

This document is the accepted manuscript version of the following article: Pouteau, R., Thuiller, W., Hobohm, C., Brunel, C., Conn, B. J., Dawson, W., ... van Kleunen, M. (2021). Climate and socio-economic factors explain differences between observed and expected naturalization patterns of European plants around the world. *Global Ecology and Biogeography*, 30(7), 1514-1531. <https://doi.org/10.1111/geb.13316>

Research paper

Title

Climate and socio-economic factors explain differences between observed and expected naturalization patterns of European plants around the world

Running head

Observed versus expected naturalizations

Authors

Robin Pouteau*, Wilfried Thuiller, Carsten Hobohm, Caroline Brunel, Barry J. Conn, Wayne Dawson, Michele de Sá Dechoum, Aleksandr L. Ebel, Franz Essl, Ori Fragman-Sapir, Trevor Fristoe, Nejc Jogan, Holger Kreft, Bernd Lenzner, Carsten Meyer, Jan Pergl, Petr Pyšek, Alla Verkhovzina, Patrick Weigelt, Qiang Yang, Elena Zykova, Svetlana Ačić, Emiliano Agrillo, Fabio Attorre, Ariel Bergamini, Christian Berg, Erwin Bergmeier, Idoia Biurrun, Steffen Boch, Gianmaria Bonari, Zoltán Botta-Dukát, Helge Bruehlheide, Juan Antonio Campos, Andraž Čarni, Laura Casella, Maria Laura Carranza, Milan Chytrý, Renata Čušterevska, Michele De Sanctis, Jürgen Dengler, Panayotis Dimopoulos, Rasmus Ejrnæs, Jörg Ewald, Federico Fernández-González, Rosario G. Gavilán, Jean-Claude Gegout, Rense Haveman, Maike Isermann, Ute Jandt, Florian Jansen, Borja Jiménez-Alfaro, Ali Kavgacı, Larisa Khanina, Ilona Knollová, Anna Kuzemko, Maria Lebedeva, Jonathan Lenoir, Tatiana Lysenko, Corrado Marcenò, Vasiliy Martynenko, Jesper Erenskjold Moeslund, Ricarda Pätsch, Remigiusz Pielech, Valerijus Rašomavičius, Iris de Ronde, Eszter Ruprecht, Solvita Rūsiņa, Pavel Shirokikh, Jozef Šibík, Urban Šilc, Angela Stanisci, Zvezdana Stančić, Jens-Christian Svenning, Grzegorz Swacha, Pavel Dan Turtureanu, Milan Valachovič, Kiril Vassilev, Sergey Yamalov, Mark van Kleunen

***Corresponding author**

Zhejiang Provincial Key Laboratory of Plant Evolutionary Ecology and Conservation, Taizhou University, Taizhou 318000, China¹

E-mail: robin.pouteau@ird.fr

Affiliations

Wilfried Thuiller: Laboratoire d'Écologie Alpine (LECA), University Grenoble Alpes, CNRS, University Savoie Mont Blanc, Grenoble, France – wilfried.thuiller@univ-grenoble-alpes.fr

Carsten Hobohm: Ecology and Environmental Education Working Group, University of Flensburg (EUF), 24943 Flensburg, Germany – hobohm@uni-flensburg.de

Caroline Brunel: Zhejiang Provincial Key Laboratory of Plant Evolutionary Ecology and Conservation, Taizhou University, Taizhou, China – caroline.brunel@ird.fr

Barry J. Conn (orcid: 0000-0002-7156-4802): School of Life and Environmental Sciences, University of Sydney, Australia – barry.conn@sydney.edu.au

Wayne Dawson: Department of Biosciences, Durham University, Durham, UK – wayne.dawson@durham.ac.uk

Michele de Sá Dechoum:

- Department of Ecology and Zoology, Graduate Course in Ecology, Federal University of Santa Catarina, Florianopolis, SC, Brazil
- The Horus Institute for Environmental Conservation and Development. Florianopolis, SC, Brazil – michele.dechoum@ufsc.br

Aleksandr L. Ebel (orcid: 0000-0002-7889-4580):

¹ Current address: AMAP, Univ Montpellier, CIRAD, CNRS, INRAE, IRD, Montpellier, France

- Department of Botany, Tomsk State University, Tomsk, Russia
- Central Siberian Botanical Garden, Novosibirsk, Russia – alex-08@mail2000.ru

Franz Essl: Bioinvasions, Global Change, Macroecology-Group, Department of Botany and Biodiversity Research, University of Vienna, Vienna, Austria – franz.essl@univie.ac.at

Ori Fragman-Sapir: Jerusalem Botanical Gardens, The Hebrew University of Jerusalem, Jerusalem, Israel – ofragman@013.net

Trevor Fristoe: Ecology, Department of Biology, University of Konstanz, Konstanz, Germany – trevor.fristoe@uni-konstanz.de

Nejc Jogan: Department of Biology, Biotechnical Faculty, University of Ljubljana, Ljubljana, Slovenia – jernej.jogan@bf.uni-lj.si

Holger Kreft:

- Biodiversity, Macroecology & Biogeography, University of Göttingen, Göttingen, Germany
- Centre of Biodiversity and Sustainable Land Use (CBL), University of Goettingen, Germany – hkreft@uni-goettingen.de

Bernd Lenzner: Bioinvasions, Global Change, Macroecology-Group, Department of Botany and Biodiversity Research, University of Vienna, Vienna, Austria – bernd.lenzner@univie.ac.at

Carsten Meyer (orcid: 0000-0003-3927-5856):

- German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany
- Institute of Biology, Leipzig University, Leipzig, Germany

- Institute for Geosciences and Geography, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany – carsten.meyer@idiv.de

Jan Pergl: Czech Academy of Sciences, Institute of Botany, Department of Invasion Ecology, Průhonice, Czech Republic – jan.pergl@ibot.cas.cz

Petr Pyšek (orcid: 0000-0001-8500-442X):

- Czech Academy of Sciences, Institute of Botany, Department of Invasion Ecology, Průhonice, Czech Republic – pysek@ibot.cas.cz
- Department of Ecology, Faculty of Science, Charles University, Prague, Czech Republic

Alla Verkhozna: Siberian Institute of Plant Physiology & Biochemistry, Siberian Branch of Russian Academy of Sciences, Irkutsk, Russia – allaverh@list.ru

Patrick Weigelt: Biodiversity, Macroecology & Biogeography, University of Göttingen, Göttingen, Germany – patrick.weigelt@forst.uni-goettingen.de

Qiang Yang: Ecology, Department of Biology, University of Konstanz, Konstanz, Germany – qiang.yang@uni-konstanz.de

Elena Zykova: Central Siberian Botanical Garden, Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia – elena.yu.zykova@gmail.com

Svetlana Ačić (orcid: 0000-0001-6553-3797): University of Belgrade, Faculty of Agriculture, Department of Botany, Belgrade, Serbia – acic@agrif.bg.ac.rs

Emiliano Agrillo (orcid: 0000-0003-2346-8346): Operational Center for Environmental Monitoring - Institute for Environmental Protection and Research, Rome, Italy – emiliano.agrillo@isprambiente.it

Fabio Attorre: Department of Environmental Biology, Sapienza University of Rome, Italy –
fabio.attorre@uniroma1.it

Ariel Bergamini (orcid: 0000-0001-8816-1420): WSL Swiss Federal Research Institute,
Zürcherstrasse 111, 8903 Birmensdorf, Switzerland – ariel.bergamini@wsl.ch

Christian Berg: Universität of Graz, Institute for Biology, Holteigasse 6, 8010 Graz, Austria –
christian.berg@uni-graz.at

Erwin Bergmeier (orcid: 0000-0002-6118-4611):

University of Göttingen, Department of Vegetation Science and Phytodiversity, 37073
Göttingen, Germany – erwin.bergmeier@bio.uni-goettingen.de

Centre of Biodiversity and Sustainable Land Use (CBL), University of Göttingen, Germany

Idoia Biurrun (orcid: 0000-0002-1454-0433): Department of Plant Biology and Ecology,
University of the Basque Country UPV/EHU, P.O. Box 644, 48080 Bilbao, Spain –
idoia.biurrun@ehu.es

Steffen Boch (orcid: 0000-0003-2814-5343) : WSL Swiss Federal Research Institute,
Zürcherstrasse 111, 8903 Birmensdorf, Switzerland – steffen.boch@wsl.ch

Gianmaria Bonari (orcid: 0000-0002-5574-6067):

- Department of Botany and Zoology, Faculty of Science, Masaryk University, Brno,
Czech Republic
- Faculty of Science and Technology, Free University of Bolzano, Bozen, Italy –
gianmaria.bonari@unibz.it

Zoltán Botta-Dukát (orcid: 0000-0002-9544-3474): Institute of Ecology and Botany, MTA
Centre for Ecological Research, Vácrátót, Hungary – botta-dukát.zoltan@okologia.mta.hu

Helge Bruelheide:

- Martin Luther University Halle-Wittenberg, Institute of Biology/Geobotany and Botanical Garden, Am Kirchtor 1, 06108 Halle, Germany
- German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany – helge.bruehlheide@botanik.uni-halle.de

Juan Antonio Campos (orcid: 0000-0001-5992-2753): Department of Plant Biology and Ecology, University of the Basque Country UPV/EHU, P.O. Box 644, 48080 Bilbao, Spain – juanan.campos@ehu.eus

Andraž Čarni (orcid: 0000-0002-8909-4298) : Research Center of the Slovenian Academy of Sciences and Arts, Novi trg 2, 1000 Ljubljana, Slovenia, University of Nova Gorica, Vipavska 13, 5000 Nova Gorica, Slovenia – carni@zrc-sazu.si

Laura Casella (orcid: 0000-0003-2550-3010): Italian National Institute for Environmental Protection and Research (ISPRA), Department for Biodiversity Conservation and Monitoring, Rome, Italy – laura.casella@isprambiente.it

Maria Laura Carranza (orcid: 0000-0001-5753-890X): Envixlab, Department of Biosciences and Territory, University of Molise, C. da Fonte Lappone, 86090 Pesche, Italy – carranza@unimol.it

Milan Chytrý (orcid: 0000-0002-8122-3075): Department of Botany and Zoology, Faculty of Science, Masaryk University, Brno, Czech Republic – chytry@sci.muni.cz

Renata Čušterevska: Institute of Biology, Faculty of Natural Sciences and Mathematics, University of Ss. Cyril and Methodius, 1000 Skopje, Republic of North Macedonia – renatapmf@yahoo.com

Michele De Sanctis: Department of Environmental Biology, Sapienza University of Rome, P.le Aldo Moro 5, 00185, Rome, Italy – michele.desanctis@uniroma1.it

Jürgen Dengler (orcid: 0000-0003-3221-660X):

- Vegetation Ecology, Institute of Natural Resource Sciences (IUNR), Zurich
University of Applied Sciences (ZHAW), Grüentalstr. 14, 8820 Wädenswil,
Switzerland
- Plant Ecology, Bayreuth Center of Ecology and Environmental Research
(BayCEER), University of Bayreuth, Universitätsstr. 30, 95447 Bayreuth,
Germany
- German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig,
Deutscher Platz 5e, 04103 Leipzig, Germany – dr.juergen.dengler@gmail.com

Panayotis Dimopoulos: University of Patras, Department of Biology, Laboratory of Botany,
26504 Patras, Greece – pdimopoulos@upatras.gr

Rasmus Ejrnæs: Section for Biodiversity, Department of Bioscience, Aarhus University,
Grenåvej 14, 8410 Rønde, Denmark – rasmus@bios.au.dk

Jörg Ewald (orcid: 0000-0002-2758-9324): Institute for Ecology and Landscape,
Weihenstephan-Triesdorf University of Applied Sciences, Hans-Carl-von-Carlowitz-Platz 3,
85354 Freising, Germany – joerg.ewald@hswt.de

Federico Fernández-González: Institute of Environmental Sciences, University of Castilla-La
Mancha, 45071 Toledo, Spain – Federico.Fdez@uclm.es

Rosario G. Gavilán (orcid: 0000-0002-1022-445X): Department of Pharmacology,
Pharmacognosy and Botany, Complutense University, E-28040 Madrid, Spain –
rgavilan@ucm.es

Giuliano Fanelli: Department of Environmental Biology, University of Rome Sapienza, Rome,
Italy – giuliano.fanelli@gmail.com

Jean-Claude Gegout: Université de Lorraine, AgroParisTech, INRAE, Silva, F-54000, Nancy,
France – jean-claude.gegout@agroparistech.fr

Rense Haveman: Central Government Real Estate Agency, Ministry of the Interior and Kingdom Relations, Wageningen, the Netherlands – rense.haveman@wur.nl

Maike Isermann:

- Institute of Ecology, Bremen University, FB 2, 28359 Bremen, Germany
- Lower Saxon Wadden Sea National Park Authority, Nature Conservation, 26382 Wilhelmshaven, Germany – maike.isermann@uni-bremen.de

Ute Jandt: (orcid: 0000-0002-3177-3669)

- Martin Luther University Halle-Wittenberg, Institute of Biology/Geobotany and Botanical Garden, Am Kirchtor 1, 06108 Halle, Germany
- German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany – ute.jandt@botanik.uni-halle.de

Florian Jansen (orcid: 0000-0002-0331-5185): Faculty of Agricultural and Environmental Sciences, University of Rostock, 18059 Rostock, Germany – florian.jansen@uni-rostock.de

Borja Jiménez-Alfaro (orcid: 0000-0001-6601-9597): Department of Organisms and Systems Biology and Research Unit of Biodiversity (UMIB, UO-CSIC-PA), University of Oviedo, Oviedo, Spain – jimenezalfaro.borja@gmail.com

Ali Kavgacı: Department of Forest Botany, Faculty of Forestry, University of Karabük, Karabük, Turkey – alikavgaci1977@yahoo.com

Larisa Khanina: Institute of Mathematical Problems of Biology RAS, Branch of the M.V. Keldysh Institute of Applied Mathematics of the Russian Academy of Sciences, Moscow, Russia – khanina.larisa@gmail.com

Ilona Knollová: Department of Botany and Zoology, Faculty of Science, Masaryk University, Brno, Czech Republic – ikuzel@sci.muni.cz

Anna Kuzemko (orcid: 0000-0002-9425-2756): Department of Geobotany and Ecology, M.G. Kholodny Institute of Botany NAS of Ukraine; 2, Tereshchenkivska str., 01601, Kyiv, Ukraine – anyameadow.ak@gmail.com

Maria Lebedeva: South-Ural botanical garden-institute, Ufa Federal Research Centre, Russian Academy of science; Mendeleev str., 195/3, Ufa, Russia – lebedevamv@mail.ru

Jonathan Lenoir: UR “Ecologie et Dynamique des Systèmes Anthropisés” (EDYSAN, UMR 7058 CNRS), Université de Picardie Jules Verne, 1 Rue des Louvels, 80090 Amiens Cedex 1, France – jonathan.lenoir@u-picardie.fr

Tatiana Lysenko (orcid: 0000-0001-6688-1590):

- Komarov Botanical Institute of the Russian Academy of Sciences, Prof. Popov str. 2, 197376 Saint-Petersburg, Russia
- Institute of Ecology of the Volga River Basin of the Russian Academy of Sciences – Branch of the Samara Scientific Center of the Russian Academy of Sciences, Komzin str. 10, 445003 Togliatti, Russia – ltm2000@mail.ru

Corrado Marcenò (orcid: 0000-0003-4361-5200): Department of Botany and Zoology, Faculty of Science, Masaryk University, Brno, Czech Republic – marcenocorrado@libero.it

Vasiliy Martynenko (orcid: 0000-0002-9071-3789): Laboratory of Geobotany and Plant Resources, Ufa Institute of Biology of Ufa Federal Scientific Centre of Russian Academy of Sciences, prospect Oktyabrya 69, 450054, Ufa, Republic of Bashkortostan, Russia – vasmar@anrb.ru

Jesper Erenskjold Moeslund (orcid: 0000-0001-8591-7149): Section for Biodiversity, Department of Bioscience, Aarhus University, Grenåvej 14, 8410 Rønne, Denmark – jesper.moeslund@bios.au.dk

Ricarda Pätsch (orcid: 0000-0002-3349-0910): Wagner Ecology Lab, Department of Biological Sciences, University of Alberta, Biological Sciences Building, Edmonton, Alberta, T6G 2E9, Canada – ricarda.paetsch@gmail.com

Remigiusz Pielech (orcid: 0000-0001-8879-3305): Department of Forest Biodiversity, Faculty of Forestry, University of Agriculture, al. 29 Listopada 46, 31-425 Kraków, Poland – remekpielech@gmail.com

Valerijus Rašomavičius: Institute of Botany, Nature Research Centre, Vilnius, Lithuania – valerijus.rasomavicius@botanika.lt

Iris de Ronde: Central Government Real Estate Agency, Ministry of the Interior and Kingdom Relations, Wageningen, the Netherlands – iris.deronde@wur.nl

Eszter Ruprecht: Hungarian Department of Biology and Ecology, Babeş-Bolyai University, Cluj-Napoca, Romania – eszter.ruprecht@ubbcluj.ro

Solvita Rūsiņa: Faculty of Geography and Earth Sciences, University of Latvia, Jelgavas iela 1, Rīga LV-1004, Latvia – Solvita.Rusina@lu.lv

Pavel Shirokikh (orcid: 0000-0003-1864-4878): Laboratory of Geobotany and Plant Resources, Ufa Institute of Biology of Ufa Federal Scientific Centre of Russian Academy of Sciences, prospect Oktyabrya 69, 450054, Ufa, Republic of Bashkortostan, Russia – shirpa@mail.ru

Jozef Šibík (orcid: 0000-0002-5949-862X): Institute of Botany, Plant Science and Biodiversity Center, Slovak Academy of Sciences, Dúbravská cesta 9, SK-84523 Bratislava, Slovakia – jozef.sibik@savba.sk

Urban Šilc: Research Center of the Slovenian Academy of Sciences and Arts, Novi trg 2,
1000 Ljubljana, Slovenia – urban@zrc-sazu.si

Angela Stanisci (orcid: 0000-0002-5302-0932): Envixlab, Department of Biosciences and
Territory, University of Molise, Via Duca degli Abruzzi - 86039 Termoli (CB), Italy –
stanisci@unimol.it

Zvezdana Stančić (orcid: 0000-0002-6124-811X): Faculty of Geotechnical Engineering,
University of Zagreb, Hallerova aleja 7, 42000 Varaždin, Croatia – zvezdana.stancic@gfv.hr

Jens-Christian Svenning (orcid: 0000-0002-3415-0862): Center for Biodiversity Dynamics in
a Changing World (BIOCHANGE) and Section for Ecoinformatics and Biodiversity,
Department of Biology, Aarhus University, Ny Munkegade 114, DK-8000 Aarhus C, Denmark
– svenning@bio.au.dk

Grzegorz Swacha (orcid: 0000-0002-6380-2954): Botanical Garden, University of Wrocław,
Wrocław, Poland – grzegorz.swacha@uwr.edu.pl

Pavel Dan Turtureanu (orcid: 0000-0002-7422-3106): A. Borza Botanical Garden, Emil G.
Racoviță Institute, Center for Systematic Biology, Biodiversity and Bioresources - 3B, Babes-
Bolyai University, Cluj-Napoca, Romania – pavel.turtureanu@ubbcluj.ro

Milan Valachovič (orcid: 0000-0001-5296-5240): Institute of Botany, Plant Science and
Biodiversity Center, Slovak Academy of Sciences, Dúbravská cesta 9, SK-84523 Bratislava,
Slovakia – milan.valachovic@savba.sk

Kiril Vassilev: Institute of Biodiversity and Ecosystem Research, Bulgarian Academy of
Sciences, Sofia, Bulgaria – kiril5914@abv.bg

Sergey Yamalov: South-Ural botanical garden-institute, Ufa Federal Research Centre,
Russian Academy of science; Mendeleev str., 195/3, Ufa, Russia – yamalovsm@mail.ru

Mark van Kleunen:

- Zhejiang Provincial Key Laboratory of Plant Evolutionary Ecology and Conservation, Taizhou University, Taizhou, China
- Ecology, Department of Biology, University of Konstanz, Konstanz, Germany – mark.vankleunen@uni-konstanz.de

Biosketch

This work is a collaboration of two international working groups: GloNAF (Global Naturalized Alien Flora) and EVA (European Vegetation Archive), together with external collaborators.

Robin Pouteau is a plant ecologist at IRD with a strong focus on the impact of biological invasions on biodiversity. This work is part of his postdoctoral research at Taizhou University.

Author contributions

Robin Pouteau and Mark van Kleunen conceived the idea for this manuscript with input from Wilfried Thuiller. All authors provided the data, Robin Pouteau performed the analysis and wrote the first draft of the manuscript with help from Wilfried Thuiller and Mark van Kleunen, and all authors contributed substantially to revisions.

Acknowledgements

Robin Pouteau was supported by the National Natural Science Foundation of China (grant number 31901176) and Taizhou University (2018YQ001). Carsten Meyer acknowledges funding by the Volkswagen Foundation through a Freigeist Fellowship (A118199), and additional support by iDiv, funded by the German Research Foundation (DFG–FZT 118, 202548816). Petr Pyšek and Jan Pergl were supported by the Czech Science Foundation (EXPRO grant 19-28807X) and the Czech Academy of Sciences (long-term research development project RVO 67985939). Idoia Biurrun and Juan Antonio Campos were supported by the Basque Government (IT936-16). Gianmaria Bonari, Milan Chytrý, Ilona

Knollová, Anna Kuzemko and Corrado Marcenò were supported by the Czech Science Foundation (19-28491X). Rosario G. Gavilán was supported by Madrid Government project REMEDINAL (P2018/EMT-4338). Jens-Christian Svenning considers this work a contribution to his VILLUM Investigator project “Biodiversity Dynamics in a Changing World” funded by VILLUM FONDEN (16549). Mark van Kleunen was supported by the German Research Foundation DFG (264740629). We appreciate the helpful comments by three anonymous referees and the Handling Editor, Melodie A. McGeoch.

1 **Research paper**

2 Climate and socio-economic factors explain differences between observed and expected
3 naturalization patterns of European plants around the world

4 **Abstract**

5 *Aim* – The number of naturalized (i.e., established) alien species has increased rapidly over
6 the last centuries. Given the differences in environmental tolerances among species, little is
7 known about what factors determine the extent to which the observed size of the naturalized
8 range of a species and hence the extent to which the observed richness of naturalized
9 species of a region approach their full potential. Here, we asked which region- and species-
10 specific characteristics explain differences between observed and expected naturalizations.

11 *Location* – Global.

12 *Time period* – Present.

13 *Major taxa studied* – Vascular plants.

14 *Methods* – We determined the observed naturalized distribution outside Europe for 1,485
15 species endemic to Europe using the GloNAF database and their expected distributions
16 outside Europe using species distribution models. First, we investigated which of seven
17 socio-economic factors related to introduction pathways, anthropogenic pressures and
18 inventory effort best explained the differences between observed and expected naturalized
19 European floras. Second, we examined whether distributional features, economic use and
20 functional traits explain the extent to which species have filled their expected ranges outside
21 Europe.

22 *Results* – In terms of suitable area, more than 95% of expected naturalizations of European
23 plants were not yet observed. Species were naturalized in only 4.2% of their suitable regions
24 outside of Europe (range filling) and in 0.4% of their unsuitable regions (range expansion).
25 Anthropogenic habitat disturbance primarily explained the difference between observed and

26 expected naturalized European floras, as did the number of treaties relevant to invasive
27 species. Species of ornamental and economic value and with large specific leaf area
28 performed better at filling and expanding beyond their expected range.

29 *Main conclusions* – The naturalization of alien plant species is explained by climate matching
30 but also by the regional level of human development, the introduction pressure associated
31 with the ornamental and economic values of the species and their adaptation to disturbed
32 environments.

33

34 **Keywords** – Alien species, anthropogenic pressure, environmental driver, functional trait,
35 global change, introduction pathway, naturalization, ornamental plant, sampling bias, species
36 distribution model.

37 **Introduction**

38 Species that maintain self-sustaining populations outside their native range independently of
39 direct human intervention (i.e., naturalized species; Richardson *et al.*, 2000; Blackburn *et al.*,
40 2011) pose a serious concern to native biodiversity and ecosystem functioning and services
41 worldwide (IPBES, 2019). Their number has increased substantially over the last few
42 centuries and does not show any indication of saturation (Seebens *et al.*, 2017). Regions
43 differ considerably in their richness of naturalized species of vascular plants, and species
44 vary considerably in their extent of naturalization (Pyšek *et al.*, 2017, Essl *et al.*, 2019). Part
45 of this variation may result from regional differences in the size of the potential naturalized
46 alien pool and the size of the range potentially available for naturalized species. Little is
47 known about the relationship of the observed and potential richness of naturalized plant
48 species (i.e. the maximum numbers of alien species that these regions could host given
49 species' environmental tolerances; Stohlgren *et al.*, 2008).

50 The actual distribution of naturalized alien plant species results from the number of
51 introductions by humans (intentional or not) followed by the interplay of factors that
52 determine the environmental space available for a species and factors that prevent or
53 promote the extent to which an equilibrium with an environmental space is reached.
54 Environmental suitability (e.g., climate, habitat) determines the potential (expected)
55 distribution range by filtering the subset of introduced species that can maintain populations
56 over long periods by successfully reproducing, i.e., naturalized species (Richardson & Pyšek,
57 2012). Human activities can further influence the observed distribution range of species by
58 altering environmental conditions and dispersal processes that dictate opportunities for
59 introductions. For instance, human transport, trade and increasing urban development tend
60 to broaden observed ranges, while, conversely, biosecurity responses can contain this
61 expansion (Hulme, 2009).

62 An increasing body of literature is investigating what proportions of species' potential alien
63 ranges are actually occupied (Petitpierre *et al.*, 2012; Strubbe *et al.*, 2013, 2015; Hill *et al.*,
64 2017; Liu *et al.*, 2020). Indeed, most alien species are occupying only a subset of the niche
65 corresponding to suitable environments that are occupied in their native range ('niche filling').
66 Certain alien species are also able to colonize new environments that are not occupied in
67 their native range ('niche expansion'). Patterns of niche filling and niche expansion vary
68 highly depending on the type of organisms. For instance, the average niche filling was found
69 to be 80% for vascular plants considered invasive (i.e., spreading into areas away from sites
70 of introduction; Richardson *et al.*, 2000), 37% for invasive insects and 20% for invasive birds,
71 whereas the average niche expansion was 26% for birds, 20% for insects and 5% for plants
72 (Petitpierre *et al.*, 2012; Strubbe *et al.*, 2013; Hill *et al.*, 2017). Niche changes during
73 invasions appear to be primarily driven by introduction history (e.g., propagule pressure and
74 colonization processes) in birds, and by human disturbance in insects (Hill *et al.*, 2017).
75 However, little is known about what causes possible differences between observed and
76 expected ranges in naturalized plants.

77 Much of the research effort has been devoted to understanding the socio-economic drivers of
78 the number of naturalized alien plant species (Pyšek *et al.*, 2010; van Kleunen *et al.*, 2015;
79 Dawson *et al.*, 2017; Essl *et al.*, 2019), some of them focusing on the subset of invasive
80 species (Essl *et al.*, 2015; Bellard *et al.*, 2016). However, studies dealing with alien species
81 richness always face the challenge of disentangling the specific role of socio-economic
82 drivers from that of large-scale biogeographical factors also known to influence species
83 richness, irrespective of species origin (e.g., latitudinal gradient, resource availability).
84 Moreover, while it is recognized that environmental matching between native and alien
85 ranges largely controls naturalization (Richardson & Pyšek, 2012; Feng *et al.*, 2016),
86 focusing on alien species richness does not allow accounting for species-specific
87 environmental requirements. Hence, determining the expected distributional range of

88 naturalized species should be an important step in understanding the influence of socio-
89 economic factors on observed patterns of naturalization (Xu *et al.*, 2019).

90 Another advantage of analyzing species composition is that it allows for the identification of
91 the species that explain the pattern of naturalized species richness and their characteristics.
92 Many other studies have attempted to determine which species' performance or natural
93 distribution traits are associated with naturalization success by comparing naturalized and
94 non-naturalized plant species (e.g., Pyšek & Richardson, 2007; van Kleunen *et al.*, 2010;
95 Pyšek *et al.*, 2015). Conclusions from these studies delivered little consensus and often
96 appear to be context- and species-dependent. This might partly be because these studies
97 only investigated trait differences of native versus alien or non-invasive versus invasive
98 species (i.e., a binary variable), or tested for relationships between traits and the observed
99 naturalized range without considering the potential range. Therefore, we believe we can gain
100 additional insight into the drivers of naturalization patterns by searching for traits that explain
101 the extent to which species fill or expand beyond their suitable alien range (i.e., a continuous
102 variable that accounts for the potential range).

103 Europe is the second-most important donor continent of naturalized alien plant species, after
104 Asia. However, the European flora is by far the most successful in naturalizing. Of the
105 ~14,200 vascular plant species native (i.e. both endemic and non-endemic) to Europe, 3,383
106 have become naturalized somewhere in the world; this is three times more than expected by
107 chance (van Kleunen *et al.*, 2015). Furthermore, five of the 10 globally most widely
108 naturalized species are native to Europe (Pyšek *et al.*, 2017). The European flora is therefore
109 suitable for studying drivers of differences between observed and expected naturalized
110 ranges.

111 Current knowledge of which region- and plant-related characteristics influence the extent to
112 which species fill their expected naturalized ranges is poor and relies only on the search of
113 the socio-economic drivers of niche/range filling of a few invasive alien species (usually < 50;

114 e.g., Hill *et al.*, 2017; Petitpierre *et al.*, 2012; Strubbe *et al.*, 2013, 2015). However, the
115 drivers of niche/range filling have never been explored for naturalized species (a much
116 broader group of species than invasives; Richardson *et al.*, 2000); yet the processes driving
117 the ability to survive and reproduce are likely to differ from the processes driving the spread
118 and impact of invasive species (e.g., Abellán *et al.*, 2017). Moreover, the drivers of the
119 difference between observed and expected naturalized regional floras (and not between
120 observed and expected distributions of a few individual species) have never been explored.

121 In this study, we focused on 1,485 plant species endemic to Europe, an unprecedented
122 number for this type of study, which allowed us to analyse in more detail what explains
123 variation in the differences between observed and expected values. We determined both
124 their current observed distribution in 931 non-European regions using the 'Global Naturalized
125 Alien Flora' database (GloNAF; van Kleunen *et al.*, 2019) and their expected distribution
126 using biogeoclimatic ensemble species distribution models (SDMs). We also explored which
127 socio-economic factors explain why some regions have a recorded naturalized flora of
128 European origin close to their expected naturalized flora, while other regions have not.
129 Finally, we examined whether species distribution patterns, economic uses and functional
130 traits explain why some species fill a larger portion of their expected range than others do.

131 **Materials and methods**

132 *Expected naturalized ranges*

133 Species selection

134 Many plant species native to Europe are also native to other continents such as Asia and
135 Africa, for which there is usually a lower density (temporal and spatial) of species occurrence
136 records (Meyer *et al.*, 2016). Therefore, we focused on vascular plant species whose native
137 range is restricted to Europe (European endemics), as this allows us to better capture the full
138 realized niche of the species in their native range. The physiography of the European
139 continent was defined as bordered by the Arctic Ocean to the north (Iceland and Norwegian

140 Islands were included, Greenland was excluded), the Atlantic Ocean to the west (the British
141 and Irish Isles were included and the Macaronesian archipelagos were excluded), the Ural
142 Mountains, the Ural River, and the Caspian Sea to the east, and the Caucasus and the
143 Mediterranean Sea to the south (Mediterranean islands were included, Anatolia was
144 excluded).

145 The database 'Endemic vascular plants in Europe' (EvaplantE; Hobohm 2014), which
146 comprises a list of > 6,200 endemic taxa, was used as a baseline for species selection.
147 Scientific names were standardized based on a working list of all plant species (The Plant
148 List; TPL; <http://www.theplantlist.org/> version 12-10-2018). This taxonomic standardization
149 was done with the R package 'Taxonstand' (Cayuela *et al.*, 2017). Standardized infraspecific
150 taxa were excluded from the list, resulting in 4,965 species (Figure 1).

151 Compilation of species occurrence records

152 To comprehensively describe the distribution of the species in Europe (which in addition to
153 the native range might for some species also include part of their alien range), we combined
154 occurrence records from six sources (see Supporting Information Appendix S1 for details
155 and references): the 'Global Biodiversity Information Facility' (GBIF), the 'European
156 Vegetation Archive' (EVA; Chytrý *et al.*, 2016), the 'EU-Forest' dataset, the 'Atlas Florae
157 Europaeae', the 'Plant Functional Diversity of Grasslands' network (DIVGRASS) and the
158 digital atlas of the German flora.

159 When several occurrence records from these different sources were duplicated on the same
160 cell, only one occurrence record per species was kept to avoid pseudoreplication. Cell size
161 was set to $0.42^\circ \times 0.42^\circ$ (~50 km \times 50 km at the equator) to approach the resolution of the
162 source of occurrence records having the coarsest resolution (Atlas Florae Europaeae). Using
163 a higher resolution (i.e., a smaller cell size) would have required to downscale this source of
164 occurrence records, which typically introduces spatial biases and uncertainty in the model
165 predictions (Bonbi & D'Amen, 2012). Moreover, this resolution appeared to be a good

166 compromise to account for the fact that the selected environmental variables (climate, land
167 use and soil type) determine invasion potential at different scales (10,000 to 200 km, 2,000 to
168 10 km, and 10 km to 10 m, respectively; Milbau *et al.*, 2009).

169 Species with fewer than 10 deduplicated occurrence records were not further considered
170 since the resulting SDM might be inaccurate. The final dataset comprised 135,189
171 occurrence records for a total of 1,485 European plant species, belonging to 327 genera and
172 67 families (Table S1), i.e. on average 91 occurrence records per species with a maximum of
173 1,382 occurrence records for *Achillea ptarmica* (Asteraceae). We had enough occurrence
174 records to build an SDM (i.e., 10 occurrences) for 272 European species among the 407
175 already naturalized species and for 1,213 species currently not known to be naturalized
176 anywhere.

177 Environmental predictors of expected ranges

178 We defined six environmental variables to model and project species expected ranges.
179 These variables were related to climate, land use, and soil physico-chemical properties,
180 which are commonly recognized to shape the distribution of plants (Gurevitch *et al.*, 2006).
181 Annual mean temperature (°C), annual precipitation (mm) and precipitation seasonality
182 (yearly coefficient of variation) representing the period 1979-2013 were provided at a 30
183 arcsec resolution by the CHELSA climate database (Karger *et al.*, 2017). Worldwide
184 correlations between these variables and the others described below did not exceed the
185 threshold of $|r| = 0.70$ (Supporting Information Appendix Figure S1) beyond which collinearity
186 begins to severely distort model estimations and subsequent predictions (Dormann *et al.*,
187 2013). However, these variables reflecting trends in average climate conditions were
188 significantly correlated with climate extremes to which plants are recognized to be highly
189 responsive (Zimmermann *et al.*, 2009). The percentage of each grid cell with primary land
190 cover based on the Harmonized Global Land Use models was also used (Chini *et al.*, 2014).
191 Organic carbon content (g per kg) and pH in the first 15 cm of soil were extracted at a 1 km

192 resolution from the global gridded soil information database SoilGrids (Hengl *et al.*, 2014).
193 Environmental variables were aggregated (using the mean value) to the resolution of 0.42° ×
194 0.42°.

195 Species distribution modelling

196 Comparing the realized niche within the native versus alien ranges can lead to three cases:
197 1) in the alien range, the species uses a similar or smaller realized niche than in the native
198 range, e.g., because of new competitors, herbivores or pathogens, or dispersal limitation; 2)
199 the species occupies a realized niche very different from the one in the native area but within
200 the species initial fundamental niche, e.g., because of new biotic interactions, multiple sites
201 of introduction, niche differentiation or different environmental conditions; or 3) the realized
202 niche extends outside of the species initial fundamental niche, e.g., because of rapid niche
203 evolution, enemy-release or new positive interactions (see Gallien *et al.*, 2010 for further
204 details). As options 2 and 3 remain impossible to predict for a large number of species, we
205 only considered option 1 in this study.

206 The expected distribution of the 1,485 European plant species was modelled by statistically
207 relating the environmental predictors to the distribution data in Europe. Six species
208 distribution modelling (SDM) methods including generalized additive models, generalized
209 linear models, generalized boosting trees, maximum entropy, multivariate adaptive
210 regression splines and random forest were used. All of these methods require presence and
211 absence or pseudo-absence/background data (a random subset of the available
212 environmental conditions in the area, i.e., Europe in our case), whose selection can
213 significantly affect predictions if not made adequately. Consequently, we performed a
214 preliminary analysis to identify the most appropriate set of pseudo-absences for each SDM
215 method (Appendix S2). The predictive performance of SDM in Europe was assessed by
216 measuring the area under the receiver operating characteristic (ROC) curve (AUC) and the
217 true skill statistics (TSS; Allouche *et al.*, 2006). AUC values can range from 0 to 1, an AUC of

218 1 indicating a perfect fit, an AUC of 0.5 meaning that predictions from SDM do not differ from
219 random, and an AUC of 0 meaning the SDM is always incorrect. TSS ranges from -1 to 1,
220 where 1 indicates perfect agreement and 0 indicates a random prediction.

221 To combine the predictive capability of the six SDMs, their projections were subsequently
222 aggregated into an average or consensus projection. To ensure the quality of the ensemble
223 SDM, we only kept the projections for which the accuracy estimated by AUC and TSS were
224 higher than 0.8 and 0.6, respectively, and assembled the selected SDMs using a committee-
225 average approach with each SDM was weighted proportional to its TSS evaluation.

226 Probability maps obtained from the ensemble forecasting projections were then transformed
227 into binary suitable/non-suitable maps using the threshold maximizing the TSS to ensure the
228 most accurate predictions, since it is based on both sensitivity and specificity. The entire
229 species distribution modelling workflow was performed within the 'biomod2' R platform
230 (Thuiller *et al.*, 2009).

231 *Recorded naturalized ranges*

232 The current observed naturalized range of the European flora was obtained through the
233 GloNAF database version 1.2, a recently compiled database of ~14,000 naturalized alien
234 plant species covering 1,029 regions worldwide, corresponding to administrative regions
235 such as countries, states and provinces, and also including 381 islands (van Kleunen *et al.*,
236 2019). The size of a region ranged from 0.03 km² (Tauna islet of the Gambier archipelago,
237 French Polynesia) to 2,486,952 km² (the Republic of the Sudan plus South Sudan). A total of
238 407 European endemic species were found as naturalized in at least one of the non-
239 European GloNAF regions.

240 We explored various definitions of suitability by analyzing what proportion of a region should
241 be suitable for a species to become naturalized in this region. The following definitions were
242 used: at least one suitable cell, 1%, 5%, 20%, 50%, or 100% suitable cells, respectively. For
243 each region, species were classified into: true positives (TP), i.e., species that are both

244 observed and predicted as being successfully naturalized in the region; false positives (FP),
245 i.e., species that have not become naturalized, but are predicted as successfully naturalized;
246 false negatives (FN), i.e., species that are observed as successfully naturalized but not
247 predicted to be naturalized; and true negatives (TN), i.e., species that are neither predicted
248 nor observed to be naturalized in the region. Then, the predicted naturalized species pool
249 was compared to the observed naturalized pool by calculating: 1) the naturalization debt in
250 the broad sense hereafter referred to as 'naturalization debt' ($FP/(TP+FP)$), i.e., the
251 proportion of predicted naturalizations that are not yet observed either because species are
252 not yet introduced to the focal area (introduction debt) or because they were introduced, but
253 are not naturalized yet (naturalization debt in the strict sense, i.e., *sensu Rouget et al.*, 2016),
254 which can take several decades or centuries for some taxa; 2) the assemblage sensitivity
255 ($TP/(TP+FN)$), i.e., the proportion of observed naturalizations that are correctly predicted ; 3)
256 the negative predictive value ($TN/(TN+FN)$), i.e., the proportion of species that were not
257 predicted to be naturalized and have not become naturalized in the region.

258 *Socio-economic drivers of naturalization patterns*

259 Seven factors were extracted to explain naturalization debt, assemblage sensitivity and the
260 negative predictive value (Table 1). These factors were selected because they provided
261 insights into introduction pathways, human pressures and inventory effort: airport density
262 ('Airports'), seaport density ('Seaports'), number of international treaties relevant to invasive
263 alien species ('Treaties'), human population density ('Population'), proportion of croplands
264 ('Cropland'), human development index ('Development') and species inventory effort
265 ('Inventory'; for details, see Appendix S3). These socio-economic factors were independent
266 of the surface area of the regions. They were available for 526 non-European regions
267 including 17 islands or archipelagos and 509 mainland areas. Regional factors were tested
268 for multicollinearity by computing a correlation matrix based on Pearson's r . Cross-
269 correlations did not exceed $r = 0.51$ (between airport and seaport density), which is below the
270 threshold of 0.70 (Figure S3; Dormann *et al.*, 2013).

271 Identification of the drivers of naturalization patterns was based on boosted regression trees
272 (BRTs), a machine-learning method that combines a large number of relatively simple tree
273 models to optimize predictive performance (Elith *et al.*, 2008). The quality of BRT fits was
274 controlled with the coefficient of determination of the regression between estimated and
275 observed values (r^2) and the standard error (se). Results were interpreted by looking at the
276 relative influence of the seven regional socio-economic factors to the predictive models
277 (calculated based on the number of times a factor is selected in the model, weighted by its
278 improvement to the overall model) and by considering the partial dependence of the
279 predictions on each factor after accounting for the average effect of the other factors.

280 *Plant features explaining naturalization patterns*

281 To assess which species-level features affect naturalization success, we first compared
282 European species recorded as naturalized outside Europe ('naturalized') and those not
283 recorded as naturalized outside Europe ('non-naturalized'). Eight features including two
284 distributional indices, the economic use of the species and four functional traits were
285 selected: 1) species prevalence in Europe measured as the number of $0.42^\circ \times 0.42^\circ$ cells
286 occupied; 2) the surface area of the species' expected range outside Europe. These
287 distributional indices were used because more widespread species are usually hypothesized
288 to have a greater potential to spread and a higher invasion success (Gallien *et al.*, 2019); 3)
289 whether the species is used as an ornamental plant or 4) has other economic relevance
290 (e.g., food plant, medicines, materials) according to the World Economic Plants (WEP)
291 database (Wiersema & León, 2013) since the ornamental and economic relevance of a
292 species is thought to be a key predictor of its probability to be introduced and its
293 naturalization success (van Kleunen *et al.*, 2018, 2020); 5) the growth form (i.e., graminoid,
294 non-graminoid herb, shrub or tree), a major determinant of invasiveness (Pyšek &
295 Richardson, 2007); 6) mean specific leaf area (SLA), i.e., the ratio of leaf area to leaf dry
296 mass (expressed in $\text{m}^2\cdot\text{g}$); 7) mean plant height (in m); and 8) mean seed mass (in g). These
297 last three functional traits represent key axes of plant ecological strategies following the leaf–

298 height–seed (LHS) scheme of Westoby (1998). Correlation between LHS traits was $|r| <$
299 0.27. Functional traits were extracted from the TRY database (Kattge *et al.*, 2020). Linear
300 mixed-effect models (LMMs) were fitted using a phylogenetic generalized least square
301 approach (PGLS) to compare features of naturalized and non-naturalized species while
302 controlling for between-species phylogenetic distances (for details on the phylogeny used,
303 see Appendix S4). As we did not have data on all features for each species (see Figure 6),
304 we ran separate LMMs for each of the eight features. LMMs were performed using the *gls()*
305 function of the R package ‘nlme’ (Pinheiro *et al.*, 2019).

306 We used LMMs accounting for phylogeny to test whether plant-feature values explain: 1) the
307 relative extent of range filling, i.e., the proportion of suitable regions where a species has
308 become naturalized; and 2) the relative extent of range expansion, i.e., the proportion of
309 unsuitable regions where a species has nevertheless become naturalized. The indices of
310 range filling and expansion that we used are similar to the indices used by Petitpierre *et al.*
311 (2012). However, we calculated them in the geographical space instead of in the
312 environmental space because occurrence data on naturalized plants are not evenly
313 comprehensive across regions (which would have been needed to offer a reliable fit of the
314 niches), and when an alien species is reported, we do not always know whether or not it
315 reproduces in the wild (Figure 1).

316 **Results**

317 *Socio-economic drivers of naturalization patterns*

318 European endemic vascular plant species have naturalized in 319 non-European regions.
319 The naturalization debt (i.e., the proportion of species which are not yet naturalized in a
320 particular region but likely to become so, if introduced, given their environmental envelope)
321 ranged from 95 to 100% (mean = 99%) among the 931 non-European regions according to
322 the threshold of one pixel and from 71 to 100% (mean = 99% as well) according to the
323 threshold of 100% of suitable cells (Figure 2). This means that less than 5% and less than

324 29% of species for which a region is suitable according to the lowest and highest thresholds,
325 respectively, have actually been recorded there. When restricting the analysis to 272
326 European species already naturalized outside of Europe, the naturalization debt decreased
327 but remained relatively high with values in the range 81-100% (mean = 96%) with the former
328 threshold and 52-100% (95%) with the latter threshold.

329 The proportion of observed naturalizations that are correctly predicted based on
330 environmental matching (sensitivity) ranged from 0 to 100% for all thresholds and both
331 species sets. However, the average value of sensitivity was two percentage points lower for
332 the full set of species (19-64%) than for the subset of already naturalized species (21-66%).
333 The average proportion of unsuccessful naturalizations in unsuitable environments (negative
334 predictive value) was between 98.8% and 100% (mean = 99.9%) with the lowest threshold,
335 and between 97.0% and 100% (mean = 99.7%) with the highest threshold for the full set of
336 species. When considering already naturalized species only, the negative predictive value
337 decreased substantially to the range 80.0-100% (mean = 99.4%) in the former case and
338 83.4-100% (mean = 98.8%) in the latter case. The threshold of 1% of suitable cells offered
339 the best solution for maximizing sensitivity while still having a high negative predictive value
340 when comparing observed and expected naturalized floras. Therefore, we used predictions
341 from this threshold in the subsequent analyses.

342 Naturalization debt was lower in more developed regions (human development index > 0.7)
343 (Figures 3 and 4) with moist and cool climates, including coastal North America, Japan,
344 temperate Australia and New Zealand (Figure 5). Sensitivity increased progressively with the
345 level of development and inventory effort (Figures 3 and 4), and reached maximum values in
346 North-East America, the South American Atlantic Forest and the southern Andes, Turkey, the
347 East-Central African mountains, South Africa, Japan, temperate Australia and New Zealand
348 (Figure 5). The negative predictive value was lower in more developed regions (human
349 development index > 0.7), with more treaties relevant to biological invasions (number of
350 treaties > 20; Figures 3 and 4), including regions in North America, the southern tip of South

351 America and Africa, most of northern Asia excluding Mongolia, temperate Australia and New
352 Zealand (Figure 5).

353 *Plant features explaining naturalization patterns*

354 Compared to plant species that are currently not known to be naturalized somewhere else
355 (1,213), those that are naturalized (272) tend to be more widespread in Europe, to have a
356 wider expected range outside Europe, and to have an ornamental or other economic use
357 (Figure 6). There was no significant difference in functional traits between naturalized and
358 non-naturalized species.

359 On average, species were found to be naturalized in 4.2% of their potentially suitable regions
360 outside Europe (Figure S4). Thirty of the species occupied less than 0.1% of suitable regions
361 and 31 species occupied more than 10% of suitable regions. The species with the most
362 complete range filling was *Syringa vulgaris* (Oleaceae), native to the Balkan Peninsula and
363 Romania, being currently naturalized in 70% of its suitable regions (Table S2). The average
364 range expansion into regions estimated to be unsuitable was 0.4%. 118 species (43%)
365 exhibited range expansion above 0.1%, but only one species had a range expansion above
366 10%: *Verbascum virgatum* (Scrophulariaceae), native to south-western Europe and southern
367 England, and currently naturalized in 10.5% of the regions that were predicted to be
368 unsuitable for it. We found that species performing best at filling their expected range have
369 an ornamental or other economic relevance, a higher SLA, and a narrower expected range
370 than species with a less complete range filling (Table 2). Furthermore, species observed as
371 naturalized beyond their expected range tend to have an ornamental or other economic use
372 and to be less widespread in their native range than species with no range expansion (Table
373 2).

374 **Discussion**

375 The global geographic pattern of richness in naturalized plant species from Europe is non-
376 random as supported by our results, originating from a complex interplay of at least four

377 mechanisms: 1) environmental matching between the native range of individual species and
378 their expected alien range, which largely controls unsuccessful naturalization; 2) the
379 alteration of the environment in the alien ranges by socio-economic activities (after
380 accounting for sampling effort); 3) introduction pressure associated with the economic use of
381 the species; and 4) differences in functional traits (e.g., SLA), which cause some species to
382 more completely fill their expected alien ranges than others.

383 *Accounting for environmental matching to decipher patterns of naturalization*

384 We modelled environmental tolerance of species and then the influence of extrinsic socio-
385 economic drivers (Figure 1), instead of both together as done in previous studies (Essl *et al.*,
386 2015; Dawson *et al.*, 2017). This allowed us to separate the effect of environmental matching
387 in explaining the global pattern of richness in naturalized plant species from Europe. We
388 elucidated that the environment alone is able to correctly predict up to two thirds of
389 successful naturalizations (i.e., sensitivity) and 99.9% of unsuccessful naturalizations (i.e.,
390 negative predictive value), depending on the method used to compare local expected
391 distributions with regional observed distributions ('upscaling method'; Figure 2). Based on a
392 similar approach, Bellard *et al.* (2013) showed that even the so-called "world's 100 worst
393 invasive alien species" have a quite restricted expected range, covering mainly Europe and
394 the areas along the Atlantic coast of North America. Thus, it clearly appears that knowing the
395 available suitable environmental space of species is critical to assess the factors that
396 determine their naturalization success.

397 Our approach indicated that non-European regions are currently occupied by less than 5% of
398 the endemic European plants for which the area, or parts of it, would be suitable (Figure 2).
399 Therefore, all regions have an enormous naturalization debt. This debt suggests that many
400 endemic European plants have not been introduced outside of Europe yet or that they still
401 have not overcome the biotic barriers provided by herbivores, pathogens and native
402 competitors or the lack of important mutualists. Nevertheless, the large environmentally

403 suitable ranges outside of Europe confirm that European plants have a considerable
404 potential of naturalization outside their native range (van Kleunen *et al.*, 2015; Pyšek *et al.*,
405 2017), once they have overcome dispersal barriers and biotic resistance. Our results
406 therefore indicate that it is unlikely that the number of new naturalizations will soon have
407 reached its peak.

408 SDMs are increasingly used as a basis to implement biosecurity policies, e.g., by drawing up
409 watch lists or lists of undesirable species (e.g., Padayachee *et al.*, 2019). As biosecurity
410 measures are usually taken at a regional scale while SDMs predict expected distributions at
411 smaller scales, it is critical to determine which portion of a region should be suitable for an
412 alien species to become naturalized, and hence, of concern. Our retrospective approach
413 comparing SDM outputs with the current naturalized flora from Europe in 931 non-European
414 regions showed that a threshold in the order of 1% of a focal region offers a good trade-off
415 between maximizing correctly predicted successful and unsuccessful naturalizations (Figure
416 2). Nevertheless, it may be preferable to choose a lower threshold with the precautionary
417 principle, and this threshold might change when using a different spatial resolution.

418 *Socio-economic drivers explaining differences from expected patterns of naturalization*

419 Our results revealed that naturalization debt mostly decreases with the level of development
420 (Figures 3 and 4), which means that a higher proportion of species with the ability to become
421 naturalized are observed as such in more developed regions. In these regions, the
422 distribution of naturalized species is typically closer to an equilibrium with the environment.
423 This result is in line with Pyšek *et al.* (2010) and Dawson *et al.* (2017), who found that human
424 impact (represented by proxies like wealth and per capita GDP, respectively) play a leading
425 role in driving the global patterns of animal and plant naturalizations. One explanation is that
426 economic wealth is often associated with alien species introductions and with profound
427 alteration of natural habitats, which benefit alien species establishment and growth
428 (MacDougall & Turkington, 2005; Seebens *et al.*, 2015).

429 Surprisingly, unlike other studies (Dawson *et al.*, 2017; Tingley *et al.*, 2018), proxies of
430 international transport (i.e., density of airports and seaports in our study) did not explain
431 naturalization debt (Figures 3 and 4). Nevertheless, we found that regions with the lowest
432 naturalization debt are all either coastal or island regions (coastal North America, Algeria,
433 Japan, temperate Australia and New Zealand; Figure 5). As a result, although naturalization
434 debt seems little influenced by the density of seaports, it is likely that the volume of maritime
435 traffic, or at least the presence of seaports, does play a significant role by increasing the
436 rates of species introductions. One might argue that this pattern also reflects that coastal
437 regions are usually more strongly developed than more inland regions as a consequence of
438 the increasing demand for infrastructures to sustain residential, commercial and tourist
439 activities (Figure S2), which offers more opportunities for introduction and naturalization of
440 alien plants in coastal regions (Gallardo *et al.*, 2015). We found that naturalization debt was
441 higher for the full set of 1,485 European species (either already naturalized or not yet
442 naturalized) than for the subset of 272 species already naturalized, which are more likely to
443 have overcome dispersal barriers (Figure 2). This suggests that introduction pathways
444 influence naturalization debt although, unfortunately, it remains difficult to assess by how
445 much.

446 In addition to the influence of development and potentially of introduction pathways on
447 naturalized range filling, sensitivity was also found to be affected by the magnitude of
448 inventory effort (Figures 3 and 4). Specifically, regions with the highest quality record of
449 native plant richness also tend to have more records of alien plants. This is also corroborated
450 by Dawson *et al.* (2017), and indicates that the geographic coverage of reported plant
451 naturalizations remains unevenly exhaustive across regions so that our findings might be
452 significantly influenced by biases in recording effort. In contrast, the negative predictive value
453 decreased in regions with more treaties relevant to invasive alien species (Figures 3 and 4).
454 This most likely does not indicate that treaties promote invasions, but that it is more likely
455 that a region adopts a treaty when it has many invasive species. The effectiveness of these
456 relatively recent treaties in preventing new invasions can only be assessed in the future.

457 Unexpectedly, human population density and the proportion of cropland had a relatively low
458 contribution in explaining the global pattern of naturalization. Contrary to this, high human
459 population densities are thought to influence the likelihood of an introduced species
460 becoming naturalized and spreading through human impacts on the environment (Essl *et al.*,
461 2011; Pyšek *et al.*, 2010). Similarly, intensive agriculture (associated with historical sowing or
462 planting of now naturalized species and associated weed-seed contaminants) is largely
463 recognized to be a major cause of new introductions of naturalized and invasive species
464 (Bellard *et al.*, 2016; Perrings *et al.*, 2005; Seebens *et al.*, 2015). A potential explanation for
465 the low contributions of population density and the proportion of cropland in driving patterns
466 of naturalization could be its redundancy with the proportion of primary land cover used as
467 environmental variable in SDMs since urbanization and agriculture are largely responsible for
468 land conversion (Chini *et al.*, 2014). Although this redundancy makes it difficult to isolate the
469 effect of human population density and agriculture, we believe the introduction of primary
470 land cover in SDMs was critical in this study to distinguish plants that are drivers and
471 passengers of human-induced disturbances (MacDougall & Turkington, 2005). Another
472 possible cause for the low contribution of agriculture is the ancient human colonization
473 history of Europeans who used to transport cultivated plants across continents while other
474 regions only recently opened up to international movements of plants (di Castri, 1989). In this
475 perspective, most of the plants used in agriculture and associated weeds endemic to Europe
476 with the potential to become naturalized might have already largely spread worldwide
477 (Monnet *et al.*, 2020).

478 *Plant features explaining differences from expected patterns of naturalization*

479 The average score of alien geographic range filling of naturalized plants from Europe was
480 4.2%. This might appear to be very low compared to percentages of environmental niche
481 filling reported in other studies (Hill *et al.*, 2017; Petitpierre *et al.*, 2012; Strubbe *et al.*, 2013,
482 2015; Liu *et al.*, 2020). However, it should be noted that it is easier to fill a niche than to fill all
483 locations that have that niche. None of our species had a range filling above the commonly

484 used threshold of 90%. Similarly, no invasive birds showed more than 90% of their native
485 niche filled in the invasive range (Strubbe *et al.*, 2013). However, no less than 52% of
486 invasive plants (Petitpierre *et al.*, 2012), 32% of 22 insects (Hill *et al.*, 2017) and 10% of
487 vertebrates (Strubbe *et al.*, 2015) had a niche filling above 90%. Moreover, only one
488 naturalized plant species from Europe out of 272 had a range expansion above the
489 commonly used threshold of 10%, while it was the case for the niche expansion of 55% of
490 invasive insects (Hill *et al.*, 2017), 29% of birds (Strubbe *et al.*, 2013), 17% of vertebrates
491 (Strubbe *et al.*, 2015) and 14% of invasive plants (Petitpierre *et al.*, 2012).

492 We found that SLA was positively correlated with range filling of naturalized plants from
493 Europe (Table 2). According to the LHS scheme of Westoby (1998), SLA is a proxy for rapid
494 growth and an acquisitive resource-use strategy, which would be typical of resource-rich
495 and/or disturbed environments, both of which have become more common in the last
496 centuries due to human activities. In the competitor/stress-tolerator/ruderal (CSR) framework
497 of the universal adaptive strategy theory (Grime, 1977), rapid growth is a key characteristic of
498 strong competitors, suggesting that strong competitors for light and nutrients are also filling
499 their naturalized ranges faster. Indeed, Guo *et al.* (2018) recently showed that species with
500 high scores on the competitor axis were more likely to naturalize. Thus, it seems very
501 consistent that having a high SLA represents a key strategy in regions subject to intensive
502 development associated with rapid, deep and widespread anthropogenic perturbations. SLA
503 should therefore also indicate potential for faster expansion towards the equilibrium range in
504 alien regions. This corroborates previous results showing that SLA is higher for naturalized
505 and invasive species compared to non-naturalized or native species (van Kleunen *et al.*,
506 2010; Pyšek & Richardson, 2007; Pyšek *et al.*, 2015).

507 Interestingly, seed mass was not significantly related to range filling (Table 2). This trait can
508 have contrasting effects on patterns of naturalization since large seeds have greater nutrient
509 reserves, which increases the chances to become naturalized, but small seeds are beneficial
510 for long-distance dispersal and therefore favour spread (Crawley *et al.*, 1996; Moodley *et al.*,

511 2013). This suggests that it is not a species' dispersal ability or the ability to settle in new
512 regions that mainly drive naturalization but the ability to be transported by humans for
513 aesthetic or economic values. Our results showed that species used by humans are indeed
514 more frequently naturalized (Figure 6), and perform better at filling (e.g., *Syringa vulgaris*)
515 and expanding (e.g., *Verbascum virgatum*) their potential range than species not used by
516 humans (Table 2). Ornamental horticulture, agriculture and forestry are recognized as major
517 pathways of alien plant introduction. For instance, it has been estimated that at least 75%
518 and 93%, respectively, of the naturalized alien plants worldwide are grown in domestic and
519 botanical gardens (van Kleunen *et al.*, 2018), and plants with a known economic importance
520 are 18 times more likely to be naturalized somewhere in the world (van Kleunen *et al.*, 2020).
521 Possible reasons include: economic importance increases the number of releases and/or the
522 number of individuals released (i.e., propagule pressure), while human selection favours
523 traits related to invasiveness (e.g., ease to grow and reproduce, early and long flowering
524 period, low susceptibility to insect pests or pathogens).

525 Moreover, we found that the same set of distributional features explains whether or not
526 species are naturalized and the extent of naturalized range filling and geographic expansion
527 of species range (range filling and expansion were significantly correlated; $r = 0.61$, P -value
528 < 0.001), but in opposite directions. Indeed, naturalized species tend to be more widespread
529 in their native range and to have wider expected alien ranges than non-naturalized species,
530 i.e., they are typically generalist species (Gallien *et al.*, 2019). In contrast, naturalized
531 species with the most complete range filling and broadest range extension tend to have more
532 limited expected alien ranges and to be less widespread in their native range, respectively.

533 This result needs to be interpreted with caution since it can simply arise from a numerical
534 artefact as it may be easier to fill a small expected range than a large one. Another possible
535 cause could be that SDMs tend to be more accurate for specialist species than for
536 generalists (Connor *et al.*, 2018). Furthermore, range filling depends on opportunities to
537 disperse, so species with a wider expected alien range may basically have less chance to fill

538 it. Finally, more narrowly distributed endemics can surprisingly have a greater range
539 expansion if their observed distribution in the native range results from biogeographical (e.g.,
540 the Mediterranean Basin, the Alps in Europe) or from biological interactions rather than from
541 climatic barriers. In that case SDMs would fail to fit their fundamental climatic envelope due
542 to niche truncation (Bush *et al.*, 2018). It could also be that their current distribution is
543 matching their climatic niche, but that this distribution is not correctly reflected in data
544 sources, maybe because they mostly inhabit under-sampled regions. Consequently, the
545 projected expected alien ranges may underestimate the true potential ranges.

546 **Conclusions**

547 Our findings indicate that the environment largely controls the expected naturalized range of
548 European plants. We demonstrated that increased anthropogenic disturbance associated
549 with human development as well as increased probability of introductions associated with
550 international exchange primarily explain the extent to which species spread within their
551 expected naturalized range. We showed that plants selected for ornamental and other
552 economic purposes perform better at filling and expanding their range, most likely because of
553 increased introduction pressure and a preference for cultivating plants with traits that also
554 make them more likely to naturalize. We revealed that species with functional traits indicative
555 of rapid growth and acquisitive resource use tend to fill their range more completely than
556 species with slower responses. These findings give a new overall picture of the drivers of
557 naturalization that can help plan future studies on the macroecology of alien species as well
558 as in designing future biosecurity plans. As harmful invasive species emerge from the pool of
559 naturalized species, the analysis of which species could naturalize is a first step in identifying
560 species that could ultimately become invasive in a region.

561 **Data availability statement**

562 The data used to build the species distribution models (occurrence records and
563 environmental variables) came from openly accessible repositories cited in the manuscript

564 (for details on the EVA database, see <http://euroveg.org/eva-database-obtaining-data>). All
565 1,485 generated habitat suitability raster files are available at
566 <https://dataverse.ird.fr/dataset.xhtml?persistentId=doi:10.23708/RNGS8Z>.

567 **References**

- 568 1. Abellán, P., Tella, J. L., Carrete, M., Cardador, L., Anadón, J. D. (2017). Climate
569 matching drives spread rate but not establishment success in recent unintentional
570 bird introductions. *Proceedings of the National Academy of Sciences USA*, 114,
571 9385–9390.
- 572 2. Allouche, O., Tsoar, A., Kadmon, R. (2006). Assessing the accuracy of species
573 distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of*
574 *Applied Ecology*, 46, 1223–1232.
- 575 3. Bellard, C., Thuiller, W., Leroy, B., Genovesi, P., Bakkenes, M., Courchamp, F.
576 (2013). Will climate change promote future invasions? *Global Change Biology*, 19,
577 3740–3748.
- 578 4. Bellard, C., Leroy, B., Thuiller, W., Rysman, J.-F., Courchamp, F. (2016). Major
579 drivers of invasion risks throughout the world. *Ecosphere*, 7, e01241.
- 580 5. Blackburn, T. M. *et al.* (2011). A proposed unified framework for biological invasions.
581 *Trends in Ecology and Evolution*, 26, 333–339.
- 582 6. Bombi, P., D’Amen, M. (2014). Scaling down distribution maps from atlas data: a test
583 of different approaches with virtual species. *Journal of Biogeography*, 39, 640–651.
- 584 7. Bush, A., Catullo, R. A., Mokany, K., Thornhill, A. H., Miller, J. T., Ferrier, S. (2018).
585 Truncation of thermal tolerance niches among Australian plants. *Global Ecology and*
586 *Biogeography*, 27, 22–31.
- 587 8. di Castri, R. (1989). History of biological invasions with special emphasis on the Old
588 World. In Drake, J. A. *et al.* (eds) *Biological Invasions: a Global Perspective*, John
589 Wiley, 1–30.

- 590 9. Cayuela, L., Stein, A., Oksanen, J. (2017). *Taxonstand: taxonomic standardization of*
591 *plant species names v.2.1*. R Foundation for Statistical Computing. Available at
592 <https://cran.r-project.org/web/packages/Taxonstand/index.html>. Accessed on 21
593 March 2020.
- 594 10. Chini, L. P., Hurtt, G. C., Frohking, S. (2014). Harmonized Global Land Use for Years
595 1500 – 2100, V1. Data set. Oak Ridge National Laboratory Distributed Active Archive
596 Center, USA. Available at <http://daac.ornl.gov>. Accessed on 21 March 2020.
- 597 11. Chytrý, M. *et al.* (2016). European Vegetation Archive (EVA): an integrated database
598 of European vegetation plots. *Applied Vegetation Science*, 19, 173–180.
- 599 12. Connor, T. *et al.* (2018). Effects of grain size and niche breadth on species
600 distribution modeling. *Ecography*, 41, 1270–1282.
- 601 13. Crawley, M. J., Harvey, P. H., Purvis, A. (1996). Comparative ecology of the native
602 and alien floras of the British Isles. *Philosophical Transactions of the Royal Society B*,
603 351, 1251–1259.
- 604 14. Dawson, W. *et al.* (2017). Global hotspots and correlates of alien species richness
605 across taxonomic groups. *Nature Ecology & Evolution*, 1, 0186.
- 606 15. Dormann, C. F. *et al.* (2013). Collinearity: A review of methods to deal with it and a
607 simulation study evaluating their performance. *Ecography*, 36, 27–46.
- 608 16. Elith, J., Leathwick, J. R., Hastie, T. (2008). A working guide to boosted regression
609 trees. *Journal of Animal Ecology*, 77, 802 – 813.
- 610 17. Essl, F. *et al.* (2011). Socioeconomic legacy yields and invasion debt. *Proceedings of*
611 *the National Academy of Sciences USA*, 108, 203–207.
- 612 18. Essl, F., Dullinger, S., Moser, D., Steinbauer, K., Mang, T. (2015). Macroecology of
613 global bryophyte invasions at different invasion stages. *Ecography*, 38, 488–498.
- 614 19. Essl, F. *et al.* (2019). Drivers of the relative richness of naturalized and invasive plant
615 species on Earth. *AoB PLANTS*, 11, plz051.

- 616 20. Feng, Y., Maurel, N., Wang, Z., Ning, L., Yu, F.-H., van Kleunen, M. (2016).
617 Introduction history, climatic suitability, native range size, species traits and their
618 interactions explain establishment of Chinese woody species in Europe. *Global*
619 *Ecology and Biogeography*, 25, 1356–1366
- 620 21. Gallardo, B., Zieritz, A., Aldridge, D. C. (2015). The importance of the human footprint
621 in shaping the global distribution of terrestrial, freshwater and marine invaders. *PLoS*
622 *ONE*, 10, e0125801.
- 623 22. Gallien, L., Münkemüller, T., Albert, C. H., Boulangeat, I., Thuiller, W. (2010).
624 Predicting potential distributions of invasive species: where to go from here? *Diversity*
625 *and Distributions*, 16, 331–342.
- 626 23. Gallien, L., Thornhill, A. H., Zurell, D., Miller, J. T., Richardson, D. M. (2019). Global
627 predictors of alien plant establishment success: combining niche and trait proxies.
628 *Proceedings of the Royal Society B: Biological Sciences*, 286, 20182477.
- 629 24. Grime, J. P. (1977). Evidence for the existence of three primary strategies in plants and
630 its relevance to ecological and evolutionary theory. *American Naturalist*, 111, 1169–
631 1194.
- 632 25. Guo, W.-Y. *et al.* (2018). The role of adaptive strategies in plant naturalization. *Ecology*
633 *Letters*, 21, 1380–1389.
- 634 26. Gurevitch, J., Scheiner, S. M., Fox, G. A. (2006). *The ecology of plants*, 2nd edition.
635 Sunderland, Sinauer.
- 636 27. Hengl, T. *et al.* (2014). SoilGrids1km — Global Soil Information Based on Automated
637 Mapping. *PLoS ONE*, 9, e105992.
- 638 28. Hill, M. P., Gallardo, B., Terblanche, J. S. (2017). A global assessment of climatic niche
639 shifts and human influence in insect invasions. *Global Ecology and Biogeography*, 26,
640 679–689.
- 641 29. Hobohm, C. (2014). *Endemism in Vascular Plants*. *Plant and Vegetation*, 9, 348 pp.,
642 Springer, Dordrecht.

- 643 30. Hulme, P. (2009). Trade, transport and trouble: managing invasive species pathways in
644 an era of globalization. *Journal of Applied Ecology*, 46, 10–18.
- 645 31. IPBES (Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem
646 Services) (2019). *Summary for policymakers of the global assessment report on*
647 *biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on*
648 *Biodiversity and Ecosystem Services*. Díaz, S. *et al.* (eds.) Bonn: IPBES secretariat.
- 649 32. Karger, D. N. *et al.* (2017). Climatologies at high resolution for the earth's land surface
650 areas. *Scientific Data*, 4, 170122.
- 651 33. Kattge, J. *et al.* (2020). TRY plant trait database – enhanced coverage and open access.
652 *Global Change Biology*, 26, 119–188.
- 653 34. van Kleunen, M., Weber, E., Fisher, M. (2010). A meta-analysis of trait differences
654 between invasive and non-invasive plant species. *Ecology Letters*, 13, 235–245.
- 655 35. van Kleunen, M. *et al.* (2015). Global exchange and accumulation of non-native plants.
656 *Nature*, 525, 100–103.
- 657 36. van Kleunen, M. *et al.* (2018). The changing role of ornamental horticulture in alien plant
658 invasions. *Biological Reviews*, 93, 1421–1437.
- 659 37. van Kleunen, M. *et al.* (2019). The Global Naturalized Alien Flora (GloNAF) database.
660 *Ecology*, 100, e02542.
- 661 38. van Kleunen, M. *et al.* (2020). Economic use of plants is key to their naturalization
662 success. *Nature Communications*, 11, 3201.
- 663 39. Liu, C., Wolter, C., Xian, W., Jeschke, J. M. (2020). Most invasive species largely
664 conserve their climatic niche. *Proceedings of the National Academy of Sciences USA*,
665 117, 23643-23651.
- 666 40. MacDougall A. S., Turkington, R. (2005). Are invasive species the drivers or passengers
667 of change in degraded ecosystems? *Ecology*, 86, 42–55.

- 668 41. Meyer, C., Weigelt, P., Kreft, H. (2016). Multidimensional biases, gaps and uncertainties
669 in global plant occurrence information. *Ecology Letters*, 19, 992–1006.
- 670 42. Milbau, A., Stout, J. C., Graae, B. J., Nijs, I. (2009). A hierarchical framework for
671 integrating invasibility experiments incorporating different factors and spatial scales.
672 *Biological Invasions*, 11, 941–950.
- 673 43. Monnet, A. C., Vorontsova, M. S., Govaerts, R. H. A., Svenning, J.-C., Sandel, B.
674 (2020). Historical legacies and ecological determinants of grass naturalizations
675 worldwide. *Ecography*, 43, 1373–1385.
- 676 44. Moodley, D., Geerts, S., Richardson, D. M., Wilson, J. R. U. (2013). Different traits
677 determine introduction, naturalization and invasion success in woody plants: Proteaceae
678 as a test case. *PLOS ONE*, 8, e75078.
- 679 45. Padayachee, A. L., Proches, S., Wilson, J. R. U. (2019). Prioritising potential incursions
680 for contingency planning: pathways, species, and sites in Durban (eThekweni), South
681 Africa as an example. *Neobiota*, 47, 1–21.
- 682 46. Perrings, C., Dehnen-Schmutz, K., Touza, J., Williamson, M. (2005). How to manage
683 biological invasions under globalization. *Trends in Ecology & Evolution*, 20, 212–215.
- 684 47. Petitpierre, B., Kueffer, C., Broennimann, O., Randin, C., Daehler, C., Guisan, A. (2012).
685 Climatic niche shifts are rare among terrestrial plant invaders. *Science*, 335, 1344–1348.
- 686 48. Pinheiro *et al.* (2019). *nlme: Linear and Nonlinear Mixed Effects Models v.3.1-141*. R
687 Foundation for Statistical Computing. Available at [https://cran.r-](https://cran.r-project.org/web/packages/nlme/index.html)
688 [project.org/web/packages/nlme/index.html](https://cran.r-project.org/web/packages/nlme/index.html). Accessed on 21 March 2020.
- 689 49. Pyšek, P., Richardson, D. M. (2007). Traits associated with invasiveness in alien plants:
690 where do we stand? In: *Biological Invasions* (ed Nentwig, W.). Springer-Verlag, Berlin &
691 Heidelberg, 97–125.
- 692 50. Pyšek, P. *et al.* (2010). Disentangling the role of environmental and human pressures on
693 biological invasions across Europe. *Proceedings of the National Academy of Sciences*
694 *USA*, 107, 12157–12162.

- 695 51. Pyšek, P. *et al.* (2015). Naturalization of central European plants in North America:
696 species traits, habitats, propagule pressure, residence time. *Ecology*, 96, 762–774.
- 697 52. Pyšek, P., *et al.* (2017). Naturalized alien flora of the world: species diversity, taxonomic
698 and phylogenetic patterns, geographic distribution and global hotspots of plant invasion.
699 *Preslia*, 89, 203–274.
- 700 53. Richardson, D. M., Pyšek, P. (2012). Naturalization of introduced plants: ecological
701 drivers of biogeographical patterns. *New Phytologist*, 196, 383–396.
- 702 54. Richardson, D. M. *et al.* (2000). Naturalization and invasion of alien plants: concepts and
703 definitions. *Diversity and Distributions*, 6, 93-107.
- 704 55. Rouget, M. *et al.* (2016). Invasion debt – quantifying future biological invasions. *Diversity*
705 *and Distributions*, 22, 445–456.
- 706 56. Seebens, H. *et al.* (2015). Global trade will accelerate plant invasions in emerging
707 economies under climate change. *Global Change Biology*, 21, 4128–4140.
- 708 57. Seebens, H. *et al.* (2017). No saturation in the accumulation of alien species worldwide.
709 *Nature Communications*, 8, 14435.
- 710 58. Stohlgren, T. J., Barnett, D. T., Jarnevich, C. S., Flather, C., Kartesz, J. (2008). The myth
711 of plant species saturation. *Ecology Letters*, 11, 313–322.
- 712 59. Strubbe, D., Broennimann, O., Chiron, F., Matthysen, E. (2013). Niche conservatism in
713 non-native birds in Europe: niche unfilling rather than niche expansion. *Global Ecology*
714 *and Biogeography*, 22, 962–970.
- 715 60. Strubbe, D., Beauchard, O., Matthysen, E. (2015). Niche conservatism among non-
716 native vertebrates in Europe and North America. *Ecography*, 38, 321–329.
- 717 61. Thuiller, W., Lafourcade, B., Engler, R., Araújo, M. B. (2009). BIOMOD - a platform for
718 ensemble forecasting of species distributions. *Ecography*, 32, 369–373.
- 719 62. Tingley, R., García-Díaz, P., Rocha Arantes, C. R., Cassey, P. (2018). Integrating
720 transport pressure data and species distribution models to estimate invasion risk for
721 alien stowaways. *Ecography*, 41, 635–646.

- 722 63. Westoby, M. (1998). A leaf-height-seed (LHS) plant ecology strategy scheme. *Plant and*
723 *Soil*, 199, 213–227.
- 724 64. Wiersema, J. H., León, B. (2013). *World Economic Plants : A Standard Reference*. CRC
725 Press, Boca Raton. Interactive database available online at [https://npgsweb.ars-](https://npgsweb.ars-grin.gov/gringlobal/taxon/taxonomysearcheco.aspx)
726 [grin.gov/gringlobal/taxon/taxonomysearcheco.aspx](https://npgsweb.ars-grin.gov/gringlobal/taxon/taxonomysearcheco.aspx). Accessed on 21 March 2020.
- 727 65. Xu, W.-B. *et al.* (2019). Human activities have opposing effects on distributions of
728 narrow-ranged and widespread plant species in China. *Proceedings of the National*
729 *Academy of Sciences USA*, 116, 26674–26681.
- 730 66. Zimmermann, N. E. *et al.* (2009). Climatic extremes improve predictions of spatial
731 patterns of tree species. *Proceedings of the National Academy of Sciences USA*, 106,
732 19723–19728.

Table 1 Summary of the seven socio-economic factors used to assess the drivers of naturalization success of the European vascular plants including the year and spatial resolution at which the original data was collected/calculated, the range of values for the regions used in analyses, and which transformation was applied to these values prior to analyses. Minimum and maximum values are those, after transformation, over the 526 non-European regions for which all socio-economic factors were available.

Variable	Year	Resolution	Unit	Min	Max	Transformation
<i>Introduction pathways</i>						
Airports	2019	-	100,000 km ²	0	11	log
Seaports	2014	-	100,000 km ²	0	15	log
Treaties	2016	-	#	6	30	-
<i>Human pressures</i>						
Population	2000	1 km	Inhabitant.km ⁻²	0	10	log
Cropland	< 2007	5 arcmin	%	0	95	-
Development	2015	5 arcmin	Dimensionless	.27	.93	-
<i>Sampling effort</i>						
Inventory	2015-16	110 km	%	0	249	-

Table 2 Regression coefficients of the relationship between distributional, economic and functional features of naturalized plants from Europe and their range filling (proportion of suitable regions where species are observed as naturalized) and range expansion (proportion of unsuitable regions where species are observed as naturalized). Continuous variables were standardized to allow comparison. Significance levels of the relationships fitted with linear mixed-effect models accounting for between-species phylogenetic distances are indicated by asterisks ($.05 < P\text{-value} \leq .01$: *, $.01 < P\text{-value} \leq .001$: **, $P\text{-value} < .001$: ***).

	Range filling	Range expansion	<i>n</i>
Prevalence in Europe	0.13	-0.40*	272
Potential range outside Europe	-0.25**	0.08	272
Ornamental	0.22***	0.14*	272
Other use	0.21***	0.22***	272
Growth form	0.02	0.01	193
Specific leaf area (SLA)	0.27*	-0.01	117
Height	0.05	0.01	175
Seed mass	0.02	-0.01	156

Figure captions

Figure 1 Workflow of this study used to determine which features of both plants and recipient regions explain possible discrepancies between observed and expected extents of naturalization.

Figure 2 Comparison of predicted and observed naturalized floras in 931 non-European regions based on various upscaling methods used to transform local naturalization predictions into regional naturalization predictions. The left column comprises our full set of European plants (either already naturalized or not yet naturalized outside of Europe) and the right column comprises only the subset of those plants already naturalized outside of Europe. Naturalization debt refers to the proportion of predicted naturalizations that are not yet observed, sensitivity to the proportion of correctly predicted successful naturalizations and negative predictive value to the proportion of correctly predicted unsuccessful naturalizations. A region was considered suitable if it contained at least one suitable pixel, 1% suitable cells, 5%, 20%, 50%, and 100%.

Figure 3 Relative influence of seven socio-economic factors in explaining differences between observed and expected European naturalized floras calculated by boosted regression trees. Naturalization debt refers to the proportion of predicted naturalizations that are not yet observed, sensitivity to the proportion of correctly predicted successful naturalizations and negative predictive value to the proportion of correctly predicted unsuccessful naturalizations. Socio-economic factors are sorted by mean relative influence over the three metrics. These results are for the upscaling method '1 %' i.e., a region was considered suitable for a species if it contained at least 1% suitable cells.

Figure 4 Joint partial dependence plots of interactions fitted by boosted regression trees between the most contributing regional socio-economic factors and differences between observed and expected European naturalized floras. Naturalization debt refers to the proportion of predicted naturalizations that are not yet observed, sensitivity to the proportion

of correctly predicted successful naturalizations and negative predictive value to the proportion of correctly predicted unsuccessful naturalizations. Black dots represent empirical data points. These results are for the upscaling method '1 %' i.e., a region was considered suitable for a species if it contained at least 1% suitable cells.

Figure 5 Global maps of the differences between observed and expected European naturalized floras. Regions with no GloNAF data are displayed in white. Naturalization debt refers to the proportion of predicted naturalizations that are not yet observed, sensitivity to the proportion of correctly predicted successful naturalizations (there are many NA values as many regions have no observed naturalizations by European endemic plants) and negative predictive value to the proportion of correctly predicted unsuccessful naturalizations. These results are for the upscaling method '1 %' i.e., a region was considered suitable for a species if it contained at least 1% suitable cells.

Figure 6 Difference in distributional, economic and functional features between European vascular plants already recorded as naturalized outside Europe ('Natur.') and those not recorded as naturalized outside Europe ('Non-natur.'). 1,472 species in total. 'Prevalence' refers to the number of $0.42^\circ \times 0.42^\circ$ cells in Europe occupied, 'Potential range' to the surface area of the species' expected range outside Europe, 'Ornamental' and 'Other use' indicate whether species are used as an ornamental plant or has other economic relevance (e.g., food plant, medicines, materials), respectively, 'Growth form' reflects the morphology of a plant, especially its physiological adaptation to the environment, 'SLA' for 'specific leaf area' is the ratio of leaf area to leaf dry mass, 'Height' and 'Seed mass' are the plant height and seed mass. Significance levels of the relationships fitted with linear mixed-effect models accounting for between-species phylogenetic distances are indicated by asterisks ($.05 < P\text{-value} \leq .01$: *, $.01 < P\text{-value} \leq .001$: **, $P\text{-value} < .001$: ***).