



Travel Tales of a Worldwide Weed: Genomic Signatures of *Plantago major* L. Reveal Distinct Genotypic Groups With Links to Colonial Trade Routes

OPEN ACCESS

Edited by:

Lisa Pokorný,
Botanical Institute of Barcelona
(CSIC), Spain

Reviewed by:

Helena Freitas,
University of Coimbra, Portugal
Graham McCulloch,
University of Otago, New Zealand

*Correspondence:

Natalie Iwanycki Ahlstrand
natalie.iwanycki@snm.ku.dk

†These authors share senior
authorship

Specialty section:

This article was submitted to
Plant Systematics and Evolution,
a section of the journal
Frontiers in Plant Science

Received: 17 December 2021

Accepted: 11 May 2022

Published: 09 June 2022

Citation:

Iwanycki Ahlstrand N,
Gopalakrishnan S, Vieira FG,
Bieker VC, Meudt HM, Dunbar-Co S,
Rothfels CJ, Martinez-Swatson KA,
Maldonado C, Hassemmer G,
Shipunov A, Bowers MD, Gardner E,
Xu M, Ghorbani A, Amano M,
Grace OM, Pringle JS, Bishop M,
Manzanilla V, Cotrim H, Blaney S,
Zubov D, Choi H-K, Yesil Y, Bennett B,
Vimolmangkang S, El-Seedi HR,
Staub PO, Li Z, Boldbaatar D,
Hislop M, Caddy LJ, Muasya AM,
Saslis-Lagoudakis CH, Gilbert MTP,
Zerega NJC and Ronsted N (2022)
Travel Tales of a Worldwide Weed:
Genomic Signatures of *Plantago*
major L. Reveal Distinct Genotypic
Groups With Links to Colonial Trade
Routes. *Front. Plant Sci.* 13:838166.
doi: 10.3389/fpls.2022.838166

Natalie Iwanycki Ahlstrand^{1*}, Shyam Gopalakrishnan^{2,3}, Filipe G. Vieira⁴,
Vanessa C. Bieker⁵, Heidi M. Meudt⁶, Stephanie Dunbar-Co⁷, Carl J. Rothfels⁸,
Karen A. Martinez-Swatson⁹, Carla Maldonado¹⁰, Gustavo Hassemmer¹¹,
Alexey Shipunov¹², M. Deane Bowers¹³, Elliot Gardner^{14,15}, Maonian Xu¹⁶,
Abdolbaset Ghorbani¹⁷, Makoto Amano¹⁸, Olwen M. Grace¹⁹, James S. Pringle²⁰,
Madonna Bishop²¹, Vincent Manzanilla²², Helena Cotrim²³, Sean Blaney²⁴,
Dimitri Zubov²⁵, Hong-Keun Choi²⁶, Yeter Yesil²⁷, Bruce Bennett²⁸,
Sornkanok Vimolmangkang²⁹, Hesham R. El-Seedi³⁰, Peter O. Staub³¹, Zhu Li³²,
Delgerbat Boldbaatar³³, Michael Hislop³⁴, Laura J. Caddy³⁵, A. Muthama Muasya³⁶,
C. Haris Saslis-Lagoudakis³⁷, M. Thomas P. Gilbert², Nyree J. C. Zerega^{14,15†} and
Nina Ronsted^{1,38†}

¹ Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark, ² Globe Institute, University of Copenhagen, Copenhagen, Denmark, ³ Bioinformatics, Department of Health Technology, Technical University of Denmark, Lyngby, Denmark, ⁴ Center for Genomic Medicine, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark, ⁵ Department of Natural History, NTNU University Museum, Norwegian University of Science and Technology, Trondheim, Norway, ⁶ Museum of New Zealand Te Papa Tongarewa, Wellington, New Zealand, ⁷ The Nature Conservancy, Kaunakakai, HI, United States, ⁸ University Herbarium and Department of Integrative Biology, University of California, Berkeley, Berkeley, CA, United States, ⁹ Havforskningssinstituttet, His, Norway, ¹⁰ Herbario Nacional de Bolivia, Universidad Mayor de San Andres, La Paz, Bolivia, ¹¹ Federal University of Mato Grosso do Sul, Três Lagoas, Brazil, ¹² Department of Biology, Minot University, Minot, ND, United States, ¹³ Ecology and Evolutionary Biology, University of Colorado at Boulder, Boulder, CO, United States, ¹⁴ Negaunee Institute for Plant Conservation Science and Action, Chicago Botanic Garden, Chicago, IL, United States, ¹⁵ Plant Biology and Conservation, Northwestern University, Evanston, IL, United States, ¹⁶ Faculty of Pharmaceutical Sciences, University of Iceland, Reykjavik, Iceland, ¹⁷ Systematic Biology, Department of Organismal Biology, Uppsala University, Uppsala, Sweden, ¹⁸ Natural History Museum and Institute, Chiba, Japan, ¹⁹ Comparative Plant and Fungal Biology, Royal Botanic Gardens, Kew, Richmond, United Kingdom, ²⁰ Royal Botanic Gardens, Hamilton, ON, Canada, ²¹ Memorial University Botanical Garden, St. John's, NL, Canada, ²² Baseclear B.V., Leiden, Netherlands, ²³ Centre for Ecology, Evolution and Environmental Changes, University of Lisbon, Lisbon, Portugal, ²⁴ Atlantic Canada Conservation Data Centre, Sackville, NB, Canada, ²⁵ Gryshko's National Botanic Garden, Kiev, Ukraine, ²⁶ Department of Life Sciences, Ajou University, Suwon, South Korea, ²⁷ Department of Pharmaceutical Botany, Istanbul University, Istanbul, Turkey, ²⁸ Yukon Conservation Data Centre, Yukon Territory, YT, Canada, ²⁹ Department of Pharmacognosy and Pharmaceutical Botany, Chulalongkorn University, Bangkok, Thailand, ³⁰ Pharmacognosy Group, Department of Medicinal Chemistry, Uppsala University, Uppsala, Sweden, ³¹ Department of Biomedical Science, University of Cagliari, Cagliari, Italy, ³² Chinese Academy of Sciences, Beijing, China, ³³ Department of Liver Center, National University of Mongolia, Ulaanbaatar, Mongolia, ³⁴ Western Australia Herbarium, Perth, WA, Australia, ³⁵ Botanical Garden, The University of British Columbia, Vancouver, BC, Canada, ³⁶ Department of Biological Sciences, University of Cape Town, Cape Town, South Africa, ³⁷ Forage Crete, Heraklion, Greece, ³⁸ National Tropical Botanic Garden, Kauai, HI, United States

Retracing pathways of historical species introductions is fundamental to understanding the factors involved in the successful colonization and spread, centuries after a species' establishment in an introduced range. Numerous plants have been introduced to regions outside their native ranges both intentionally and accidentally by European voyagers

and early colonists making transoceanic journeys; however, records are scarce to document this. We use genotyping-by-sequencing and genotype-likelihood methods on the selfing, global weed, *Plantago major*, collected from 50 populations worldwide to investigate how patterns of genomic diversity are distributed among populations of this global weed. Although genomic differentiation among populations is found to be low, we identify six unique genotype groups showing very little sign of admixture and low degree of outcrossing among them. We show that genotype groups are latitudinally restricted, and that more than one successful genotype colonized and spread into the introduced ranges. With the exception of New Zealand, only one genotype group is present in the Southern Hemisphere. Three of the most prevalent genotypes present in the native Eurasian range gave rise to introduced populations in the Americas, Africa, Australia, and New Zealand, which could lend support to the hypothesis that *P. major* was unknowingly dispersed by early European colonists. Dispersal of multiple successful genotypes is a likely reason for success. Genomic signatures and phylogeographic methods can provide new perspectives on the drivers behind the historic introductions and the successful colonization of introduced species, contributing to our understanding of the role of genomic variation for successful establishment of introduced taxa.

Keywords: introduced species, weed phylogeography, human mediated dispersal, historical introduction, introduction pathways

INTRODUCTION

Retracing pathways of species introductions to new lands is a fundamental part of understanding what ecological and evolutionary factors are involved in the successful establishment and spread of species into new ranges (Wilson et al., 2009; Estoup and Guillemaud, 2010; Seebens et al., 2015; van Kleunen et al., 2015; Chapman et al., 2017). The global movement and spread of introduced species have received much attention in recent decades, with particular focus being placed on studying contemporary species invasions that impact socioeconomic wellbeing and threaten indigenous biodiversity (Mack et al., 2000; Chapman et al., 2016; Puckett et al., 2016; Guo et al., 2017). However, humans have long been mediating the transoceanic dispersal and spread of species both intentionally and accidentally. Retracing pathways of historical invasions is equally important in advancing our understanding of species' adaptations to new lands and reasons for their success, independent of the ecological or economic effects of such invasions (Mack, 2003; Preston C. D. et al., 2004; Wilson et al., 2009; van Kleunen et al., 2015; Chapman et al., 2017).

The human-mediated dispersal of plants from Europe to other continents became particularly prevalent around the year 1500, a time coinciding with European exploration, colonialism, and the start of wide-scale changes in human demography, land use, trade and industrial development (Godwin, 1944; Pyšek, 1998; Preston C. D. et al., 2004; Mancall, 2006; Hulme, 2009; Banks et al., 2015; Turbelin et al., 2017). In the absence of records describing the native flora before the arrival of Europeans, written accounts from early voyagers and colonists dating back to the sixteenth and seventeenth centuries provide evidence of the introduction of European species to continents such as North America;

however, such records are scarce (Fernald, 1900). This lack of information hinders our ability to retrace the origins of historical introductions (Wilson et al., 2009; Estoup and Guillemaud, 2010; Puckett et al., 2020).

Molecular methods have helped to disentangle the origins of a number of introduced plants, by inferring source populations and retracing putative pathways of invasion, and comparing genetic variation between native and introduced ranges (Ortiz et al., 2008; Estoup and Guillemaud, 2010; Lambertini et al., 2012; Canavan et al., 2017; Zhu et al., 2017; Smith A. L. et al., 2020). However, multiple colonization events, hybridization, and genetic changes within populations over time attributed to admixture or loss of genetic diversity due to founder effects can pose challenges in identifying source populations and inferring dispersal routes for historical introductions (Dlugosch and Parker, 2007; Gaskin et al., 2013; Dormontt et al., 2014; Martin et al., 2014). Advances in genome-wide sequencing technologies now make it possible to genotype individuals using thousands of markers and therefore provide a promising solution to identifying the source of historical introduction events even if genetic variation is very low or genetic changes between populations are pronounced (Estoup and Guillemaud, 2010; Elshire et al., 2011; Lu et al., 2013; Narum et al., 2013; Vigueira et al., 2013; Cornille et al., 2016; Puckett et al., 2016).

Here we use the worldwide weed *Plantago major* L. (Plantaginaceae), also known as common, greater or broadleaf plantain, as a case for understanding human-mediated plant dispersals and potential reasons for successful establishment in new ranges. Native to Eurasia, the species grows in a wide range of disturbed habitats, and although it is widespread in both the native and introduced ranges, is only rarely considered a problematic invasive species (Hawthorn, 1974; Holm et al., 1979;

van Dijk, 1984). The species is considered commensal with humans and, based on its medicinal properties, has had a long history of human use in both native and introduced ranges (Bennett and Prance, 2000; Samuelsen, 2000; Stepp and Moermann, 2000; Palmer, 2004). The plant was thought to have been introduced to North America by European voyagers or possibly even earlier with Norse (Samuelsen, 2000), and to other parts of introduced ranges during colonial times, including Australasia, South America, and southern Africa. Historical written records and herbarium specimens collected in the introduced ranges before the nineteenth century are limited and offer limited insights into elucidating the plant's arrival and early spread outside of its native distribution. *Plantago major* is known today from every continent except Antarctica (Figure 1) and, due to its human-mediated dispersal, is arguably one of the world's most prevalent weeds (Rousseau, 1966; Holm et al., 1979; van Dijk and Wolff, 1992; Rahn, 1996; Hodgkinson and Thompson, 1997). By the seventeenth century, the plant had already been noted to be well-established in New England (northeastern United States), where indigenous peoples referred to the plant as "Englishman's foot" because it followed colonists wherever they went (Josslyn, 1672). Although the species is considered introduced throughout North America, there are reports suggesting that it, or at least a variety of the species, is native to northern North America, north of 50° latitude, based on the species' presence in isolated habitats that were considered undisturbed by early Europeans, though this remains unconfirmed (Hawthorn, 1974).

Plantago major has been extensively studied in its native range and a wealth of ecological, physiological and genetic information is available for the species (Mølgaard, 1976; Kuiper and Bos, 1992; Wolff and Morgan-Richards, 1998;

Morgan-Richards and Wolff, 2002). It possesses many of the traits that are common amongst the most successful introduced plants (Pyšek et al., 2009, 2015; Hejda et al., 2015), including high phenotypic plasticity, large ecological amplitudes, a high tolerance to human disturbance, rapid growth rates, and the production of propagules with specialized adaptations for long-distance dispersal such as mucilaginous seeds and wind-dispersed pollen (Young and Evans, 1973; Hawthorn, 1974; van Dijk, 1984; Rahn, 1996; Samuelsen, 2000; Kreitschitz et al., 2016; Iwanycki Ahlstrand et al., 2018, 2019). Furthermore, the species is predominantly self-fertilizing with an outcrossing rate of 10.3%, meaning that a single propagule can putatively establish sexually reproducing populations in new ranges (Baker, 1965, 1974; Wolff, 1991; Morgan-Richards and Wolff, 2002). As with other highly selfing species, thousands of individuals of *P. major* can persist in a small area but low genetic diversity within such populations is found (van Dijk and van Delden, 1981; Wolff et al., 1994; Wolff and Morgan-Richards, 1998). The low genetic diversity and low heterozygosity associated with highly selfing species suggests that every population of *P. major* is presumed to be a highly inbred lineage (Wolff, 1991; Wolff and Morgan-Richards, 1998).

Despite the species being extensively studied in the native parts of its range in the United Kingdom, Denmark, and The Netherlands, the genetic variation between and among populations elsewhere in its native and introduced ranges have seldom been investigated, and dispersal pathways to its introduced ranges have never been inferred using genomic methods. Although *P. major* is highly self-fertilizing, the genus *Plantago* includes species exhibiting a range of different breeding systems, and thus has been an excellent model for studying the effects of a plant's breeding system on ecology and genetic variation (van Dijk et al., 1988; Wolff, 1991).

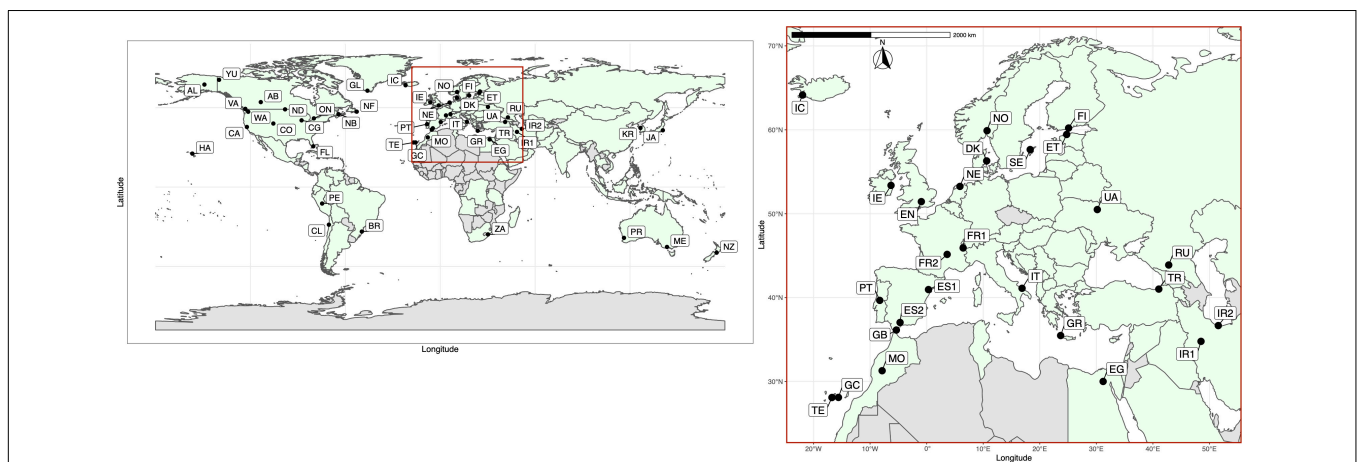


FIGURE 1 | Locations of the 50 global sampling sites for *Plantago major*. See Table 1 for population details. Population codes (native range populations in bold): Alaska = AL, Alberta = AB, Brazil = BR, California = CA, Chicago = CG, Chile = CL, Colorado = CO, **Egypt = EG, England = EN, Estonia = ET, Finland = FI**, Florida = FL, **France1 = FR1, France2 = FR2, Gibraltar = GB, Gran Canaria = GC, Greece = GR**, Greenland = GL, Hawaii = HA, Iceland = IC, **Iran1 = IR1, Iran2 = IR2, Ireland = IE, Italy = IT, Japan = JA**, Melbourne = ME, **Morocco = MO, The Netherlands = NE**, New Brunswick = NB, Newfoundland = NF, New Zealand = NZ, North Dakota = ND, **Norway = NO, Denmark = DK**, Ontario = ON, Perth = PR, Peru = PE, **Portugal = PT, Russia = RU**, South Africa = ZA, **South Korea = KR, Spain1 = ES1, Spain2 = ES2, Sweden = SE, Tenerife = TE, Turkey = TR, Ukraine = UA**, Vancouver = VA, Washington = WA, Yukon = YU. Countries colored in light green denote the global distribution of *Plantago major* based on 633, 356 georeferenced occurrences registered in GBIF.org (GBIF, 2021, accessed 30 April, 2021) and locality data accessed from CAB International (CABI, 2021, accessed 30 April, 2021).

The genotypic variation of *Plantago lanceolata* L., a strictly outcrossing but equally cosmopolitan weedy congener of *P. major*, was recently investigated by sampling populations across the native and introduced ranges; this global weed was found to have higher genetic variation in the non-native ranges compared to the native ranges due to multiple introductions largely from the Mediterranean parts of its native range, and subsequent admixture in introduced populations (Smith A. L. et al., 2020). Unlike the highly selfing *P. major*, *P. lanceolata* is self-incompatible, and therefore one would expect a higher degree of admixture in both native and introduced ranges (van Dijk et al., 1988).

Globally distributed selfing species such as *P. major* with very low chance of admixture provide unique cases to improve our understanding of the relative importance, or lack thereof, of genetic variability in the successful establishment of introduced species since repeated introductions in the introduced range would not necessarily contribute an increase in genetic variability, thereby showing that the success of establishing in new ranges with widely differing environmental conditions is not dependent on genetic variation (Smith A. L. et al., 2020). Highly selfing species with low admixing between populations also offer the potential for a higher level of confidence in retracing introduction pathways using genomics methods (Estoup and Guillemaud, 2010). We sampled and sequenced *P. major* from 50 populations across its global range and taking a population-genomic approach we analyze thousands of genome-wide single-nucleotide polymorphisms (SNPs) generated by genotyping-by-sequencing (GBS) to investigate: (i) the genomic variation among populations of *P. major* subsp. *major* at a global scale (ii) the ancestry of native vs. introduced populations, and (iii) what the genomic patterns we observe indicate about the pathways of introduction from native Eurasia to introduced ranges around the world. We discuss what might be the role of colonial history in explaining current genetic variation among populations worldwide. This knowledge can help us understand the importance of landscape-scale genomic variation in the dispersal and establishment of new potential plant introductions, and provides a starting point for future research in understanding how introduced plants cope under changing environmental conditions when genomic variation is low due to high levels of inbreeding.

MATERIALS AND METHODS

Global Sample Collection

Due in part to its extensive distribution and phenotypic diversity, *Plantago major* has a complex taxonomic history with over 50 taxonomic divisions and synonyms having been proposed and a number of ecotypes being recognized (Pilger, 1937; Mølgaard, 1976; Peruzzi and Passalacqua, 2003; POWO, 2019). At least three subspecies are recognized in its native range, two of which, *P. major* subsp. *major* and *P. major* subsp. *intermedia* (Gilib.) Lange, have been widely studied in Europe and even recognized as separate species based on genetic structure and differentiation (*P. major* and *P. intermedia* Gilib.; Wolff and

Morgan-Richards, 1998; Morgan-Richards and Wolff, 2002; POWO, 2019). The full extent of the geographic ranges occupied by each subspecies, and overlap between them, is poorly known, particularly outside of northern Europe. Our sampling FOCUSED on *P. major* subsp. *major* which is considered more widely distributed and more tolerant of a wider range of environmental variation and anthropogenic disturbance, as well as having higher outcrossing rates and a longer lifespan (Mølgaard, 1976; Morgan-Richards and Wolff, 2002). *Plantago major* subsp. *major* can be discriminated from *P. major* subsp. *intermedia* based on the number of seeds per capsule and seed size (Mølgaard, 1976; van Dijk, 1984; Wolff, 1991). Only specimens fitting these characters were selected for our study.

Fifty populations of naturally occurring *Plantago major* subsp. *major* plants sampled from 34 countries worldwide between 2014 and 2015 are included in our study, including 27 populations from across the native range (Europe, Asia, North Africa) and 22 populations from the introduced range (North and South America, Iceland, Greenland, Hawai'i, Australia, New Zealand, and southern Africa) (Figure 1 and Table 1). Population sampling was targeted in as many locations as possible across the known range of *P. major*. Populations sampled in several locations were determined to be other species in the genus *Plantago* and were therefore excluded from the current study [i.e., plants sampled in Malaysia, Thailand, Mongolia, China (Beijing, Wuhan), United States (North Carolina), and Bolivia]. Six to ten individuals were sampled from each population (only three for Chicago), amounting to 385 individuals in total, and dried and stored in silica gel. A herbarium voucher was collected from each population to confirm species identity and deposited in Herbarium C at the Natural History Museum of Denmark, University of Copenhagen, in Copenhagen, Denmark. Herbarium vouchers were not obtainable for Peru; however, digital photos were taken in lieu of vouchers.

DNA Extraction and Sequencing

Dried leaf tissue was pulverized with ceramic beads in a Qiagen TissueLyser (Qiagen, Germany) and DNA extractions were prepared using a modified Qiagen DNEasy® Mini Plant Kit (Qiagen, Germany) or a CTAB protocol (conducted by ADNid, Montpellier, France). DNEasy Mini Plant Kit extractions were modified by adding 50 µL of proteinase K to the cell lysis solution after 10 min incubation at 65°C, followed by 2 h of incubation at 45°C. Double extractions were made for each sample, then pooled, to ensure the quantity and quality of genomic DNA was sufficient for genotyping-by-sequencing (GBS).

In total, DNA extracts from 385 individuals were sequenced (Table 1). GBS library preparation was performed at the Genomic Diversity Facility at Cornell University, following Elshire et al. (2011). DNA from each sample was digested using the restriction enzyme *Pst*I (CTGCA[^]G), and both a sample-specific barcoded adapter and a common adapter were ligated to the sticky ends of fragments. Samples were pooled and fragment libraries cleaned using a QIA-quick PCR purification kit (Qiagen, United States). Libraries were amplified using an 18-cycle PCR with long primers complementary to the barcoded and common adapters, purified again using QIAquick, and quantified using an intercalating dye

TABLE 1 | *Plantago major* populations sampled for genotyping, including number of individuals included per population and genotype group each population was found to cluster with based on NGSADMIX analyses ($K = 7$).

| No. | Population name | Country code | Site name | Collector | Collection number | Collection date | Lat | Long | Number of individuals | Genotype group |
|-----|-----------------|--------------|-------------------------------------|----------------------|-------------------|-----------------|------------|------------|-----------------------|----------------|
| 1. | Alaska | AL | Fairbanks, Alaska, United States | A Shipunov | | 10/08/2014 | 64.5742 | -149.1195 | 8 | Group II |
| 2. | Alberta | AB | Edmonton, Alberta, Canada | CJ Rothfels | 4,505 | 07/26/2015 | 53.52398 | -113.47652 | 8 | Group II |
| 3. | Brazil | BR | Urubici, Brazil | G Hassemer | | 06/01/2015 | -28.014664 | -49.593561 | 8 | Group V |
| 4. | California | CA | Berkeley, California, United States | CJ Rothfels | 4,606 | 11/09/2015 | 37.87245 | -122.26438 | 8 | Group IV |
| 5. | Chicago | CG | Chicago, United States | NJC Zerega | | 17/10/2015 | 42.028325 | -87.687983 | 3 | Group II |
| 6. | Chile | CL | Antofagasta, Chile | G Hassemer | | 07/01/2016 | -23.681014 | -70.4126 | 6 | Group VI |
| 7. | Colorado | CO | Boulder, Colorado, United States | D Bowers | | 19/09/2015 | 40.0316667 | -105.53722 | 8 | Group II |
| 8. | Denmark | DK | Øjesø, Jutland, Denmark | N Iwanycki Ahlstrand | | 25/07/2015 | 56.2930556 | 10.611111 | 10 | Group I |
| 9. | Egypt | EG | Cairo, Egypt | H El-seedi | | 29/08/2015 | 30.0166667 | 31.2 | 8 | Group VI |
| 10. | England | EN | Reading, England | N Iwanycki Ahlstrand | | 06/08/2015 | 51.4418583 | -0.9410444 | 8 | Group II |
| 11. | Estonia | ET | Talinn, Estonia | N Iwanycki Ahlstrand | | 29/06/2015 | 59.4422222 | 24.7505556 | 8 | Group II |
| 12. | Finland | FI | Helsinki, Finland | N Iwanycki Ahlstrand | | 29/06/2015 | 60.2108333 | 25.0975 | 8 | Group II |
| 13. | Florida | FL | Florida, United States | E Gardiner | | 17/06/2015 | 25.682456 | -80.278343 | 8 | Group VI |
| 14. | France1 | FR1 | Haute-Savoie, France | K Martinez | | 20/07/2014 | 45.9222222 | 6.4480556 | 7 | Group I |
| 15. | France2 | FR2 | Paulhaguet, France | TMP Gilbert | | 30/07/2014 | 45.1558333 | 3.6247222 | 8 | Group II |
| 16. | Gibraltar | GB | Gibraltar, United Kingdom | N Iwanycki Ahlstrand | | 30/04/2015 | 36.1423806 | -5.3592028 | 6 | Group V |
| 17. | GranCanaria | GC | Gran Canaria, Canary Islands | N Iwanycki Ahlstrand | | 07/05/2015 | 28.115 | -15.593889 | 8 | Group V |
| 18. | Greece | GR | Crete, Greece | K Martinez | PM290415_1 | 29/04/2015 | 35.4869444 | 23.688055 | 8 | Group V |
| 19. | Greenland | GL | Qaqortoq, Greenland | K Høegh | | 17/08/2014 | 60.7188889 | -46.0375 | 8 | Group II |
| 20. | Hawai'i | HA | Molokai, Hawai'i | S Dunbar-Co | | 09/07/2015 | 21.0833333 | -156.76667 | 8 | Group VI |
| 21. | Iceland | IC | Reykjavik, Iceland | M Xu | | 06/09/2014 | 64.1431148 | -21.935278 | 9 | Group I |
| 22. | Iran1 | IR1 | Hamedan, Iran | A Ghorbani | | 09/30/2015 | 34.7933611 | 48.5037778 | 7 | Group IV |
| 23. | Iran2 | IR2 | West Azarbaijan, Iran | A Ghorbani | | 10/15/2015 | 36.6742222 | 51.5581389 | 4 | Group II |
| 24. | Ireland | IE | Dublin, Ireland | N Iwanycki Ahlstrand | | 28/08/2014 | 53.3772222 | -6.3 | 9 | Group I |
| 25. | Italy | IT | Bari, Italy | K Martinez | | 21/06/2014 | 41.1166667 | 16.8666667 | 7 | Group II |
| 26. | Japan | JA | Chiba, Japan | M Amano | | 08/10/2015 | 35.7737 | 140.7528 | 5 | Group III |
| 27. | Melbourne | ME | Melbourne, Australia | T Schultz | | 04/04/2015 | -37.864222 | 144.06006 | 8 | Group VI |
| 28. | Morocco | MO | Amegdale, Morocco | K Martinez | | 22/05/2014 | 31.2955556 | -7.8719444 | 7 | Group VI |
| 29. | Netherlands | NE | Oentsjerk, Netherlands | OM Grace | | 03/09/2015 | 53.253704 | 5.886092 | 8 | Group II |
| 30. | NewBrunswick | NB | St. John, New Brunswick, Canada | S Blaney | | 21/10/2014 | 45.907548 | -64.368625 | 8 | Group I |

(Continued)

TABLE 1 | (Continued)

| No. | Population name | Country code | Site name | Collector | Collection number | Collection date | Lat | Long | Number of individuals | Genotype group |
|-----|-----------------|--------------|--|----------------------|-------------------|-----------------|------------|------------|-----------------------|----------------|
| 31. | NewZealand | NZ | Wellington, New Zealand | HM Meudt | | 30/04/2015 | -41.29774 | 174.76721 | 8 | Group I |
| 32. | Newfoundland | NF | St John's, Newfoundland, Canada | M Bishop | | 29/07/2015 | 47.495 | -52.794722 | 8 | Group II |
| 33. | NorthDakota | ND | North Dakota, United States | A Shipunov | | 24/08/2014 | 48.9629 | -98.1247 | 8 | Group I |
| 34. | Norway | NO | Oslo, Norway | V Manzanillo | | 20/10/2014 | 59.894956 | 10.675849 | 8 | Group II |
| 35. | Ontario | ON | Hamilton, Ontario, Canada | JS Pringle | | 19/08/2014 | 43.2691667 | -79.904167 | 8 | Group II |
| 36. | Perth | PR | Perth, Australia | M Hislop | | 21/11/2015 | -32.034167 | 115.914444 | 8 | Group VI |
| 37. | Peru | PE | Oxapampa, Peru | G Hassemer | | 13/06/2016 | -10.573199 | -75.403121 | 8 | Group VI |
| 38. | Portugal | PT | Ribeira, Portugal | H Cotrim | | 16/08/2015 | 39.6753583 | -8.2887583 | 7 | Group VI |
| 39. | Russia | RU | Kislovodsk, North Caucasus, Russian Federation | D Zubov | | 26/08/2014 | 43.8919861 | 42.7937833 | 9 | Group I |
| 40. | SouthAfrica | ZA | Byrne, South Africa | CJ Rothfels | 4693 | 20/10/2015 | -29.796712 | 30.182844 | 8 | Group VI |
| 41. | SouthKorea | KR | Gyeonggi-do, South Korea | H-K Choi | | 07/08/2015 | 37.2316111 | 126.607389 | 8 | Group III |
| 42. | Spain1 | ES1 | Horta de Sant Joan, Spain | K Martinez | | 03/05/2015 | 40.963195 | 0.322406 | 9 | Group V |
| 43. | Spain2 | ES2 | Bobadilla, Spain | N Iwanycki Ahlstrand | | 05/05/2015 | 37.0447222 | -4.703611 | 8 | Group V |
| 44. | Sweden | SE | Visby, Sweden | N Iwanycki Ahlstrand | | 16/06/2015 | 57.6438889 | 18.3013889 | 8 | Group II |
| 45. | Tenerife | TE | Tenerife, Canary Islands | N Iwanycki Ahlstrand | | 09/05/2015 | 28.1055556 | -16.7425 | 6 | Group VI |
| 46. | Turkey | TR | Güroluk, Turkey | Y Yesil | | 20/08/2015 | 41.0208889 | 41.0410278 | 7 | Group II |
| 47. | Ukraine | UA | Kiev, Ukraine | D Zubov | | 02/08/2014 | 50.4973583 | 30.1693833 | 9 | Group I |
| 48. | Vancouver | VA | Vancouver, British Columbia, Canada | CJ Rothfels | 4543 | 25/07/2014 | 49.263 | -123.2503 | 8 | Group II |
| 49. | Washington | WA | Washington, United States | CJ Rothfels | 4533 | 16/07/2014 | 47.44539 | -121.42326 | 8 | Group I |
| 50. | Yukon | YU | Old Crow, Yukon, Canada | B Bennett | | 09/07/2014 | 67.56652 | -139.84999 | 9 | Group III |

(PicoGreen[®]; Invitrogen, Carlsbad, CA, United States) following Elshire et al. (2011). Samples were run on seven plates, and one blank was included per plate. Plates were run on lanes of a 100 single-end Illumina HiSeq 2000, at the Cornell Core Laboratories Center (Cornell University, New York, United States).

Assembly of and Mapping to the Pseudo-Reference Genome

STACKS, a software pipeline designed for restriction-enzyme based data for organisms without a reference genome, was used to generate a catalog or pseudo-reference genome. The sequencing reads for 392 samples (385 samples plus a blank from each of the seven flow batches of samples) were demultiplexed

using the `process_radtags` function in STACKS v.1.45 (Catchen et al., 2013). The demultiplexing process was run in an iterative manner, with the discarded reads at each step being fed as input to the next `process_radtags` step with a smaller barcode size, to retain the maximum number of reads per sample. During the demultiplexing process, we rescued barcodes (using the `-r` option), and retained reads that did not match any barcode of the given length in a separate file. The parameter `-adapter_1` was used to identify common adapter sequence, the maximum adapter mismatch was set to 2, while we required a perfect match on the barcodes.

We selected 20 samples with the maximum number of reads, while maximizing the number of sampling locations spanned, to build a reference catalog using CSTACKS in STACKS v.1.45 (see

Table 1. These samples were processed using USTACKS, with the developing algorithm enabled (-d option), while retaining unused reads (-r option) and requiring a minimum of three reads to build a stack (-m 3). The resulting output was used to run CSTACKS with four mismatches required between different loci (-n 4) when building the catalog. We also allowed gapped alignment between the loci, with a maximum gap length of four (-gapped, -max_gaps 4). This results in a minimum Hamming distance of 4 between any pair of loci in the reference catalog. From the loci identified in the reference catalog, we created a pseudo-reference genome by collapsing these loci into 10 chromosomes, with a string of 150 Ns inserted between consecutive loci.

Raw reads were demultiplexed using the demultiplexer function in GBSX v.1.3 (Herten K. et al., 2015), allowing for one mismatch in the enzyme, and one mismatch in the barcode (-me 1 and -mb 1). Demultiplexing with GBSX retained a larger fraction of reads than STACKS, which in turn allowed us to obtain a larger set of variants for downstream analyses. Therefore, the demultiplexed reads obtained from GBSX were used to map reads to the catalog assembled in CSTACKS. The mapping was performed using the PALEOMIX pipeline, which was run with default parameters (Schubert et al., 2014). As part of the PALEOMIX pipeline, the reads were first filtered for adapter sequences using ADAPTERREMOVAL2 v. 2.2.0 (Schubert et al., 2016), and these trimmed sequences were mapped against the pseudo-reference genome using the backtrack algorithm in BWA v. 0.7.15 (Li and Durbin, 2009). The blanks from each of the seven plates also were included to ensure that they did not map to the pseudo-reference genome. Blanks were then excluded from our downstream analyses.

Single-Nucleotide Polymorphism-Calling and Genotype Likelihoods

The alignments generated by BWA were used to identify SNPs in our samples. Since GBS data inherently has very high variance in terms of coverage of loci and depth of coverage among and within loci, there is high uncertainty associated with the calling of genotypes at variant sites. In order to carry this uncertainty to downstream analyses, genotype likelihood-based methods were used, which allow us to account for this uncertainty in our analyses (Nielsen et al., 2013). The SNP locations and genotype likelihoods for the samples at these locations were computed using ANGSD v. 0.921 (Korneliussen et al., 2014). A total of 385 samples were used to identify variant positions (we excluded blanks in all further analyses). As a pre-processing step, we assessed the distribution of the depth of coverage of a randomly chosen subset of 100 samples to assess the cut-off for the maximum number of reads that can cover a single position. We discarded reads that mapped to multiple positions and low-quality bases and reads (quality score 20). Using the distribution of depths obtained from ANGSD, we chose a maximum average depth cut-off of 70 coverage per sample. Variant discovery and computation of genotype likelihoods was performed using ANGSD with the same parameters as above using the following additional parameters: a depth cut-off of 75 per sample

(-maxDepth NumberOfSamples*75), a minimum of 50% of the samples must have at least one read covering the site (-minInd NumberOfSamples*0.5), a minimum SNP p -value of 10^{-6} (-SNP_pval 1e-6), and a penalty for poorly aligning reads that have a lot of mismatches with the reference locus (-C 50). 7594 SNPs were retained for analyses. The genotype likelihoods were calculated using the model described in SAMTOOLS (-GL 1; v. 1.4; Li, 2011).

Genomic Covariance, Admixture Analyses, and Inference of Population Splits and Migration Events

To remove SNPs in linkage disequilibrium (LD), PLINK v. 1.90b4.4 (Chang et al., 2015) was run with a window size of 50, step size of 5 and threshold of 0.5, removing 1,545 of 7,593 variants. The LD pruned dataset was used to generate a covariance matrix with PCAANGSD (Meisner and Albrechtsen, 2018) with a minimum minor allele frequency of 0.1. This resulted in a total of 2,099 SNPs used for principal component analyses (PCA). The covariance matrix was used to perform a PCA analysis of the first six coordinates in R v. 4.1.1 using the *prcomp* function.

To identify population structure and identify patterns of admixture among our samples an individual-based assignment test was performed using NGSADMIX (Skotte et al., 2013; Fumagalli et al., 2014) on the LD pruned data set and a minimum minor allele frequency of 0.5. NGSADMIX is a maximum likelihood method that uses the genotype likelihood data obtained from ANGSD. Our analyses were run with the number of ancestral populations, K , set from two to 12 ($K = 2$ to $K = 12$). For $K = 2$ to $K = 8$, analyses were replicated 200 times, and for $K = 9$ to $K = 12$ the analyses were replicated 500 times to ensure convergence to the global maximum. The replicate with the highest log-likelihood among replicates was chosen for each K . A test statistic for determining the optimal K -value is not an available option in NGSADMIX. Therefore, results for each value of K were reviewed and compared with results from the PCA plots to select a K -value that had the most biological relevance based on the genomics of the species being studied (Hansen et al., 2018; Peñnerová et al., 2021).

Heterozygosity for each sample was estimated using ANGSD version 0.917 (Korneliussen et al., 2014). First, site allele frequency likelihoods (SAF) was estimated with a minimum base quality of 20, minimum mapping quality of 20 and a maximum sequencing depth of 60 using the samtools model (-gl 1). Base quality around indels was adjusted [-baq 1, (Li, 2011)] and not, failure, duplicate reads and those with multiple best hits were removed (-remove_bads 1 and -uniqueOnly 1). Per sample SAF were estimated using the realSFS tool within ANGSD. SAF was used to calculate heterozygosity, significant differences in heterozygosity between native and introduced populations were tested using a Mann-Whitney U -test, and ANOVA was performed to test for differences in heterozygosity between genotype groups as well as populations.

TREEMIX v. 1.13 (Pickrell and Pritchard, 2012) was used to construct a maximum likelihood (ML) tree and to visualize how well the relationships can be represented by a bifurcating tree using population allele frequencies. In addition to admixture, TREEMIX also provides some information about the directionality of gene flow by allowing for the modeling of migration events (Patterson et al., 2012). A script in PYTHON was written to convert Beagle files generated in ANGSD to TREEMIX input files (see **Supplementary Material**). Before running migration events, it is important to set the position of the root (Pickrell and Pritchard, 2012). In the absence of an outgroup, several different TREEMIX analyses were run using populations from Spain (1 and 2), France (1 and 2), Turkey, Iran1, and Japan as the root (-root). Weeds that are successful of achieving global distributions are the most difficult to place origins (Baker, 1974). The populations we selected to test as a root were chosen based on the premise that many European weeds have their origins in the Mediterranean basin and the Irano-Turanian region, or have origins in temperate regions of Europe (Baker, 1974; Preston C. D. et al., 2004; Jabbour and Renner, 2019), and based on the findings that closest extant relatives to *P. major* have native distributions in temperate Eurasia and putative origins in southwestern Eurasia (Hassemer et al., 2019; Iwanycki Ahlstrand et al., 2019). One hundred iterations were run for each rooted analysis, and each ML tree search was made using the following parameters: a round of rearrangements performed after all populations were added to the tree (-global), bootstrap (-bootstrap) replicates were run sampling blocks of 1,000 SNPs (-k), and standard errors (-se) and were used to evaluate the confidence of the inferred tree topologies. The variance explained by each ML tree was calculated using the get_f() R function supplied in the R plotting scripts in the TREEMIX suite and ln(likelihood) values were reviewed for each tree. The best trees resulting from each of the different roots we modeled were compared. The analyses resulted in trees of similar topologies in which relationships between populations and the amount of drift were unchanged regardless of the root selected; however, trees rooted with Spain1 explained the highest variance of our data and had the highest non-positive ln(likelihood) value (**Supplementary Table 2**). The best-fitting tree was used as a starting point to model gene flow by sequentially adding 15 migration events (-m). We modeled 15 migrated to determine the extent to which additional migration events improves variance for our data. Five independent runs with different random seeds (-seeds) were carried out to examine the consistency of the migration events. The TREEMIX plotting script was used to visualize the trees and the proportion and direction of gene flow events between the 50 sampled populations.

RESULTS

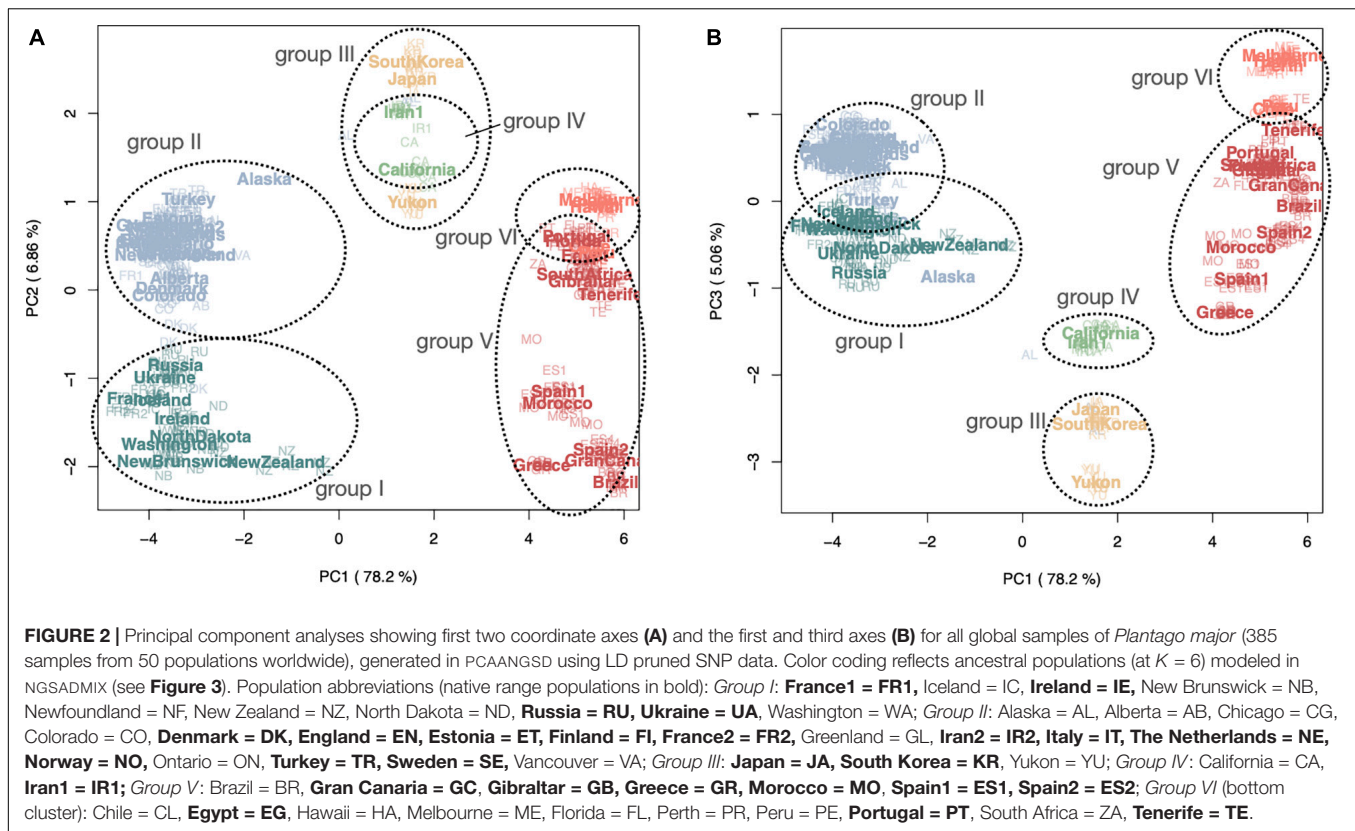
Covariance With Principal Component Analyses

PCA plots based on a covariance matrix generated in PCAANGSD using the LD pruned SNP data reveal differentiation between global samples, especially along the first, second and third

coordinate axes (**Figure 2** and **Supplementary Table 1**). Six clusters or genotype groups (groups I–VI) are clearly distinguishable. PCA axis 1 (represented 78.2% of the variance) divides three broad clusters, while PCA axes 2 and 3 (representing 6.86 and 5.06% of the variance, respectively) show further divisions. All groups consist of populations from both native and introduced ranges; however, only three of the genotype groups from the native range are found to occur extensively in both native and introduced ranges (**Figure 3B**). All individuals collected from the same population are genetically uniform such that they form part of the same genotype group, except in the case of Denmark and Alaska, where individuals sampled from these populations are split between two different genotype groups, providing evidence of multiple introductions and subsequent admixture for these populations. Some of the populations sampled from locations less than one thousand kilometers apart in the native range were resolved in different groups (i.e., France1 belongs to group I and France2 belongs to group II; Iran1 belongs to group IV and Iran2 belongs to group II). With the exception of New Zealand, genotype groups I and II are found above the 35–40° N latitude range, and groups V and VI were the only groups detected in the southern hemisphere (**Figure 3B**). The three most prevalent genotypes we identified in Europe (two in the northern latitudes, and one in the southern latitudes) are also found to be most widespread in the introduced ranges; very little genetic differentiation is seen between native and introduced populations of the same genotype group.

Population Admixture, Shared Ancestry and Heterozygosity

Shared ancestry was estimated on genotype likelihoods in NGSADMIX for values of K (ancestral populations) between two and 12, and results for value $K = 3$ and $K = 6$ are illustrated as a structure graph in **Figure 3A**. The colors represent estimated shared ancestry for all sampled populations for each value of K . Each population comprised 6–10 individuals for which genotyping was performed (three for Chicago). Results for additional values of K are shown in **Supplementary Figure 2**. Our analyses revealed very clear genetic differentiation between the sample populations; however, we found low levels of admixture, as well as a high degree of genetic uniformity between native and introduced populations assigned to the same group at $K = 3$ and $K = 6$, despite wide geographic distances between populations. At a K -value of 3, populations from eastern Asia (Japan, South Korea), Yukon, Alaska (in part), California, and Iran1 are comprised of similar ancestry components. Populations from North America (except for Florida), Iceland, Greenland, and New Zealand share ancestry with north-central and eastern Europe and western Asia (Russia, Turkey, Iran2). Populations from South America, Florida, Hawai'i, Africa, and Australia share ancestry with southern European populations (**Figure 3A**). These three groups resemble the split observed along the first PCA axis. Individual populations from both introduced and native ranges are very homogenous such that very little admixture is seen within each population, except for populations from Alaska, California, and New Zealand. The population from New Zealand,



even at the lowest K -value ($K = 2$), shows admixture between populations from Northern Europe/North America (group II) with lineages from Southern Europe, South America, Hawai'i, Africa, and Australia (groups V and VI).

At a K -value of 6, the groupings based on shared ancestry resemble the groupings seen in the PCA plots taking the first three PCA axes into account and therefore are presented here as a likely ancestry scenario on which to base discussions. Populations belonging to groups V and VI in the PCA plots (Figure 2) differ slightly from the shared ancestry inferred from our admixture analyses. Spain1, Spain2, Greece, and Brazil are identified as having shared ancestry and are found to be almost pure lines, and the same is true for populations from Chile, Peru, Hawai'i, Melbourne, and Perth. The remaining populations are admixed between these two lines (Portugal, Brazil, Florida, Canary Islands, and African populations). Denmark, Alaska, and New Zealand also show notable admixture between two groups at $K5$ and higher levels of K (Supplementary Material).

Levels of heterozygosity were found to be higher in South Korea, California, Russia, North Dakota, New Brunswick, New Zealand, Greece, Egypt, and Brazil (Figure 3C). These populations may experience a greater extent of outcrossing than others. No significant differences were found between the levels of heterozygosity between native and introduced populations, and similarly, no significant differences were found between the six genotype groups (Supplementary Figure 3). However, certain genotype groups have higher proportions of populations with higher levels of heterozygosity suggesting greater levels of

outcrossing within some genotype groups (e.g., group I), and higher degrees of selfing in others (e.g., group II).

Phylogeny and Migration Events

The best ML tree resulting from the TREEMIX analyses explained 94.3 percent (%) of the variance in the relatedness between populations (Figure 4). The topology of the ML tree shows a pattern similar to the genotype groups seen in the PCA plots and NGSADMIX results for $K = 6$. Several of the populations are shown to have long branches corresponding to the amount of drift, including Greece, Colorado, South Korea, Alberta, and Greenland. Residuals from the fit of the model indicate that the tree could not completely explain the ancestry of a few populations, including the relationship between New Zealand and Ireland, New Brunswick, North Dakota, and Washington, suggesting that these are populations where gene flow occurred (Supplementary Figure 3A). Adding fifteen migration events improved the model by explaining 95.8% of the variance. This shows only a small improvement from the tree without migration and suggests that additional events exist in the data (Pickrell and Pritchard, 2012). We found migration events and associated p -values, as well as the variance explained by each level of m , to be identical between runs for the first five (1–5) migration events; however, subsequent migration events ($m = 6$ –15) produced inconsistent results from iteration to iteration, despite all migrations being statistically significant in each iteration. The variance explained by each level of m (1–15) is shown in Supplementary Figure 3C. Based on the lack of consistency in

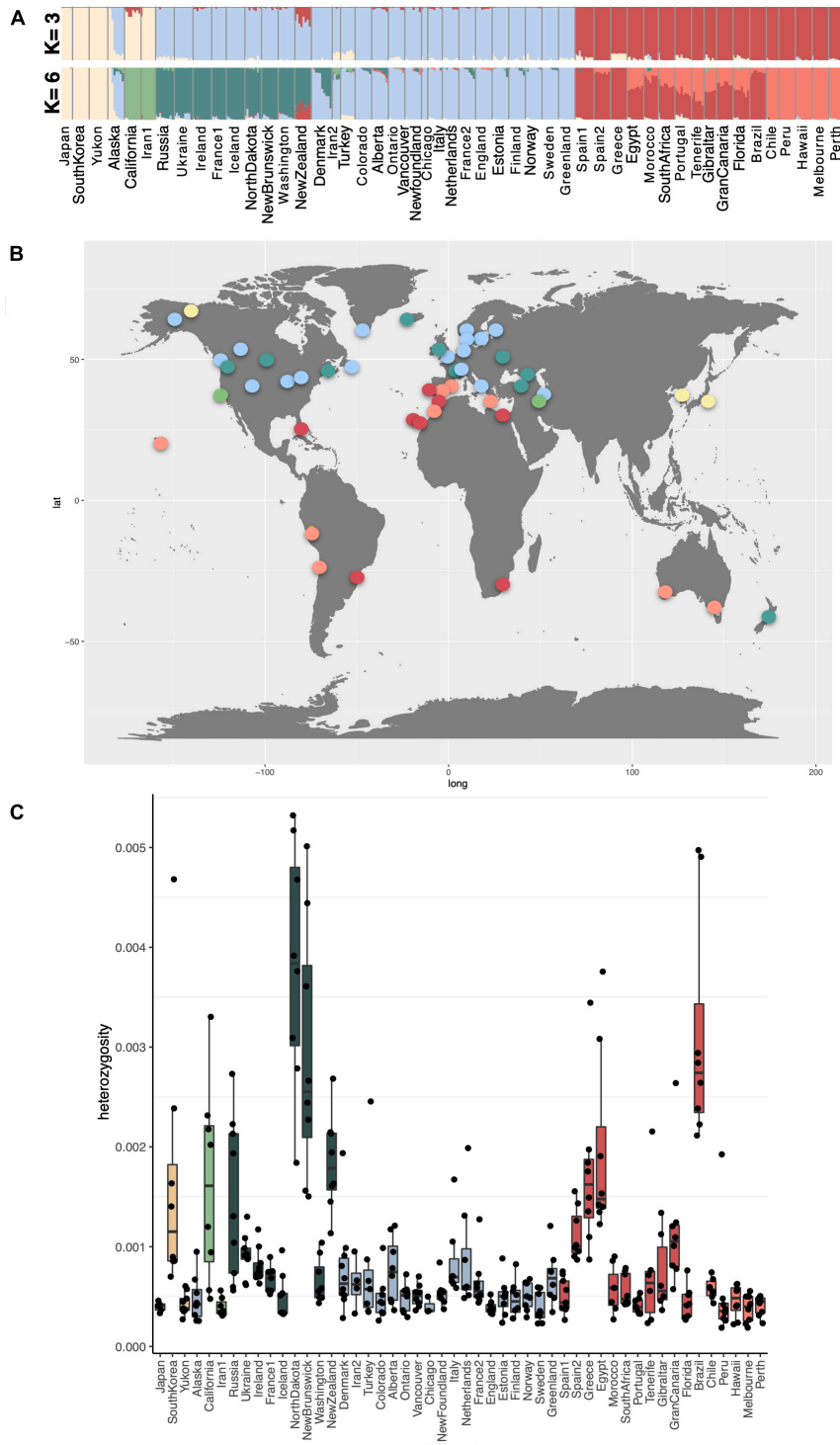


FIGURE 3 | (A) NGSADMIX results for K -values 3 and 7, based on the highest likelihood runs for all samples of *Plantago major* from 50 global populations on LD pruned data. The probability of each individual belonging to a population is indicated by differing colors. **(B)** Global distribution of shared ancestry groups at $K = 7$. **(C)** Heterozygosity for each population computed in ANGSD.

migrations inferred for $m = 6$ and greater, we present and discuss results for the first five migration events only. The direction, weight and significance of each inferred migration event is

presented in **Table 2** (residuals are shown in **Supplementary Figure 3B**). Gene flow from Greece to New Zealand was inferred as the first migration edge ($m = 1$, p -value $\leq 2.2 e^{-308}$). In

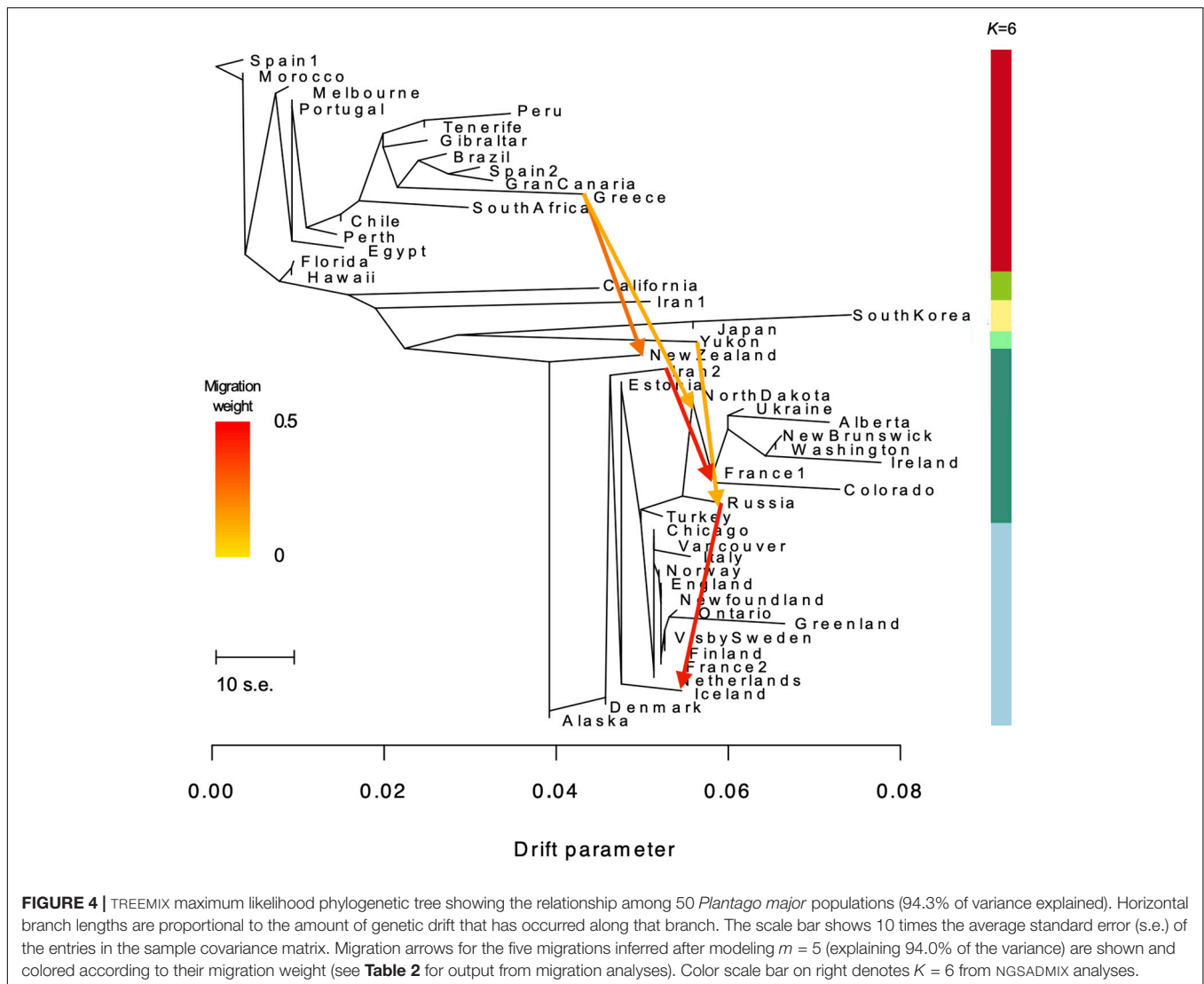


FIGURE 4 | TREEMIX maximum likelihood phylogenetic tree showing the relationship among 50 *Plantago major* populations (94.3% of variance explained). Horizontal branch lengths are proportional to the amount of genetic drift that has occurred along that branch. The scale bar shows 10 times the average standard error (s.e.) of the entries in the sample covariance matrix. Migration arrows for the five migrations inferred after modeling $m = 5$ (explaining 94.0% of the variance) are shown and colored according to their migration weight (see **Table 2** for output from migration analyses). Color scale bar on right denotes $K = 6$ from NGSADMIX analyses.

addition, gene flow is also inferred from Greece to North Dakota ($m = 2$, p -value $\leq 2.2 \times 10^{-308}$), from Yukon to Russia ($m = 3$, p -value $\leq 1.1 \times 10^{-16}$), Iran2 to France1 ($m = 4$, p -value $\leq 2.2 \times 10^{-308}$), and from Russia to Iceland ($m = 5$, p -value $\leq 2.2 \times 10^{-308}$).

DISCUSSION

Global Distribution of Genotype Groups

Our study is the first to sample populations of *Plantago major* subsp. *major* so extensively across its global range and to provide an overview of the global distribution of this prolific weed's genotypic variation. One of our main results was that individuals from across the global range cluster into several distinct genotype groups, based on genetic covariance, shared ancestry, and ML phylogenetic analyses. Genotypes from a given population were found to be homogenous, such that one can predict the genotype group of an individual from those of the other individuals in a given population, based on where in the world it is growing.

Patterns in genotype distribution we observed in the native range may be determined by environmental factors. In the recent worldwide study of *Plantago lanceolata*, populations in the native European range had strong spatial genetic structure associated with geographic distance and precipitation seasonality, while non-native populations had weaker spatial genetic structure that was not associated with environmental gradients but with higher within-population genetic diversity (Smith A. L. et al., 2020). Environmental factors therefore deserve further investigation for their role in shaping the current distribution of genotypes of our study species.

Even though our sampling was focused on populations that fit the morphological descriptions for *Plantago major* subsp. *major*, it is possible that the six genotype groups we identified reflect the complexity of several intraspecific taxa—or hybrids of them resulting from introgression. Two well-accepted subspecies, *P. major* subsp. *major* and *P. major* subsp. *intermedia*, have been the focus of past genetic and ecological studies (Wolff et al., 1994; Wolff and Morgan-Richards, 1998). Although capable of

TABLE 2 | Migration events inferred by treemix fitting 5 gene flow events, with 94.1% of the variance explained (W = weight of the inferred migration edge; W_j = jackknife estimate of the weight; SE_j = jackknife estimate of the standard error).

| No. | From | To | From group | To group | W | W _j | SE _j | p-value |
|-----|--------|---|------------|----------|------|----------------|-----------------|------------------------|
| 1 | Greece | New Zealand | Group V | Group II | 0.22 | 0.22 | 0.02 | <2.2 e ⁻³⁰⁸ |
| 2 | Greece | NorthDakota [(Alberta, Colorado), (Washington (NewZealand, NewBrunswick), Russia (France, Ukraine), Ireland)] | Group V | Group II | 0.10 | 0.10 | 0.01 | <2.2 e ⁻³⁰⁸ |
| 3 | Yukon | Russia | Group III | Group II | 0.49 | 0.49 | 0.01 | <1.1 e ⁻¹⁶ |
| 4 | Iran2 | France1 | Group II | Group II | 0.27 | 0.27 | 0.01 | <2.2 e ⁻³⁰⁸ |
| 5 | Russia | Iceland | Group II | Group II | 0.21 | 0.21 | 0.01 | <2.2 e ⁻³⁰⁸ |

Starting ln (likelihood) = -23643.4; Exiting ln (likelihood) = -11561.8.

interbreeding, high selfing rates in both subspecies is thought to limit introgression (Mølgaard, 1976; Morgan-Richards and Wolff, 2002). El-Bakatoushi (2011) found populations in Egypt to be genetically intermediate between the two subspecies, and that some populations exhibited morphological characters resembling *P. major* subsp. *major* despite genetic evidence of introgression with *P. major* subsp. *intermedia*. The same may be true for other poorly studied subspecies of *P. major*. Regardless of what taxonomic level the genotype groups we identified represent, our findings demonstrate that introgression is limited, but gene flow does occur between genotype groups and is therefore discussed below.

Admixture and Genetic Flow

We anticipated finding lower genetic variation in introduced populations compared to native populations of *P. major* due to founder effects and lack of admixture found in other studies of selfing weedy species (Barrett et al., 2008; Zhang et al., 2010; Ferrero et al., 2015; Zhu et al., 2017). In fact, our admixture analyses revealed a surprising level of genetic uniformity among native and introduced populations assigned to the same genotype group. A small level of admixture was detected in only a few populations, both within the native range (Denmark, admixed between group I and II) and the introduced range (New Zealand, admixed between group II and V; and Alaska, admixed between group I and III). This is in contrast to what has been recently found for *P. lanceolata*, a strictly outcrossing but equally cosmopolitan congener of *P. major*, that was found to have higher genetic variation in the non-native ranges compared to the native ranges due to multiple introductions and subsequent admixture (Smith A. L. et al., 2020). Our data show that ongoing introductions do not shape genetic diversity in introduced ranges of *P. major* as is found for *P. lanceolata*. The lack of admixture indicates that a plant could potentially remain the same, genetically speaking, for years or centuries after arriving in a new region, regardless of additional introductions. Our results further show that levels of heterozygosity do not differ between native and introduced populations; however, we found that outcrossing is higher in some populations and/or genotype groups than others based on higher observed levels of heterozygosity (South Korea, California, Russia, North Dakota, New Brunswick, New Zealand, Greece, Egypt, and Brazil). This includes only one of the populations where admixture was modeled (i.e., New Zealand). A study including the phenological assessment of each population or

genotype group would help inform whether overlaps in flowering time could facilitate outcrossing.

Modeling migration events between populations in TREEMIX provided insight into gene flow from and to specific populations, including those shown to admix in our analyses. For example, admixture observed in New Zealand was best explained by a gene flow event from Greece to New Zealand. Gene flow was also modeled from Yukon to Russia; however, the opposite direction of gene flow was expected (from the native range in Eurasia to Northern Canada). TREEMIX modeling errors are most often in directionality of gene flow (Pickrell and Pritchard, 2012). However, we cannot rule out the possibility that the plants in Yukon are genetically closer to a common ancestor that was either missed in our sampling (by under-representing Asian populations) or is no longer present in the genomic landscape today. We acknowledge that the inference of migration events from our TREEMIX modeling may have been further influenced by the lack of a true outgroup.

Pathways of Introduction

We show that multiple genotype lineages from the native Eurasian range successfully colonized introduced ranges. Three separate introduction events occurred from Europe to North America, one event from the Middle East (Iran) to western North America (California), two introduction events to the Australasian continent (Australia and New Zealand), and at least one introduction event to each of South America, South Africa, Canary Islands, Hawai'i, Greenland, and Iceland.

Despite our extensive sampling across native and introduced ranges, low genetic variation observed within each *P. major* genotype group and the absence of divergence time estimates for our data make it difficult to infer precise origins for the introduced populations. Moreover, one would expect that multiple introduction events of the same genotype occurred historically, particularly given the extraordinary dispersal abilities and commensal nature of *P. major* and other species in the genus *Plantago* (Iwanycki Ahlstrand et al., 2019; Smith A. L. et al., 2020). However, because outcrossing and admixture for the species is so low, it is possible that the genotypic variation in introduced ranges does not vary much over time. Such a scenario makes it possible to make inferences regarding putative historical dispersal pathways despite the absence of historical plant material or records and the lack of divergence time estimates. The distribution of genotype groups and shared ancestry inferred between plants in native and introduced ranges does coincide

with European colonial human movements and/or European colonies thereby supporting previous hypotheses made by others (Samuelson, 2000). For example, the genotype group found in southern Europe (Spain, Portugal) is also found in the former Spanish or Portuguese colonies in Florida (United States), Peru, Chile, and Brazil, and plant propagules could have followed colonial Spanish and/or Portuguese voyagers and settlers between the fifteenth and eighteenth centuries (Mancall, 2006).

Similarly, the two genotypes in central and northern Europe (group I and II) likely gave rise to the majority of populations in North America, as well as Greenland, Iceland, and New Zealand. As previously hypothesized, plants could have traveled with early colonial explorers/settlers from France and England in the sixteenth and seventeenth centuries where these countries had colonial power (Mancall, 2006). Despite hypotheses that the Norse played a role in the movement of *Plantago major* in their travels across the northern Atlantic (Samuelson, 2000), our data do not reveal any specific shared ancestry between plants in Scandinavia, Greenland, and northeastern Canada (Newfoundland). Because gene flow was inferred between southern Europe (Greece) to the plants of New Zealand, rather than from Australia or South Africa (from the same genotype group V), we postulate other populations from southern Europe were dispersed to New Zealand, admixing with plants originating from the northerly latitudes, and potentially did not survive in the more temperate climate.

Some populations of *Plantago major* do not show links to European colonial patterns. For example, although populations from both Melbourne and Perth (Australia) share ancestry with the southern European populations, there is no evidence of direct voyages between the Spanish or Portuguese during early colonial times, although Spanish and Portuguese were both in southern Africa and southeast Asia, and voyagers from the United Kingdom made stops in South Africa *en route* to Australia. Plants could have been introduced to southern Africa and then further dispersed to Australia. Alternatively, plants could have arrived with the Portuguese to Timor in the sixteenth century (Mancall, 2006), approximately 650 km from the Australian coast, and later dispersed by birds to Australia (Iwanycki Ahlstrand et al., 2019). The first record for the species in Australia is in 1770 (GBIF.org, 2020), around the same period the English made their first voyages there, which indicates that plant may have arrived with earlier explorers such as the Dutch, who arrived in Australia (and New Zealand) in the seventeenth century (Mancall, 2006). The plants we sampled in Hawai'i also belong to the southern European genotype group. Given that English explorers were the first to reach Hawai'i (Captain James Cook in 1778), after stops in Tahiti and South Africa, it is also plausible that plant genotypes from southern Europe were picked up and dispersed along the way, and possibly, genotypes from more northern European latitudes also arrived in Australia, South Africa, and Hawai'i, but that the climate or environment did not allow their persistence. Even though the historical herbarium specimens of *P. major* were collected up to centuries after its putative introduction to new ranges, DNA analyses of these specimens may shed light on historical introduction pathways due to their highly selfing nature and are

therefore worthy of future investigation (Dormontt et al., 2014; Martin et al., 2014).

The genotypes found in eastern Asia (South Korea and Japan) were not found within any of the introduced ranges we sampled other than in Yukon, Canada, and gene flow originating from these Asian populations was not inferred in our models. This could be an artifact of poor sampling across Asia, or could represent a true biological scenario in which *P. major* genotypes from eastern Asia are not as successful as genotypes we identified in Europe in colonizing and spreading to other regions, despite ecological similarities between habitats and long history of trade between eastern Asia and introduced parts of the range (i.e., between Japan and western North America or Australia and New Zealand). At least two ecotypes are recognized in Japan—one restricted to sandy shores and brackish waters (which is represented in our sampling and considered native), the other considered introduced and weedy (pers. comm. M. Amano, 2020). Interestingly, populations sampled elsewhere in eastern and Southeast Asia were identified as other *Plantago* species with strong morphological resemblance to *P. major*. Reports of the worldwide distribution of *P. major* may be obscured by lookalikes and mistaken identity, even for trained field botanists.

Our data is too sparse to draw definitive conclusions on origins or putative pathways on dispersal to the Yukon/northernmost populations in North America, yet our findings, and Particularly the PCA plots which show Yukon as a distinct cluster, do not exclude the notion that the northerly latitudes in North America consist of native populations of *P. major* that predate the arrival of early European colonists (Hawthorn, 1974). Further investigation of the link between Asian plants and those in northernmost North America is needed to unravel relationships and migration patterns and clarify any possible links between the movement of eastern Asian plants to northern United States and Canada (for example by Russian colonists or by earlier human migrations from Beringia Erlandson and Braje, 2011).

One unexpected finding was that the population sampled in California was found to differ from plants elsewhere in North America, sharing ancestry with the population from Iran1. Other weedy species, including other *Plantago* species, with native distributions in the Middle East have been introduced to and have become well established in California possibly also based on climatic similarities (i.e., Meyers and Liston, 2008; Ortiz et al., 2008). Pathways of introduction between these biogeographic regions remain poorly studied but the dispersal of weedy species, including *Plantago*, could be linked to the movement of plants for horticultural trade with which weeds are also moved (Chapman et al., 2017; Dullinger et al., 2017). We cannot rule out non-anthropogenic introduction events, particularly since the genus *Plantago* is well adapted for long-distance dispersal, and because North American populations of *Plantago ovata* Forrsk. were found to have arrived in America long before the arrival of Europeans based on molecular dating (Meyers and Liston, 2008; Iwanycki Ahlstrand et al., 2019). More intensive population sampling around the Middle East and the Mediterranean regions and California would be needed to further narrow down the native origins and further resolve ancestry and distribution for this Middle eastern genotype of *P. major*.

CONCLUSION

The six genotype groups we identify serve as an excellent starting point for future ecological and genomic studies, and the distribution of genomic diversity across the globe provides a glimpse into the complex interactions between the environment and the genome that influence the distribution of selfing plant species and mediate phenotypic adaptation to local conditions (Bragg et al., 2015). Our findings indicate that *Plantago major* can establish in a wide variety of new ecological conditions with very little genotypic change and that repeated introductions will not necessarily result in novel genotypes that are better adapted to new environmental conditions. Therefore, our phylogeographic findings can advance our ecological and evolutionary understanding of successful introductions particularly for species with low outcrossing rates.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repository and accession number(s) can be found below: <https://www.ncbi.nlm.nih.gov/>, PRJNA641850.

AUTHOR CONTRIBUTIONS

NIA, NR, and NZ conceptualized the study. NIA designed the global sampling and performed the DNA extractions. NIA,

HMM, SDC, CJR, KMS, CM, GH, AS, DB, EG, MX, AG, MA, OMG, JSP, MB, VM, HC, SB, DZ, HKC, YY, BB, SV, HES, PS, ZL, DB, MH, LJC, AM, CHSL, MG, and NZ performed sampling of populations in different localities. NIA, SG, and VCB processed and analyzed the sequence data and made the figures. NIA wrote the manuscript with SG, VCB, and NR. All authors read and commented on the manuscript and approved the final version.

FUNDING

Funding for the project was provided Marie Curie Actions of the 7th European Community Framework Programme, Grant/Award No. FP7/2007-2013/, REA grant agreement no. 606895-Med, no. 656853.

ACKNOWLEDGMENTS

We thank ADNid, Montpellier (France) for DNA extractions, and the Global Diversity Facility (GDF) of Cornell University for genotyping-by-sequencing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.838166/full#supplementary-material>

REFERENCES

- Baker, H. G. (1965). "Characteristics and modes of origin of weeds," in *Genetics of Colonizing Species*, eds H. G. Baker and G. L. Stebbins (New York, NY: Academic Press). doi: 10.1002/ps.3789
- Baker, H. G. (1974). The evolution of weeds. *Ann. Rev. Ecol. Systems* 5, 1–24.
- Banks, N. C., Paini, D. R., Bayliss, K. L., and Hodda, M. (2015). The role of global trade and transport network topology in the human-mediated dispersal of alien species. *Ecol. Lett.* 18, 188–199. doi: 10.1111/ele.12397
- Barrett, S. C. H., Colautti, R. I., and Eckert, C. G. (2008). Plant reproductive systems and evolution during biological invasion. *Mol. Ecol.* 17, 373–383. doi: 10.1111/j.1365-294X.2007.03503.x
- Bennett, B. C., and Prance, G. T. (2000). Introduced plants in the indigenous pharmacopoeia of Northern South America. *Econ. Botany* 54, 90–102. doi: 10.1016/j.jep.2019.111891
- Bragg, J. G., Supple, M. A., Andrew, R. L., and Borevitz, J. O. (2015). Genomic variation across landscapes: insights and applications. *New Phytol.* 207, 953–967. doi: 10.1111/nph.13410
- Canavan, K., Paterson, I. D., and Hill, M. P. (2017). Exploring the origin and genetic diversity of the giant reed, *Arundo donax* in South Africa. *Invasive Plant Sci. Manag.* 10, 53–60. doi: 10.1017/inp.2016.5
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., and Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Mol. Ecol.* 22, 3124–3140. doi: 10.1534/g3.111.000240
- Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger, and richer datasets. *GigaScience* 4:7. doi: 10.1186/s13742-015-0047-8
- Chapman, D. S., Makra, L., Albertini, R., Bonini, M., Paldy, A., Rodinkova, V., et al. (2016). Modelling the introduction and spread of non-native species: international trade and climate change drive ragweed invasion. *Global Change Biol.* 22, 3067–3079. doi: 10.1111/gcb.13220
- Chapman, D., Purse, B. V., Roy, H. E., and Bullock, J. M. (2017). Global trade networks determine the distribution of invasive non-native species. *Global Ecol. Biogeography* 26, 907–917. doi: 10.1111/geb.12599
- Cornille, A., Salcedo, A., Kryvokhyzha, D., Glemin, S., Holm, K., Wright, S. I., et al. (2016). Genomic signature of successful colonization of Eurasia by the allopolyploid shepherd's purse (*Capsella bursa-pastoris*). *Mol. Ecol.* 25, 616–629. doi: 10.1111/mec.13491
- Dlugosch, K. M., and Parker, I. M. (2007). Molecular and quantitative trait variation across the native range of the invasive species *Hypericum canariense*: evidence for ancient patterns of colonization via pre-adaptation? *Mol. Ecol.* 16, 4269–4283. doi: 10.1111/j.1365-294X.2007.03508.x
- Dormontt, E. E., Gardner, M. G., Breed, M. F., Rodger, J. G., Prentis, P. J., and Lowe, A. J. (2014). Genetic bottlenecks in time and space: reconstructing invasions from contemporary and historical collections. *PLoS One* 9:e106874. doi: 10.1371/journal.pone.0106874
- Dullinger, I., Wessely, J., Bossdorf, O., Dawson, W., Essl, F., Gatringer, A., et al. (2017). Climate change will increase the naturalization risk from garden plants in Europe. *Global Ecol. Biogeography* 26, 43–53. doi: 10.1111/geb.12512
- El-Bakatoushi, R. (2011). Introgressive hybridization between *Plantago major* L. *taxa. Flora* 206, 1045–1051. doi: 10.1016/j.flora.2011.07.010
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:e19379. doi: 10.1371/journal.pone.0019379
- Erlandson, J. M., and Braje, T. J. (2011). From Asia to the Americas by boat? paleogeography, paleoecology, and stemmed points of the northwest Pacific. *Quaternary Int.* 239, 28–37. doi: 10.1016/j.quaint.2011.02.030
- Estoup, A., and Guillemaud, T. (2010). Reconstructing routes of invasion using genetic data: why, how and so what? *Mol. Ecol.* 19, 4113–4130. doi: 10.1111/j.1365-294X.2010.04773.x

- Fernald, M. L. (1900). Some Jesuit influences upon our Northeastern flora. *Rhodora* 2, 133–142.
- Ferrero, V., Barrett, S. C. H., Castro, S., Caldeirinha, P., Navarro, L., Loureiro, J., et al. (2015). Invasion genetics of the Bermuda buttercup (*Oxalis pes-caprae*): complex intercontinental patterns of genetic diversity, polyploidy and heterostyly characterize both native and introduced populations. *Mol. Ecol.* 24, 2143–2155. doi: 10.1111/mec.13056
- Fumagalli, M., Vieira, F. G., Linderoth, T., and Nielsen, R. (2014). ngsTools: methods for population genetics analyses from next-generation sequencing data. *Bioinformatics* 30, 1486–1487. doi: 10.1093/bioinformatics/btu041
- Gaskin, J. F., Schwarzaender, M., Kinter, C. L., Smith, J. F., and Novak, S. J. (2013). Propagule pressure, genetic structure, and geographic origins of *Chondrilla juncea* (Asteraceae): an apomorphic invader on three continents. *Am. J. Botany* 100, 1871–1882. doi: 10.3732/ajb.1200621
- Gbif.org (2020). *GBIF Occurrence Download*. Copenhagen: GBIF.org.
- Godwin, H. (1944). Neolithic forest clearance. *Nature* 153, 511–512. doi: 10.1038/153511a0
- Guo, Q. F., Iannone, B. V., Nunez-Mir, G. C., Potter, K. M., Oswald, C. M., and Fei, S. L. (2017). Species pool, human population, and global versus regional invasion patterns. *Landscape Ecol.* 32, 229–238. doi: 10.1007/s10980-016-0475-6
- Hansen, C. C. R., Hvilsom, C., Schmidt, N. M., Aastrup, P., Van Coeverden, de Groot, P. J., et al. (2018). The muskox lost a substantial part of its genetic diversity on its long road to Greenland. *Curr. Biol.* 28, 4022–4028. doi: 10.1016/j.cub.2018.10.054
- Hassemer, G., Hassemer, G., Bruun-Lund, S., Shipunov, A. B., Briggs, B. G., Meudt, H. M., et al. (2019). The application of high-throughput sequencing for taxonomy: the case of *Plantago* subg. *Plantago* (Plantaginaceae). *Mol. Phylogenet. Evol.* 138, 156–173. doi: 10.1016/j.ympev.2019.05.013
- Hawthorn, W. R. (1974). Biology of canadian weeds .4. *Plantago major* and *Plantago rugelii*. *Canadian J. Plant Sci.* 54, 383–396. doi: 10.4141/cjps74-059
- Hejda, M., Chytrý, M., Pergl, J., and Pyšek, P. (2015). Native-range habitats of invasive plants: are they similar to invaded-range habitats and do they differ according to the geographical direction of invasion? *Diversity Distribution* 21, 312–321. doi: 10.1111/ddi.12269
- Herten, K., Hestand, M. S., Vermeesch, J. R., and Van Houdt, J. K. J. (2015). GBSX: a toolkit for experimental design and demultiplexing genotyping by sequencing experiments. *BMC Bioinform.* 16:73. doi: 10.1186/s12859-015-0514-3
- Hodkinson, D. J., and Thompson, K. (1997). Plant dispersal: the role of man. *J. Appl. Ecol.* 34, 1484–1496.
- Holm, L., Pancho, J., Herberger, J., and Plucknett, D. (1979). *A Geographical Atlas of World Weeds*. New York, NY: John Wiley & Sons.
- Hulme, P. E. (2009). Trade, transport and trouble: managing invasive species pathways in an era of globalization. *J. Appl. Ecol.* 46, 10–18. doi: 10.1111/j.1365-2664.2008.01600.x
- Iwanycki Ahlstrand, N., Markussen, B., Havskov Reghev, N., Eiriksson, F., Thorsteinsdóttir, M., Rønsted, N., et al. (2018). Untargeted metabolic profiling reveals geography as a stronger predictor of metabolic phenotypes than environment or herbivore damage of a cosmopolitan weed. *Ecol. Evol.* 8, 6812–6826. doi: 10.1002/ece3.4195
- Iwanycki Ahlstrand, N., Verstraete, B., Hassemer, G., Dunbar-Co, S., Hoggard, R., and Rønsted, N. (2019). Ancestral range reconstruction of remote oceanic island species of *Plantago* (Plantaginaceae) reveals differing scales of modes of dispersal. *J. Biogeography* 46, 706–722. doi: 10.1111/jbi.13525
- Jabbour, F., and Renner, S. S. (2019). *Consolida* and *aconitella* are an annual clade of delphinium (Ranunculaceae) that diversified in the Mediterranean basin and the Irano-Turanian region. *Taxon* 60, 1029–1040. doi: 10.1002/tax.604007
- Josslyn, J. (1672). *New-Englands Rarities, Discovered in Birds, Beasts, Fishes, Serpents, and Plants of that Country*. London: GD Widdowes.
- Korneliusson, T. S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. *BMC Bioinform.* 15:356. doi: 10.1186/s12859-014-0356-4
- Kreitschitz, A., Kovalev, A., and Gorb, S. N. (2016). "Sticky invasion" - the physical properties of *Plantago lanceolata* L. seed mucilage. *Beilstein J. Nanotechnol.* 7, 1918–1927. doi: 10.3762/bjnano.7.183
- Kuiper, P. J. C., and Bos, M. (1992). "Ecological studies analysis and synthesis," in *Plantago: A Multidisciplinary Study*, eds P. J. C. Kuiper and M. Bos (Berlin: Springer Verlag).
- Lambertini, C., Mendelssohn, I. A., Gustafsson, M. H. G., Olesen, B., Riis, T., Sorrell, B. K., et al. (2012). Tracing the origin of Gulf Coast Phragmites (Poaceae): a story of long-distance dispersal and hybridization. *Am. J. Bot.* 99, 538–551. doi: 10.3732/ajb.1100396
- Li, H. (2011). Improving SNP discovery by base alignment quality. *Bioinformatics* 27, 1157–1158. doi: 10.1093/bioinformatics/btr076
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Lu, F., Lipka, A. E., Glaubitz, J., Elshire, R., Cherney, J. H., Casler, M. D., et al. (2013). Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLoS Genetics* 9:e1003215. doi: 10.1371/journal.pgen.1003215
- Mack, R. N. (2003). Plant naturalizations and invasions in the eastern United States: 1634–1860. *Ann. Missouri Bot. Garden* 90, 77–90. doi: 10.2307/3298528
- Mack, R. N., Simberloff, D., Lonsdale, W. M., Evans, H., Clout, M., and Bazzaz, F. A. (2000). Biotic invasions: causes, epidemiology, global consequences, and control. *Ecol. Appl.* 10, 689–710.
- Mancall, P. C. (2006). *Travel Narratives from the Age of Discovery: an Anthology*. CityOxford: Oxford University Press.
- Martin, M. D., Zimmer, E. A., Olsen, M. T., Foote, A. D., Gilbert, M. T. P., and Brush, G. S. (2014). Herbarium specimens reveal a historical shift in phylogeographic structure of common ragweed during native range disturbance. *Mol. Ecol.* 23, 1701–1716. doi: 10.1111/mec.12675
- Meisner, J., and Albrechtsen, A. (2018). Inferring population structure, and admixture proportions in low-depth NGS Data. *Genetics* 210, 719–731. doi: 10.1534/genetics.118.301336
- Meyers, S. C., and Liston, A. (2008). The biogeography of *Plantago ovata* forssk. (Plantaginaceae). *Int. J. Plant Sci.* 169, 954–962. doi: 10.1086/589699
- Mølgaard, P. (1976). *Plantago major* ssp. *major* and ssp. *pleiosperma*. morphology, biology and ecology in Denmark. *Botany Tidsskrift Bd* 71, 31–56.
- Morgan-Richards, M., and Wolff, K. (2002). Genetic structure and differentiation of *Plantago major* reveals a pair of sympatric sister species. *Mol. Ecol.* 8, 1027–1036. doi: 10.1046/j.1365-294x.1999.00665.x
- Narum, S. R., Buerkle, C. A., Davey, J. W., Miller, M. R., and Hohenlohe, P. A. (2013). Genotyping-by-sequencing in ecological and conservation genomics. *Mol. Ecol.* 22, 2841–2847. doi: 10.1111/mec.12350
- Nielsen, R., Paul, J. S., Albrechtsen, A., and Song, Y. S. (2013). Genotype and SNP calling from next-generation sequencing data. *Nat. Rev. Genet.* 12, 443–451. doi: 10.1038/nrg2986
- Ortiz, M. A., Tremetsberger, K., Terrab, A., Stuessy, T. F., Garcia-Castano, J. L., Urtubey, E., et al. (2008). Phylogeography of the invasive weed *Hypochaeris radicata* (Asteraceae): from Moroccan origin to worldwide introduced populations. *Mol. Ecol.* 17, 3654–3667. doi: 10.1111/j.1365-294X.2008.03835.x
- Palmer, C. (2004). *Plantago* spp. and *Bidens* spp.: a case study of change in Hawaiian herbal medicine. *J. Ethnobiol.* 24, 13–31.
- Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., et al. (2012). Ancient admixture in human history. *Genetics* 192, 1065–1093. doi: 10.1534/genetics.112.145037
- Peñeroová, P., Garcia-Erill, G., Liu, X., Nursyifa, C., Waples, R. K., Santander, C. G., et al. (2021). High genetic diversity and low differentiation reflect the ecological versatility of the African leopard. *Curr. Biol.* 31, 1862–1871. doi: 10.1016/j.cub.2021.01.064
- Peruzzi, L., and Passalacqua, N. G. (2003). *Plantago sinuata* Lam. (Plantaginaceae), a misinterpreted unit, typical of moist places. morphological and karyological evidence. *J. Plant Taxonomy Geography* 58, 441–450. doi: 10.1080/00837792.2003.10670757
- Pickrell J. K., and Pritchard J. K., (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 8:e1002967. doi: 10.1371/journal.pgen.1002967
- Pilger, R. (1937). "Plantaginaceae," in *Engelmann*, eds A. Engler and L. Diels (Leipzig: Engelmann), 466.

- POWO (2019). *Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew*. Available online at: <http://www.plantsoftheworldonline.org/> (accessed 10 April 2020).
- Preston, C. D., Pearman, D. A., and Hall, A. R. (2004). Archaeophytes in Britain. *Botan. J. Linnean Soc.* 145, 257–294. doi: 10.1111/j.1095-8339.2004.00284.x
- Puckett, E. E., Magnussen, E., Khlyap, L. A., Strand, T. M., Lundkvist, Å., and Munshi-South, J. (2020). Genomic analyses reveal three independent introductions of the invasive brown rat (*Rattus norvegicus*) to the Faroe Islands. *Heredity* 124, 15–27. doi: 10.1038/s41437-019-0255-6
- Puckett, E. E., Park, J., Combs, M., Blum, M. J., Bryant, J. E., Caccione, A., et al. (2016). Global population divergence and admixture of the brown rat (*Rattus norvegicus*). *Proc. R. Soc. B: Biol. Sci.* 283:20161762. doi: 10.1098/rspb.2016.1762
- Pyšek, P. (1998). Alien and native species in Central European urban floras: a quantitative comparison. *J. Biogeography* 25, 155–163.
- Pyšek, P., Jarošík, V., Pergl, J., Randall, R., Chytrý, M., Kühn, I., et al. (2009). The global invasion success of Central European plants is related to distribution characteristics in their native range and species traits. *Diversity Distrib.* 15, 891–903. doi: 10.1111/j.1472-4642.2009.00602.x
- Pyšek, P., Manceur, A. M., Alba, C., McGregor, K. F., Pergl, J., Štajerová, K., et al. (2015). Naturalization of central European plants in North America: species traits, habitats, propagule pressure, residence time. *Ecology* 96, 762–774. doi: 10.1890/14-1005.1
- Rahn, K. (1996). A phylogenetic study of the plantaginaceae. *Botan. J. Linnean Soc.* 120, 145–198. doi: 10.1111/j.1095-8339.1996.tb00484.x
- Rousseau, J. (1966). [Au Query: Provide the city name and publisher name for the reference “Rousseau, 1966”.] “Movement of plants under the influence of man,” in *The Evolution of Canada's Flora*, eds R. L. Taylor and R. A. Ludwig 81–99.
- Samuelsen, A. B. (2000). The traditional uses, chemical constituents and biological activities of *Plantago major* L. a review. *J. Ethnopharmacol.* 71, 1–21. doi: 10.1016/S0378-8741(00)00212-9
- Schubert, M., Ermini, L., Der Sarkissian, C., Jónsson, H., Ginolhac, A., Schaefer, R., et al. (2014). Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat. Protocols* 9:1056. doi: 10.1038/nprot.2014.063
- Schubert, M., Lindgreen, S., and Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* 9:88. doi: 10.1186/s13104-016-1900-2
- Seebens, H., Essl, F., Dawson, W., Fuentes, N., Moser, D., Pergl, J., et al. (2015). Global trade will accelerate plant invasions in emerging economies under climate change. *Glob. Change Biol.* 21, 4128–4140.
- Skotte, L., Korneliusen, T. S., and Albrechtsen, A. (2013). Estimating individual admixture proportions from next generation sequencing data. *Genetics* 195, 693–702. doi: 10.1534/genetics.113.154138
- Smith, A. L., Hodkinson, T. R., Vilellas, J., Catford, J. A., Csörgő, A. M., Blomberg, S., et al. (2020). Global gene flow releases invasive plants from environmental constraints on genetic diversity. *Proc. Natl. Acad. Sci. U S A.* 117, 4218–4227. doi: 10.1073/pnas.1915848117
- Stepp, J. R., and Moermann, D. E. (2000). The importance of weeds in ethnopharmacology. *J. Ethnopharmacol.* 75, 19–23. doi: 10.1016/S0378-8741(00)00385-8
- Turbelin, A. J., Malamud, B. D., and Francis, R. A. (2017). Mapping the global state of invasive alien species: patterns of invasion and policy responses. *Global Ecol. Biogeography* 26, 78–92. doi: 10.1111/geb.12517
- van Dijk, H. (1984). Genetic variability in *Plantago* species in relation to their ecology: 2. quantitative characters and allozyme loci in *P. major*. *THEOR. APPL. GENET.* 68:43. doi: 10.1007/BF00252310
- van Dijk, H., and van Delden, W. (1981). Genetic variability in *Plantago* species in relation to their ecology. *Theoretical Appl. Genet.* 60, 285–290. doi: 10.1007/BF00263720
- van Dijk, H., and Wolff, K. (1992). “Allozyme variation and genetic structure in *Plantago* species,” in *Plantago: A Multidisciplinary Study. Ecological Studies*, Vol. 89, eds P. J. C. Kuiper and M. Bos (Berlin: Springer Verlag), 184–192.
- van Dijk, H., Wolff, K., and De Vries, A. (1988). Genetic variability in *Plantago* species in relation to their ecology. 3. structure of populations of *P. major*, *P. lanceolata*, and *P. coronopus*. *Theoretical Appl. Genet.* 75, 518–528.
- van Kleunen, M., Dawson, W., Essl, F., Pergl, J., Winter, M., Weber, E., et al. (2015). Global exchange and accumulation of non-native plants. *Nature* 525, 100–103. doi: 10.1038/nature14910
- Vigueira, C. C., Olsen, K. M., and Caicedo, A. L. (2013). The red queen in the corn: agricultural weeds as models of rapid adaptive evolution. *Heredity* 110, 303–311. doi: 10.1038/hdy.2012.104
- Wilson, J. R., Dormontt, E. E., Prentis, P. J., Lowe, A. J., and Richardson, D. M. (2009). Something in the way you move: dispersal pathways affect invasion success. *Trends Ecol. Evol.* 24, 136–144. doi: 10.1016/j.tree.2008.10.007
- Wolff, K. (1991). Genetic analysis of morphological variability in three *Plantago* species with different mating systems. *Theoretical Appl. Genet.* 81, 111–118. doi: 10.1007/BF00226120
- Wolff, K., and Morgan-Richards, M. (1998). PCR markers distinguish *Plantago* major subspecies. *Theoretical Appl. Genet.* 96, 282–286. doi: 10.1007/s001220050737
- Wolff, K., Rogstad, S. H., and Schaal, B. A. (1994). Population and species variation of minisatellite DNA in *Plantago*. *Theoretical Appl. Genet.* 87, 733–740. doi: 10.1007/BF00222899
- Young, J. A., and Evans, R. A. (1973). Mucilaginous seed coats. *Weed Sci.* 21, 52–54. doi: 10.1111/j.1438-8677.2012.00684.x
- Zhang, Y. Y., Zhang, D. Y., and Barrett, S. C. (2010). Genetic uniformity characterizes the invasive spread of water hyacinth (*Eichhornia crassipes*), a clonal aquatic plant. *Mol. Ecol.* 19, 1774–1786. doi: 10.1111/j.1365-294X.2010.04609.x
- Zhu, B. R., Barrett, S. C., Zhang, D. Y., and Liao, W. J. (2017). Invasion genetics of *Senecio vulgaris*: loss of genetic diversity characterizes the invasion of a selfing annual, despite multiple introductions. *Biol. Invasions* 19, 255–267. doi: 10.1007/s10530-016-1277-0

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Iwanycki Ahlstrand, Gopalakrishnan, Vieira, Bieker, Meudt, Dunbar-Co, Rothfels, Martinez-Swatson, Maldonado, Hassemmer, Shipunov, Bowers, Gardner, Xu, Ghorbani, Amano, Grace, Pringle, Bishop, Manzanilla, Cotrim, Blaney, Zubov, Choi, Yesil, Bennett, Vimolmangkang, El-Seedi, Staub, Li, Boldbaatar, Hislop, Caddy, Muasya, Saslis-Lagoudakis, Gilbert, Zerega and Rønsted. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.