3 (*82*).

Environmental convergence—We started by regressing each environmental variable separately against geographic distance from the urban center, to understand how environments changed along our sampling transects (Fig. S2, Table S2, Fig. S7). Distance was standardized between 0 and 1, where 0 represented the geographic center (i.e., most built up central area of a city) of the city, 1 was the most distant rural population, and the transition between urban/suburban areas to periurban and rural areas typically occurred at a distance of ~0.5. Standardizing distance in this way allowed all transects to be on an equivalent scale, regardless of city size. Environmental variables were standardized to a mean of zero and unit variance within each city, so that the differences in scale among cities would be removed.

Regressions were performed using robust regression fit with iteratively re-weighted least squares (IRLS) using the *rlm* function in the MASS package, which reduces the influence of potential outliers (*83*). In the absence of outliers, the standard ordinary least-squares regression (OLS) is the same as in the robust procedure, and comparisons of the two models yielded very

ACCEPTED MANUSCRIPT - CLEAN COPY

similar results. To maximize our ability to describe variation between urban and rural environments for our measures of urban environmental change, we used the predicted values for environmental variables from the robust regressions for the urban center (distance = 0) and furthest rural (distance = 1) locations.

To test whether urbanization leads to environmental convergence across the 160 cities, we calculated three multivariate attributes of environmental change between urban and rural habitats: (i) the magnitude of multivariate environmental change (Fig. 2A), (ii) the direction of environmental change (Fig. 2A), and (iii) variance in the environment among sampling sites within urban or rural habitats (Fig. S3). Together these metrics provide a comprehensive test of environmental convergence in response to urbanization. Due to missing environmental data from two cities, these analyses used 158 of the 160 cities studied.

The first attribute measured the magnitude of change between urban and rural habitats using a multivariate approach modified from Collyer and Adams (*84*). We conducted a PCA on a matrix containing the predicted values for the urban and rural habitats (i.e., 2× 158 cities) across all nine environmental variables. If convergence was strong, we expected urban and rural sites to form distinct clusters regardless of their cities. To test for the differences in multivariate mean values between urban and rural sites, we used the F-value from a MANOVA on the same matrix to test whether urban and rural sites significantly differed in their mean environmental conditions. To estimate the *P*-value, we used a permutation test as described in Collyer and Adams (*84*) based on bootstrapped residuals from the MANOVA model.

The second attribute measured the direction of change (i.e., angle) between urban and rural environments. For each city we estimated the multivariate angle between urban and rural sites across all environmental variables as in Collyer and Adams (*84*). Pairwise Euclidean distances

ACCEPTED MANUSCRIPT - CLEAN COPY

between cities in their angles were calculated and the total sum of distances was used as a test statistic. To estimate the *P*-value under the null hypothesis that cities did not differ in their direction of environmental change (i.e., they exhibited parallel environmental change), we estimated the MANOVA F-value and used the same bootstrapped-based permutation procedure from the MANOVA model described above. For each permutation, we re-calculated the multivariate angles and the sum of the Euclidean distances. Under the null hypothesis, this sum is smaller than if the alternative hypothesis is true that cities differ in the direction of change between urban to rural environments.

The third attribute was designed to measure changes in environmental variance between urban populations versus rural populations. For each city, we used the predicted values for the five sites closest to the urban center, and the five furthest rural sites across all nine environmental variables. We then created two matrices, one for the five urban sites and another for the five rural sites (i.e., 5 sites × 158 cities = 1580 by 9 variables per habitat). Each variable in the matrix was standardized (i.e., mean = 0, sd = 1) and converted into absolute values. The two matrices (i.e., urban and rural) were then combined into a single matrix of 3160 rows. A PCA was performed on the resultant matrix and if variances differed between urban and rural sites, we expected greater multivariate dispersion in one habitat relative to the other. To estimate the *P*-value under the null hypothesis that cities did not differ in their multivariate dispersion, we used a MANOVA of the matrix of absolute values with urban and rural habitats as the two groups. This is equivalent to a multivariate Levene's test for homogeneity of variance (*85*).

Parallel clines in HCN—To test for parallel evolution of HCN clines across cities, we first fit a generalized linear mixed-effects model to the population HCN frequencies of all cities. For each population in each city, we estimated the proportion of plants producing HCN, and used this as a

ACCEPTED MANUSCRIPT - CLEAN COPY

response variable in a generalized linear mixed-effects model (family = binomial) with standardized distance, continent, and the distance × continent interaction as fixed effect predictors (Table S3). In this model, we used the total number of plants in the population assigned as observation weights during fitting. The model additionally included random effects that allowed both baseline HCN frequencies (i.e., random intercept by city) and the slope of HCN frequencies versus distance to vary across cities (i.e., random slope by city), in which the latter quantifies the extent to which cities varied in the strength and direction of HCN clines. We obtained parameter estimates (i.e., β coefficients) directly from the output of *glmer* in the *lme4* package (*86*), while *P*-values for fixed effects were obtained using the *Anova* function from the *car* package (*87*), fit with type III sums-of-squares because of interactions being present in the model, and for the random effects using a likelihood ratio test implemented in the *anova* function of *lmerTest (88)*. We ran similar models using percent impervious surface (global manmade impervious surface–GMIS) or the Human Influence Index (HII, *89*) as predictors instead of distance to the city center. HII is a global dataset of 1 km grid cells based on data describing human population density, human land use, and infrastructure. The results from all three sets of models were qualitatively identical and models that included distance explained the most variation in HCN frequencies. We did not run an analysis including all three predictors in a single model (i.e., distance, GMIS and HII), because these urban metrics were correlated ($|r_{Pearson}|$ range = 0.35-0.45). We therefore only interpret models with distance as a predictor in the main text, and include results using distance, GMIS or HII as predictors individually in Table S3.

In addition to the mixed model described above, we performed separate binomial regressions quantifying the change in HCN frequency across each city's urbanization transect. For each city, we fit a binomial regression with population-mean HCN frequencies as the response variable,

ACCEPTED MANUSCRIPT - CLEAN COPY

and standardized distance to the city center as the sole predictor. Binomial regressions were performed using the *glm* function in the *stats* package (*82*) with the total number of plants in the population assigned as observation weights during fitting. We used these models to evaluate the extent to which cyanogenesis clines are consistent in magnitude and direction across cities (Table S4). We used the $\log_e$(odds) slopes extracted from the full GLMM model run in *glmer* when evaluating the environmental predictors of variation in clines across cities, as described below.

Diversity and population differentiation—To contrast the effects of genetic drift between urban and rural habitats, we estimated genome-wide nucleotide diversity (π) at 4-fold degenerate sites separately for the urban and rural plants from each city using the folded, one-dimensional SFS files as priors to the empirical Bayes estimation of π implemented in *ANGSD* (*77*). The number of sites used in estimating 4-fold π varied across cities and ranged from 191,599 ($\text{Hiroshima}_{\text{rural}}$) to 236,280 ($\text{Toronto}_{\text{urban}}$), with a mean of 229,900. To test whether the urban-rural differences in π deviate from null expectations, we implemented a permutation test by randomly reshuffling individuals between habitats in proportion to their observed population sizes, re-estimating the urban-rural difference in π, and repeating this procedure 100 times. We calculated *P*-values as the number of permuted values that were larger or smaller than the observed difference in π (i.e., the *P*-value was the quantile). This analysis tests whether the observed urban-rural difference in π is greater or less than expected if all individuals were from a single population (Fig. S4A).

To assess how urbanization might affect genetic differentiation, we estimated urban-rural F_{ST} for each city from the folded, two-dimensional urban-rural joint SFS using Hudson's estimator of F_{ST} (*90, 91, realSFS fst index -whichFst 1*). We used Hudson's estimator because Weir and Cockerham's F_{ST} (-whichFst 0) can be inflated at low sample sizes (*6*) and biased when

ACCEPTED MANUSCRIPT - CLEAN COPY

replication is uneven (*90*). The number of sites used in estimating F_{ST} ranged from 186,401 (Hiroshima) to 235,166 (Toronto) with a mean of 225,785. We performed a similar permutation test as above to assess whether the observed urban-rural F_{ST} is greater or less than expected if all individuals were randomly assigned to urban or rural habitats in proportion to the original sample (Fig. S4B). The *P*-value was again calculated as the quantile of the obverserved F_{ST} compared to the permuted neutral distribution.

As an alternative measure of neutral genetic differentiation, we calculated the Euclidean distance between urban and rural centroids for each city from a PCA (Fig. S5). Using the same randomly-selected 4-fold sites as above, we used *ANGSD* to first estimate genotype likelihoods for all samples, with minor allele frequencies estimated directly from these likelihoods (-doMajorMinor 1, *92*). We used the same site and read filters as above with the following additions: (1) we only included sites with MAF > 0.05; (2) we only included SNPs with a *P*-value less than 1e-6 (-SNP_pval 1e-6); (3) we excluded sites where >50% of samples were missing data (-minInd 0.5*N); and (4) we excluded sites with a total depth greater than 4200× (i.e., 2 × mean coverage × # samples), which excluded approximately 2% of sites likely corresponding to highly repetitive genomic elements. Following filtering, we retained 1,616 four-fold degenerate SNPs (genome-wide) and used these as input to PCAngsd v0.99 (*93*), which estimates a covariance matrix of allele frequencies across input samples directly from genotype likelihoods. We imported this covariance matrix into R, performed a PCA using the *princomp* function, and extracted eigenvectors using the *eigen* function. For each city, we then calculated the urban and rural centroids as the mean of the within-habitat sample positions across the first two PCs, which together explained ~25% of the variation in allele frequencies (PC 3 explained

ACCEPTED MANUSCRIPT - CLEAN COPY

0.5% of the variance), and calculated the Euclidean distance between the urban and rural centroids for each city as:

$$Distance = \sqrt{(PC1_{urban} - PC1_{rural})^2 + (PC2_{urban} - PC2_{rural})^2}$$

This approach is similar to that implemented by Stuart et al. (see Fig. 4D in *94*)

Finally, we estimated admixture proportions using PCAngsd as a complementary analysis of urban-rural differentiation (Fig. S6). For each city, we estimated genotype-likelihoods in *ANGSD* using the same filtering criteria as above with the following minor modifications: (1) we did not impose a maximum depth filter due to the relatively fewer number of samples per city; (2) we excluded really low coverage samples in each city (< 0.2X) since these resulted in lots of missing data; and, (3) we excluded sites when >50% of remaining individuals within a city were missing data. The number of sites used in estimating admixture proportions ranged from 1,085 (Hiroshima) to 11,369 (Toronto), with a mean of 4,640. We then used these genotype-likelihoods as input in PCAngsd with the *-admix* option to estimate per-sample admixture proportions for each city. PCAngsd implements a minimum partial average (MAP) test to determine the number of PCs (D) required to estimate individual allele frequencies, and the optimal K used for estimating admixture proportions can be estimated as D + 1 (*59*). Optimal K was 2 for every city.

Differentiation of HCN relative to neutral expectation— We examined whether clinal cities showed on average greater urban-rural differentiation in HCN than expected if the HCN phenotype was encoded by two randomly-selected neutral sites from across the genome. The null hypothesis is that urban-rural differentiation observed in HCN could be driven by neutral differentiation in the genome. However, the dominant epistatic genetic architecture of HCN means that there may not be differentiation in both of the underlying *Ac* and *Li* loci where clines are observed. Therefore, we estimated the distribution of HCN differentiation that would be

ACCEPTED MANUSCRIPT - CLEAN COPY

expected in HCN if a random pair of neutral SNPs controlled the HCN phenotype and we compared the null distribution to the observed differentiation in HCN. First, we estimated urban and rural HCN frequencies using the predicted values from the binomial regressions run separately for each city. We then used these frequencies to estimate observed urban-rural differentiation in the HCN phenotype as:

$$HCN_{diff} = \frac{H_S - H_T}{1 - H_T}$$

In this way, HCN_{diff} is analogous to F_{ST}. H_T is the fraction of times a pair of samples from the total population (city) are the same phenotype. H_S is the mean fraction of pairs sharing a phenotype from within the same sub-population (urban or rural) (*95, 96*). H_T and H_S were estimated as:

$$H_T = \sum_{i=1}^{I} (\frac{1}{K} \sum_{k=1}^{K} p_{ki})^2$$

$$H_S = \frac{1}{K} \sum_{k=1}^{K} \sum_{i=1}^{I} p_{ki}^2$$

where K is the number of subpopulations (two in our case—urban and rural), I is the number of distinct phenotypes (also two in our case – HCN+ and HCN–), and p_{ki} is the frequency of phenotype i in subpopulation k.

To generate a null distribution against which we could compare our observed differentiation in HCN, we first used *ANGSD* to identify 4-fold SNPs across all samples within a city. We then extracted these SNP positions and estimated their frequencies separately in each urban and rural habitat for each city. This approach ensures that we include sites that are fixed in one habitat but variable in the other. We estimated allele frequencies (-doMaf 1) using the *samtools* genotype likelihood model (-GL 1) and forced the major allele to be the reference base in both habitats (-

ACCEPTED MANUSCRIPT - CLEAN COPY

doMajorMinor 4 with -ref). We only considered reads with a minimum phred-scaled mapping quality of 30 (-minMapQ 30), recalibrated base qualities using the extended samtools model (-baq 2), and only considered sites with a minimum phred-scaled base quality of 20 (-minQ 20) for sites with high probability of being polymorphic (-SNP_pval 1e-6). For each city, we only retained sites with frequencies estimated for the same minor allele in both habitats and proceeded to generate a null distribution.

To compare our observed estimate of HCN_{diff} to a null distribution of HCN_{diff} values generated from genome-wide 4-fold sites with a similar genetic architecture to HCN, we randomly sampled two 4-fold sites and assigned one to represent "*Ac*" and the other "*Li*". We then estimated what urban and rural "HCN" frequencies would be from these pseudo-*Ac* and *Li* frequencies as:

$$p_{HCN} = (2p_{Ac}q_{Ac} + p_{Ac}^2) \times (2p_{Li}q_{Li} + p_{Li}^2)$$

where p_{Ac} and p_{Li} are the pseudo-*Ac* and *Li* frequencies, respectively (and q_{Ac} and q_{Li} are their complements). From these urban and rural pseudo-HCN frequencies, we calculated the urban-rural null HCN_{diff} using the same approach as above and repeated this process 1000 times to generate a null distribution of HCN_{diff}. For each city, we assessed whether HCN was more differentiated than expected under neutrality by calculating the proportion of null HCN_{diff} values that were greater or equal to the observed urban-rural HCN_{diff} in HCN. Finally, we performed a χ^2–test to examine whether cities with significant clines in HCN were more likely to show significant divergence in HCN_{diff} relative to neutral expectations.

Allele frequencies at Ac and Li—We developed a novel genotyping method to estimate allele frequencies at *Ac* and *Li* for low coverage data. The method is based on the knowledge that HCN production requires the combination of functional proteins encoded by the *Li* and *CYP79D15* loci, whereas acyanogenic plants carry large deletions of both copies at one or both loci. The

ACCEPTED MANUSCRIPT - CLEAN COPY

upstream boundaries of these deletions are unknown, which has prevented the development of simple genotyping assays for their presence/absence (*18, 43*). We developed an assay to assign genotype likelihoods to individual plants and estimate allele frequencies based on the sequence read counts in the *Li* and *CYP79D15* genes and their flanking regions.

Individuals with two copies of the intact locus (+/+) should have a standardized read count consistent with the mean standardized read count of the rest of the genome (1.0×), while those carrying one (+/-) or two (-/-) deleted alleles would have 0.5× and 0.0×, respectively. Because the boundaries of the deletions are unknown, we defined "diagnostic regions" for the two loci by comparing samples known to be homozygous for the deletion to samples known to carry at least one intact allele. For each locus, a region surrounding the gene was at low coverage in the homozygous deletion samples relative to the other samples (Fig. S9). We therefore defined the diagnostic region for each locus that would be used to determine the genotype likelihoods of each sample. For *Ac*, the region was CM019103.1:19559221-19573344, where the 5' end abutted a scaffolded region in the genome comprised of ambiguous 'N' nucleotides (Fig. S9 A, B). For *Li* we used the region CM019108.1:30218214-30247247 after masking a repeat at CM019108.1:30229250-30230911 (Fig S9 C, D).

The number of reads a given sample has in the diagnostic region is a function of its genotype, the total number of reads sequenced, the size of the focal region, alignment errors, and sampling error associated with sequencing. Initial observations indicated that the amount of sampling error was greater than that expected under a Poisson distribution (i.e., it was overdispersed). We therefore modelled read counts using a negative binomial distribution, where we could estimate the mean and variance of the number of reads expected for a given genotype. We fit these parameters using the following procedure: for each plant we counted the number of reads falling

ACCEPTED MANUSCRIPT - CLEAN COPY

in the focal region and normalized the read counts by the total unique aligned reads in that sample. When these read counts were plotted as a histogram, three peaks represented the read counts for plants that are -/-, +/-, +/+. The mean of each peak reflected the number of reads expected in a given genotype. The height of each peak reflected the genotype frequencies in the sample and the width (variance) of the peaks represented sampling noise around the expected number of reads for a given genotype. Using non-linear least-squares regression, we fit this trimodal distribution as a combination of three negative binomial distributions that correspond to the three genotypes with 0.0×, 0.5× and 1× probability of a read mapping in the diagnostic region. Our fitted distribution estimates the mean and variance in the number of reads expected to map for each genotype as well as the frequency of the deletion in the population, which is used to calculate the number of samples in each peak assuming Hardy-Weinberg equilibrium.

Using the parameters from the fitted model, and the read counts from each sample, we calculated the likelihood that each sample was -/-, +/-, and +/+ at each locus. The likelihoods were normalized such that they add to 1.0 for each sample, allowing us to use the normalized likelihoods as estimates of genotype or allele frequencies for groups of samples.

Differentiation at Ac and Li relative to neutral expectations—We used the allele frequencies at *Ac* and *Li* above to examine whether individual loci underlying HCN production were significantly differentiated relative to genome-wide neutral expectations. We first estimated observed urban-rural differentiation at *Ac* and *Li* using Hudson's F_{ST} (*90*) separately for each city. We then compared these observed values to a null distribution of Hudson's F_{ST} values for all 4-fold degenerate SNPs along the same chromosomes as the *Ac* and *Li* loci (CM019103.1 for *Ac* and CM019108.1 for *Li*) and considered the locus a significant outlier if it was in the top 2.5% of this distribution. Finally, we performed a χ^2-test of independence to assess whether cities with

ACCEPTED MANUSCRIPT - CLEAN COPY

significant clines in HCN are more likely to show differentiation in at least one locus, which is appropriate given that differentiation at a single locus is sufficient to drive clines in HCN production because of its epistatic genetic architecture (*25, 26*).

Environmental predictors of clines in HCN—To evaluate the environmental predictors of variation in the strength of clines, we used regularized regression (elastic net) to identify the combination of environmental variables that best predict the strength of clines. We fit a single multivariate model with the slopes of urban-rural HCN clines across all cities as the response variable (i.e., one slope value per city, extracted from the full GLMM model) and the following predictors: regional means (i.e., mean environmental condition across all urban and rural sites sampled) of all nine environmental variables, slopes depicting the change in these nine environmental variables across urban-rural transects (i.e., slopes of each environmental variable vs distance from the urban center, calculated for each city), all two-way interactions between environmental slopes, and all two-way interactions between environmental slopes and regional environmental means. We did not include two-way interactions between different regional mean environmental variables because we were primarily interested in understanding how changes in the environment across urban-rural transects drive changes in HCN evolution. From this saturated model, we used elastic net regularization to identify predictors with little to no effect on the response relative to remaining predictors. Elastic net regression combines LASSO regression and ridge regression into a single method of analysis that allows for parameter estimation and model selection (*97*). This approach penalizes large coefficients, shrinking them to zero to avoid model overfitting and high variance in parameter estimates, with the strength of shrinkage determined by the tuning parameter λ; when λ equals 0, the shrinkage penalty has no effect, and the estimates are the same as those obtained from OLS regression. One additional "mixing"

ACCEPTED MANUSCRIPT - CLEAN COPY

parameter α, determines whether the shrinkage method is more like LASSO or ridge regression. λ and α were estimated using 10-fold cross-validation implemented in the *caret* package (*98*), which selects λ and α values that minimize the root mean-squared error of the model fit, with HCN slopes as the response variable and the matrix of environmental variables described above as predictors.

Because independent runs of elastic net models can provide variable coefficient estimates, especially in the presence of collinearity, we repeated the elastic net model selection 100 times, each with a different random seed. We then estimated the number of models in which each environmental predictor was non-zero; important predictors should appear in many models, while uninformative predictors should be shrunk to zero in most models. We then estimated the average coefficient for each environmental predictor across the 100 models. Because all predictors were standardized prior to analysis, comparison of these model coefficients provides a direct estimate of the relative importance of different environmental predictors. Of the 136 initial environmental predictors, 11 were retained in at least one elastic net model, while the rest were consistently shrunk to zero. Our final model retained 6 predictors that appeared in >75% of the 100 models (Table S5).

As a complement to the elastic net model selection approach, we performed a principal component (PC) regression to evaluate the environmental predictors of variation in the strength of clines across cities (Table S6). This approach reduces the dimensionality of the multivariate data into PCs based on the covariance among predictor variables, and then tests how the PCs predict the slope of HCN clines. We began by performing two separate PCAs: the first was a PCA on the regional means (i.e., mean environmental condition across all urban and rural sites sampled per city) of all nine environmental variables (Fig. S7A), while the second was a PCA on

ACCEPTED MANUSCRIPT - CLEAN COPY

the slopes depicting the change in these nine environmental variables across urban-rural transects (i.e., slopes of each environmental variable vs distance from the urban center, calculated for each city) (Fig. S7B). For both PCAs, we retained the optimal number of PCs based on the broken-stick model, which selects PCs that explain more variation in the component variables than expected under neutrality (*99*). This resulted in retention of two PCs (variance explained: PC1: 42%; PC2: 22%) for the environmental-means PCA (Fig. S7A), and three PCs for the environmental-slopes PCA (variance explained: PC1: 36%; PC2: 21%; PC3: 19%) (Fig. S7B).

Using the PCs retained above, we created a model that included each PC as main effect continuous predictors, and two-way interactions between environmental slopes PCs and between environmental mean and slope PCs. We did not include two-way interactions between environmental-mean PCs because we were primarily interested in understanding how changes in the environment across urban-rural transects affect changes in HCN evolution. To obtain our final model, we used a corrected Akaike Information Criterion (AIC_c)-based multi-model selection and averaging process, whereby models with differing combinations of predictors were ranked by AIC_c using the *dredge* function from the *MuMIn* package (*100*), and all models within 2 AIC_c units of the best fitting model were averaged using *model.avg*. We interpreted the "full" model coefficients from the averaging process, which are more conservative and reported in Table S6 (*100*).

City characteristics— To test how city characteristics affected the slope of HCN clines, we performed multiple regression. Four of the five predictor variables (i.e. city area, relative city age, human population size, human population density) were highly skewed, so we log-transformed these variables, which substantially improved the homogeneity of variance of the full regression model and marginally improved normality compared to no transformations. We

ACCEPTED MANUSCRIPT - CLEAN COPY

also inspected the predictor variables for collinearity and removed population size because it was highly correlated ($r_{Pearson} > 0.79$) with city area and relative city age. City area and relative city age were also highly correlated ($r_{Pearson} = 0.74$), so we ran alternative models with one or the other variable included. The full model was estimated as: HCN slope = log(city area) + no. cities + log (population density) + error, implemented using type II sums-of-squares since there were no interactions in the model. We substituted log(city age +1) for log(city area) and reran the analyses. Because a large number of cities (N = 52) had a relative city age of 0, we excluded these data in a third model and reran the analyses. No significant relationships were detected between the slopes of the HCN clines and any of the city characteristics in the models described above (Table S7).

ACCEPTED MANUSCRIPT - CLEAN COPY

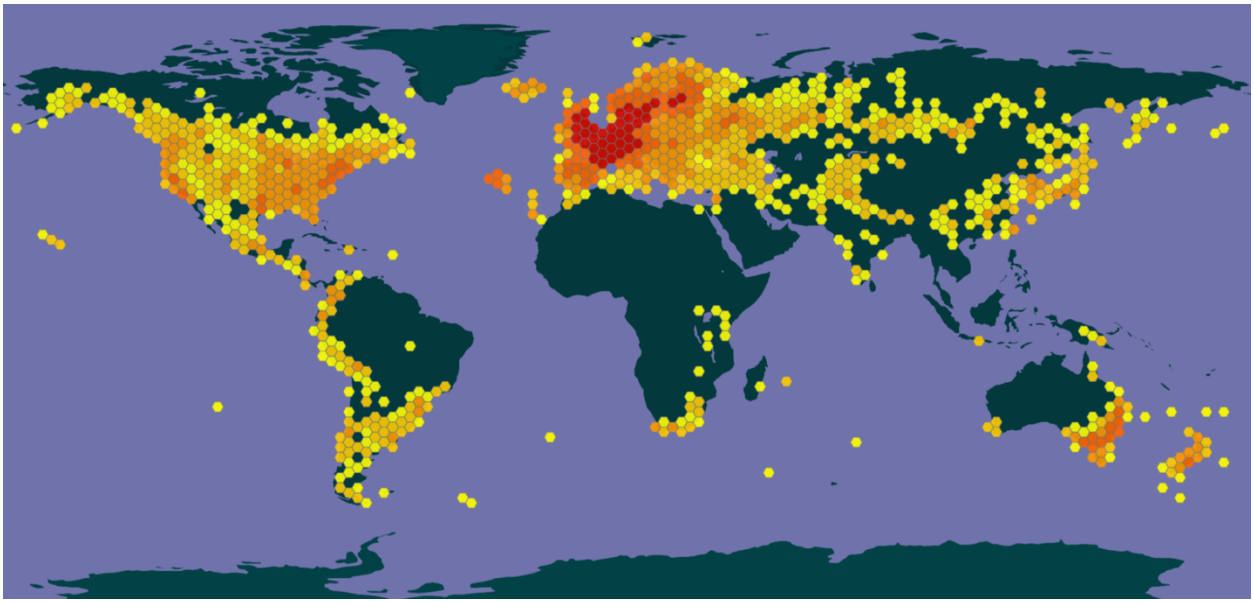


Fig. S1. Distribution of white clover (*Trifolium repens* L.) as reported by GBIF (*49*). The distribution is shown as a heat map (yellow = fewest records, red = most records) based on 946,694 georeferenced records of *T. repens* plants between 1756 and 2021 (downloaded July 15, 2021).

ACCEPTED MANUSCRIPT - CLEAN COPY

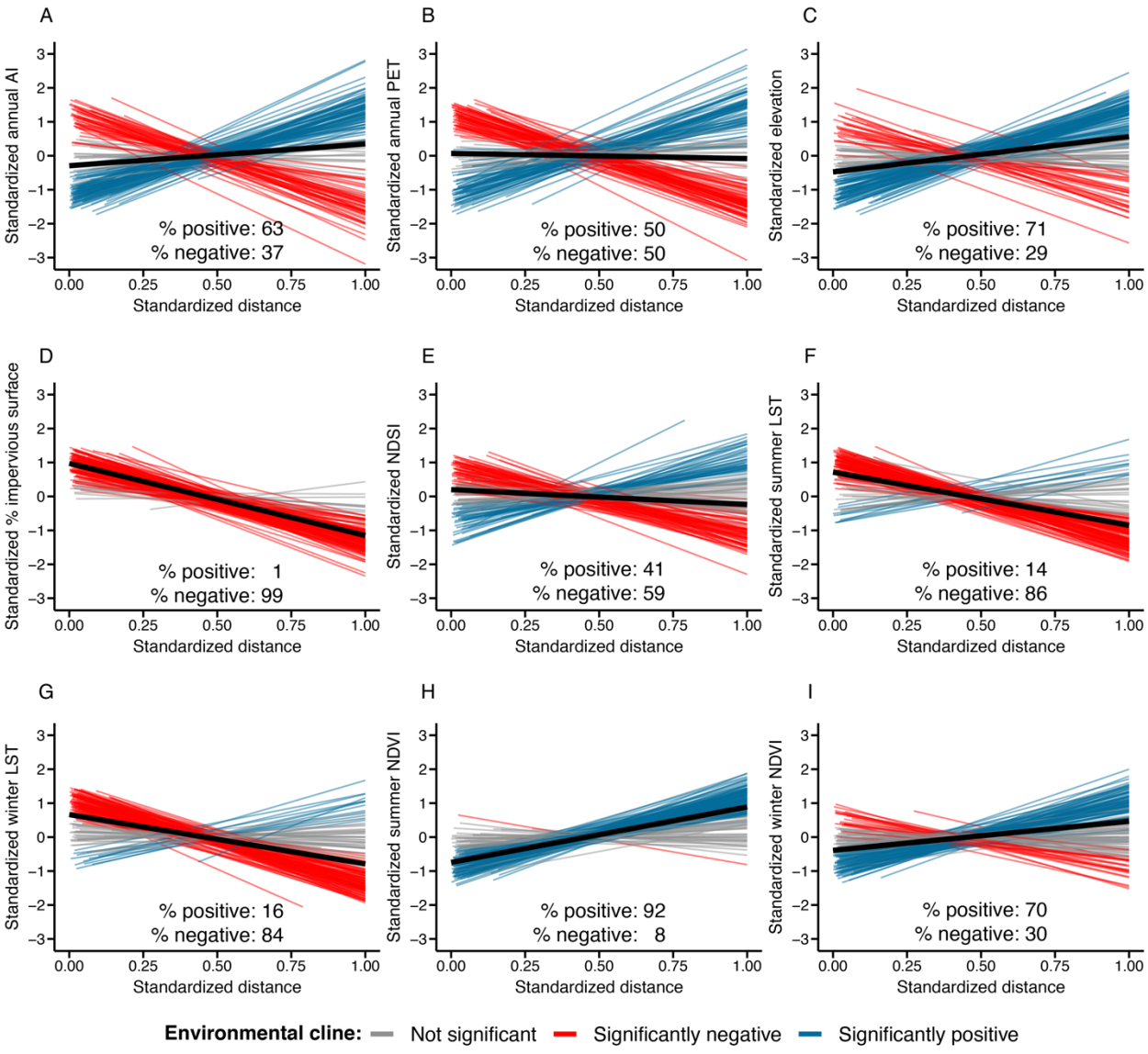


Fig. S2. Changes in urban environments along urban-rural transects. Thin lines show the estimated relationship between: (**A**) annual aridity index (AI-lower values correspond to higher aridity), (**B**) annual potential evapotranspiration (PET), (**C**) digital elevation model above sea level (DEM), (**D**) global manmade impervious surface (GMIS or % impervious surface), (**E**) normalized difference snow index (NDSI), (**F**) land surface temperature in summer (LST), (**G**) LST in winter, (**H**) normalized difference vegetation index (NDVI) in summer, (**I**) NDVI in winter, and distance from the city center for each of the 160 cities. Lines are coloured with respect to the direction and significance of linear relationships (P<0.05) as shown in the figure legend; the thick black line in each panel shows the mean slope across all cities. Distance was standardized to vary between 0 (urban center) and 1 (furthest rural population) as described in the Methods and environmental variables were standardized to a mean of 0 and variance of 1. Results for each city and environmental variable are reported in Table S2.

ACCEPTED MANUSCRIPT - CLEAN COPY

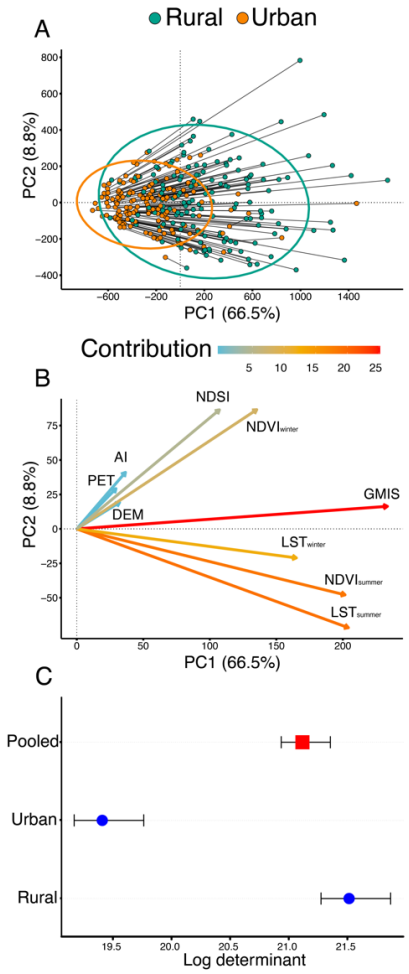


Fig. S3. Environmental variance among populations within urban and rural habitats. Multivariate environmental variance was calculated among five populations within each habitat for every city, and the equality of the environmental variance was compared among habitats. (**A**) Principal components of the environmental variance among urban sites and among the furthest rural sites along each urban-rural transect; ellipses and connecting lines as in Fig. 2A. (**B**) The contribution of each environmental variable to environmental variance within each habitat, with loadings coloured by their contribution to PC1, which was associated with the greatest differences in environmental variance among urban populations and among rural populations. The environmental variables included winter ($NDVI_{winter}$) and summer ($NDVI_{summer}$) normalized different vegetation index, snow accumulation quantified using the normalized difference snow index (NDSI), winter (LST_{winter}) and summer (LST_{summer}) land surface temperature, aridity index (AI; lower values denote greater aridity), annual potential evapotranspiration (PET), percent global manmade impervious surface (GMIS), and elevation above sea level determined from the digital elevation model (DEM). (**C**) In addition to the multivariate Levene's test, we calculated the log-determinant; higher values correspond to greater among-population environmental variance within a habitat. The points show the multivariate mean ± 95% CI. The environmental variance within rural habitats is significantly larger than urban habitats based on non-overlapping CI.

ACCEPTED MANUSCRIPT - CLEAN COPY

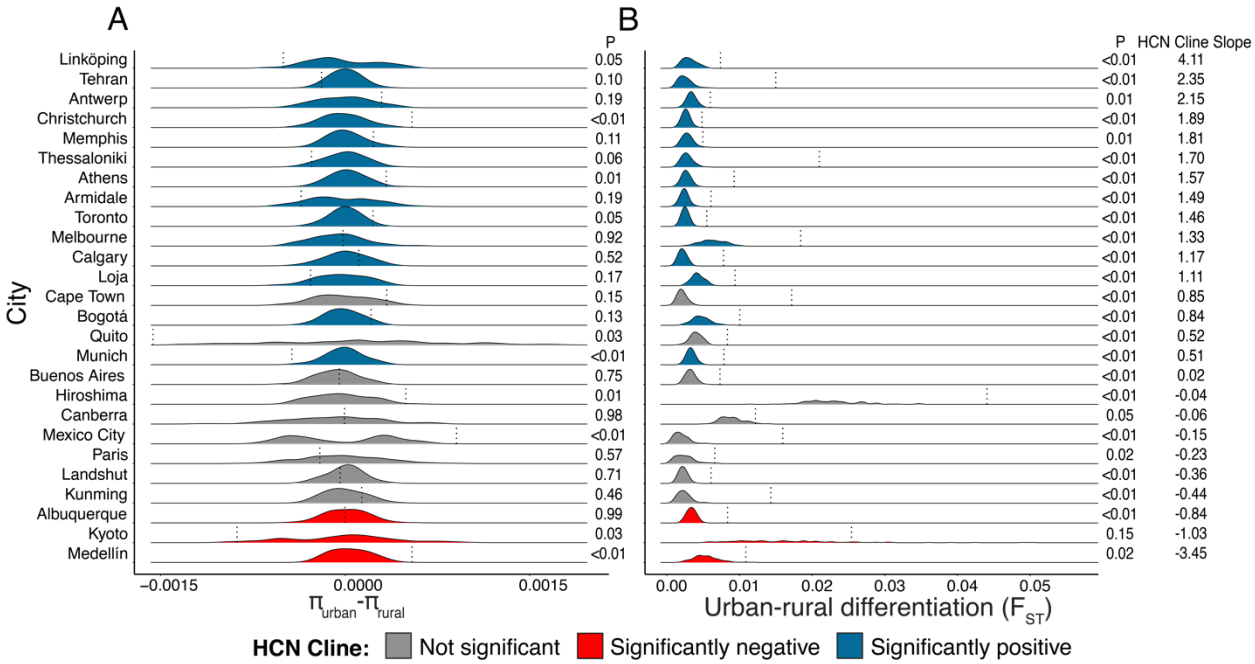


Fig. S4. Genetic variation within and between urban and rural habitats. (**A**) The observed difference in π (dashed vertical line) between habitats versus the null distribution of the difference in π within each city. Nucleotide diversity was estimated using 250,000 putatively neutral 4-fold degenerate SNPs across the genome. (**B**) The observed F_{ST} (dashed vertical line) between habitats versus the null distribution of F_{ST}. The *P*-values indicate the quantile of the observed (**A**) difference in π ,or (**B**) F_{ST} between habitats compared to the null distribution that individuals are sampled from the same population based on 100 permutations. Cities are arranged in both panels from top to bottom with decreasing values of the $\log_e$(odds) slope of the regression of HCN frequency versus distance. The colours of distributions indicate the statistical significance (P<0.05) and direction of each cline as indicated in the legend.

ACCEPTED MANUSCRIPT - CLEAN COPY

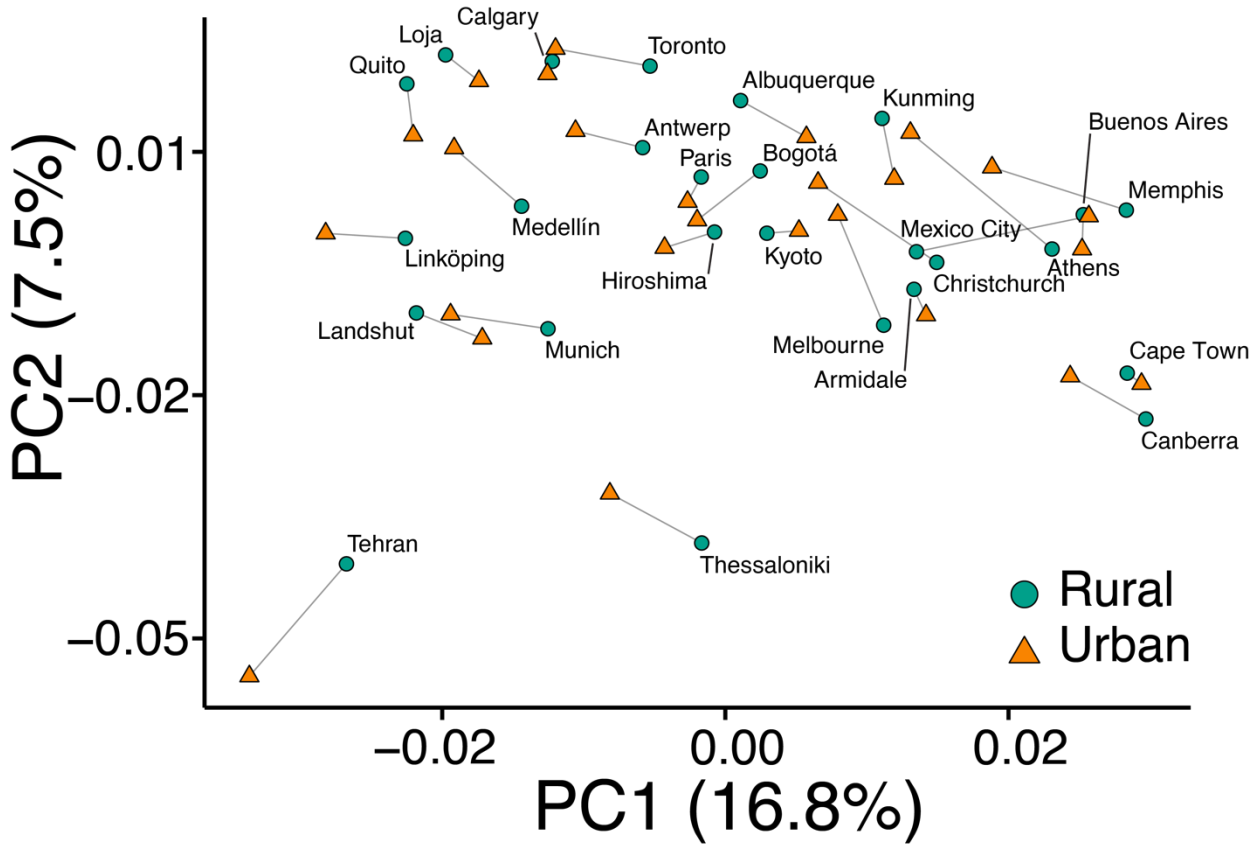


Fig. S5. Principal components depicting genomic variation across all cities and habitats. Genetic differences among individuals were determined using PCA (~80 individuals/city, split equally between urban and rural habitats), and the centroids of the ~40 individuals within each habitat are shown. The urban and rural centroids from the same city are connected by lines, and each city is labeled with respect to its rural habitat. The first two axes explained 24.3% of the variance in the global sample of genomic variation; PC3 and PC4 explained 4.5% and 4.0% of the variance, respectively, and is not depicted.

ACCEPTED MANUSCRIPT - CLEAN COPY

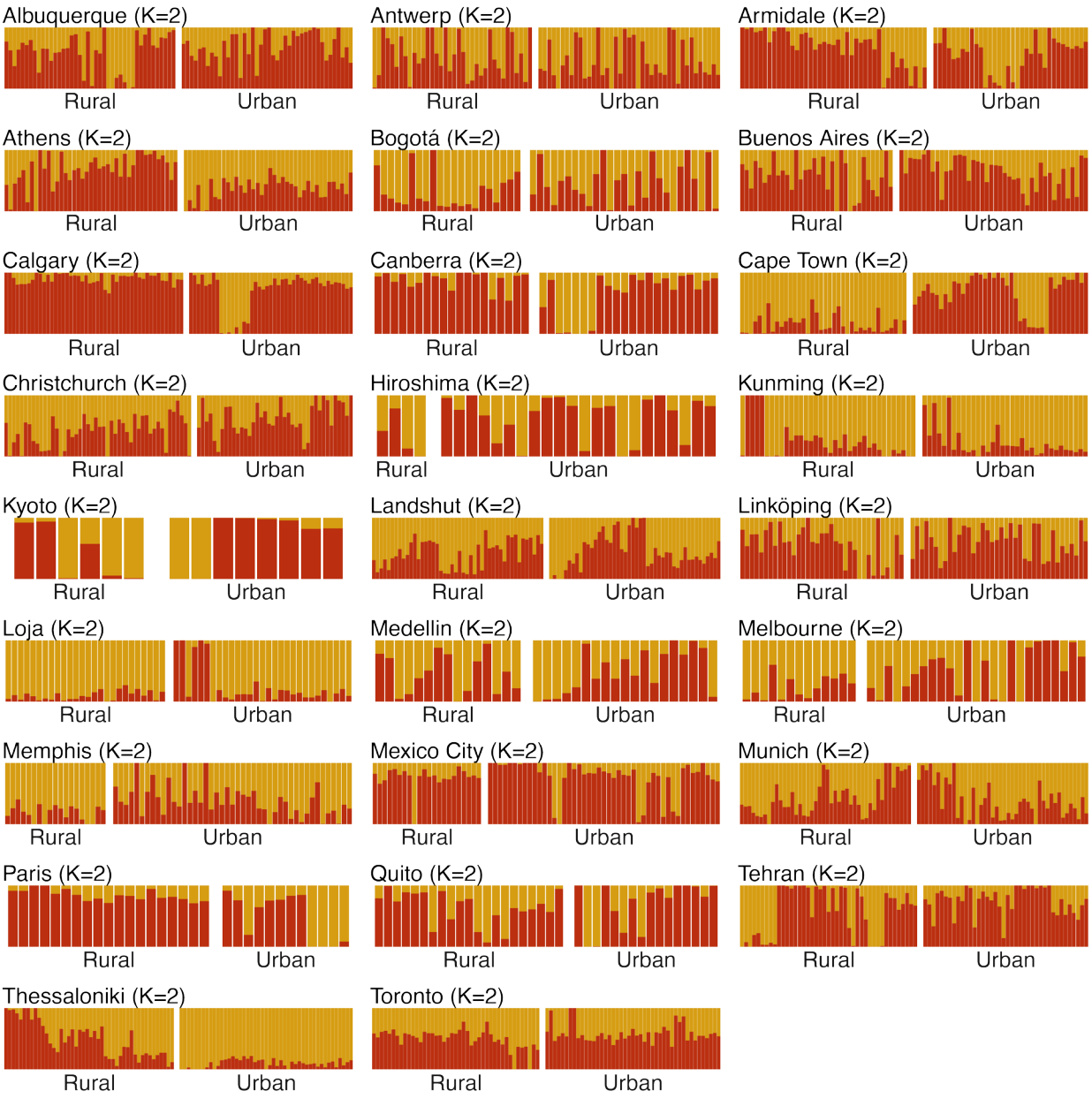


Fig. S6. Admixture proportions between urban and rural populations for each city estimated using PCAngsd. Optimal K for each city was selected as one plus the number of principal components used to estimate individual allele frequencies; optimal K was 2 for every city. Each coloured bar represents an individual, with the height of the colours indicating the proportion of an individual's genome derived from ancestral populations within a given city.

ACCEPTED MANUSCRIPT - CLEAN COPY

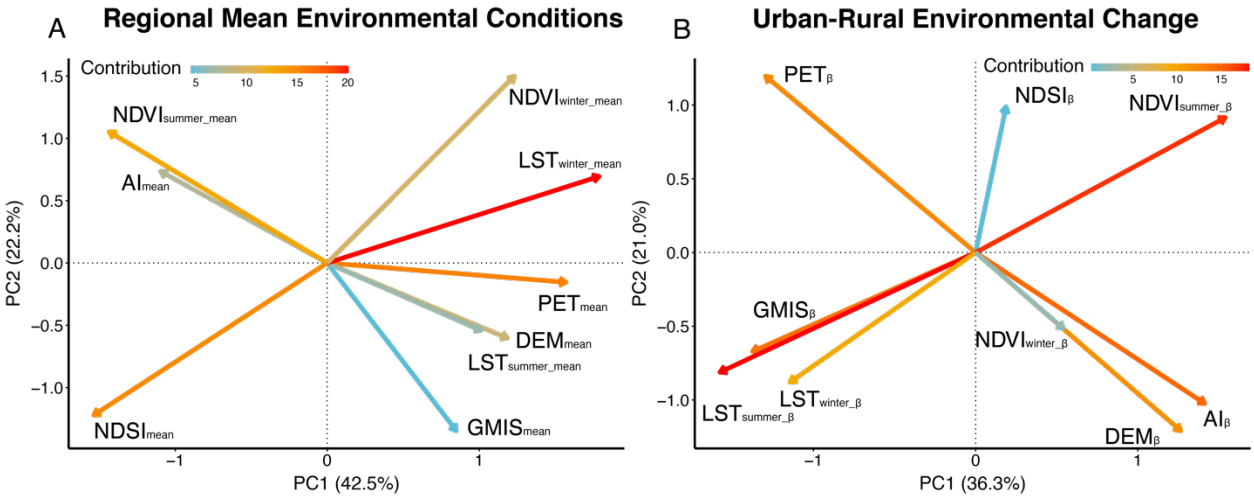


Fig. S7. Eigenvectors from principal components analysis (PCA) of the regional mean environmental conditions and urban-rural environmental change. (**A**) Eigenvectors of regional mean environmental conditions using data from all cities. The regional mean environmental conditions for each environmental factor was determined by taking the mean of that factor across all populations sampled along an urban-rural transect in a given city. The regional means were then subject to PCA to reduce the dimensionality of the data, with PC1 and PC2 retained according to a broken-stick model. (**B**) Eigenvectors of the direction and rate of change in environmental conditions along urban-rural gradients using data from all cities. The direction and rate of change of each environmental variable was calculated as the slope of the relationship between each environmental factor versus standardized distance, estimated individually for all 160 urban-rural transects. PCA was then performed to reduce the dimensionality of the data and PC 1 through 3 were retained according to a broken-stick model. The contribution of each environmental variable in (A) and (B) are coloured by their contribution to PC1. The city scores along the retained PC axes from (A) and (B) were all used in a single multiple regression to identify the multivariate environmental factors that explain variation in HCN clines as described in the Methods. The results from multi-model averaging are reported in Table S6.

ACCEPTED MANUSCRIPT - CLEAN COPY

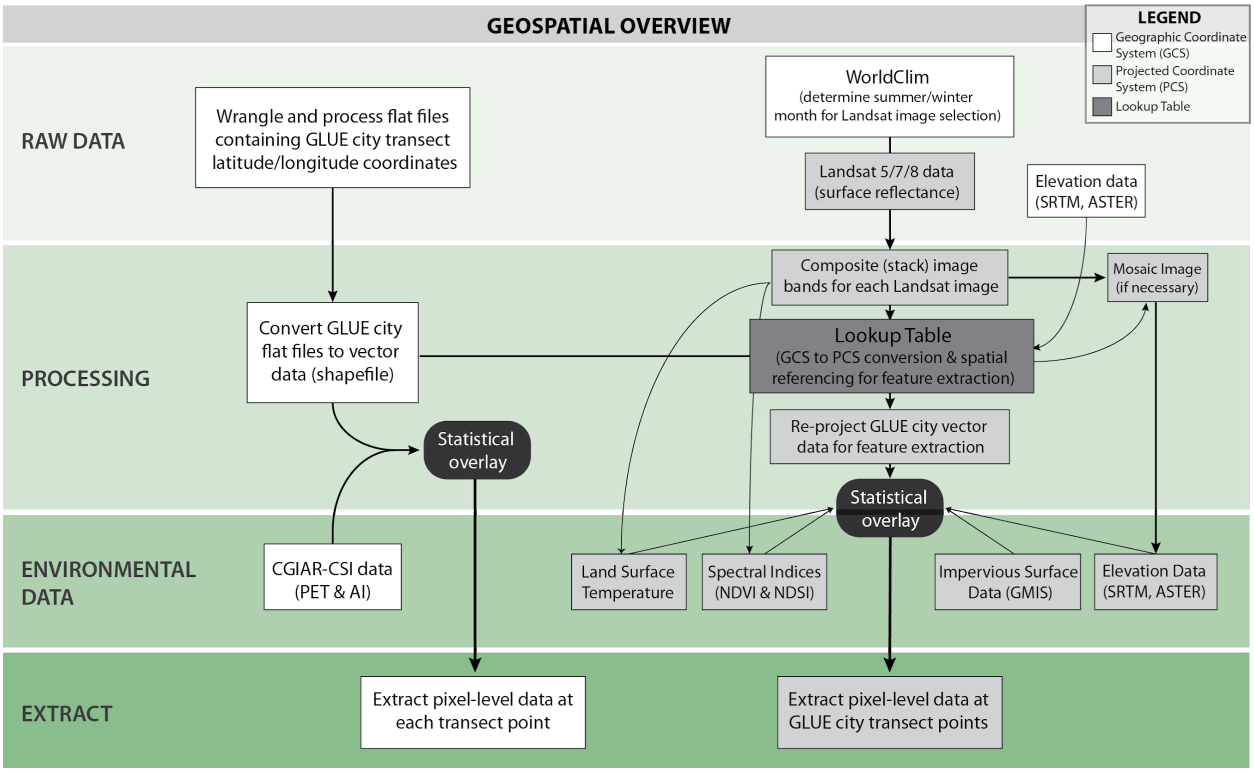


Fig. S8. Overview of geospatial environmental data collection. Detailed workflow used to process vector and raster datasets for feature extraction of environmental data.

ACCEPTED MANUSCRIPT - CLEAN COPY

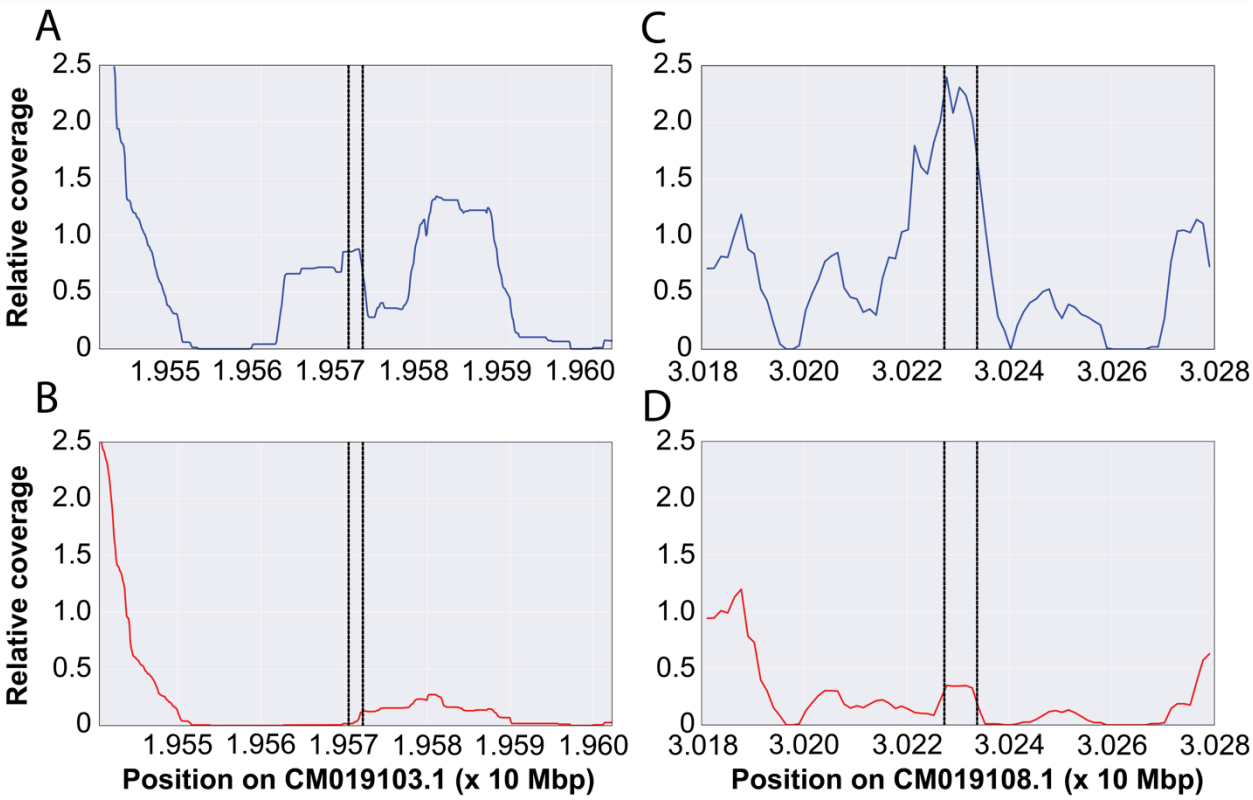


Fig. S9. Relative sequence coverage in the genomic regions including the *Ac* and *Li* loci for representative plants subject to whole genome sequencing. (**A**) Shows a plant with the presence of at least one copy of *Ac*, whereas (**B**) shows a plant with the absence of this region resulting from a homozygous deletion of the region. *Ac* is a 138 kb three gene metabolic cluster responsible for the production of cyanogenic glycosides, including the rate limiting cytochrome P450 enzyme encoded by *CYP79D15* (*16*). (**C**) Shows a representative plant with at least one copy of the *Li* gene, which encodes the enzyme linamarase, and (**D**) shows a plant in which *Li* is absent due to the fixation of the homozygous recessive allele *li*. Linamarase is required to cleave the sugar moiety of cyanogenic glycosides to produce hydrogen cyanide. For both loci, we show the position in the *T. repens* genome (*31*). The vertical axis represents relative coverage of sequence reads in a sliding window (window size 10 kbp with 100 bp steps), calculated as coverage divided by the mean coverage in the surrounding ± 2 Mb. Dashed vertical lines represent the 5' and 3' flanking regions of the *CYP79D15* (A,B) and *Li* (C and D) genes as annotated in the *T. repens* reference genome.

ACCEPTED MANUSCRIPT - CLEAN COPY

Tables S1-S9 are provided in supplementary file Tables S1_S9.xlsx

Table S1. 160 focal cities sampled in the Global Urban Evolution Project, their general characteristics and sampling statistics. We indicate the continent, country and coordinates of each city, the city's area (km^2) covered by urban and suburban habitats, human population size, human population density, and relative city age (i.e., the number of years prior to 2020 that the city reached a population of 150,000). We also include the number of populations sampled, number of plants sampled, length of the transect (km), $\log_e$(odds) intercept and slope of the HCN cline determined by binomial (logistic) regression, whether the slope was statistically significant at $P<0.05$, the overall mean of HCN in the city, and the individuals that sampled the city. Cities in bold were used for whole genome sequencing.

Table S2. Urban environmental change along urban-rural transects and regional mean of environmental conditions for each of the 160 cities. For each city we show regional mean values taken across all sampling sites for annual mean aridity (AI_{mean}), annual mean potential evapotranspiration (PET_{mean}), elevation above sea-level (digital elevation model – DEM_{mean}), percent impervious surface (global manmade impervious surface – $GMIS_{mean}$), normalized difference snow index ($NDSI_{mean}$), land surface temperature in summer (LST_{summer_mean}) and winter (LST_{winter_mean}), normalized difference vegetation index in summer ($NDVI_{summer\_mean}$) and winter (LST_{winter_mean}). The slopes (β) and *P*-values of environmental change (with distance from the city center) for each environmental factor are also shown and correspond to the results shown in Fig. S2. When no value is included the environmental variable could not be obtained due to the lack of clear Landsat images for the location.

Table S3. Binomial generalized linear models testing how distance from the city center, impervious surface and the Human Influence Index predict variation in HCN among populations. (**A**) Analysis of variance table showing results of how distance from the city center (Distance), continent, city, and the interactions of Distance with continent and city predict variation in HCN among populations (also see Fig. 2). (**B**) Same as A, but using the global manmade impervious surface (GMIS) as a predictor. (**B**) Same as A, but using the Human Influence Index as a predictor (*89*). All analyses are identical in structure; random effects in the model are italicized and their significance was tested using a log-likelihood ratio test.

Table S4. The change in the frequency of HCN along urban-rural transects across 160 cities. For each city we show the $\log_e$(odds) intercept (β_0) and slope (β) coefficients, and the *P*-value of the slope from binomial (logistic) regressions performed for each city, which correspond to Fig. 2.

Table S5. Environmental predictors of HCN urban-rural clines. We show the predictor effects, their coefficients, and the number of permuted models (out of 100) the predictors appear

ACCEPTED MANUSCRIPT - CLEAN COPY

in using the elastic net method of regularized regression. The terms are arranged by the number of permuted models in which coefficients occurred within them. The main results are depicted in Fig. 4.

Table S6. Environmental predictors of HCN urban-rural clines using multivariate environmental predictors from PCA. We show the predictor effects, their coefficient estimates, standard errors (SE) of the estimates, z-values and *P*-values for each predictor calculated using multi-model averaging of multiple regressions (see Methods). The original predictors in the full model included the scores from the first two PC axes of the mean regional environmental conditions for each city, and the first three PC axes from the slope of the urban-rural change in environmental conditions for each city (Fig. S7). The response variable was the $\log_e$(odds) slope of HCN versus distance from the urban center extracted from the full binomial GLMM model. All predictors shown were retained in the final model

Table S7. Multiple regression of HCN clines versus city characteristics. HCN clines were quantified as the slope of HCN frequency versus distance from the city center for each of the 160 cities. The characteristics of cities included city area (km^2), human population size, human population density, number of cities within 50 km (i.e. no. cities), and relative city age (i.e., number of years before 2020 that the city reached a human population of 150,000). All predictor variables except no. cities were log-transformed to reduce skew and increase normality of the predictors, which also improved homogeneity of variance and normality in the multiple regression analysis. We performed three separate analyses to examine whether city characteristics predicted the strength of HCN clines, measured as the $\log_e$(slope) of HCN frequency versus distance from the city center. Log(population size) was excluded from all analyses because it was highly correlated with city area and city age ($r_{\text{Pearson}} > 0.78$). (**A**) The first model included data from all cities (N = 160), and log(population density), no. cities, and log(city age + 1) as predictors. The model excluded log(city area) because it was highly correlated with log (city age + 1). (**B**) The second model was similar to the first one, but included log(city area) instead of log(city age) as a predictor. (**C**) The final model included N = 108 cities, and excluded all cities with a city age of 0. Results shown are from type II sums-of-squares for the full model. All models lead to the same conclusion that city characteristics do not predict the strength of HCN clines.

Table S8. Landsat images for each city in summer. For each city we provide the Landsat scene from the warmest month, named following standard NASA/USGS naming conventions, the country, city, and estimated values of transmissivity, upwelling radiance and downwelling radiance. For some cities, multiple overlapping images from the same date were required to capture the entire transect. These landsat images were used for environmental feature extraction as depicted in Fig S8.

ACCEPTED MANUSCRIPT - CLEAN COPY

Table S9. Landsat images for each city in winter. For each city we provide the Landsat scene for the coldest month, and the name of the image following standard NASA/USGS naming conventions, the country, city, and estimated values of transmissivity, upwelling radiance and downwelling radiance. For some cities multiple overlapping images from the same date were required to capture the entire transect. These landsat images were used for environmental feature extraction as depicted in Fig S8.

ACCEPTED MANUSCRIPT - CLEAN COPY

Supplementary Text S1—Methods for collecting environmental data

Collection of environmental data involved pixel-level time series analyses. These analyses involved extracting image pixel data at each sampled population from image datasets. To carry out this workflow, Python was used to automate data wrangling and analysis, geoprocessing (i.e., data conversion/projection, spatial operations and calculations of vector and raster datasets), data scraping and map production tasks. An overview of this workflow is shown in Fig. S8.

Vector and raster datasets

GPS coordinates (latitude, longitude) of each population were converted to decimal degrees (DD) notation to generate vector datasets (shapefiles) using the World Geodetic System 1984 (WGS 84) coordinate system. A separate shapefile was created for each focal city.

To understand how urbanization affected the physical environment, and how such changes could then influence plant evolution, nine environmental variables were collected as raster datasets, which included: land surface temperature in winter, land surface temperature in summer, vegetation cover in winter, vegetation cover in summer, snow cover, percent impervious surface cover, annual potential evapotranspiration, annual aridity, and elevation.

Environmental variables were extracted from NASA's Landsat images to create remote sensing datasets. We selected two Landsat images covering the sampling locations for each city: one image represented the summer period, and the other represented the winter period, including snow cover where applicable. A summer/winter period was determined as the month with the warmest/coolest temperatures, respectively, determined from the WorldClim global climate database for the city's coordinates (*50*). The Landsat image selection workflow involved selecting suitable images using the United States Geological Survey (USGS) Earth Resources Observation and Science Center (EROS) interface (https://earthexplorer.usgs.gov/), which

ACCEPTED MANUSCRIPT - CLEAN COPY

involved a visual assessment of each image to determine if they were clear of any atmospheric effects (e.g., cloud or haze cover) in the city of interest. This assessment process was assisted by using the quality assessment (QA) band provided for Landsat 8 images; Landsat 5 and 7 images did not have a QA band, so assessment was done after the image was retrieved. Images were taken from 2010 or later (Table S8, Table S9), except in one case (Cincinnati OH, USA in winter) where recent images contained atmospheric interference (e.g., cloud cover), so we took the most recent clear image we could find (i.e., from 2003). When no images were available since 2000, we took an image obtained in the month preceding or following the target month (e.g., if the warmest month was July, then we would consider an image from June or August) from the past 10 years. When no clear images were available using these criteria, we took a clear image from any month; all such cases were from equatorial regions where cloud cover was frequent and seasonal climatic variation was small. Once clear images were selected, a bulk order request was submitted for Landsat Level-2 products using the USGS-EROS Center Science Processing Architecture (ESPA) on demand interface (https://espa.cr.usgs.gov/), and finally downloaded using the USGS EROS bulk downloader (https://github.com/USGS-EROS/espa-bulk-downloader). A total of 330 Landsat images were retrieved for analysis.

Level-2 Landsat images were used because they are radiometrically calibrated and corrected, orthorectified and geometrically corrected, atmospherically corrected using a radiative transfer model (RTM), and processed to surface reflectance data. Furthermore, only Landsat images designated as L1TP and as Tier 1 were selected as they are considered most suitable for time-series analysis. However, a few images given the designation of L1TP and Tier 2 were also selected after careful analysis of the geo-registration error and other properties. This only affected a few images, especially ones that had substantial snow coverage (e.g., Trondheim,

ACCEPTED MANUSCRIPT - CLEAN COPY

Norway). Finally, all Landsat datasets come projected in the Universal Transverse Mercator (UTM) system.

Land surface temperature—Land surface temperature (LST) was determined using the single-channel method as proposed by Jiménez-Muñoz and Sobrino (*101*) and modified by Jiménez-Muñoz et al. (*102*). This method uses a single thermal infrared sensor (TIRS) band or channel, and requires a representation of the atmospheric profile to describe three atmospheric functions - atmospheric transmissivity, atmospheric upwelling radiance and atmospheric downwelling radiance - within the thermal infrared (TIR) window at the time of image acquisition. Atmospheric functions were determined using the National Centers for Environmental Prediction (NCEP) atmospheric profiles that use the MODTRAN 4.0 RTM code (*103, 104*), and are implemented in the web-based tool http://atmcorr.gsfc.nasa.gov. Additionally, the single-channel method requires resolving the non-unity of land surface emissivity (LSE) of ground features within an image (*105*). This was done using the NDVI-based emissivity method (NBEM) (*106*). Accurate measures of LST using the NBEM method require the following components: 1) at-sensor radiance of a TIRS band; 2) at-sensor brightness temperature (temperature value near to the LST value) of a TIRS band; 3) NDVI surface (for LSE calculation); 4) LSE surface (emissivity values given by effective wavelength); and 5) atmospheric functions (transmissivity, upwelling radiance, downwelling radiance). For Landsat 8 images, Band 10 was selected since it is located in the lower atmospheric absorption region resulting in higher atmospheric transmissivity values (*107*), and it further has lower reported uncertainty of temperature estimates (± 1 K) compared to other bands (https://landsat.usgs.gov/landsat-8-data-users-handbook-appendix-a). When Landsat 5 and 7 images were used, Band 6 was used as it was the only TIR band available.

ACCEPTED MANUSCRIPT - CLEAN COPY

The revised single-channel method by Jiménez-Muñoz et al. (*102*) for LST was developed from Planck's Law and its derivatives and therefore uses the concept of an effective wavelength that is calculated using a certain channel's width. The equation computes LST in K, which we converted to degrees Celsius (° C) as follows:

$$LST = \gamma[\varepsilon^{-1}(\psi_1 L_{sensor} + \psi_2) + \psi_3] + \delta - 273.15$$

The (γ, δ) parameters were calculated as:

$$\gamma \approx \frac{T_{sensor}^2}{b_\gamma L_{sensor}}$$

$$\delta \approx T_{sensor} - \frac{T_{sensor}^2}{b_\gamma}$$

Where:

L_{sensor} = at-sensor radiance (W m^{-2} sr^{-1} μm^{-1}) at a given wavelength (channel) given by λ (thermal band effective wavelength)

T_{sensor} = at-sensor brightness temperature (K) at a given wavelength (channel) given by λ (thermal band effective wavelength); included with USGS-EROS ESPA Landsat Level 2 products.

$b_\gamma = c_2/\lambda$ 1324 K for Landsat 8 Band 10 (*107*);

1277 K for Landsat 7 Band 6 (*102*);

1256 K for Landsat 5 Band 6 (*102*)

To derive L_{sensor}, the at-sensor digital number (DN) values of the TIRS band were converted to at-sensor radiance as follows:

$$L_{sensor} = M_L Q_{cal} + A_L$$

Where:

L_{sensor} = at-sensor radiance (W m^{-2} sr^{-1} μm^{-1})

ACCEPTED MANUSCRIPT - CLEAN COPY

M_L = Band-specific multiplicative rescaling factor from image metadata (RADIANCE_MULT_BAND_x, where x is the band number) (Table S10)

A_L = Band-specific additive rescaling factor from image metadata (RADIANCE_ADD_BAND_x, where x is the band number) (Table S10)

Q_{cal} = Quantized and calibrated standard product pixel values (DN)

Table S10. **Values of M_L and A_L for different Landsat image types.**

Landsat Sensor Band	**M_L**	**A_L**
Landsat 5 Band 6	5.5375E-02	1.18243
Landsat 7 Band 6 (low)	6.7087E-02	-0.06709
Landsat 7 Band 6 (high)	3.7205E-02	3.16280
Landsat 8 Band 10	3.3420E-04	0.10000

The atmospheric functions (ψ_1, ψ_2, ψ_3) are given by:

$$\psi_1 = \frac{1}{\tau}$$

$$\psi_2 = -L\downarrow - \frac{L\uparrow}{\tau}$$

$$\psi_3 = L\downarrow$$

Where:

τ = atmospheric transmissivity

$L\downarrow$ = downwelling atmospheric radiance

$L\uparrow$ = upwelling atmospheric radiance

$\psi 1$, $\psi 2$, $\psi 3$ as calculated here, relies on an atmospheric model to derive atmospheric transmittance (τ), downwelling atmospheric radiance ($L\downarrow$), and upwelling atmospheric radiance

ACCEPTED MANUSCRIPT - CLEAN COPY

(L↑). These values were calculated by means of the Atmospheric Correction Parameter Calculator (ACPC) (*103, 104*). Input data into ACPC was provided from local meteorological stations for each city.

Surface emissivity (ε) was derived using NBEM, which involved the calculation of a NDVI surface (see *NDVI* below) and was used as a conditional input for calculating an emissivity surface for LST. A pixel was considered to correspond to water/ice/snow when NDVI < 0, bare soil when 0 < NDVI < 0.2, mixed soil and vegetation given $0.2 \leq NDVI \leq 0.5$, and vegetation when $NDVI > 0.5$. The following equations were used:

$$NDVI = \frac{NIR - Red}{NIR + Red}$$

$$P_v = \left[\frac{NDVI - NDVI_{min}}{NDVI_{max} - NDVI_{min}}\right]^2$$

$$C_i = \left(1 - \varepsilon_{s,i}\right)\varepsilon_{v,i} \cdot F' \cdot (1 - P_v)$$

$$\varepsilon_i = \begin{cases} \varepsilon_{w,i} & NDVI < 0 \\ \varepsilon_{s,i} & 0 < NDVI < 0.2 \\ \varepsilon_{v,i}P_v + \varepsilon_{s,i}(1 - P_v) + C_i & 0.2 \leq NDVI \leq 0.5 \\ \varepsilon_{v,i} + C_i & NDVI > 0.5 \end{cases}$$

Where:

NDVI = normalized difference vegetation index

P_v = proportion of vegetation

C_i = cavity effect (surface roughness)

F' = geometrical factor ranging from 0 to 1; typically given as 0.55

$\varepsilon_{w,i}$ = emissivity of water at $\lambda_{effective}$ of channel *i*

$\varepsilon_{s,i}$ = emissivity of soil at $\lambda_{effective}$ of channel *i*

ACCEPTED MANUSCRIPT - CLEAN COPY

$\varepsilon_{v,i}$ = emissivity of vegetation $\lambda_{effective}$ of channel *i*

The emissivity values were calculated at the effective wavelength of the TIRS band, which was defined as:

$$\lambda_{effective} = \frac{\int \lambda f_\lambda d_\lambda}{\int f_\lambda d_\lambda}$$

f_λ is the filter function or spectral response for a specific TIRS channel *i*. The following effective wavelengths for the respective Landsat TIRS channels were used:

Landsat 8 Band 10: $\lambda_{effective} = 10.904\ \mu m$ (*107*)

Landsat 7 Band 6: $\lambda_{effective} = 11.269\ \mu m$ (*102*)

Landsat 5 Band 6: $\lambda_{effective} = 11.457\ \mu m$ (*102*)

Our application of NBEM applied global emissivity constants reported in the literature. We used 0.97 for soil, and 0.99 for vegetation and water (*102, 106, 108-112*)

NDVI—We used the normalized difference vegetation index (NDVI) as a measure of vegetation cover (*113*). NDVI operates on targeting the red absorption curve of leaves due to chlorophyll in the visible spectrum, whilst exhibiting strong reflectance in the near-infrared (NIR) spectrum owing to the leaf structure (spongy mesophyll) (*114*), and was calculated as follows:

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

NDVI values range between -1 and 1. Values from 0 to 1 indicate increasing vegetation cover, whereas negative values correspond to the absence of vegetation (e.g., soil, water, snow, etc.).

ACCEPTED MANUSCRIPT - CLEAN COPY

NDSI—The normalized difference snow index (NDSI) was used to quantify snow cover. NDSI uses the high reflectance of snow in the visible spectrum, and the low reflectance (high absorption) in the shortwave-infrared (SWIR) spectrum, to distinguish snow cover from clouds and haze (*115-118*). NDSI was calculated as:

$$NDSI = \frac{Green - SWIR}{Green + SWIR}$$

NDSI ranges between -1 and 1, with values greater than zero corresponding to increased snow cover; values less than zero indicate the absence of snow.

Impervious Surface (GMIS)—Impervious surface was recorded from each sampled population using the global man-made impervious surface (GMIS) dataset created by the Socioeconomic Development and Application Center (SEDAC) (*52*). We first downloaded Country-level GMIS datasets (N = 26) from SEDAC (https://sedac.ciesin.columbia.edu/data/set/ulandsat-gmis-v1/data-download) and referenced the datasets to the WGS 84 coordinate system. Each GMIS dataset provided percent impervious surface for a given country at 30 m resolution. For each city, we first created a *RasterLayer* object from the appropriate country's GMIS raster dataset using the *raster* function from the "raster" R package (*119*) and converted population latitude and longitude coordinates to a spatial points data frame using the *SpatialPointsDataFrame* function in the "sp" R package (*120, 121*). Finally, we used this data frame together with the *extract* function from the "raster" R package to extract population-level GMIS values from the GMIS *RasterLayer* object. GMIS extraction used a 100-metre buffer around each population, and final percent impervious surface values were taken as the mean of all cells within this buffer.

ACCEPTED MANUSCRIPT - CLEAN COPY

Aridity and potential evapotranspiration—Aridity and potential evapotranspiration were extracted for each population using the Global Aridity Index (AI) and Global Potential Evapotranspiration (PET) geospatial datasets created and distributed by the Consultative Group for International Agriculture Research – Consortium for Spatial Information (CGIAR-CSI) (*51*). The Global-Aridity dataset is provided as a single raster representing aridity based on annual mean aridity from 1950-2000, and it is calculated as a function of precipitation, temperature and PET. A generalized climate classification scheme (*122*) can be used to quantify the pixel values from the AI dataset as shown in Table S11.

Table S11. Climate classification with respect to the aridity index.

Aridity Index Value	Climate Class
< 0.03	Hyper Arid
0.03 – 0.2	Arid
0.2 – 0.5	Semi-Arid
0.5 – 0.65	Dry sub-humid
> 0.65	Humid

The PET dataset is provided as a single raster representing an annual average of PET for the period of 1950-2000, from which we extracted values for each sampling point. Datasets came unprojected in the WGS 84 coordinate system, and data for each population was extracted as described below (see Feature Extraction).

ACCEPTED MANUSCRIPT - CLEAN COPY

Elevation—Elevation was estimated from each population using data collected from the Shuttle Radar Topography Mission (SRTM) (between 60°N latitude and 54°S latitude) and the Global Digital Elevation Model Version 2 (DEM) from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER-GDEM V2) (above 60°N latitude). The former dataset was preferred when data were available for individual cities because SRTM is theoretically more accurate given that it uses radar to penetrate into canopies to retrieve accurate topographic measurements. By contrast, ASTER GDEM V2 uses stereo imagery and photogrammetric techniques to derive the vertical measurement. These datasets came unprojected in the WGS 84 coordinate system and had to be projected to the UTM system for downstream analysis as described below (see Feature Extraction).

Feature extraction

Once the vector and raster datasets were processed, we performed feature extraction to obtain environmental data for each population. Our workflow involved overlaying a vector layer (i.e., shapefile) on a raster layer (i.e., raster dataset) and extracting the pixel data for each population. Feature extraction was straightforward for raster datasets that were unprojected in the WGS 84 coordinate system (e.g., WorldClim, CGIAR-CSI), since the vector datasets were also in the same coordinate system, while the discrepancy of raster datasets using the UTM system had to be reconciled. We accomplished this using a lookup table approach to allow for conversion of graphical coordinate system to projection coordinate system, and on-demand spatial referencing for feature extraction. The lookup table enabled the re-projection of vector and raster datasets to the appropriate UTM zone of a Landsat image for each city. Additionally, the lookup table ensured that the vector and raster dataset were spatially coincident for feature extraction.

ACCEPTED MANUSCRIPT - CLEAN COPY

Supplementary Text S2—Protocol for creating urban-rural transects

Supplies and Materials

- 60 small sealable plastic bags
- Paper to make labels that go in bags with samples
- Black Sharpie marker
- Notebook
- Cooler
- 4-8 kg of ice (place in bottom of cooler)
- GPS or phone with GPS capability
- A long (6-8”) nail or similar pointed tool to help lift stolons. Scissors also work well.
- Pencil
- 1.5 mL microcentrifuge tubes (enough for one per plant sample)
- Cryoboxes

Procedure

Once a city has been selected, we used the following procedure to design a sampling transect:

1. Use Google Earth or Google Maps to identify a transect that represents a clear and fairly continuous gradient in urbanization. Transects should satisfy the following criteria:
    - Urbanization gradients should include downtown areas dominated by impervious surface (buildings and roads), residential suburban areas, and rural or otherwise non-(sub)urban areas.
    - Avoid transects that have environmental gradients unrelated to urbanization, such as elevation, varying proximity to an ocean or a major lake.
    - For safety, avoid transects through areas with high crime.

ACCEPTED MANUSCRIPT - CLEAN COPY

2. Measure the radius from the city center to the edge of the residential area (i.e., where it transitions to rural, grassland or forested habitat); this represents half the length of the total transect.
3. Measure a straight line that is 2× the length of the radius of the city, so half of the transect is in (sub)urban (city and high-density suburbs) habitat and the other half is in non-urban (rural, grassland, forest, low density residential [periurban]) habitat.
4. Identify a driving/biking/walking/transit route that follows the transect (Fig. S10). Routes that pass through areas with an abundance of mowed lawns will have a greater abundance of white clover. When available, Google Street View can be used to identify suitable habitat and parking locations.
5. Aim for roughly equal spacing between sampled populations along the entire transect length, ensuring there is a minimum of 200 m between populations. Aim for >40 sampling sites per city, so the shortest transect should be 8 km. In exceptional circumstances (e.g., clover is rare along a transect), sampling 20 populations is the absolute minimum threshold to enable robust estimation of a cline. Clover populations can be difficult to find in city centers due to high impervious surface, so try to collect samples from 3-5 populations in the downtown core of each city, even if they are off the transect. Elsewhere, when populations are difficult to find directly on a transect, sample populations in a similar habitat near the transect, and record the population coordinates.

ACCEPTED MANUSCRIPT - CLEAN COPY

Fig. S10. Example of sampling transect along an urban-rural gradient. Transect from Guelph, Ontario, Canada, shown as a thin red line: 8.8 km transect; 4.4 km in city; 200 m between each of 44 collection sites.

ACCEPTED MANUSCRIPT - CLEAN COPY

Supplementary Text S3—Protocol for Sampling Plants

Standardization of sampling protocols across all cities was achieved by distributing a written protocol to all collaborators, accompanied by a video, distributed at www.globalurbanevolution.com.

Identification of white clover (*Trifolium repens*)

White clover (*T. repens*) is a distinctive herbaceous perennial plant. While there are ca. 255 species in the genus (*123*), there are only a few *Trifolium* and related Fabaceae species that it regularly co-occurs with, and all can be easily distinguished in the field. To prevent misidentification, collaborators were provided with detailed written descriptions and photographs of *T. repens*, as well as the most commonly co-occurring species that it might be confused with, which included: *T. hybridum*, *T. fragiferum*, *T. pratense* and *Medicago lupulina*.

Procedure

1. If driving, park the car in a safe and secure location, ideally adjacent to where you will sample.
2. Do a brief scan for plants to make sure white clover is present.
3. Once plants have been found, record the following information for the site:
    - Latitude and longitude in decimal degrees (decimal degrees: 40.446° N 79.982° W)
    - Address or nearest major intersection
    - Population number
    - Whether the habitat is: i) urban (mostly impervious surface, lots of city buildings), ii) suburban (high density and continuous residential), iii) periurban (low density, discontinuous residential), iv) rural

ACCEPTED MANUSCRIPT - CLEAN COPY

4. Collect the target number of 20 plant samples (minimum 10) at each population; each sample should contain a stolon and 3-4 leaves (Fig. S11).
   - Take a minimum of 3 large paces between plants sampled to avoid sampling the same clone.
   - Sample along a route that does not double-back on itself.
   - You may have to cover an area as large as 100 m × 100 m to find an adequate sample.

Fig. S11. *Trifolium repens* stolon with 3-4 leaves.

5. Record the population number and address/intersection on a small piece of paper and place inside the bag; seal the bag.
6. Label the outside of the bag with a letter city code (e.g., PDX for Portland) and the population number using a Sharpie marker.
7. Store the sample in a cooler containing ice.
8. Move onto the next population, repeat steps 1 through 7.
9. In the lab, transfer samples individually to labeled 1.5 mL microcentrifuge tubes within three days of collection. The samples should be stored in their sealable bags in a

ACCEPTED MANUSCRIPT - CLEAN COPY

refrigerator at 4°C prior to transfer. Label each tube with the city 3-letter code, population and plant number.

10. Each individual stolon containing leaves is placed in its own individual 1.5 mL microcentrifuge tube using forceps. Extra stolon tissue can be removed, and for samples with large leaves, leaves can be placed directly in the tube, discarding the stolon. You do not need to be gentle when placing the tissue in the tube; the tissue is typically very crammed within the tube.
11. Keep tubes organized within cryoboxes by population, ordered sequentially with respect to plant number, and place them into cryoboxes. You should be able to fit four populations per cryobox. If fewer than 20 plants are sampled from a population, leave the remaining places in the row empty.
12. When the cryobox is full, clearly label the lid with the city, population numbers, (e.g., populations 15-18), date of collection, collectors, and place the box in the -80°C freezer. When a -80°C freezer was not available, samples were stored at -20°C and trials showed variation in storage method did not affect assay results (data not shown).

ACCEPTED MANUSCRIPT - CLEAN COPY

Supplementary Text S4—Assaying HCN production in plants

The presence of HCN is indicative of plants having at least one functional copy of *Ac* and *Li*, whereas the absence of HCN indicates that plants are homozygous recessive for a deletion at either *Ac* and/or *Li*. Plants were assayed for the presence/absence of HCN using the Feigl-Anger assay as described in the Methods (*see* HCN Feigl-Anger Assays). Protocols were standardized among collaborators by following a single written protocol and video distributed via the project website www.globalurbanevolution.com. All assay papers were made in the lab of M. Johnson following the methods of Gleadow et al (*58*) adapted to a 96-well format. Feigl-Anger assay papers were distributed with 96-well plates to ensure consistency in assay methods. Our assay protocol was as follows:

Supplies and materials to prepare Feigl-Anger assay papers

- Filter papers (Whatman grade 3; precut to fit 96-well plates)
- Incubator
- Watch glass
- Beaker (100 mL)
- Tweezers
- Retort stand
- Fume hood
- Hotplate stirrer
- Heating block with spinner
- Ice
- Distilled water
- Tetrabase (4-4'–methylene-bis-N,N-dimethyl-aniline) (Sigma-Aldrich, M44451)

ACCEPTED MANUSCRIPT - CLEAN COPY

- Ethanol (Commercial Alcohols, Brampton, Canada)
- Cupric sulphate (Sigma-Aldrich, cat #12849)
- Ethyl acetoacetate (18% w/v) (Sigma-Aldrich, cat# 688983)
- Sodium acetate (18% w/v) **(**Sigma-Aldrich, cat# S2889)
- Chloroform (ACP, cat #C-3300)

Preparation of copper ethyl acetate reagent

1. Weigh 4 g cupric sulphate into a beaker and add 100 mL distilled water to make a 4% (w/v) solution.
2. Gently heat 30 mL cupric sulphate (4% w/v) with 10 mL ethyl acetoacetate and 20 mL sodium acetate for 10 min in a small beaker on a hot plate in the fume hood. Set plate to 90°C and spin bar set to medium; cover beaker with foil. **Note:** The temperature does not usually have time to reach 90°C.
3. Cool the mixture in the beaker by placing it on ice.
4. Collect the crystals that form: Pour mixture through filter paper then use spatula to collect crystals onto watch glass.
5. Redissolve the crystals in 2 mL of ethanol in a watch glass.
6. Recrystallize the copper ethylacetoacetate crystals by placing the watch glass in an oven at 50°C and dry overnight.
7. Collect and store crystals in brown glass jar.

Preparation of Feigl-Anger test papers

1. Dissolve 0.5 g of copper ethyl acetate reagent in 50 mL of chloroform.
2. Dissolve 0.5 g of tetrabase in 50 mL of chloroform.

ACCEPTED MANUSCRIPT - CLEAN COPY

3. Make a solution containing equal volumes of copper ethyl acetate and tetrabase in a beaker in a fume hood (i.e., mixing 50 mL of each). Cover beaker with foil and allow to dissolve for 2-3 hours with spin bar set to medium-high.
4. Tip solution into an open dish such as a Petri dish.
5. Using tweezers, dip precut filter paper in solution and hang to air dry on a retort stand or horizontally oriented string/wire with clips.
6. Store papers in a brown glass jar (or a jar covered with aluminium foil). Place a small amount of silica gel or other desiccant in the jar. Shelf life is extended beyond one year, if papers are stored at 4°C.

Supplies and materials for performing HCN assay

- Feigl-Anger assay papers (provided by M. Johnson lab)
- Data template Excel sheet (downloaded from www.globalurbanevolution.com)
- 96-well microtiter plate (Brandtech Scientific, VWR cat. #89093-588) (provided by M. Johnson lab)
- Elastic bands (provided by M. Johnson lab)
- Multichannel pipette trough (provided by M. Johnson lab)
- Samples in -80ºC cryofreezer
- Heavy weight (e.g., water bottle, brick, etc.)
- Incubator set to 37ºC
- 600 pipette tips (200 µL)
- Cryoboxes
- Distilled water
- Forceps

ACCEPTED MANUSCRIPT - CLEAN COPY

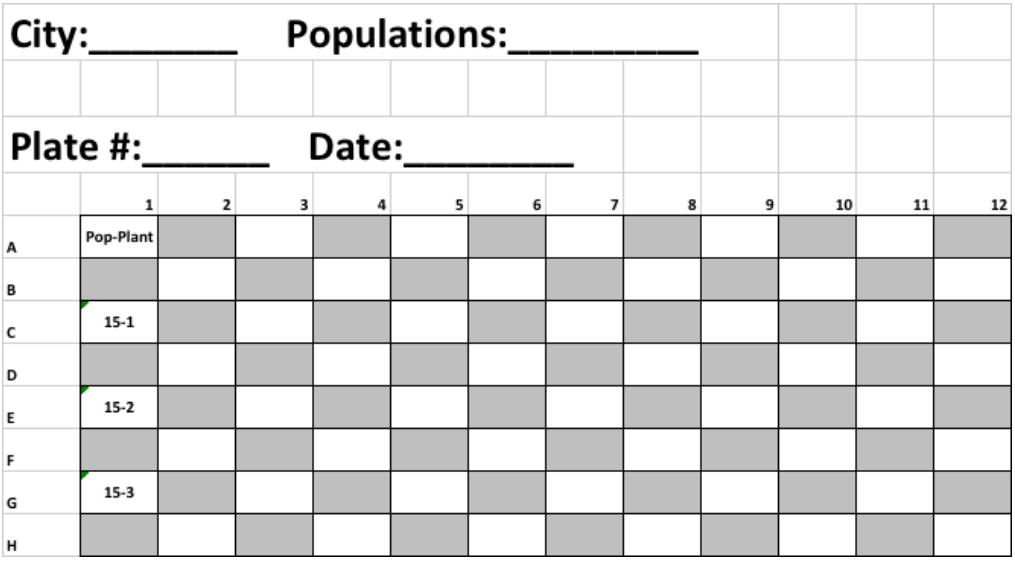
- Kimwipes
- Multichannel or single channel pipettor capable of dispensing 80 µL of water

Procedure

1. Remove samples from the freezer 20-30 min prior to starting the assay.
2. While tissue is defrosting, download data template Excel sheet from www.globalurbanevolution.com. This sheet will be used to record and track which plant is in which well of the 96-well microtiter plate.
   - The assay uses 48 of the 96 wells; empty wells between samples containing tissue reduces contamination and false positives in the detection of HCN.
   - Begin in the top left of the plate (cell A1), working your way down the 8 wells (towards cell H1) before moving to column 2 (i.e., cell B2).
   - Avoid splitting populations across multiple plates. If not all plants from a population fit on the same plate, begin a new plate. This reduces errors.
   - Each unshaded cell in the data template Excel sheet represents a single sample (Fig. S12). For each plate, record:
     i. In the header:
        a. city name
        b. population number
        c. plate number (e.g. 1 of x, 2 of x, 3 of x, etc.)
        d. date assayed
     ii. In the spreadsheet, record:
        a. population number
        b. plant sample number within a population

ACCEPTED MANUSCRIPT - CLEAN COPY

Fig. S12. Data template used to record samples assayed and phenotyping results. In each cell we recorded the population number (e.g., 15) and the plant number (e.g., 1), written as 15-1. Each plant in a given city had a unique *population #-plant #* combination.



3. Once the tissue is defrosted (20-30 min), add leaf material from plants to wells in the 96-well plate, following the template created in step 2.
   - Place a single large leaf (1-2 cm across, containing all 3 leaflets, minus the petiole) in every other well; 2 leaves can be placed in a well for plants with small leaves (<1cm across; see Fig. S13); one or two leaflets can be placed in a well for very large leaves (>2 cm diameter).

ACCEPTED MANUSCRIPT - CLEAN COPY

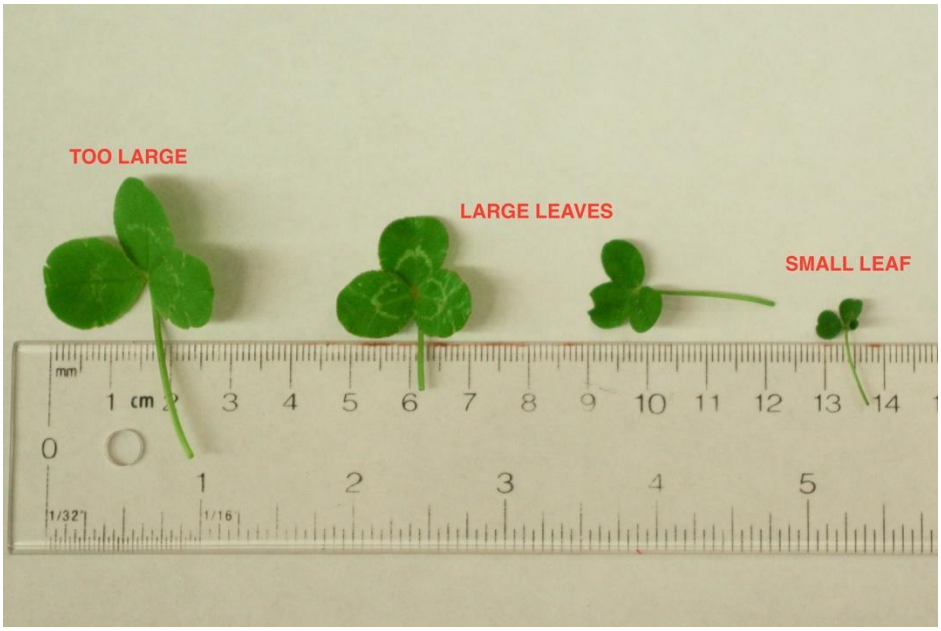


Fig. S13. Leaf sizes used for assays. The target sized leaf was 1-2 cm in diameter. When smaller or larger leaves were taken we adjusted the amount of tissue used as described in step 3.

4. Pour distilled water into the plastic multichannel trough provided and add 80 µL of water to each well. It is easiest to use a multichannel pipette for this step, but a single channel pipette works too. It is fine to add water to empty cells. The same pipette tip can be used to dispense all water across a single plate.
5. Using 200 µL pipette tips, macerate the tissue in the wells until well blended. The water should be a homogeneous green and an abundance of small bubbles will form indicating the tissue is properly macerated (Fig. S14). A separate pipette tip must be used for each plant to avoid contamination.
6. Once all plants are macerated, wipe down the top of the plate with Kimwipes to remove all liquid outside of wells; water on the plate surface will interfere with the assay.
7. Use a fine-tipped Sharpie marker or pencil to label the top and left edges of the upper surface of the Feigl-Anger assay paper with corresponding column numbers and row

ACCEPTED MANUSCRIPT - CLEAN COPY

letters, respectively; this will help minimize errors when scoring the plates once the assay is complete. Place the Feigl-Anger assay paper on top of the plate so that the paper covers all of the wells and the row and column labels are properly aligned. Add the 96-well plate cover on top of the plate and secure with elastic bands. The assay paper should be sandwiched between the plates and the cover (Fig. S15)

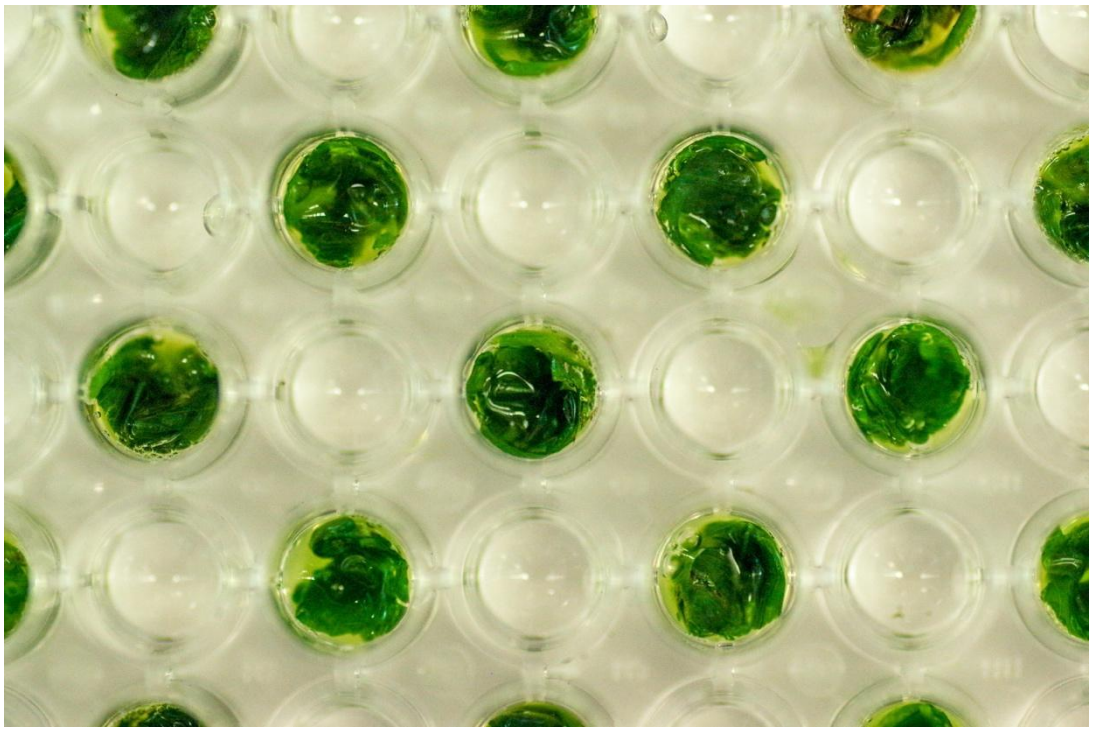


Fig. S14. **Macerated tissue displaying a homogeneous green hue to the water and the presence of bubbles on the surface.**

8. Place the plate in an incubator set to 37°C for 3 hours. Add a weight on top of the plate to ensure the assay paper maintains a firm and even contact with the top of the plate (Fig. S15). While in the incubator, the water in the wells will begin to evaporate–this is normal. If there is HCN in the solution, it will react with the copper ethylacetoacetate on the assay paper and create a blue circle over top of the wells containing cyanogenic plant samples (HCN+, positive).

ACCEPTED MANUSCRIPT - CLEAN COPY

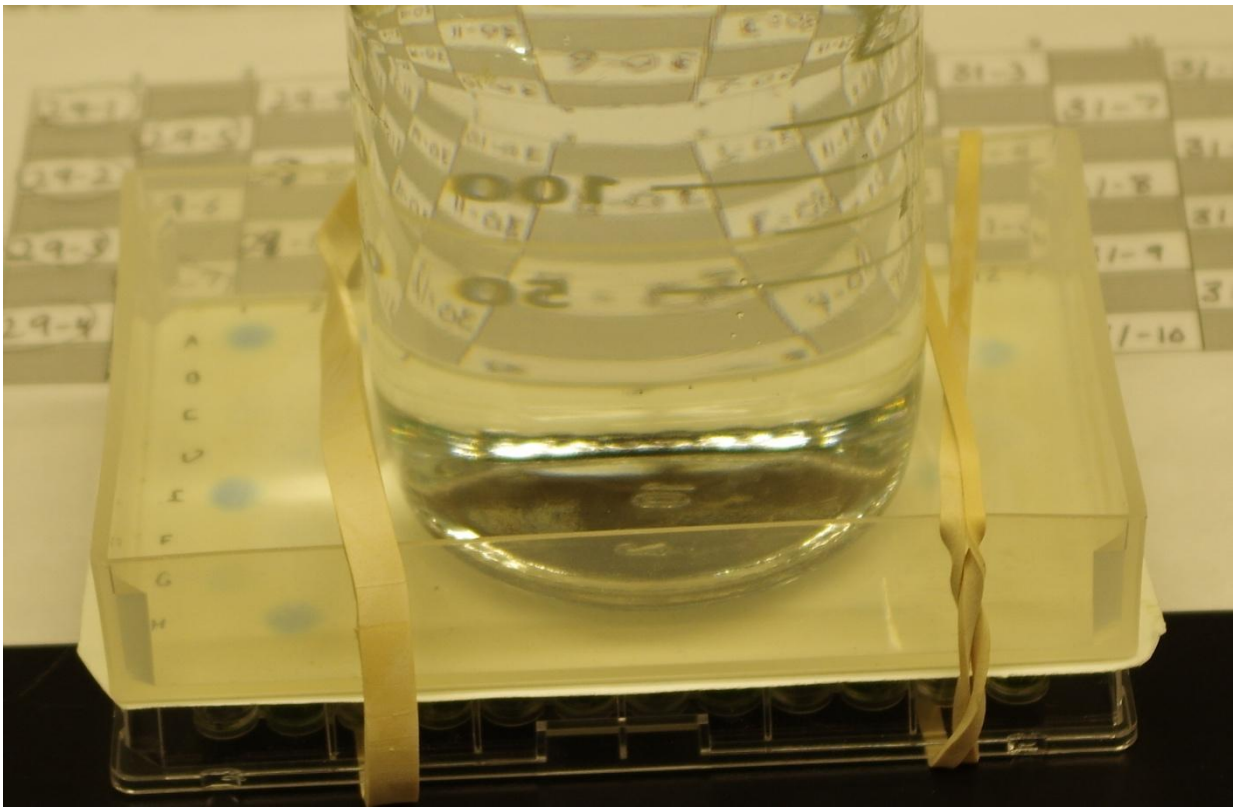


Fig. S15. Feigl-Anger assay paper sandwiched between 96-well microtiter plate and plate cover with elastics securing cover to plate and weight ensuring assay paper is in firm contact with the plate.

9. After 3 hours, gently remove the plate from the incubator. With the 96-well plate placed on top of the template paper in the correct orientation (Fig. S16), take a photo of the plate and assay paper from above. Turn the assay paper over, by turning it away from yourself, place the 96-well plate on top of the template paper again in the correct orientation, and take a second picture.
    - Photographs were used to double-check cyanogenesis scores.
    - Photos are taken of the upper and lower surfaces because sometimes cyanogenic plants can only be distinguished from the lower-side of the paper (i.e. the side in contact with the microtiter plate).

ACCEPTED MANUSCRIPT - CLEAN COPY

10. Immediately following removal, use the lower-side of the assay paper (i.e. side in contact with the plate during the assay) to record which plants are cyanogenic based on the presence of a blue or green dot (even if faint) over the well. Circle on the assay paper samples that are positive for the presence of HCN so it can be referenced later when entering the data. The filter paper will remain white above plants that are HCN-.
11. After scoring the presence or absence of HCN, add the HCN scores to the Excel data sheet in the column labeled "HCN (1 = presence, 0 = absent)". Cyanogenic plants (HCN+ = blue circle) should be coded as "1"; acyanogenic plants (HCN– = white) as "0". Upload data and photographs to www.globalurbanevolution.com.

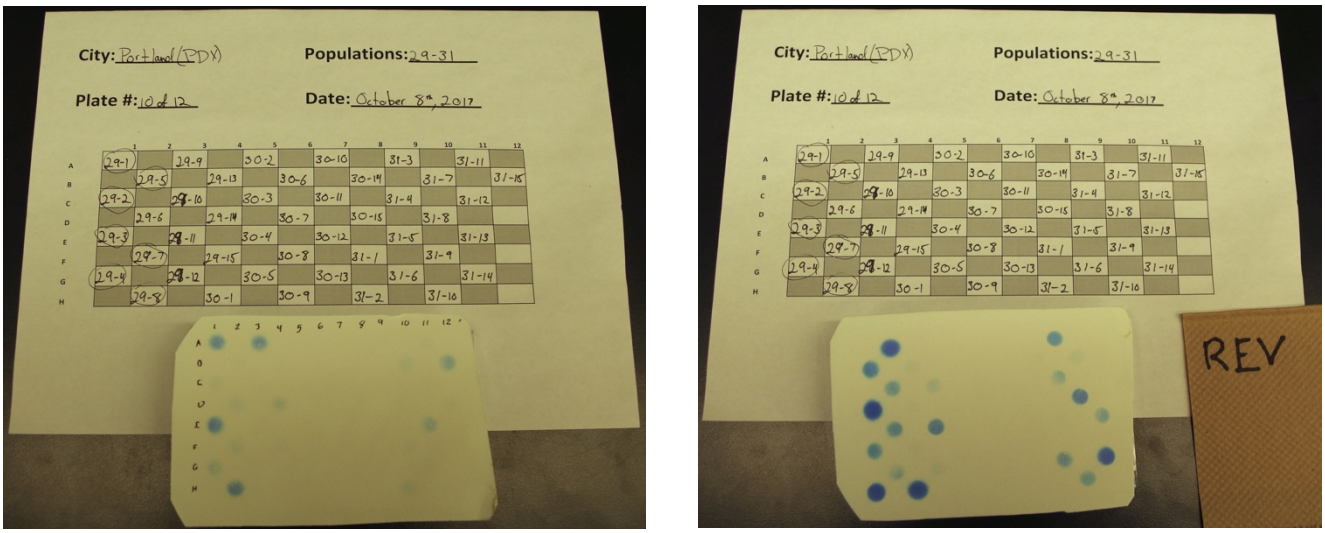


Fig. S16. Scoring presence/absence of HCN from plant samples. (left) Photo of top side of Feigl-Anger assay paper along with associated template sheet. (right) Example of lower side photo of Feigl-Anger assay paper turned away.

ACCEPTED MANUSCRIPT - CLEAN COPY

Supplementary Text S5—CTAB-Phenol/Chloroform Genomic Extraction Procedure

This protocol implements a modified phenol:chloroform extraction as in Arruda et al. (*124*) to extract genomic DNA in 96-well format. The tissue was freeze-dried and homogenized to a fine powder in 2 mL tubes prior to extraction.

Supplies and materials

- CTAB buffer (see recipe)
- β-mercaptoethanol (Sigma-Aldrich, cat #M6250)
- Phenol-chloroform-isoamyl alcohol (25:24:1) (Sigma-Aldrich, cat # P2069-400ML)
- Chloroform (ACP, cat #C-3300)
- RNAse A 10 mg/mL (Fisher Scientific, cat. #FEREN0531)
- 2-Isopropanol (Caledon Laboratory Chemicals, cat #8601-7-40)
- Tris-EDTA (TE) (Sigma-Aldrich, cat no. T4415-4L)
- 2 mL cluster tubes + strip caps (VWR, cat. #82006-694 (tubes), cat. #93000-046 (caps))
- 96-well assay blocks (VWR, cat. #82051-228)
- 96-well aluminum sealing film (Axygen Scientific, PCR-AS-200)

CTAB recipe

This buffer is a combination of the recipes from Arruda et al. (*124*) and Inglis et al. (*125*).

- 1M Tris buffer pH 8.0 (Fisher Scientific, cat. #15568025)
- 0.02 M EDTA (made from 0.5M solution; VWR, cat. #97062-656)
- 2.5 M NaCl (made from 5M stock; Lonza, VWR, cat. #CA12001-384)
- 3% (m/v) Hexadecyltrimethylammonium bromide (CTAB) (Sigma-Aldrich, cat #H6269
- 2% (m/v) Polyvinylpyrrolidone (PVP) (Bioshop, Burlington Canada, cat #PVP504)
- 1% (v/v) β-mercaptoethanol (added immediately before use)

ACCEPTED MANUSCRIPT - CLEAN COPY

Procedures

1. Heat water bath to 60°C.
2. Add 1% β-mercaptoethanol to the working CTAB solution. Invert to mix.
3. Add 624 µL CTAB mixture to each 2 mL tube. Cap tubes and carefully invert plates to mix powdered sample and buffer.
4. Incubate plates in a 60°C water bath for 30 min. Invert the tubes every 10 min.
5. Remove plates from the water bath; let cool for 5 min.
6. Centrifuge plates for 10 min at maximum speed (4060 rcf, Beckman Coulter Allegra X-14R with deep well plate rotor attachment, Beckman Coulter Life Sciences, California, USA) at room temperature.
7. Transfer ~500 µL of liquid into new 2 mL tubes.
8. Add 500 µL of phenol:chloroform:IAA (25:24:1) to tubes. Invert plates for 10 min or vortex for 1–2 min.
9. Centrifuge plates for 20 min at 4060 rcf at room temperature.
10. Pipette 400 µL of aqueous upper layer and transfer to a new set of 2 mL tubes
11. Add 2.0 µL RNAse A (10 mg/mL) to tubes. Incubate for 30 min at 37 °C
12. Add 400 µL of chloroform to tubes. Invert tubes for 10 min or vortex for 1–2 min.
13. Centrifuge plates for 20 min at 4060 rcf at room temperature.
14. Remove ~250 µL of aqueous (top) layer being careful not to take up chloroform (bottom) layer; transfer to 96-well assay block.
15. Add 1 volume of chilled isopropanol to wells to precipitate DNA. Seal plates with 96-well plate sealing adhesive cover and affix with roller; incubate at –20°C overnight.

ACCEPTED MANUSCRIPT - CLEAN COPY

16. Centrifuge plates for 20 min at maximum speed (1788 rcf, Eppendorf 5430R with plate rotor, Eppendorf, Hamburg, Germany) at room temperature to pellet DNA.
17. Pour off supernatant.
18. Add 200 µL of cold 70% ethanol and invert.
19. Centrifuge plates for 20 min at 1788 rcf at room temperature to pellet DNA.
20. Pour off supernatant. Dry plates at 37 °C for 60 min or until dry.
21. Elute in 50 µL TE.

ACCEPTED MANUSCRIPT - CLEAN COPY

Supplementary Text S6—Genomic Library Preparations

This protocol was used to create dual-indexed, Illumina-compatible gDNA libraries for whole genome sequencing. It does not rely on any commercial kits, but instead uses homemade master mixes for all reactions. Briefly, the protocol uses an "on-bead" library preparation (*62*) with "home-brewed" master mixes (*62, 65, 66*) and SPRI beads (*63*), and incorporates i5 and i7 Illumina indices into adapters through PCR (*64*).

Supplies and materials

Reagents:

- 0.2 mL Bioruptor tubes (0.2 mL, Diagenode, cat. # C30010020)
- PCR tubes (0.2 mL, VWR, cat. #93001-118)
- Pipette tips 10 µL and 200 µL (Axygen, T-300-L and TR-222-C-L, respectively)
- dNTP (ThermoFisher, R1121)
- ATP (ThermoFisher, R0441)
- Tango buffer 10X (ThermoFisher, BY5)
- T4 Polynucleotide Kinase (ThermoFisher, EK0032)
- T4 DNA polymerase (ThermoFisher, EP0062)
- Nuclease free ultrapure water (Invitrogen, cat. # 10977015)
- dATP 100 mM (Invitrogen, cat #10216018
- taq DNA polymerase (Invitrogen, cat #18038067)
- iTrusR2-stubRCp (IDT)
- iTrusR1-stub (IDT)
- T4 DNA ligase (ThermoFisher, EL0011)
- Cytiva Sera-Mag SpeedBeads (ThermoFisher, cat no. 09-981-123)

ACCEPTED MANUSCRIPT - CLEAN COPY

- PEG-8000 (VWR, cat. # 97061-098)
- EDTA (VWR, cat. #97062-656)
- 1M Tris buffer pH 8.0 (Fisher Scientific, cat. #15568025)
- Tween 20 (VWR, cat. #97062-332)
- 5M NaCl (Lonza, VWR, cat. # CA12001-384)
- 50 bp DNA Ladder (ThermoScientific, cat. # FERSM0371)
- ITru5_Group1_8-Indices (IDT)
- ITru7_Group1_12-Indices (IDT)
- Phusion HiFi polymerase (New England BioLabs, cat no. M0530S)

Equipment:

- Bioruptor Pico Sonicator (Diagenode, Denville, NJ, USA)
- PCR thermocycler with heated lid
- Gel rigs
- 96-well plate magnet (e.g., Dynamag-96 well #12331D)
- Racks for 0.2 mL PCR tubes or plates (VWR, cat. #93001-118)
- PCR plates or tubes
- 10 and 100 µL multichannel or single channel pipettes
- Qubit 3.0 Fluorometer (Invitrogen, ThermoFisher, Mississauga, Canada)
- Qubit dsDNA High Sensitivity Assay Kit (ThermoFisher Scientific, Cat. No. Q33231)

Protocol

This protocol assumes that you have:

ACCEPTED MANUSCRIPT - CLEAN COPY

1. Hydrated all index-containing forward and reverse PCR primers (described in supplementary file: *Protocol for preparing 1.25 nmol iTru primer aliquots for use during library preparation* in Glenn et al. (*64*)). These are diluted to 5 µM working solutions.
2. Annealed the adapter stubs (i.e., iTrusR2-stubRCp, iTrusR1-stub) to form the Y-yoke adapter (supplementary file: *Protocol for preparing double-stranded iTru adapters for use during library preparation* in Glenn et al. (*64*)). These are diluted to 5 µM working solutions. This is referred to as the *adaptor mix* throughout the protocol.
3. Prepared the SPRI bead solution, which replaces Ampure XP beads for cleaning enzymatic reactions (*63*); reproduced in supplementary file: *Protocol for preparing an inexpensive substitute for AMPure* in Glenn et al. (*64*)). Be sure to also prepare a PEG solution lacking the actual SPRI beads (described in the same file), which will be used throughout the protocol. Addition of 0.05% Tween-20 to this PEG/NaCl solution will greatly facilitate the bead wash.
4. Prepared a fresh aliquot of 80% ethanol.

DNA Shearing:

1. Normalize DNA to 10 ng/µL (quantified by Qubit BR) in 25 µL of TE (10 mM Tris-HCl pH8 and 0.1 mM EDTA) in Bioruptor tubes. Note, this volume allows for two library preps per sample.
2. Shear DNA by sonication to mean fragment size of ~500 bp.
   - For our samples we used 3 cycles of 15s “ON”, 30s “OFF” on a Bioruptor Pico Sonicator
3. After shearing, move 25 µL of DNA to empty PCR tube and proceed to post-shearing cleanup.

ACCEPTED MANUSCRIPT - CLEAN COPY

Post-shearing cleanup:

1. Add 20 µL of SPRI bead solution to each sample. This is a 0.8× bead:DNA ratio, which is meant to remove fragment < ~250 bp.
2. Incubate at room temperature for 5 min.
3. Place samples on magnetic rack and allow the beads to bind to the side of the tube. The supernatant should be clear.
4. Remove and discard the supernatant.
5. Add 80 µL of freshly prepared 80% ethanol. Let sit on magnetic rack for 30 to 60 s and discard ethanol using a pipette.
6. Perform a second ethanol wash as in step 5.
7. Remove any residual ethanol using a 10 µL pipette. Let samples sit uncovered at room temperature for 5 to 10 min or until all ethanol has evaporated.
    - **Important:** Do not let beads dry out too much as this will reduce yield. Beads should be matte brown (not glossy). Dry beads will begin to crack.
8. Elute DNA in 20 µL TE for 5 min. Pipette mix to homogenize beads and TE. The DNA will now be in solution.

Note: The beads from the above reaction will be kept through all subsequent enzymatic reactions and reused for all SPRI bead cleanups (except PCR). The beads do not interfere with the reactions (*62*).

End-repair:

1. Prepare the end-repair master mix below shown in Table S12. Each reaction requires 10 µL.

ACCEPTED MANUSCRIPT - CLEAN COPY

Table S12. End-repair master mix from Meyer and Kircher (*65*). Final concentrations determined from the total master mix (10 µL) + sample (20 µL) volume.

Reagent	Units	Initial conc.	Vol. (µL)	Final conc.
Tango buffer	X	10	3	1
ATP	mM	10	3	1
dNTP	mM	10	0.3	0.1
T4 PNK	U/µL	10	1.5	0.5
T4 POL	U/µL	5	0.6	0.1
Water	NA	NA	1.6	NA
Total	NA	NA	**10**	NA

2. Pipette 10 µL of master mix into 20 µL of each DNA sample. Mix thoroughly by pipetting or flicking.
3. Incubate reaction in thermocycler using the following conditions (*62*):
    - 15 min at 12°C followed by 15 min at 25°C (30 min total).

Post end-repair cleanup:

1. Add 84 µL of 20% PEG-8000/2.5M NaCl (i.e., SPRI bead solution without the beads) to samples. This is a 2.8× bead:DNA ratio (*62*).
2. Perform bead cleanup as in *post-shearing cleanup*, eluting DNA in 20 µL of TE for 5 min.

A-tailing:

1. Prepare the A-tailing master mix shown in Table S13. Each sample requires 10 µL of the master mix.

ACCEPTED MANUSCRIPT - CLEAN COPY

Table S13: A-tailing master mix from Kobs (*66*) and wiki from the Rausher lab at Duke. Final concentrations determined from the total master mix (10 µL) + sample (20 µL) volume.

Reagent	Units	Initial conc.	Vol. (µL)	Final conc.
Taq buffer	X	10	3	1
dATP	mM	10	0.6	0.2
Taq	U	5	0.75	0.125
MgCl2	mM	50	1.5	2.5
Water	NA	NA	4.15	NA
Total	NA	NA	**10**	NA

2. Add 10 µL of master mix to 20 µL of each DNA sample.
3. Incubate in thermocycler for 30 min at 70°C.

A-tailing cleanup:

1. Add 66 µL of 20% PEG-8000/2.5M NaCl (i.e., SPRI bead solution without the beads) to samples. This is a 2.2× bead:DNA ratio (*62*).
2. Perform bead cleanup as in *post-shearing cleanup*, eluting DNA in 20 µL of TE for 5 min.

Adapter ligation:

1. Prepare the adapter ligation master mix shown in Table S14. Each sample requires 10 µL.

ACCEPTED MANUSCRIPT - CLEAN COPY

Table S14: Adapter ligation master mix from Meyer and Kircher (*65*). Final concentrations determined from the total volume (35 µL) of the combined adapter mix (5 µL), master mix (10 µL) and sample (20 µL).

Reagent	**Units**	**Initial conc.**	**Vol. (µL)**	**Final conc.**
Ligase buffer	X	10	3.5	1
PEG-4000	mM	50	3.5	5
T4 Ligase	U	5	0.88	0.125
Water	NA	NA	2.12	NA
Adapter mix		5	5	0.143
Total	NA	NA	**35**	NA

2. Add 5 µL of 5 µM adapter mix (see point 2 of Protocol introduction about preparing Y-yoke adapter) to each DNA sample.
3. Add 10 µL of adapter ligation master mix (distinct from "adapter mix") to each sample. Pipette mix thoroughly. The total volume should now be 35 µL.
4. Incubate in thermocycler for 30 min at 25°C

Adapter ligation cleanup:

1. Add 31.5 µL of 20% PEG/2.5M NaCl (i.e., speedbeads without the beads) to samples. This is a 0.9× bead:DNA ratio. This will get rid of smaller DNA fragments (e.g., < ~250 bp) like adapter dimers that may have formed during the ligation reaction.
2. Perform bead cleanup as in *post-shearing clean-up*, eluting DNA in 20 µL of TE for 5 min.

ACCEPTED MANUSCRIPT - CLEAN COPY

3. After elution, place beads on magnet and carefully transfer the supernatant to new tube/plate by slowly drawing up the supernatant containing the DNA with a pipette. This sample will be used for indexing PCR.

Indexing PCR:

1. Prepare the PCR master mix shown in Table S15. Each reaction requires 12.5 µL.

Table S15. Indexing PCR master mix. Final concentrations calculated from total sample volume (i.e., master mix + primers + sample = 25 µL)

Reagent	**Units**	**Initial conc.**	**Vol. (µL)**	**Final conc.**
Phusion buffer	×	5	5	1
dNTP	mM	10	0.5	0.2
Phusion polymerase	U/µL	2	0.25	0.02
Water	NA	NA	6.75	NA
Total	NA	NA	**12.5**	NA

2. Add 12.5 µL of master mix to empty PCR tubes/plates corresponding to the number of samples to be amplified and indexed.
3. Add 10 µL of DNA sample to each well containing PCR master mix.
4. Add 1.25 µL of forward primer and 1.25 µL of reverse primer to each sample.
    - *Note:* Each sample should receive a unique combination of forward and reverse primers since these will incorporate the indices into the libraries. 8 forward + 12 reverse primers can uniquely tag 96 individuals.
    - See Glenn et al. (*64*) for details on these primers.
5. Perform PCR using the following conditions:

ACCEPTED MANUSCRIPT - CLEAN COPY

98 °C for 30s

15 cycles of:

98 °C for 30s

65 °C for 30s

72 °C for 60s

72 °C for 1 min

2nd indexing PCR and pooling (optional):

Because only 10 µL of the pre-PCR (i.e., ligated) library is used in the indexing PCR, each library prep allows for a total of two PCR reactions. If you find that final library concentrations (i.e., after cleaning, see *post-PCR cleanup*) are too low, you can perform a second PCR reaction on the remaining pre-PCR library using the same conditions as above. These two reactions can then be pooled prior to cleaning, and then eluted in a lower volume to increase concentrations (see *post-PCR cleanup*).

Post-PCR cleanup:

1. Add 0.8× volume of the SPRI bead solution to each sample of the post-PCR DNA. If you only performed a single PCR reaction, this should be approximately 25 µL. If you pooled two 25 µL PCR reactions, this should be approximately 50 µL of SPRI bead solution. This should remove primer dimers and other small fragments remaining from PCR.
2. Perform bead cleanup as in *post-shearing cleanup*, eluting for 10 min in 23 µL (i.e., 20 µL for sequencing and 3 µL for post-cleaning QC)
3. Quantify library concentration using Qubit HS assay. Further QC library by visualizing on a gel using a 50 bp ladder or by passing through an Agilent Bioanalyzer with a DNA 1000 high sensitivity chip.

ACCEPTED MANUSCRIPT - CLEAN COPY

Pooling for sequencing

Dual-indexed libraries were pooled into equimolar ratio pools to ensure approximately equal sequencing coverage. This was done in two steps: (1) Creating equimolar dilutions for each sample, and (2) creating an equimolar pool for all samples to be multiplexed on a single lane.

Creating equimolar dilutions:

1. Create equimolar dilutions for each sample, diluting the concentration down to that of the least concentrated sample (0.81 ng/µL was used as our minimum concentration). For highly concentrated samples, serial dilution was required. The script in step 1 outputs the volume required for each library, in addition to the volume of TE required to create the equimolar pool. It additionally indicates whether serial dilutions are necessary and any required volumes.
    - *Note:* This script can facilitate determining the library volumes required to create equimolar pools for sequencing.

Creating equimolar pool:

1. Pipette the same volume from each sample (which are now normalized to the same concentration from step 1) into the appropriate tube. You should have 1 tube for each sequencing lane.
    - *Note:* To avoid pipetting error associated with using very small volumes, it is best to ensure that at least 2 µL is being taken from each sample when creating the equimolar pool.

Quality control of final libraries

We performed the following QC steps prior to sending samples for sequencing.

ACCEPTED MANUSCRIPT - CLEAN COPY

1. Quantify the concentration of the final library pools to be sequenced using a Qubit fluorometer. Quantify each lane separately.
2. Visualize the library on a gel
   - Use 50 bp DNA ladder
   - Use a 1.5% agarose gel with 3 μL of 1% Ethidium Bromide (EtBr) added to gel
   - Run gel for 60 min at 100 V with 3 μL EtBr added to buffer
3. (optional) Visualize fragment size distribution using an Agilent Bioanalyzer with a DNA 1000 High Sensitivity chip

ACCEPTED MANUSCRIPT - CLEAN COPY