

# Travel tales of a worldwide weed: genomic signatures reveal colonial trade routes and prior adaptation are key to the success of *Plantago major*

Natalie Iwanycki Ahlstrand<sup>1</sup>, Shyam Gopalakrishnan<sup>1</sup>, Filipe Vieira<sup>2</sup>, Heidi Meudt<sup>3</sup>, Stephanie Dunbar-Co<sup>4</sup>, Carl Rothfels<sup>5</sup>, Karen Martinez-Swatson<sup>6</sup>, Carla Maldonado<sup>7</sup>, Gustavo Hassemer<sup>8</sup>, Alexey Shipunov<sup>9</sup>, Deanne Bowers<sup>10</sup>, Elliot Gardner<sup>11</sup>, Maonian Xu<sup>12</sup>, Abdolbaset Ghorbani<sup>13</sup>, Makoto Amano<sup>14</sup>, Olwen Grace<sup>15</sup>, James Pringle<sup>16</sup>, Madonna Bishop<sup>17</sup>, Vincent Manzanilla<sup>18</sup>, Helena Cotrim<sup>19</sup>, Sean Blaney<sup>20</sup>, Dmitri Zubov<sup>21</sup>, Hong-Keun Choi<sup>22</sup>, Yester Yesil<sup>23</sup>, Bruce Bennett<sup>24</sup>, Sornkanok Vimolmangkang<sup>25</sup>, Hesham El-Seedi<sup>13</sup>, Peter Staub<sup>26</sup>, Zhu Li<sup>27</sup>, Delgerbat Boldbaatar<sup>28</sup>, Michael Hislop<sup>29</sup>, Laura Caddy<sup>30</sup>, Muthama Muasya<sup>31</sup>, Haris Saslis-Lagoudakis<sup>32</sup>, M. Tomas P. Gilbert<sup>33</sup>, Nyree Zerega<sup>11</sup>, and Nina Rønsted<sup>1</sup>

<sup>1</sup>Københavns Universitet

<sup>2</sup>Rigshospitalet

<sup>3</sup>Museum of New Zealand Te Papa Tongarewa

<sup>4</sup>University of Hawaii

<sup>5</sup>University of California Berkeley

<sup>6</sup>Havforskninginstituttet

<sup>7</sup>Herbario Nacional De Bolivia

<sup>8</sup>Federal University of Mato Grosso do Sul

<sup>9</sup>Minot State University

<sup>10</sup>University of Colorado at Boulder

<sup>11</sup>Northwestern University

<sup>12</sup>University of Iceland

<sup>13</sup>Uppsala Universitet

<sup>14</sup>Natural History Museum and Institute Chiba

<sup>15</sup>Royal Botanic Gardens Kew

<sup>16</sup>Royal Botanical Gardens Ontario

<sup>17</sup>Memorial University of Newfoundland

<sup>18</sup>University of Oslo

<sup>19</sup>University of Lisbon

<sup>20</sup>Atlantic Canada Conservation Data Centre

<sup>21</sup>Gryshko's National Botanic Garden

<sup>22</sup>Ajou University

<sup>23</sup>Istanbul Universitesi

<sup>24</sup>Yukon Conservation Data Centre

<sup>25</sup>Chulalongkorn University

<sup>26</sup>University of Cagliari

<sup>27</sup>Chinese Academy of Sciences  
<sup>28</sup>National University of Mongolia  
<sup>29</sup>Western Australia Herbarium  
<sup>30</sup>The University of British Columbia  
<sup>31</sup>University of Cape Town  
<sup>32</sup>Forage Crete  
<sup>33</sup>University of Copenhagen

June 26, 2020

## Abstract

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 are thought to have been introduced to regions outside their native ranges 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 test hypotheses that the plant was brought to new regions during colonial times. We further investigate how patterns in genomic diversity facilitate the success of this global weed. Although genomic differentiation among populations is found to be low, we identify six unique ecotypes showing very little sign of admixture. Three of the most prevalent of these ecotypes present in the native range gave rise to introduced populations in the Americas, Africa, Australia and New Zealand, indicating that more than one successful ecotype colonized and spread. The distribution of ecotypes is found to have links to colonial history, and ecotypes are further found to be restricted by latitude. Dispersal of multiple successful ecotypes and prior adaptation in the native range to latitudinally dependent environmental factors (such as climate) are likely reasons for the success of this prolific, global weed. Genomic signatures can provide new perspectives on the drivers behind the historic introductions and the successful colonization of introduced species in an era of global change.

## Keywords

Introduced species, human mediated dispersal, historical introduction, introduction routes, prior adaptation, genotyping-by-sequencing

## 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 & Guillemaud, 2010; van Klunen et al., 2015; Seebens 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 well-being and threaten indigenous biodiversity (Mack et al., 2000; Puckett et al., 2016; Chapman 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, Perlman, & Hall 2004; van Klunen 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, the birth of colonialism, and the start

of wide-scale changes in human demography, land use, trade and industrial development (Godwin 1944; Pysek, 1988; Preston, Perlman, & Hall, 2004; Mancall 2006; Hulme 2009; Banks et al., 2015; Turbelin et al., 2016). Written accounts from early voyagers and colonists dating back to the 16<sup>th</sup> and 17<sup>th</sup> centuries can provide evidence of the introduction of European species to continents such as North America; however, such records are scarce (Fernald, 1900). This, and the poor documentation of native biodiversity prior to arrival of Europeans, hinders our ability to retrace the origins of historical introductions (Estoup & 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 differentiation between native and introduced ranges (Ortiz et al, 2008; Estoup & Guillemaud, 2010; Lambertini et al. 2012; Zhu et al., 2017; Canavan et al. 2017). However, genetic changes within populations over time attributed to admixture or loss of genetic diversity due to founder effects or genetic drift can pose challenges in identifying source populations and inferring dispersal routes for historical introductions (Dlugosch and Parker 2008; 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 differentiation is very low or genetic changes between populations are pronounced (Estoup & Guillemaud, 2010; Elshire et al., 2011; Lu et al., 2013; Vigueira et al., 2013; Narum 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 Europe and Asia, the species grows in a wide range of disturbed habitats (Hawthorn, 1974; Holm et al., 1979; van Dijk, 1984). The species is considered commensal and, based on its medicinal properties, has had a long history of human use in both native and introduced ranges (Samuelsen, 2000; Stepp & Moerman, 2000; Bennett & Prance, 2000; Palmer, 2004). It was introduced to North America by European voyagers or possibly earlier with Norse voyagers (Samuelsen, 2000), and to other parts of the introduced ranges during colonial times, including Australasia, South America, and southern Africa. It is an excellent example of a poorly documented historical plant introduction for which genomic analysis can help unravel. Historical written records and herbarium specimens collected in the introduced ranges before the 19<sup>th</sup> 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 & Wolff, 1992; Rahn, 1996). By the 17<sup>th</sup> century, the plant had already been noted to be well-established in New England (northeastern USA), where indigenous peoples referred to the plant as “Englishman’s foot” because it followed colonists wherever they went (Josselyn, 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 & Bos, 1992; Morgan-Richards & Wolff, 1999). It possesses many of the traits that are common amongst the most successful introduced plants (Pysek et al., 2009; Hejda et al., 2014; Pysek 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 pollination (Young & Evans, 1973; Hawthorn, 1974; van Dijk, 1984; Rahn, 1996; Samuelsen, 2000; Kreitschitz, Kovalev, & Gorb, 2016; Iwanycki Ahlstrand et al, 2018; Iwanycki Ahlstrand et al, 2019). Furthermore, the species is capable of self-fertilizing, meaning that a single propagule can putatively establish sexually reproducing populations in new ranges (Baker, 1974; Wolff, 1991). As with other selfing or clonal species, thousands of individuals of *P. major* can persist in a small area, but low genetic diversity within such populations has been found (van Dijk & van Delden, 1981; Wolff et al., 1994; Wolff & 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 line (Wolff, 1991; Wolff and Morgan-Richards, 1998).

Despite the species being extensively studied in the UK, Denmark and The Netherlands, the genetics of populations elsewhere in its native and introduced ranges have seldom been investigated. The origins and genetic diversity of non-crop plants that have been introduced across the globe, and how genomic variation contributes to the success of global weeds have rarely been studied (but see Ortiz et al, 2008; Cornille et al., 2016; Zhu et al., 2017). To address these gaps, we take a population-genomic approach and analyse thousands of genome-wide single-nucleotide polymorphisms (SNPs), generated by genotyping-by-sequencing (GBS) of *Plantago major* sampled from 50 populations across its global range to (1) infer ancestral origins of introduced populations, and (2) retrace pathways of introduction to new ranges including North and South America, Greenland, Iceland, Africa, Hawai'i, the Canary Islands, Australia and New Zealand. GBS was chosen based on its application to population genomics in non-model plants (Lu et al., 2013). We test the hypotheses that *P. major* was dispersed by Europeans during colonial times and that a high level of genetic differentiation was not necessary for the successful establishment and spread of introduced plants. This knowledge can help us understand the dispersal and establishment of new potential plant introductions as well as how introduced plants cope under climate change.

## 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 lower taxonomic divisions and synonyms having been proposed (Pilger, 1937; Mølgaard, 1976; Peruzzi & Passalacqua, 2003; POWO 2019). At least three subspecies are recognized in Europe, two of which, *P. major* subsp. *major* and *P. major* subsp. *intermedia* (Gilib.) Lange, have been widely studied and even recognized as separate species (*P. major* and *P. intermedia* DC.; Morgan-Richards & Wolff, 1999; 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. However, we focused our sampling on *P. major* subsp. *major* which is thought to be more widely distributed and tolerant of a wider range of environmental variation and anthropogenic disturbance (Mølgaard, 1976). *Plantago major* subsp. *major* can be discriminated from *P. major* subsp. *intermedia* based on the number of seeds per capsule (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 were sampled worldwide, including 22 populations from across the native range (Europe, Asia) and 28 populations from the introduced range (North and South America, Iceland, Greenland, Hawai'i, Australia, New Zealand, and Africa) (**Figure 2 ; Table 1**). Six to ten individuals were sampled from each population, amounting to 385 individuals in total. A herbarium voucher was sampled 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.

### DNA extractions

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 hours 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).

## Sequencing

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 PstI (CTGCA<sup>^</sup>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, USA). 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 a PicoGreen assay (Molecular Probes, Carlsbad, CA, USA). Samples were run on seven different plates (96-wells), and one blank was included per plate. Plates were run on lanes of a 100-basepair single-end Illumina HiSeq 2000, at the Cornell Core Laboratories Center (Cornell University, New York, USA).

## Assembly of pseudo-reference genome (catalogue)

stacks, a software pipeline designed for restriction-enzyme based data for organisms without a reference genome, was used to generate a catalogue 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 catalogue 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 catalogue. 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 catalogue. From the loci identified in the reference catalogue, we created a pseudo-reference genome by collapsing these loci into 10 chromosomes, with a string of 150 Ns inserted between consecutive loci.

## Mapping to the pseudo-reference genome (catalogue)

Raw reads were demultiplexed using the demultiplexer function in gbsx v.1.3 (Herten 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 catalogue 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 & 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.

## SNP-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 propagate 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 less than 20). Using the distribution of depths obtained from *angsd*, we chose a maximum average depth cut-off of 70 X coverage per sample. Variant discovery and computation of genotype likelihoods was performed using *angsd* with the same parameters as above, but additionally a depth cut-off of 70 per sample (`-maxDepth NumberOfSamples*70`), 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`). The genotype likelihoods were calculated using the model described in *samtools* (`-GL 1`; v. 1.4; Li, 2011).

## Phylogenomic inference

Pairwise genetic distances between individuals were examined using *ngsdist* (Vieira, Lassalle, Korneliussen, & Fumagalli, 2015), a distance-based phylogenetic inference method that takes genotype-likelihoods into account and uses the genotype posterior probabilities estimated in *angsd*. Pairwise genetic distances among the global samples were computed using SNPs that had a minor allele frequency greater than 10% in the sample, resulting in 2807 total sites. In addition, for each pair of samples, all sites that did not have any reads covering them in either of the samples were discarded. Pairwise distances were visualized as multidimensional scaling (MDS) plots using the *cmdscale* function in base *R* (v 3.6.2) to compute the first six coordinates.

## Admixture analysis

To identify population structure and identify patterns of admixture among our samples an individual-based assignment test was performed using *ngsadm* (Skotte et al., 2013; Fumagalli et al., 2014) using 7594 SNPs. *ngsadm* 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 *ngsadm*. Therefore, results for each value of  $K$  were reviewed and compared with results from the MDS plots to select a  $K$  value that had the most biological relevance based on the genomics of the species being studied.

## Inference of population splits and migration events

*treemix* v. 1.13 (Pickrell and Pritchard, 2012) was used to visualize how well the relationships can be represented by a bifurcating tree using population allele frequencies. An in-house script in python was written to convert Beagle files generated in *angsd* to *treemix* input files (See Supplementary Materials). In addition to admixture, *treemix* also provides some information about the directionality of gene flow (Patterson et al, 2012). Trees were rooted (`-root`) using populations from Yukon, Japan, and South Korea (based on output from admixture analyses suggesting that these populations are outgroups to the others), and we estimated the likelihood of graphs with 0 – 4 migration events added in order to visualize the proportion and direction of gene flow events between the 50 sampled populations.

# Results

## Phylogenomic analyses

Multidimensional scaling (MDS) plots based on a distance matrix generated in *ngsdist* of 2807 of the most informative SNPs reveal differentiation between global samples, especially along the first and second coordinate axes (**Figure 3**). Genetic differentiation among individuals in each of the 50 populations detected in our SNP data based on MDS plots was low as expected for a selfing plant. Six distinct genotype groups (groups I – VI) are clearly distinguishable. 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 are split between two different groups and/or include a few outlying individuals. Some of the closest geographic populations sampled from 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). From left to right in Figure 3, the first genotype group (group I) includes populations from France1, Iceland, Ireland, Denmark (in part), Russia, Ukraine, North Dakota (USA), Washington (USA), New Brunswick (CAN), and New Zealand. The second (group II) includes France2, England, Netherlands, Denmark (majority of samples), Sweden, Norway, Finland, Estonia, Italy, Turkey, Iran2, Greenland, and Alberta (CAN), Colorado (USA), Ontario (CAN), Vancouver (CAN), Chicago (USA), Newfoundland (CAN), and Alaska (USA, majority of samples). These two genotype groups primarily consist of populations that originate above the 35–40° N latitude range (except for New Zealand; **Figure 4b**). Individuals collected from Yukon (CAN) form their own genotype group (group III). Populations from Japan, South Korea, Iran1, California (USA), and the remaining individuals from Alaska form a fourth group (group IV). A fifth genotype group (group V) consists of populations from Spain (Spain1, Spain2), Greece, Gran Canaria, Morocco, Brazil. The final group (group VI) consists of populations from Portugal, Egypt, Tenerife, South Africa, Australia (Melbourne and Perth), Peru, Chile, Hawai'i (USA) and Florida (USA). Populations from groups V and VI all occur below the 35–40° N latitude range.

## Demographic history & admixture

Shared ancestry was estimated on genotype likelihoods using 7594 SNPs in *ngsadm* for values of  $K$  (ancestral populations) between two and 12, and results for value  $K = 3$  and  $K = 8$  are illustrated as a structure graph in **Figure 4a**. The colours represent estimated shared ancestry for all sampled populations for each value of  $K$ . Each population comprised 6 – 10 individuals for which genotyping was performed. Results for additional values of  $K$  are shown in Supplementary Materials. At a  $K$  value of 3, populations from eastern Asia (Japan, South Korea), Yukon, Alaska (in part), California, and Iran1 comprise 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 4a**). 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 more admixture than other populations, and although individuals from New Zealand share most of their ancestry with populations from Northern Europe and North America, there is evidence of admixture with lineages from Southern Europe, South America, Hawai'i, Africa, and Australia.

At a  $K$  value of 6, the groupings of shared ancestry resemble the groupings seen in the MDS plot (**Figure 3**) and therefore represent a likely ancestry scenario to base discussions on. Two notable differences are found here compared to the MDS plot. First, Japan and South Korea are inferred to have a different ancestry than do California and Iran1 further splitting group IV as shown in the MDS plot. Secondly, populations belonging to groups V and VI in the MDS plot are not split into unique ancestries. Only three of the 50 populations sampled (Denmark, Alaska and New Zealand) showed signs of admixture between two groups at this value of  $K$ .

## Population splits and migration

The maximum likelihood tree resulting from the treemix analyses (**Figure 5**) shows a pattern similar to the genotype groups seen under MDS plots and ngsadm results. The three most basal groups are similar to the groupings seen in admixture analyses for  $K = 3$ , except for the positions of California and Iran which are shown to be sister to genotype groups V and VI (occurring below the 35–40° N latitude range) rather than more closely related to Japan, South Korea and Yukon. Our treemix results suggest that gene flow occurred between most of the genotype groups. Of the four migration events modelled, the strongest migration weight involves gene flow between populations from Alberta to populations in North Dakota (genotype group I to group II). The other migration events inferred include migration from Spain2 to New Zealand (genotype group V to group II), Yukon to Ukraine and Russia (genotype group III to group I), and Spain1 to Iran2 (genotype group V to group I).

## Discussion

Our genomic study of 50 global populations of a worldwide, selfing weed revealed several important findings. First, populations were found to form six unique genotype groups or ecotypes showing very little sign of admixture and are therefore considered reproductively isolated. Second, three of the most prevalent of these ecotypes present in Europe are found most commonly throughout the introduced ranges, indicating that more than one aggressive genotype was introduced and established. Third, the multiple independent introduction events we detected and the distribution of ecotypes have links to colonial history. Finally, the global distribution of ecotypes was found to be restricted based on latitude (and therefore environmental factors related to latitude such as climatic). We therefore propose that multiple successful genotypes and prior adaptation to climate in the native ranges are likely reasons for the success of this global weed, such that a the right (prior adapted) ecotype must have arrived in an introduced location with a corresponding environment in order to persist and thrive.

## Global distribution of ecotypes

Our study is the first to sample populations of *Plantago major* so extensively across its global range and to provide an overview of the global distribution of the species' genotypic variation. The low genetic differentiation we found within populations worldwide is in keeping with what is expected for a highly selfing plant, and with the findings of past genetic studies of *P. major* (Mølgaard, 1976; Wolff et al., 1994; Wolff & Morgan-Richards, 1998; Morgan-Richards & Wolff, 1999; Barrett, Colautti, and Eckert, 2008). One of our main findings was that based on genetic distances and shared ancestry tests, individuals from across the global range cluster into six genotype groups rather than by population despite extraordinary geographic distances between them. However, genotypes from a given population were homogenous, such that you 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. A commensal species such as *P. major* that is known to flower continuously, to produce wind-borne pollen and thousands of propagules (seeds) from a single individual, has undoubtedly been moved around and is still being moved around especially given today's global trade networks (McNeill et al, 2011; Early et al., 2016). However, the lack of admixture or limited evidence of gene flow among genotype groups, even where populations occur in close proximity, suggests very low rates of out-crossing (Wolff, 1991), and provides evidence that the groups we identified are largely reproductively isolated.

The six distinct genotype groups may represent different taxonomic subspecies or varieties of *P. major* and/or intermediates between taxa that may have arisen by historical introgression. Two well-accepted subspecies, *P. major* subsp. *major* and *P. major* subsp. *intermedia*, have been the focus of past genetic and ecological studies. For example, Wolff et al. (1994) demonstrated low levels of polymorphism between populations of



both *P. major* subsp. *major* and *P. major* subsp. *intermedia*, but found notable genetic differentiation between subspecies. Furthermore, a comparison of Dutch and Scottish populations for the same two subspecies revealed that genetic divergence between subspecies was stronger than divergence within populations of the same subspecies between the two countries (Wolff & Morgan-Richards, 1998). These well-studied subspecies are sympatric and capable of interbreeding, yet genetic, ecological, and morphological differentiation between them was found to be pronounced enough for them to be resolved as separate species/subspecies likely because high selfing rates limits interbreeding (Mølgaard, 1976; Morgan-Richards & Wolff, 1999).

Even though our sampling was focused on populations that fit the morphological descriptions for subsp. *major*, it is possible that our genotype groups reflect the complexity of several intraspecific taxa—or hybrids of them resulting from historical introgression and being maintained due to reproductive isolation. El-Bakatoushi (2011) found populations in Egypt to be intermediate between two subspecies, and that some populations exhibited morphological characters resembling *P. major* subsp. *major* despite evidence of introgression with *P. major* subsp. *intermedia*. Because the goals of our work were not taxonomic in nature, and because very little is known about the numbers and global distribution of intraspecific taxa and the extent of their introgression, we consider the six groups identified in our study to be ecotypes of *P. major sensu lato*.

Three of the ecotypes are found to occur in both native and introduced ranges, and one unique ecotype (from the populations in Yukon and Alaska [in part]) is found in the introduced ranges. Selfing or clonal plants have been found to exhibit low genetic variation due to founder effects and lack of admixture in introduced populations compared to native populations (Zhang et al, 2010; Ferraro et al., 2015). The three most prevalent ecotypes 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 ecotype.

*Plantago major* is a prime example of a species that increased drastically in numbers and spread within its native range after the widespread clearing of land begun during the Neolithic, between 4900–2000 BCE as evidenced by palynological records (Godwin, 1944; Pysek, 1998; Preston et al., 2004; Kaplan et al., 2016). Weedy and aggressive ecotypes—which are commonly reported in successful introduced or invasive plants (i.e. Hawthorn, 1974; Taylor & Keller, 2007; Lambertini et al., 2012)—could have evolved and spread widely in Europe before being introduced to other continents. We show that not just one but at least three ecotypes have successfully colonized introduced ranges and may represent lineages that have evolved increased abilities to withstand human disturbance and human-altered habitats (Mølgaard, 1986; El-Bakatoushi et al., 2011; Hufbauer et al., 2012; Vigueira et al., 2013). The nature of our genomic data and lack of historical records make it impossible to determine when and where in the native range these ecotypes of *P. major* evolved and when their ranges expanded; however, as is hypothesized for other weedy and commensal European species, the rapid onset of agriculture and the profound modification of the native landscape in Europe around the Neolithic likely facilitated the spread (Godwin, 1944; Hicks, 1971; Preston et al., 2004; Brun, 2009). 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 (Dormontt et al., 2007; Martin et al., 2014).

## Multiple introduction events with colonial ties

Multiple source origins from the native range are inferred for this global weed. The three most widespread ecotypes found in Europe and the unique ecotype found in the Middle East (Iran1) were all detected in the introduced ranges we sampled. Therefore, at least 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, and at least one introduction event to each of South America, South Africa, Canary Islands, Hawai'i, Greenland and Iceland.

The introduction of more than one lineage from the native range has been reported for other introduced species. For example, Guo et al. (2017) found two introduced lineages in North America for the invasive

grass *Phragmites australis* (Cav.) Trin. ex Steud. – one lineage native to north-central Europe and Asia occurs in the Great Lakes region of North America, and another lineage from the Mediterranean region is naturalized in the Gulf Coast (USA) and in Central and South America. These patterns conspicuously match patterns we see with ecotypes of *Plantago major* from Europe and their distributions in North and South America.

Unfortunately, the very low genetic differentiation within each *Plantago major* ecotype makes it difficult to infer more precise origins for the introduced populations, despite our extensive sampling across native and introduced ranges. Furthermore, we cannot rule out that multiple introduction events of the same ecotype occurred historically, particularly given the extraordinary dispersal abilities and commensal nature of the plant. Our data does however support hypotheses that *P. major* followed early European colonists to new ranges (Samuelsen, 2000). The distribution of ecotypes and shared ancestry inferred between plants in native and introduced ranges reveals links to early colonial human movements and/or European colonies. For example, ecotypes found in southern Europe (Spain, Portugal) are also found in former Spanish or Portuguese colonies in Florida (USA), Peru, Chile, Brazil, and the Canary Islands, and could have followed colonial Spanish and/or Portuguese voyagers and settlers between the 15<sup>th</sup> and 18<sup>th</sup> centuries (Mancall, 2006). Plants in southern Africa could have been transported and introduced by Portuguese voyagers along their voyages to southern and eastern Africa.

Similarly, the two ecotypes prevalent in central and northern Europe and western Asia likely gave rise to the majority of populations in North America, as well as Greenland, Iceland, and New Zealand. Plants could have travelled with early colonial explorers/settlers from France and England in the 16<sup>th</sup> and 17<sup>th</sup> centuries where these countries had colonial power (Mancall, 2006). In spite of hypotheses that the Norse played a role in the movement of *Plantago major* in their travels across the northern Atlantic (Samuelsen, 2000), our data do not reveal any specific shared ancestry between plants in Scandinavia, Greenland and northeastern Canada (Newfoundland). The earliest known record of the plant in New Zealand is from 1832; however, the plant likely was introduced well before that (Webb et al., 1988).

Some populations of *Plantago major* do not show obvious colonial ties. 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 UK made stops in South Africa on 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 16<sup>th</sup> 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), 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 17<sup>th</sup> century (Mancall, 2006). The plants we sampled in Hawai'i also belong to the southern European ecotype 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 more northern European latitudes also arrived in Australia, South Africa, and Hawai'i, but that the climate or environment did not allow their persistence. DNA from historical herbarium specimens could help to shed light on this.

The ecotypes found in eastern Asia (South Korea and Japan) were not found within any of the introduced ranges we sampled. This could be an artefact of poor sampling across Asia, or could represent a true biological scenario in which *Plantago major* ecotypes from eastern Asia are not as successful in colonizing and spreading in other regions. At least two ecotypes are recognized in Japan – one restricted to sandy shores and brackish waters (which is represented in our sampling), the other considered introduced and weedy (pers. comm. M. Amano, 2020). The population sampled in Yukon, and some of the individuals from Alaska (USA) share common ancestry with eastern Asian populations, yet are found to have a unique lineage. These findings for populations we sampled in Yukon and Alaska (in part) provide support to the hypothesis that the northerly latitudes in North America consist of native populations of *P. major* that

predate the arrival of early Europeans (Hawthorn, 1974). Increasing sampling and further investigating 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 USA and Canada (for example by Russian colonists or by earlier human migrations from Beringia [Erlandson & Braje, 2011]).

One unexpected finding was that the population sampled in California was found to differ from plants elsewhere in North America, and shared ancestry with the population from Iran<sup>1</sup>. However, other weedy species with native distributions in the Middle East have been introduced to and have become well established in California (i.e. Ortiz et al., 2008; Meyers & Liston, 2008). The relationships among such populations is most likely due to similarities in climate. Pathways of migration between these biogeographic regions remain poorly studied but the dispersal of weedy species 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 also cannot rule out non-anthropogenic introduction events, particularly since the genus *Plantago* is well adapted for long-distance dispersal (Meyers & Liston, 2008; Iwanycki Ahlstrand et al., 2019). In a study by Morgan-Richards & Wolff (1999), populations sampled in Los Angeles (California, USA) and Trinidad were genetically and morphologically determined to be *P. major* subsp. *intermedia* (syn. *P. intermedia* DC.). Although it was not the goal of their study to characterize the global distribution for each of the subspecies or retrace origins of introduced populations, their sampling demonstrates that *P. major* subsp. *intermedia* is found in some regions of the introduced range, such as the more southerly latitudes in North America and the Caribbean. More intensive population sampling around the Middle East and California would be needed to further narrow down the native origins and further resolve ancestry and distribution for this Mediterranean ecotype of *P. major* and other *Plantago* species.

## Prior adaptation is key to global success

*Plantago major* is a classic example of a widely-introduced species for which its successful colonisation and spread in new lands is not compromised by low genetic variation or the lack of sexual recombination (Beaumont et al., 2009; Dormontt et al., 2014). Based on the distribution of ecotypes that have spread to the introduced parts of the range, and the division seen along the 35–40° N latitude range, we propose that prior adaptation to climate in multiple successful ecotypes may explain the distribution and persistence of this global weed.

When introduced species successfully colonize new ranges, they either do so by quickly adapting to new conditions, or they arrive with prior adaptations that arose in their ancestral native range before introduction (Rey et al., 2012; Jackson et al., 2015; Barrett, 2015; Bock et al., 2015). The pattern we see in the global ecotype distribution where the majority of the global populations we sampled suggests that the successful colonisation of new ranges by *P. major* is dependent on plants having a prior adaptation to climate before arriving to new regions worldwide. Such prior adaptation to climate has been noted for other species with successful colonists, i.e. in plants (Sexton et al., 2002; Dlugosch and Parker, 2007; Schlaepfer et al., 2010; Rosche et al., 2018), birds (Jackson et al., 2015), and beetles (Vahsen et al., 2018). Although climatic conditions are quite different between regions occupied by populations belonging to the same ecotype (i.e., semi-tropical Florida and southern Spain), it is possible that further climatic niche shifts occurred after introduction, and although such shifts are thought to be rare in terrestrial plants (Petitpierre, 2012), it has been demonstrated for other species (i.e. Jackson et al., 2015). Alternatively, prior adapted lineages could have evolved tolerance to a large climatic amplitude, and therefore climatic and environmental variation deserves further investigation.

Although very little is known about the ecological habitats of *P. major* in pre-colonial times, the ecotypes in its native range may have evolved anthropogenically-induced adaptations to invade (AIAI; Theoharides et al., 2007; Rey et al. 2012; Hufbauer et al., 2012). Plants like *P. major* that occupy a broad range of habitat conditions in their native range, including those that are human-made and disturbed habitats, can naturalize more easily in new regions (Hufbauer et al., 2012). Human activity and disturbance have created conditions

that are now more similar across broad geographic regions, and this homogenization of the environment likely facilitated the colonisation and successful spread of disturbance tolerant species like *P. major* (van Dijk & van Delden, 1981; Estoup & Guillemaud, 2010; Kalusova et al, 2017).

Asexual plants have been found to change just as often and as fast as do sexual plants when introduced to a new range (Dalrymple et al, 2015). The extraordinary phenotypic plasticity known in *P. major* is likely important for ecotypes coping with a high level of environmental and climate heterogeneity (Warwick & Briggs, 1980; Samuelsen, 2000). *Plantago major* can vary extensively in morphology even within populations, and therefore high phenotypic variation for the species is not related to genetic diversity at least within a population level, though further studies are needed (Mølgaard, 1976; Lotz and Blom, 1986; Wolff, 1991; Iwanycki Ahlstrand et al., 2018). Wide environmental tolerance and an ability to grow in a multitude of climatic and edaphic conditions as a result of well-developed phenotypic plasticity, makes *Plantago major* a classic example of a weed that possesses the “general purpose genotype strategy” as coined by Baker (1965) – in this case, it appears that as few as three general-purpose genotypes or ecotypes may have been involved in the global success of the common plantain.

## Conclusions

Population genomic approaches can provide useful data where historical records are lacking in helping to retrace migration routes of historical introductions of plants, such as *Plantago major*, and improve our understanding of the role of genomic variation in explaining success of globally distributed species. Our findings show that beyond human-mediated migration, factors such as prior adaptation to climate and anthropogenic disturbance may be key to success for a worldwide weed. That is, a genotype introduced into a new range with environmental conditions for which it is already adapted is more likely to succeed. This suggests that even if globalization creates more opportunities for species invasions, not all species with invasive potential will be as likely to succeed in a new area. The six ecotypes 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 plant species and mediate phenotypic adaptation to local conditions (Bragg et al. 2015). Introduced plants that exhibit high phenotypic plasticity are hypothesized to perform better under changing climatic conditions, despite low genetic variation and sexual recombination. Therefore, our findings can be used to advance our ecological and evolutionary understanding of successful invasions, and contribute to more accurately predicting species responses to global change (Chapman et al., 2016; Klonner et al., 2017).

## Acknowledgements

This work was supported by the Marie Curie Actions of the 7<sup>th</sup> European Community Framework Programme: FP7/2007-2013/, REA grant agreement n° 606895-MedPlant to NR, NIA, KMS, VM, PS, MX, NZ, and OG. We would like to thank ADNid, Montpellier (France) for DNA extractions, and The Global Diversity Facility (GDF) of Cornell University for genotyping-by-sequencing.

## References

- Baker, H. G. 1965. Characteristics and Modes of Origin of Weeds. In: Genetics of Colonizing Species, Academic Press, New York, 147-172
- Baker, H. G. 1974. The evolution of weeds, *Annual Review Ecological Systems* , **5** , 1–24

- Banks, N. C., Paini, D. R., Bayliss, K. L., & Hodda, M. 2015. The role of global trade and transport network topology in the human-mediated dispersal of alien species. *Ecology Letters*, **18** (2), 188-199. doi:10.1111/ele.12397
- Barrett, S. C. H. 2015. Foundations of invasion genetics: the Baker and Stebbins legacy. *Molecular Ecology*, **24** (9), 1927-1941. doi:10.1111/mec.13014
- Barrett, S. C. H., Colautti, R. I., & Eckert, C. G. 2008. Plant reproductive systems and evolution during biological invasion. *Molecular Ecology*, **17** (1), 373-383. doi:10.1111/j.1365-294X.2007.03503.x
- Beaumont, L. J., Gallagher, R. V., Thuiller, W., Downey, P. O., Leishman, M. R., & Hughes, L. 2009. Different climatic envelopes among invasive populations may lead to underestimations of current and future biological invasions. *Diversity and Distributions*, **15** (3), 409-420. doi:10.1111/j.1472-4642.2008.00547.x
- Bennett, B. C. & Prance, G. T. 2000. Introduced plants in the indigenous pharmacopoeia of Northern South America. *Economic Botany*, **54**, 90-102.
- Bock, D. G., Caseys, C., Cousens, R. D., Hahn, M. A., Heredia, S. M., Hubner, S., Turner, K. G., Whitney, K. D. & Rieseberg, L. H. 2015. What we still don't know about invasion genetics. *Molecular Ecology*, **24**, 2277-2297.
- Bragg, J. G., Supple, M. A., Andrew, R. L., Borevitz J. O. 2015. Genomic variation across landscapes: insights and applications. *New Phytologist* **207**, 953-967.
- Canavan, K., Paterson, I. D., Hill, M. P. 2017 Exploring the Origin and Genetic Diversity of the Giant Reed, *Arundo donax* in South Africa. *Invasive Plant Science and Management* **10**, 53-60.
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., Cresko, W. A. 2013. Stacks: an analysis tool set for population genomics. *Molecular Ecology* **22**, 3124-3140.
- Chapman, D., Purse, B. V., Roy, H. E., & Bullock, J. M. 2017. Global trade networks determine the distribution of invasive non-native species. *Global Ecology and Biogeography*, **26** (8), 907-917. doi:10.1111/geb.12599
- Chapman, D. S., Makra, L., Albertini, R., Bonini, M., Paldy, A., Rodinkova, V., Bullock, J. M. 2016. Modelling the introduction and spread of non-native species: international trade and climate change drive ragweed invasion. *Global Change Biology*, **22** (9), 3067-3079. doi:10.1111/gcb.13220
- Cornille A., Salcedo A., Kryvokhyzha D., Glemin, S., Holm, K., Wright, S. I. & Lascoux, M. 2016. Genomic signature of successful colonization of Eurasia by the allopolyploid shepherd's purse (*Capsella bursa-pastoris*). *Molecular Ecology* **25**, 616-629.
- Dalrymple R. L., Buswell J. M., Moles A. T. 2015. Asexual plants change just as often and just as fast as do sexual plants when introduced to a new range. *Oikos* **124**, 196-205.
- Dlugosch, K. M., & 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? *Molecular Ecology*, **16** (20), 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. & Lowe, A. J. 2014. Genetic Bottlenecks in Time and Space: Reconstructing Invasions from Contemporary and Historical Collections *PloS One*, **9**
- Dullinger, I., Wessely, J., Bossdorf, O., Dawson, W., Essl, F., Gatringer, A., Klöner, G., Kreft, H., Kuttner, M., Moser, D., Pergl, J., Pysek, P., Thuiller, W., van Kleunen, M., Weigelt, P., Winter, M. & Dullinger, S. 2017. Climate change will increase the naturalization risk from garden plants in Europe. *Global Ecology and Biogeography*, **26**, 43-53
- Early R., Bradley B. A., Dukes J. S., Lawler, J. J., Olden, J. D., Blumenthal, D. M., Gonzalez, P., Grosholz, E. D., Ibanez, I., Miller, L. P., Sorte, C. J. B. & Tatem, A. J. 2016. Global threats from invasive alien

- species in the twenty-first century and national response capacities. *Nature Communications* **7** : 12485. doi: 10.1038/ncomms12485.
- El-Bakatoushi R. 2011. Introgressive hybridization between *Plantago major* L. taxa. *Flora* **206** , 1045-1051.
- Ellegren H. 2014. Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution* **29** , 51-63.
- Elshire, R. J., Glaubitz J. C., Sun Q., Poland J. A., Kawamoto K., Buckler E. S., Mitchell S. E. 2011. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *Plos One* 6(5).
- Erlandson, J. M., Braje T. J. 2011. From Asia to the Americas by boat? Paleogeography, paleoecology, and stemmed points of the northwest Pacific. *Quaternary International* **239** , 28-37.
- Estoup, A. & Guillemaud, T. 2010. Reconstructing routes of invasion using genetic data: why, how and so what? *Molecular Ecology* **19** , 4113-4130.
- 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., & Rodriguez-Echeverria, S. 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. *Molecular Ecology*, **24** (9), 2143-2155. doi:10.1111/mec.13056
- Fumagalli, M., Vieira F. G., Linderoth T., Nielsen R. 2014. ngsTools: methods for population genetics analyses from next-generation sequencing data. *Bioinformatics* 30(10): 1486-1487.
- Gaskin, J. F., Schwarzlaender, M., Kinter, C. L., Smith, J. F., & Novak, S. J. 2013. Propagule pressure, genetic structure, and geographic origins of *Chondrilla juncea* (Asteraceae): An apomictic invader on three continents. *American Journal of Botany*, **100** (9), 1871-1882. doi:10.3732/ajb.1200621
- Godwin, H. 1944. Neolithic forest clearance. *Nature* **153** , 511-512.
- Guo, Q. F., Iannone, B. V., Nunez-Mir, G. C., Potter, K. M., Oswalt, C. M., & Fei, S. L. 2017. Species pool, human population, and global versus regional invasion patterns. *Landscape Ecology*, **32** (2), 229-238. doi:10.1007/s10980-016-0475-6
- GBIF.org (21 January 2020) GBIF Occurrence Download <https://doi.org/10.15468/dl.fmh0zt>
- Hawthorn, W. R. 1974. Biology of Canadian weeds .4. *Plantago major* and *Plantago rugelii* . *Canadian Journal of Plant Science* **54** , 383-396.
- Hejda, M., Chytry, M., Pergl, J., & Pysek, 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 and Distributions*, **21** (3), 312-321. doi:10.1111/ddi.12269
- Herten, K., Hestand, M. S., Vermeesch, J. R., Van Houdt, J. K. J. 2015. GBSX: a toolkit for experimental design and demultiplexing genotyping by sequencing experiments. *BMC Bioinformatics* **16** .
- Hicks, S. P. 1971. Pollen-analytical evidence for effect of prehistoric agriculture on vegetation of north Derbyshire. *New Phytologist* **70** , 647-667.
- Hodkinson, D. J., Thompson, K. 1997. Plant dispersal: the role of man. *Journal of Applied Ecology* **34** , 1484-1496.
- Holm, L., Pancho, J., Herberger, J., Plucknett, D. 1979. A Geographical Atlas of World Weeds. New York, USA: John Wiley & Sons.
- Hufbauer, R.A., Facon, B., Ravigne, V., Turgeon, J., Foucaud, J., Lee, C. E., Rey, O. & Estoup, A. 2012. Anthropogenically induced adaptation to invade (AIAI): contemporary adaptation to human-altered habitats within the native range can promote invasions. *Evolutionary Applications* **5** , 89-101.

- Hulme, P. E. 2009. Trade, transport and trouble: managing invasive species pathways in an era of globalization. *Journal of Applied Ecology*, 46 (1), 10-18. doi:10.1111/j.1365-2664.2008.01600.x
- Iwanycki Ahlstrand, N., Markussen, B., Havskov Reghev, N., Eiríksson, F., Thorsteinsdóttir, M., Rønsted, N., Barnes, C. J. 2018. Untargeted metabolic profiling reveals geography as a stronger predictor of metabolic phenotypes than environment or herbivore damage of a cosmopolitan weed. *Ecology and Evolution* , 6812-6826.
- Iwanycki Ahlstrand, N., Verstraete, B., Hassemer, G., Dunbar-Co, S., Hoggard, R., Meudt, H.M., Rønsted, N. 2019. Ancestral range reconstruction of remote oceanic island species of *Plantago*(Plantaginaceae) reveals differing scales of modes of dispersal.*Journal of Biogeography* , 46, 706-722.
- Iwanycki Ahlstrand, N., Gopalakrishnan, S., Vieira, F. G., Meudt, H. M., Dunbar-Co, S., Rothfels, C. J., Martinez-Swatson, K. A., Maldonado, C., Hassemer, G., Shipunov, A., Bowers, D., Gardner, E., Høegh, K., Xu, M., Ghorbani, A., Amano, M., Schultz, T.M., Grace, O. M., Pringle, J. P., Bishop, M., Manzanilla, V., Cotrim, H. C., Blaney, S., Zubov, D., Choi, H.-K., Yesil, Y., Bennett, B., Vimolmangkang, S., El-Seedi, H. R., Staub, P. O., Li, Z., Boldbaatar, D., Hislop, M., Caddy, L., J., Muasya, A. M., Saslis-Lagoudakis, C. H., Gilbert, M. T. P., Zerega, N. J. C. and Rønsted, N. 2020. Travel tales of a worldwide weed: genomic signatures reveal colonial trade routes and prior adaptation to climate are key to the success of *Plantago major* . Genotyping-by-sequencing (GBS) data: demultiplexed reads. Sequence Read Archive (SRA) <https://www.ncbi.nlm.nih.gov/sra>; BioProject ID: PRJNA641850.
- Jackson, H., Strubbe, D., Tollington, S., Prys-Jones, R., Matthysen, E. & Groombridge, J. J. 2015. Ancestral origins and invasion pathways in a globally invasive bird correlate with climate and influences from bird trade. *Molecular Ecology* **24** , 4269-4285.
- Josslyn, J. 1672. New-Englands Rarities, Discovered in Birds, Beasts, Fishes, Serpents, and Plants of that Country. GD Widdowes, London (accessed on-line biodiversitylibrary.org, August 2017).
- Kalusova, V., Chytry, M., van Kleunen, M., Mucina, L., Dawson, W., Essl, F., Pysek, P. 2017. Naturalization of European plants on other continents: The role of donor habitats. *Proceedings of the National Academy of Sciences of the United States of America*, 114 (52), 13756-13761. doi:10.1073/pnas.1705487114
- Kaplan, J. O., Pfeiffer, M., Kolen, J. C. A., Davis, B. A. S. 2016. Large Scale Anthropogenic Reduction of Forest Cover in Last Glacial Maximum Europe. *PloS One* **11** .
- Klonner, G., Dullinger, I., Wessely, J., Bossdorf, O., Carboni, M., Dawson, W., Essl, F., Gattringer, A., Haeuser, E., van Kleunen, M., Kreft, H., Moser, D., Pergl, J., Pysek, P., Thuiller, W., Weigelt, P., Winter, M., Dullinger, S.. 2017. Will climate change increase hybridization risk between potential plant invaders and their congeners in Europe? *Diversity and Distributions* **23** , 934-943.
- Korneliussen, T.S., Albrechtsen, A., Nielsen, R. 2014. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15** , 356.
- Kreitschitz A., Kovalev, A., Gorb, S. N. 2016. "Sticky invasion" - the physical properties of *Plantago lanceolata* L. seed mucilage.*Beilstein Journal of Nanotechnology* **7** , 1918-1927
- Kuiper PJC, & Bos M (Eds.) 1992. Ecological studies analysis and synthesis, vol. 89. *Plantago: A multidisciplinary study* (379 pp). Berlin, Germany: Springer Verlag.
- Lambertini, C., Mendelssohn, I. A., Gustafsson, M. H. G., Olesen, B., Riis, T., Sorrell, B. K., & Brix, H. 2012. Tracing the origin of Gulf Coast *Phragmites* (Poaceae): a story of long-distance dispersal and hybridization. *American Journal of Botany*, 99 (3), 538-551. doi:10.3732/ajb.1100396
- Li, H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatic s*, **1** :27, 2987-93.

- Li, H. and Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* , **25** :1754-60.
- Lotz, L. A. P., Blom, C. 1986. Plasticity in life-history traits of *Plantago major* L ssp *pleiosperma* Pilger. *Oecologia***69** , 25-30.
- Lu F., Lipka A. E., Glaubitz J., Elshire, R., Cherney, J. H., Casler, M. D., Buckler, E. S. & Costich, D. E. 2013. Switchgrass Genomic Diversity, Ploidy, and Evolution: Novel Insights from a Network-Based SNP Discovery Protocol. *PloS Genetics* **9** .
- Mancall, P.C., 2006. Travel Narratives from the Age of Discovery: an anthology. Oxford University Press US, 413 pp. ISBN 0195155971
- Mack, R. N. (2003) Plant naturalizations and invasions in the eastern United States: 1634-1860. *Annals of the Missouri Botanical Garden***90** , 77-90.
- Mack, R.N., Simberloff, D., Lonsdale, W. M., Evans, H., Clout, M. & Bazzaz, F. A. 2000. Biotic invasions: Causes, epidemiology, global consequences, and control. *Ecological Applications* **10** , 689-710.
- Martin, M. D., Zimmer, E. A., Olsen, M. T., Foote, A.D., Gilbert, M. T. P., & Brush, G. S. 2014. Herbarium specimens reveal a historical shift in phylogeographic structure of common ragweed during native range disturbance. *Molecular Ecology*, *23* (7), 1701-1716. doi:10.1111/mec.12675
- McNeill M., Phillips, C., Young, S., Shah, F., Aalders, L., Bell, N., Gerard, E. & Littlejohn, R. 2011. Transportation of nonindigenous species via soil on international aircraft passengers' footwear. *Biological Invasions* **13** , 2799-2815.
- Meyers SC & Liston A (2008) The biogeography of *Plantago ovata*Forssk. (Plantaginaceae). *International Journal of Plant Sciences***169** , 954-962.
- Morgan-Richards, M., and Wolff, K. 1999. Genetic structure and differentiation of *Plantago major* reveals a pair of sympatric sister species. *Molecular Ecology* **8** , 1027-1036.
- Mølgaard, P. 1976. *Plantago major* ssp. *major* and ssp.*pleiosperma* . Morphology, biology and ecology in Denmark. *Botany Tidsskrift Bd* , **71** , 31-56.
- Narum, S. R., Buerkle, C. A., Davey, J. W., Miller, M. R., & Hohenlohe, P. A. 2013. Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, **22** (11), 2841-2847. doi:10.1111/mec.12350
- Nielsen, R., Paul, J. S., Albrechtsen, A., Song, Y. S. 2013. Genotype and SNP calling from next-generation sequencing data. *Nature Reviews. Genetics* , **12** , 443-451.
- Ortiz, M. A., Tremetsberger, K., Terrab, A., Stuessy, T. F., Garcia-Castano, J. L., Urtubey, E., Baeza, C. M., Ruas, C. F., Gibbs, P. E. & Talavera, S. 2008. Phylogeography of the invasive weed *Hypochaeris radicata* (Asteraceae): from Moroccan origin to worldwide introduced populations. *Molecular Ecology* **17** , 3654-3667.
- Palmer, C. 2004. *Plantago* spp. and *Bidens* spp.: A case study of change in Hawai'ian herbal medicine. *Journal of Ethnobiology***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.
- Peruzzi, L. & Passalacqua, N. G. 2003. *Plantago sinuata* Lam. (Plantaginaceae), a misinterpreted unit, typical of moist places. Morphological and karyological evidence. *Journal of Plant Taxonomy and Geography*, **58** , 441-450.
- Petitpierre, B., Kueffer, C., Broennimann, O., Randin, C., Daehler, C. & Guisan, A. 2012. Climatic Niche Shifts Are Rare Among Terrestrial Plant Invaders. *Science* **335** , 1344-1348.



- Pickrell, J.K., Pritchard, J.K. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLOS Genetics* **8** : e1002967.
- Pilger, R.K.F., 1937. Plantaginaceae. In: In: Engler, H.G.A., Diels, F.L.E. (Eds.), *Das Pflanzenreich*, vol. 102. W. Engelmann, Leipzig, pp. 466.
- POWO. 2019. Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew. Published on the Internet; <http://www.plantsoftheworldonline.org/> Retrieved 10 April 2020.
- Preston, C. D., Pearman, D. A., Hall, A. R. 2004. Archaeophytes in Britain. *Botanical Journal of the Linnean Society* **145** , 257-294.
- Puckett, E.E., Park, J., Combs, M., Blum, M. J., Bryant, J. E., Caccone, A., Munshi-South, J. 2016. Global population divergence and admixture of the brown rat (*Rattus norvegicus* ). *Proceedings of the Royal Society B-Biological Sciences*, **283** (1841). doi:10.1098/rspb.2016.1762
- Puckett, E. E., Magnussen, E., Khlyap, L. A., Strand, T. M., Lundkvist, A., & Munshi-South, J. (2020). Genomic analyses reveal three independent introductions of the invasive brown rat (*Rattus norvegicus* ) to the Faroe Islands. *Heredity*, **124** (1), 15-27. doi:10.1038/s41437-019-0255-6
- Pysek, P. 1998. Alien and native species in Central European urban floras: a quantitative comparison. *Journal of Biogeography* **25** , 155-163.
- Pysek, P., Jarosik, V., Pergl, J., *et al.* 2009. The global invasion success of Central European plants is related to distribution characteristics in their native range and species traits. *Diversity and Distributions* **15** , 891-903.
- Pysek, P., Manceur, A. M., Alba, C., *et al.* 2015. Naturalization of central European plants in North America: species traits, habitats, propagule pressure, residence time. *Ecology* **96** , 762-774.
- Rahn, K. 1996. A phylogenetic study of the Plantaginaceae. *Botanical Journal of the Linnean Society* **120** , 145-198.
- Rey, O., Estoup, A., Vonshak, M., Loiseau, A., Blanchet, S., Calcaterra, L., Chifflet, L., Rossi, J.-P., Kergoat, G., Foucard, J., Orivel, J., Laponche, M., Schultz, T., Facon, B. 2012. Where do adaptive shifts occur during invasion? A multidisciplinary approach to unravelling cold adaptation in a tropical ant species invading the Mediterranean area. *Ecology Letters* **15** , 1266-1275.
- Rosche, C., Hensen, I., & Lachmuth, S. 2018. Local pre-adaptation to disturbance and inbreeding-environment interactions affect colonisation abilities of diploid and tetraploid *Centaurea stoebe* . *Plant Biology*, **20** (1), 75-84. doi:10.1111/plb.12628
- Rousseau, J. 1966. Movements of plants under the influence of man. In R.L. Taylor and R.A. Ludwig, eds. *The evolution of Canada's flora*. University of Toronto Press, Toronto, Ontario. 137pp.
- Samuelsen, A. B. 2000. The traditional uses, chemical constituents and biological activities of *Plantago major* L. A review. *Journal of Ethnopharmacology* , **71** , 1-21. [https://doi.org/10.1016/S0378-8741\(00\)00212-9](https://doi.org/10.1016/S0378-8741(00)00212-9)
- Schlaepfer, D. R., Glattli, M., Fischer, M., & van Kleunen, M. 2010. A multi-species experiment in their native range indicates pre-adaptation of invasive alien plant species. *New Phytologist*, **185** (4), 1087-1099. doi:10.1111/j.1469-8137.2009.03114.x
- Schubert, M., Ermini, L., Sarkissian, C.D., Jonsson, H., Ginolhac, A., Schaefer, R., Martin, M. D., Fernandez, R., Kircher, M., McCue, M., Willerslev, E. & Orlando, L. 2014. Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nature Protocols* **9** , 1056-1082.
- Schubert, M., Lindgreen, S., and Orlando, L. 2016. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Research Notes*, **12**;9(1):88

- Seebens H., Essl F., Dawson W., Fuentes, N., Moser, D., Pergl, J., Pysek, P., van Kleunen, M., Weber, E., Winter, M. & Blasius, B. 2015. Global trade will accelerate plant invasions in emerging economies under climate change. *Global Change Biology* **21** , 4128-4140.
- Sexton, J. P., McKay, J. K., & Sala, A. 2002. Plasticity and genetic diversity may allow saltcedar to invade cold climates in North America. *Ecological Applications*, *12* (6), 1652-1660. doi:10.2307/3099929
- Skotte, L., Korneliussen, T.S., Albrechtsen, A. 2013. Estimating individual admixture proportions from next generation sequencing data. *Genetics* , **195** : 693-702.
- Stepp, J. R. & Moermann, D. E. 2000. The importance of weeds in ethnopharmacology. *Journal of Ethnopharmacology* , **75** , 19-23.
- Taylor, D. R., & Keller, S. R. 2007. Historical range expansion determines the phylogenetic diversity introduced during contemporary species invasion. *Evolution*, *61* (2), 334-345. doi:10.1111/j.1558-5646.2007.00037.x
- Theoharides, K. A., & Dukes, J. S. 2007. Plant invasion across space and time: factors affecting nonindigenous species success during four stages of invasion. *New Phytologist*, *176* (2), 256-273. doi:10.1111/j.1469-8137.2007.02207.x
- Turbelin, A. J., Malamud, B. D., & Francis, R. A. 2017. Mapping the global state of invasive alien species: patterns of invasion and policy responses. *Global Ecology and Biogeography*, *26* (1), 78-92. doi:10.1111/geb.12517
- Vahsen, M. L., Shea, K., Hovis, C. L., Teller, B. J., & Hufbauer, R. A. 2018. Prior adaptation, diversity, and introduction frequency mediate the positive relationship between propagule pressure and the initial success of founding populations. *Biological Invasions*, *20* (9), 2451-2459. doi:10.1007/s10530-018-1713-4
- van Dijk, H., & Wolff, K. 1992. Allozyme variation and genetic structure in *Plantago* species. In *Kuiper, P. J. C. and M. Bos* (pp. 184-192).
- van Dijk, H. 1984. Genetic variability in *Plantago* species in relation to their ecology. 2. Quantitative characters and allozyme loci in *P. major* . *Theoretical & Applied Genetics* **68** , 43-52.
- van Dijk, H., & van Delden, W. 1981. Genetic variability in *Plantago* species in relation to their ecology. *Theoretical Applications in Genetics* , **60** , 285-290.
- van Dijk, H., Wolff, K. 1992. Allozyme variation and genetic structure in *Plantago* species. *Kuiper, P. J. C. and M. Bos* , eds, *Plantago: A Multidisciplinary Study*. Ecological Studies vol. 89. Berlin, Germany: Springer Verlag. pp 184-192.
- van Kleunen, M., Dawson, W., Essl, F., Pergl, J., Winter, M., Weber, E., Pysek, P. 2015. Global exchange and accumulation of non-native plants. *Nature*, *525* (7567), 100-+. doi:10.1038/nature14910
- Vieira, F. G., Lassalle, F., Korneliussen, T. S., Fumgalli, M. 2015. Improving the estimation of genetic distances from Next-Generation Sequencing data. *Biological Journal of the Linnean Society* **117** , 139-149.
- Vigueira, C. C., Olsen, K. M., Caicedo, A. L. 2013. The red queen in the corn: agricultural weeds as models of rapid adaptive evolution. *Heredity* **110** , 303-311.
- Warwick, S. I. & Briggs, D. 1979. The genealogy of lawn weeds III. Cultivation experiments with *Achillea millefolium* L., *Bellis perennis* L., *Plantago lanceolata* L., *Plantago major* L. and *Prunella vulgaris* L. collected from lawns and contrasting grassland habitats. *New Phytologist* , **83** :509-536.
- Webb, C. J., Sykes, W. R., and P. J- Garnock-Jones. 1988. Flora of New Zealand, volume IV, Naturalised Pteridophytes, Gynnosperms, Dicotyledons. Botany Division, D.S.I.R., Christchurch, New Zealand. 1365 pp.

- Wilson, J. R. U., Dormontt, E. E., Prentis, P. J., Lowe, A. J., & Richardson, D. M. 2009. Something in the way you move: dispersal pathways affect invasion success. *Trends in Ecology & Evolution*, *24* (3), 136-144. doi:10.1016/j.tree.2008.10.007
- Wolff, K. 1991. genetic-analysis of morphological variability in 3*Plantago* species with different mating systems. *Theoretical and Applied Genetics* **81** , 111-118.
- Wolff, K., Morgan-Richards, M. 1998. PCR markers distinguish *Plantago major* s subspecies. *Theoretical and Applied Genetics* **96** , 282-286.
- Wolff, K., Rogstad, S. H., Schaal, B. A. 1994. population and species variation of minisatellite DNA in *Plantago* . *Theoretical and Applied Genetics* **87** , 733-740.
- Young, J.A., and Evans, R. A. 1973. Mucilaginous seed coats. *Weed Science*, **21** :52-24.
- Zhang, Y. Y., Zhang, D. Y., & Barrett, S. C. H. 2010. Genetic uniformity characterizes the invasive spread of water hyacinth (*Eichhornia crassipes* ), a clonal aquatic plant. *Molecular Ecology*, *19* (9), 1774-1786. doi:10.1111/j.1365-294X.2010.04609.x
- Zhu, B. R., Barrett, S. C. H., Zhang, D. Y., & Liao, W. J. 2017. Invasion genetics of *Senecio vulgaris* : loss of genetic diversity characterizes the invasion of a selfing annual, despite multiple introductions. *Biological Invasions*, *19* (1), 255-267. doi:10.1007/s10530-016-1277-0

## Data Accessibility

Upon acceptance, sequences will be deposited in the short read archive (<https://www.ncbi.nlm.nih.gov/sra>, BioProject ID: PRJNA641850; Iwanycki Ahlstrand et al., 2020). A summary of our bioinformatics pipeline including codes will be available on Github (<https://github.com/shyamsg/plantago>).

## Authors' Contributions

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

NIA, HTM, SDC, CJR, KMS, CM, GH, AS, DB, EG, KH, MX, AG, MA, TMS, OMG, JSP, MB, VM, HCC, SB, DZ, HKC, YY, BB, SV, HES, PS, ZL, DB, NH, RA, LJC, MM, CHSL, TPG, & NZ performed sampling of populations in different localities. NIA performed DNA extractions. NIA and SG processed and analysed sequence data and made the figures. NIA wrote the manuscript with SG, and NR. All authors read and commented on the manuscript and approved the final version.

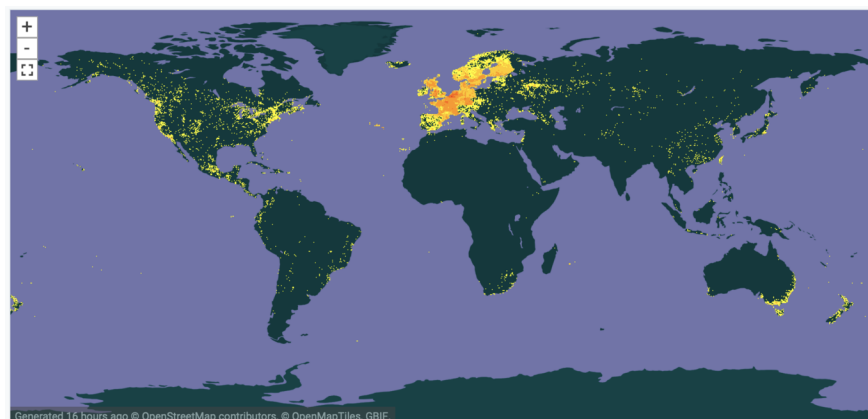
**Tables and Figures** (with captions)

**Tables and Figures** (with captions)

**Table 1** . *Plantago major* populations sampled for genotyping.

No.	Species	Population Name	Country Code	Site Name	Collector
1	<i>Plantago major</i> L.	Alaska	AL	Fairbanks, Alaska, USA	A Shipu
2	<i>Plantago major</i> L.	Alberta	AB	Edmonton, Alberta, Canada	CJ Roth
3	<i>Plantago major</i> L.	Brazil	BR	Urubici, Brazil	G Hasse
4	<i>Plantago major</i> L.	California	CA	Berkeley, California, USA	CJ Roth
5	<i>Plantago major</i> L.	Chicago	CG	Chicago, USA	NJC Ze
6	<i>Plantago major</i> L.	Chile	CL	Antofagasta, Chile	G Hasse
7	<i>Plantago major</i> L.	Colorado	CO	Boulder, Colorado, USA	D Bowe
8	<i>Plantago major</i> L.	Denmark	DK	Øjesø, Jutland, Denmark	N Iwany

No.	Species	Population Name	Country Code	Site Name	Collector
9	<i>Plantago major</i> L.	Egypt	EG	Cairo, Egypt	H El-se
10	<i>Plantago major</i> L.	England	EN	Reading, England	N Iwany
11	<i>Plantago major</i> L.	Estonia	ET	Talinn, Estonia	N Iwany
12	<i>Plantago major</i> L.	Finland	FI	Helsinki, Finland	N Iwany
13	<i>Plantago major</i> L.	Florida	FL	Florida, USA	E Gardi
14	<i>Plantago major</i> L.	France1	FR1	Haute-Savoie, France	K Marti
15	<i>Plantago major</i> L.	France2	FR2	Paulhaguet, France	TMP G
16	<i>Plantago major</i> L.	Gibraltar	GB	Gibraltar, UK	N Iwany
17	<i>Plantago major</i> L.	GranCanaria	GC	Gran Canaria, Canary Islands	N Iwany
18	<i>Plantago major</i> L.	Greece	GR	Crete, Greece	K Marti
19	<i>Plantago major</i> L.	Greenland	GL	Qaqortoq, Greenland	K Høegl
20	<i>Plantago major</i> L.	Hawai'i	HA	Molokai, Hawai'i	S Dunba
21	<i>Plantago major</i> L.	Iceland	IC	Reykjavik, Iceland	M Xu
22	<i>Plantago major</i> L.	Iran1	IR1	Hamedan, Iran	A Ghorl
23	<i>Plantago major</i> L.	Iran2	IR2	West Azarbaijan, Iran	A Ghorl
24	<i>Plantago major</i> L.	Ireland	IE	Dublin, Ireland	N Iwany
25	<i>Plantago major</i> L.	Italy	IT	Bari, Italy	K Marti
26	<i>Plantago major</i> L.	Japan	JA	Chiba, Japan	M Aman
27	<i>Plantago major</i> L.	Melbourne	ME	Melbourne, Australia	T Schult
28	<i>Plantago major</i> L.	Morocco	MO	Amegdale, Morocco	K Marti
29	<i>Plantago major</i> L.	Netherlands	NE	Oentsjerk, Netherlands	OM Gra
30	<i>Plantago major</i> L.	NewBrunswick	NB	St. John, New Brunswick, Canada	S Blane
31	<i>Plantago major</i> L.	NewZealand	NZ	Wellington, New Zealand	HM Me
32	<i>Plantago major</i> L.	Newfoundland	NF	St Jonh's, Newfoundland, Canada	M Bisho
33	<i>Plantago major</i> L.	NorthDakota	ND	North Dakota, USA	A Shipu
34	<i>Plantago major</i> L.	Norway	NO	Oslo, Norway	V Manz
35	<i>Plantago major</i> L.	Ontario	ON	Hamilton, Ontario, Canada	JS Pring
36	<i>Plantago major</i> L.	Perth	PR	Perth, Australia	M Hislo
37	<i>Plantago major</i> L.	Peru	PE	Oxapampa, Peru	G Hasse
38	<i>Plantago major</i> L.	Portugal	PT	Ribeira do Casal Fundeiro, Besteiras, Portugal	H Cotrin
39	<i>Plantago major</i> L.	Russia	RU	Kislovodsk, North Caucasus, Russian Federation	D Zubov
40	<i>Plantago major</i> L.	SouthAfrica	ZA	Byrne, South Africa	CJ Roth
41	<i>Plantago major</i> L.	SouthKorea	KR	Gyeonggi-do, South Korea	H-K Ch
42	<i>Plantago major</i> L.	Spain1	ES1	Horta de Sant Joan, Spain	K Marti
43	<i>Plantago major</i> L.	Spain2	ES2	Bobadilla, Spain	N Iwany
44	<i>Plantago major</i> L.	Sweden	SE	Visby, Sweden	N Iwany
45	<i>Plantago major</i> L.	TenerifeSouth	TE	Tenerife, Canary Islands	N Iwany
46	<i>Plantago major</i> L.	Turkey	TR	Güroluk, Turkey	Y Yesil
47	<i>Plantago major</i> L.	Ukraine	UA	Kiev, Ukraine	D Zubov
48	<i>Plantago major</i> L.	Vancouver	VA	Vancouver, British Columbia, Canada	CJ Roth
49	<i>Plantago major</i> L.	Washington	WA	Washington, USA	CJ Roth
50	<i>Plantago major</i> L.	Yukon	YU	Old Crow, Yukon, Canada	B Benne



**Figure 1** . Distribution of *Plantago major* based on 526, 497 georeferenced occurrences registered in GBIF.org (GBIF, 2020, accessed 19 June, 2020).

Figure 2. Locations of the 50 global sampling sites for *Plantago major*. See Table S1 for population details. Population codes: 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, Sweden=SE, South Korea=KR, Spain1=ES1, Spain2=ES2, Sweden=SE, Tenerife=TE, Turkey=TR, Ukraine=UA, Vancouver=VA, Washington=WA, Yukon=YU.

#### Hosted file

image3.emf available at <https://authorea.com/users/337132/articles/462777-travel-tales-of-a-worldwide-weed-genomic-signatures-reveal-colonial-trade-routes-and-prior-adaptation-are-key-to-the-success-of-plantago-major>

**Figure 3** . Multidimensional scaling plot (MDS) showing first two coordinate axes for all global samples of *Plantago major* (385 samples from 50 populations worldwide), generated by creating a distance matrix using 2807 of the most informative SNPs in ngsdist. Colour coding reflects ancestral populations (at  $K = 6$ ) modelled in ngsadm (see Figure 3). Population abbreviations: 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, Sweden=SE, South Korea=KR, Spain1=ES1, Spain2=ES2, Sweden=SE, Tenerife=TE, Turkey=TR, Ukraine=UA, Vancouver=VA, Washington=WA, Yukon=YU.

#### Hosted file

image4.emf available at <https://authorea.com/users/337132/articles/462777-travel-tales-of-a-worldwide-weed-genomic-signatures-reveal-colonial-trade-routes-and-prior-adaptation-are-key-to-the-success-of-plantago-major>

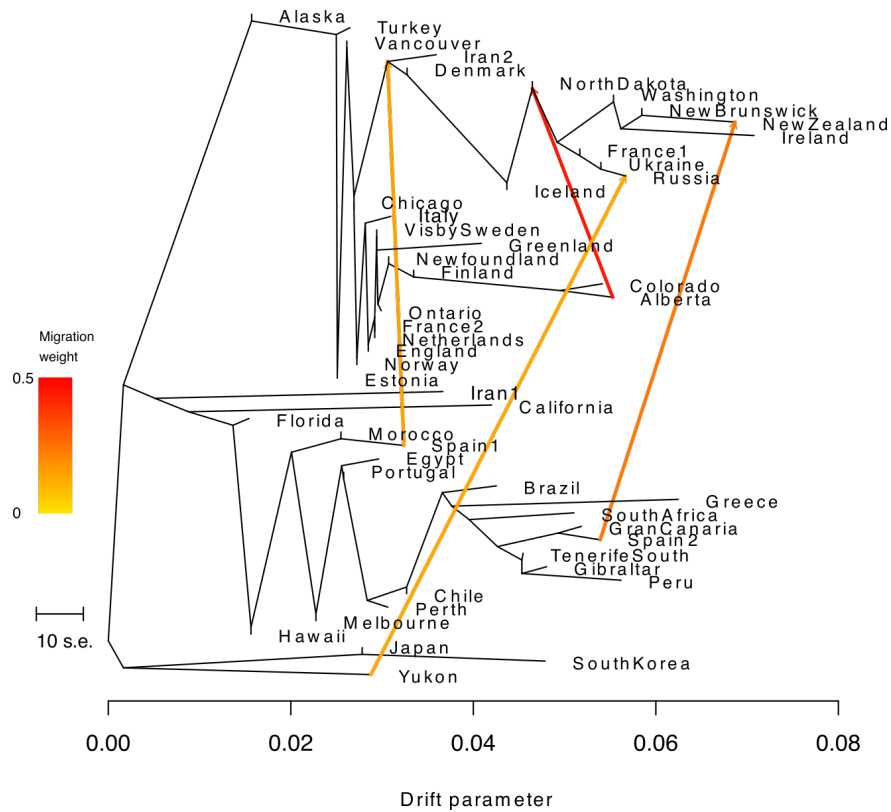
#### Hosted file

image5.emf available at <https://authorea.com/users/337132/articles/462777-travel-tales-of-a-worldwide-weed-genomic-signatures-reveal-colonial-trade-routes-and-prior-adaptation-are-key-to-the-success-of-plantago-major>

### Hosted file

image6.emf available at <https://authorea.com/users/337132/articles/462777-travel-tales-of-a-worldwide-weed-genomic-signatures-reveal-colonial-trade-routes-and-prior-adaptation-are-key-to-the-success-of-plantago-major>

**Figure 4 (a)** ngsadm results for  $K$  values 3 and 6, based on the highest likelihood runs for all samples of *Plantago major* from 50 global populations. The probability of each individual belonging to population is indicated by differing colours. **(b)** Global distribution of shared ancestry groups at  $K = 6$ .



**Figure 5 .** treemix maximum likelihood tree showing the relationship among populations, allowing four migration events. Migration arrows are coloured according to their weight. 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.