Paper II

Early germination to seed set in *Heracleum persicum*. Photos: Dilli P. Rijal

MOLECULAR ECOLOGY

Molecular Ecology (2015) **24**, 5522–5543 doi: 10.1111/mec.13411

Reconstructing the invasion history of *Heracleum persicum* (Apiaceae) into Europe

DILLI P. RIJAL,* TORBJØRN ALM,* ŠÁRKA JAHODOVÁ,† ‡ HANS K. STENØIEN§ and INGER G. ALSOS*

**Department of Natural Sciences, Tromsø Museum, University of Tromsø-The Arctic University of Norway, 9037 Tromsø, Norway, †Institute of Botany, The Czech Academy of Sciences, CZ-252 43, Průhonice, Czech Republic, ‡Department of Ecology, Faculty of Science, Charles University in Prague, Viničná 7, Prague CZ-128 44, Czech Republic, §Department of Natural History, Centre for Biodiversity Dynamics, NTNU University Museum, Norwegian University of Science and Technology, 7491 Trondheim, Norway*

Abstract

Sparse, incomplete and inappropriate historical records of invasive species often hamper invasive species management interventions. Population genetic analyses of invaders might provide a suitable context for the identification of their source populations and possible introduction routes. Here, we describe the population genetics of *Heracleum persicum* Desf. ex Fisch and trace its route of introduction into Europe. Microsatellite markers revealed a significantly higher genetic diversity of *H. persicum* in its native range, and the loss of diversity in the introduced range may be attributed to a recent genetic bottleneck. Bayesian cluster analysis on regional levels identified three and two genetic clusters in the native and the introduced ranges, respectively. A global structure analysis revealed two worldwide distinct genetic groups: one primarily in Iran and Denmark, the other primarily in Norway. There were also varying degrees of admixture in England, Sweden, Finland and Latvia. Approximate Bayesian computation indicated two independent introductions of *H. persicum* from Iran to Europe: the first one in Denmark and the second one in England. Finland was subsequently colonized by English populations. In contrast to the contemporary hypothesis of English origin of Norwegian populations, we found Finland to be a more likely source for Norwegian populations, a scenario supported by higher estimated historical migration from Finland to Norway. Genetic diversity *per se* is not a primary determinant of invasiveness in *H. persicum*. Our results indicate that, due to either pre-adaptations or rapid local adaptations, introduced populations may have acquired invasiveness after subsequent introductions, once a suitable environment was encountered.

Keywords: approximate Bayesian computation, biodiversity, genetic variation, giant hogweeds, invasive alien species, population genetics

Received 8 May 2015; revision received 3 October 2015; accepted 6 October 2015

Introduction

Invasive alien species affect biodiversity at all organizational levels from genes to ecosystems (Vitousek & Walker 1989; Vilà *et al.* 2011), and cause significant damage to the environment and economy (Pimentel 2011). Interspecies hybridization between the invasive and native species is considered a major cause for loss of native genetic distinctness (Rhymer & Simberloff 1996; Lockwood *et al.* 2013). Moreover, invasive alien species can change entire ecosystems by altering fire regimes (Pemberton & Ferriter 1998; Brooks *et al.* 2004; Watt *et al.* 2009; Simberloff 2013), hydrology (Zavaleta 2000), fauna of decomposers (Bedano *et al.* 2014) and nutrient pools (Vitousek *et al.* 1987; Wang *et al.* 2015).

Correspondence: Dilli P. Rijal, Fax: +47 776 45520;
E-mail: dilliprijal@gmail.com

Invasive alien species are considered one of the major threats to global biodiversity (CBD 2001; Genovesi *et al.* 2013). Besides, considerable concern in understanding biological invasion, management, control, and eradication of invasive species remains challenging due to sparse, incomplete and inappropriate historical records (Estoup & Guillemaud 2010). Due to this lack of historical information, many invasive species remain unnoticed until their populations explode. However, indirect methods based on molecular genetic markers have proved effective in bridging such gaps between invasion history and management by providing insight into the complex history of biological invasions (Lombaert *et al.* 2014).

Information about population genetics, introduction history and identification of source populations are crucial in understanding the invasion process (Cristescu 2015). The genetic diversity of a species indicates its evolutionary potential to adapt to a novel environment (Sakai *et al.* 2001). This may be especially important for exotic invasive species as they have to adapt and survive to novel environments. Genetic diversity of introduced populations largely depends on the number of founders and the number of introductions from the genetically differentiated (native) source populations (Kolbe *et al.* 2004; Lavergne & Molofsky 2007; Ward *et al.* 2008; Simberloff 2009). Genetically diverse populations may have higher establishment success if they contain genetic variants more suited to the new environment, thereby posing greater invasion risk (Lee 2002; Forsman 2014; Bock *et al.* 2015). Although introduced invasive species suffer from genetic bottlenecks, they often overcome adverse effects of population reduction by genetic admixture via multiple introductions from the native range (Kolbe *et al.* 2004) and/or other successful introduced populations (invasive bridgehead effect, Lombaert *et al.* 2010; Benazzo *et al.* 2015). Given that multiple introductions and genetic admixture may enhance invasibility (Kolbe *et al.* 2004; Roman & Darling 2007; Marrs *et al.* 2008; Ward *et al.* 2008), the number of introductions may indicate risk of further regional spread of a species. Better understanding of the genetic diversity of introduced populations and vital source populations along with the number of introductions may be used to prevent further introductions and/or spread of invasive species by designing monitoring and quarantine strategies targeting the source area and the important vectors (Estoup & Guillemaud 2010). Thus, genetic diversity of invasive populations can be used as a risk assessment tool.

The change in effective sizes and ranges of natural populations in the past leave signatures in their genetics (Cornuet *et al.* 2010), and this historical signature can be inferred by examining genetic variation among populations (Lawton-Rauh 2008). For example, genetic differentiation among populations is considered a product of limited dispersal and gradual genetic drift. As a result, genetic similarity becomes correlated with geographical distance (isolation by distance, Wright 1943). Introduction route of a species can be inferred using molecular data in several ways, including assessing similarity among genetic clusters (Pritchard *et al.* 2000; Besnard *et al.* 2014; Yu *et al.* 2014), assigning individuals to source populations (Rannala & Mountain 1997; Paetkau *et al.* 2004), quantifying gene flow between isolated populations (Nielsen & Wakeley 2001) and comparing plausible migration scenarios using simulation approaches (Beaumont *et al.* 2002; Cornuet *et al.* 2010; Besnard *et al.* 2014).

Invasive vascular plants constitute about 53% of the invasive species of Europe, and 49% of these plants are of non-European origin (Pyšek *et al.* 2009). Anthropogenic pressure is a main driver of European plant invasion, and a strong positive correlation is found between human population density and alien richness (Marini *et al.* 2012). Most alien plant species have deliberately been introduced into Europe, ornamentals in particular (Lambdon *et al.* 2008). Among the many terrestrial invasive plant species, a group of large hogweeds commonly known as 'giant hogweeds' are posing threats to public health and biodiversity in different parts of Europe (Nielsen *et al.* 2005; EPPO 2009). Giant hogweeds (*sensu* Nielsen *et al.* 2005) include three invasive species of *Heracleum* (Apiaceae) in Europe (i.e. *H. mantegazzianum*, *H. persicum* and *H. sosnowskyi*). The first two species were famous garden plants during the 19th century in Europe, and the latter was introduced into northwest Russia as a forage crop at the end of the 1940s (Nielsen *et al.* 2005; EPPO 2009; Alm 2013). Within <2 centuries of introduction, giant hogweeds became some of the most prominent invasive species in northern Europe. They possess some typical features of invasive species, for example early and fast growth, high stature, huge biomass production, extensive cover and abundant seed production. In addition, *H. persicum* is perennial and highly clonal, which is not the case for other two giant hogweeds. It has successfully adapted to new environmental conditions, from hot summers of Persia, with 'short' days, to the much cooler conditions and perpetual daylight in parts of its introduced range at 51–71° northern latitude. An invasive species possessing all the characteristics of the 'ideal-weed' (Baker 1965) rarely exists in nature; however, *H. persicum* seems to exhibit most of the necessary characteristics (van Kleunen *et al.* 2015). Thus, *H. persicum* represents a model to provide broader understanding of the evolution of invasiveness, especially the paradoxical role of population bottlenecks, genetic diversity of the source populations, and introduction history.

The source and introduction route of *H. persicum* in Europe are unclear. Hypotheses concerning introduction routes are based on historical accounts and limited observational data (Estoup & Guillemaud 2010). The first seed record of *H. persicum* in Europe comes from the seed list of Royal Botanic Garden Kew from 1819 (Pyšek *et al.* 2010). Historical records show that an English man planted seeds in northern Norway in 1836 (Christy 1837; Fröberg 2010; Alm 2013); however, it is unclear whether he brought seeds from naturally growing English populations or from other sources. Meanwhile, the absence of naturalized populations of *H. persicum* in the UK (Sell & Murrell 2009; Stace 2010) is surprising, as the species has proved highly invasive elsewhere in NW Europe. In addition, the taxonomy of the giant hogweeds has been a subject of controversy (Jahodová *et al.* 2007; Fröberg 2010; Alm 2013), and a variety of ill-defined Latin names have been used for Scandinavian plants, including *H. giganteum, H. laciniatum, and H. panaces*. *Heracleum persicum* may be hiding in historical accounts due to misinterpretation as *H. mantegazzianum*. Under such circumstances, population genetics of *H. persicum* may serve as a promising alternative to resolve not only introduction pathways, but also illuminate the complex invasion history (Estoup & Guillemaud 2010; Brouat *et al.* 2014).

Even though *H. persicum* is highly invasive in the introduced range, we assume that it suffered a loss of genetic diversity due to population bottlenecks during the initial introduction. To test whether introduced populations are genetically depauperate, we compared the genetic diversity of native and introduced populations. Introduced populations often overcome the effects of genetic bottlenecks due to multiple introductions or genetic admixture, and we considered the number of introductions as an indicator of propagule pressure that may enhance establishment success of *H. persicum*. We evaluated whether introduced populations were formed by multiple introductions and if there has been admixture between introduced populations. To aid management interventions, we identified respective source populations of the introduced invasive populations and tested whether genetic diversity *per se* was inherently linked with invasiveness. By tracing the routes of introduction, we evaluated whether *H. persicum* followed the route indicated by historical accounts when invading Europe.

Material and methods

Study species

The enigmatic, invasive *Heracleum* species found in northern Scandinavia has been identified as *Heracleum persicum* based on genetic similarity with Iranian species (Jahodová *et al.* 2007), which is also supported by morphological investigations (Fröberg 2010). Although earlier studies (Nielsen *et al.* 2005; EPPO 2009; Fröberg 2010) stated that *H. persicum* was native to Iran and Turkey, Ahmad (2014) has recently reported it as a new species in Iraq, at a single station close to the Iranian border. Similarly, *H. persicum* is narrowly distributed in southeast Turkey (SE Anatolia) (Ahmad 2014; Arslan *et al.* 2015) in an area bordering northwest Iran. However, it is widely distributed in north, west, northeast and central Iran (Rechinger 1987; Ahmad 2014). It was introduced to Denmark, England, Finland, Latvia, Norway, Sweden and Iceland (Fröberg 2010; Wasowicz *et al.* 2013). The plant is polycarpic and generally attains a height of 2.5 m and sometimes reaches up to 3 m (Fröberg 2010; Alm 2013). Seed germination requires stratification at 2–4 °C for two months and flowering starts after the third year post germination. Temporal variation in flower maturation promotes outcrossing. Male flowers in the primary umbel mature earlier than female flowers. In the secondary umbels, flowering occurs after seeds are set in the primary umbels, and female flowers are generally abortive (Often & Graff 1994; Fröberg 2010). Reproduction primarily occurs through seeds; however, clonal reproduction is also common in disturbed habitats where seed reproduction fails. The plant sap is phototoxic and induces photocontact allergy when exposed to ultraviolet radiations (Nielsen *et al.* 2005; EPPO 2009). In the introduced range, *H. persicum* commonly grows at seashores, roadsides, abandoned farmlands, highly disturbed areas and seminatural habitats like forest clearings. The earliest European record of the species appeared in the seed list of Royal Botanic Gardens, Kew, London, in 1819 (Pyšek *et al.* 2010). It has been recommended for regulation as a quarantine pest in Europe (EPPO 2009) and is blacklisted in Norway (Gederaas *et al.* 2012).

Plant material

Historical records of the species from the global biodiversity information facility (GBIF) (http://www.gbif.org/species/3628745), Norwegian Biodiversity Information Centre (http://www.biodiversity.no/), sampling locations reported by Jahodová *et al.* (2007) and the most recent data available for Norway (Fremstad & Elven 2006) were rigorously evaluated before starting the sampling (Fig. 1). Sampling was done throughout the species' distribution range between 2012 and 2014 (Fig. 1), except Iraq and Iceland, for which the species has only recently been found (Wasowicz *et al.* 2013; Ahmad 2014), and Turkey, from where export of plant material is now prohibited. We collected four

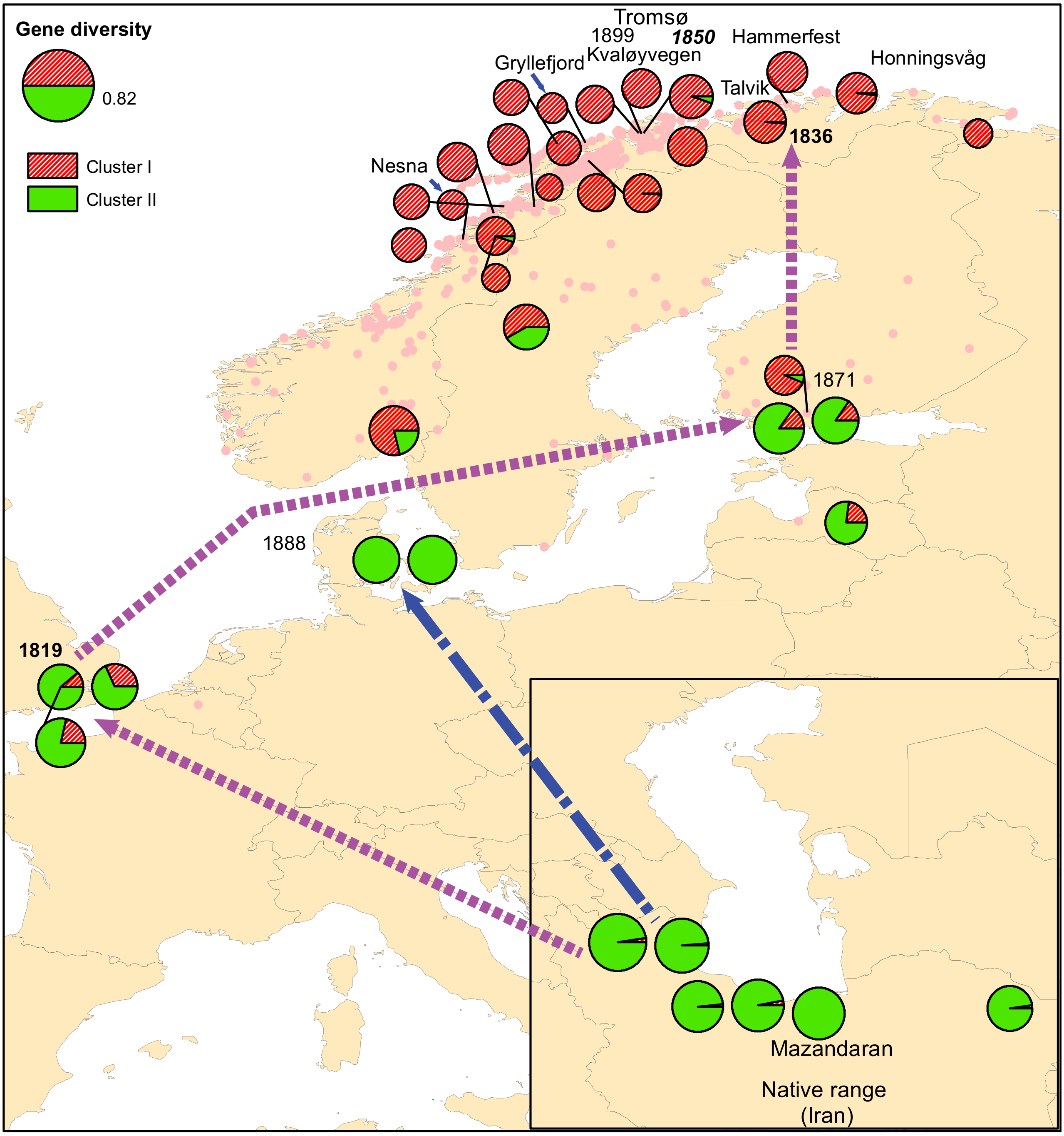


Fig. 1 Geographical locations of previous records (small circles) and genetic structure of sampled populations from native and introduced ranges of *Heracleum persicum*. Size of a pie chart reflects gene diversity (expected heterozygosity) of each population. Hatched and plain pie charts indicate proportion of genomes of each population assigned to Cluster I and Cluster II, respectively, as revealed by global structure analysis based on $K = 2$. Dates indicate the first seed and plantation record for England and Norway (bold), respectively, the first cultivation record for Tromsø (bold italic), and the earliest records of garden escapes for Scandinavia (normal). Arrow indicates inferred route of introduction of *H. persicum* into Europe based on approximate Bayesian computation analysis.

samples and one representative herbarium voucher from 5 different spots at 5–10-m intervals per population, and care was taken to avoid resampling from the same genet, resulting in 1–20 samples per population. All samples were dried on silica gel and photographed. A few populations collected during 2003–2004 were retrieved from the material of Jahodová *et al.* (2007) (see Table 1) and herbarium vouchers for those samples are

Table 1 Sampling details and genetic diversity indices for populations of *Heracleum persicum*. Populations with <4 samples (italicized) were not considered while calculating average diversity statistics across country (bold)

Country	District/Region	Location	Latitude	Longitude	Collectors	Year	N	P (%)	I	N_A	N_E	A_R	H_O	H_E	uH$_E$	F_{IS}
Denmark	Sjælland	Roskilde	55.6833	12.0333	SJ/LF	2003	15.0	81.82	0.58	2.44	1.68	1.92	0.38	0.34	0.35	−0.08
	Sjælland	Roskilde	55.6833	12.0364	TJ	2012	20.0	81.82	0.59	2.20	1.81	1.93	0.46	0.38	0.39	−0.17
							17.5	**81.82**	**0.58**	**2.32**	**1.74**	**1.93**	**0.42**	**0.36**	**0.37**	**−0.13**
England	London	Buckingham Palace	51.4984	−0.1457	SJ/OB	2004	10.0	72.73	0.51	2.04	1.68	1.81	0.46	0.33	0.35	−0.35
	London	Kensington Garden	51.5079	−0.1740	SJ/OB	2004	15.0	86.36	0.64	2.52	1.81	1.97	0.39	0.39	0.40	0.00
	London	Kensington Garden	51.5102	−0.1751	DPR	2012	4.0	68.18	0.49	1.88	1.66	NA	0.45	0.33	0.37	−0.36
							9.7	**75.76**	**0.55**	**2.15**	**1.72**	**1.89**	**0.44**	**0.35**	**0.37**	**−0.24**
Finland	Uusimaa	Helsinki	60.2558	24.9711	SJ/PU	2004	15.0	86.36	0.55	2.32	1.70	1.83	0.38	0.34	0.35	0.04
	Uusimaa	Karkkila	60.5211	24.3483	SJ/PU	2004	15.0	59.09	0.37	1.76	1.44	1.58	0.27	0.24	0.25	0.04
	Uusimaa	Tammisari	59.9836	23.4153	SJ/PU	2004	15.0	86.36	0.72	3.00	2.05	2.04	0.42	0.41	0.42	−0.06
							15.0	**77.27**	**0.55**	**2.36**	**1.73**	**1.82**	**0.36**	**0.33**	**0.34**	**0.01**
Iran	Namin	Anbaran	38.5244	48.4625	MFA	2013	19.0	95.45	1.01	4.52	2.66	2.43	0.49	0.53	0.54	0.04
	Ardabil	Fandoughlu	38.4159	48.5719	MFA	2013	19.0	90.91	0.87	3.88	2.29	2.39	0.40	0.46	0.48	0.14
	Mazandaran	Javaherdeh	36.8482	50.4710	MFA	2014	16.0	72.73	0.77	3.16	2.30	2.10	0.34	0.42	0.43	0.11
	Mashhad	Mashhad	36.3611	59.3500	SJ/RS	2005	16.0	77.27	0.53	2.24	1.66	1.75	0.34	0.33	0.34	−0.05
	Mazandaran	Mazandaran	36.1918	51.3385	AP	2013	13.0	95.45	0.75	3.00	2.05	2.14	0.31	0.43	0.46	0.33
	Mazandaran	Rudbarak	36.4520	51.0744	MFA	2014	16.0	81.82	0.83	3.60	2.32	2.14	0.37	0.44	0.45	0.09
							16.5	**85.61**	**0.80**	**3.40**	**2.21**	**2.16**	**0.38**	**0.43**	**0.45**	**0.11**
Latvia	Madona	Ergil	56.9000	25.6333	SJ/GG	2003	**15.0**	**59.09**	**0.44**	**1.72**	**1.45**	**1.73**	**0.33**	**0.28**	**0.29**	**−0.09**
Norway	Vesterålen	Andenes	69.3218	16.1277	DPR, IGA, TA	2012	19.0	59.09	0.32	1.80	1.34	1.42	0.21	0.20	0.20	0.05
	Salten	Bodø	67.2866	14.3993	DPR	2012	20.0	59.09	0.31	1.80	1.35	1.41	0.23	0.20	0.20	−0.03
	Nord-Troms	Breiviklia	69.6780	18.9766	DPR	2012	20.0	81.82	0.50	2.40	1.61	1.68	0.36	0.30	0.31	−0.02
	Helgeland	Båsmoveien	66.3368	14.1133	DPR	2013	4.0	27.27	0.19	1.28	1.14	NA	0.21	0.12	0.14	−0.52
	Salten	Fauske	67.2583	15.3842	DPR	2012	20.0	68.18	0.41	1.84	1.51	1.53	0.27	0.27	0.27	0.14
	Central Hålogaland	*Gratangen*	*68.6732*	*17.6966*	*DPR*	*2013*	*1.0*	*16.00*	*0.11*	*1.16*	*1.16*	*NA*	*0.16*	*0.08*	*0.16*	*−1.00*
	Midt-Troms	Gryllefjord	69.3626	17.0570	DPR, IGA, TA	2012	20.0	36.36	0.23	1.52	1.28	1.36	0.20	0.15	0.15	−0.22
	Vest-Finnmark	Hammerfest	70.6656	23.6985	DPR	2012	18.0	68.18	0.42	2.00	1.49	1.59	0.29	0.26	0.27	−0.05
	Vest-Finnmark	Honningsvåg	70.9944	25.9733	DPR	2012	20.0	72.73	0.41	2.04	1.53	1.53	0.28	0.26	0.27	0.10
	Stjørdalen	*Husbyvegen*	*63.471*	*10.967*	*DPR*	*2013*	*3.0*	*28.00*	*0.19*	*1.32*	*1.27*	*NA*	*0.17*	*0.13*	*0.16*	*−0.29*
	Hålogaland	Ibestad	68.7872	17.1573	DPR	2013	20.0	54.55	0.35	1.80	1.44	1.49	0.27	0.23	0.23	−0.10
	Salten	Inndyr	67.0477	14.0446	DPR	2013	6.0	54.55	0.35	1.60	1.44	NA	0.31	0.24	0.26	−0.23
	Nord-Troms	Kvaløya	69.6837	18.8113	DPR	2012	20.0	54.55	0.33	1.64	1.39	1.48	0.31	0.22	0.23	−0.28
	Nord-Troms	Kvaløyvegen	69.6651	18.9085	DPR	2012	20.0	59.09	0.36	1.80	1.43	1.45	0.28	0.24	0.25	−0.07
	Salten	*Langstranda*	*67.2714*	*14.3488*	*DPR*	*2013*	*3.0*	*48.00*	*0.30*	*1.56*	*1.37*	*NA*	*0.29*	*0.20*	*0.24*	*−0.41*
	Ofoten	Narvik	68.4398	17.4252	DPR	2013	6.0	50.00	0.32	1.56	1.42	NA	0.22	0.22	0.24	0.00
	Helgeland	Nesna	66.1951	13.0298	LUT	2012	18.0	22.73	0.21	1.36	1.29	1.36	0.20	0.14	0.15	−0.34
	Helgeland	*Nordlandsveien*	*66.316*	*14.157*	*DPR*	*2013*	*2.0*	*24.00*	*0.17*	*1.12*	*1.12*	*NA*	*0.24*	*0.12*	*0.16*	*−1.00*
	Helgeland	Novikveien	66.0068	12.5763	DPR	2013	15.0	40.91	0.30	1.68	1.41	1.41	0.25	0.19	0.19	−0.25
	Trondheim Region	*Othilienborgvegen*	*63.4072*	*10.4455*	*DPR*	*2013*	*3.0*	*32.00*	*0.19*	*1.28*	*1.19*	*NA*	*0.16*	*0.13*	*0.15*	*−0.20*
	Øst-Finnmark	Sandnesveien	69.6754	29.9626	DPR	2013	4.0	22.73	0.18	1.28	1.25	NA	0.22	0.13	0.15	−0.64

Table 1 *Continued*

Country	District/Region	Location	Latitude	Longitude	Collectors	Year	N	P (%)	I	N_A	N_E	A_R	H_O	H_E	uH$_E$	F_{IS}
	Central Hålogaland	Sandtorg	68.5675	16.3504	DPR, IGA, TA	2012	20.0	36.36	0.28	1.68	1.36	1.44	0.23	0.18	0.19	−0.08
	Helgeland	*Sjøbergs gate*	*66.022*	*12.6355*	*DPR*	*2013*	*3.0*	*36.00*	*0.24*	*1.40*	*1.32*	*NA*	*0.25*	*0.16*	*0.19*	*−0.52*
	Midt-Troms	Soleng	69.2458	19.4366	DPR	2013	10.0	54.55	0.36	1.68	1.44	1.50	0.33	0.24	0.25	−0.24
	Helgeland	Sørlandsveien	66.2998	14.1065	DPR	2013	5.0	50.00	0.35	1.64	1.45	NA	0.23	0.23	0.26	0.07
	Vest-Finnmark	Talvik	70.0470	22.9630	DPR	2012	20.0	77.27	0.46	2.16	1.55	1.64	0.29	0.28	0.29	0.14
	Salten	Tømmerneset	67.9067	15.8742	DPR	2013	4.0	13.64	0.16	1.24	1.18	NA	0.19	0.11	0.13	−0.67
	Østlandet	Tøyen Botancial Garden, Oslo	59.9181	10.7693	DPR	2012	7.0	81.82	0.64	2.40	1.90	NA	0.37	0.39	0.42	0.00
	Nord-Troms	*Åsgård-Giværbukta*	*69.6676*	*18.9118*	*DPR*	*2012*	*3.0*	*44.00*	*0.28*	*1.40*	*1.32*	*NA*	*0.25*	*0.19*	*0.25*	*−0.28*
							14.4	**52.07**	**0.34**	**1.74**	**1.42**	**1.48**	**0.26**	**0.22**	**0.23**	**−0.15**
Sweden	*Vilhelmina Ö*	*Latikberg*	*64.6443*	*17.0482*	*DPR*	*2013*	*1.0*	*40.00*	*0.28*	*1.24*	*1.24*	*NA*	*0.40*	*0.20*	*0.40*	*−1.00*
	Jämtland	Lit	63.3170	14.8387	DPR	2013	**9.0**	**86.36**	**0.50**	**2.12**	**1.64**	**1.67**	**0.42**	**0.32**	**0.34**	**−0.22**
	Lycksele	*Lycksele*	*64.6757*	*17.83*	*DPR*	*2013*	*3.0*	*44.00*	*0.33*	*1.56*	*1.46*	*NA*	*0.36*	*0.22*	*0.26*	*−0.64*
	Järpen	*Tossövägen*	*63.3416*	*13.4476*	*DPR*	*2013*	*2.0*	*56.00*	*0.36*	*1.56*	*1.46*	*NA*	*0.36*	*0.25*	*0.35*	*−0.43*
	Vännäs	*Tväråbäck*	*63.9978*	*19.7241*	*DPR*	*2013*	*1.0*	*36.00*	*0.25*	*1.32*	*1.32*	*NA*	*0.36*	*0.18*	*0.36*	*−1.00*
	Västerbotten	*Umeå*	*63.8237*	*20.2783*	*DPR*	*2013*	*2.0*	*40.00*	*0.28*	*1.44*	*1.39*	*NA*	*0.28*	*0.20*	*0.26*	*−0.43*

Sample collectors: AP, Atehfeh Pirany; DPR, Dilli Prasad Rijal; GG, Gertrude Gavrilova; IGA, Inger Greve Alsos; LF, Lars Fröberg; MFA, Mohsen Falahati-Anbaran; OB, Olaf Booy; PU, Pertti Uotila; RS, Rouhollah Sobhian; SJ, Šárka Jahodová; TA, Torbjørn Alm; TJ, Tina Jørgensen; N, number of samples; P (%), percentage of polymorphic loci; I, Shannon's information index; N_A, average number of alleles over loci; N_E, effective number of alleles; A_R, allelic richness based on three samples; H_O, observed heterozygosity; H_E, expected heterozygosity; uH$_E$, unbiased expected heterozygosity; F_{IS}, inbreeding coefficient; NA, not applicable.

deposited with original collectors. The leaf samples, DNA extracts, and herbarium vouchers of all other samples are deposited at Tromsø Museum (TROM).

DNA extraction and standardization

DNA was extracted using a DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) following manufacturer's protocol. DNA concentration of each sample was measured by NanoDrop 2000 (Thermo Scientific, Waltham, MA, USA), and all the samples were normalized to 10 ng/μL for downstream analyses.

Microsatellite genotyping

We selected 25 microsatellite markers developed by Rijal *et al.* (2015) and two markers developed by Henry *et al.* (2008), the latter two accommodated in multiplex II and III of Rijal *et al.* (2015), to genotype microsatellites of *H. persicum*. Altogether, 578 samples of *H. persicum* were screened in three multiplexes as described by Rijal *et al.* (2015). The total volume of PCR was 6 μL, which consisted of 3 μL master mix and 0.5 μL RNA-free water (Type-it Microsatellite PCR Kit; Qiagen), 1 μL primer mix and 1.5 μL template DNA. The thermal cycling conditions of each multiplex PCR were as follows: initial denaturation at 95 °C for 10 min followed by 10 cycles of 95 °C for 30 s, 60–50 °C of touch down PCR for 1 min with 1°C decrease per cycle, and 72 °C for 45 s; 25 cycles of 95 °C for 30 s, 50 °C for 1 min, 72 °C for 45 s; and a final extension of 60 °C for 15 min. A mixture of 2 μL of 1:20 diluted PCR product, 7.8 μL of HiDi Formamide and 0.2 μL of LIZ 600 (Applied Biosystems, Foster City, CA, USA) was denatured at 95 °C for 5 min, and electrophoresis was performed on 3130×L genetic analyzer (Applied Biosystems). Samples that had poor amplification or failed during fragment analysis were re-analysed. Any samples with poor chromatogram, after re-analysis, were discarded from genotyping. The genotyping error rate (Bonin *et al.* 2004) was estimated by replicating 96 samples for 7 loci from multiplex III.

Data analysis

The fragments were further analysed in GENEIOUS version 6.1.6 (Biomatters Ltd, New Zealand) following 3rd-order least squares method implemented in microsatellite plugin for allele calling. Due to stutter band in locus Hp_25, allele calling became problematic in some of the populations. The locus Hp_05 was polymorphic for only one sample from Denmark. Thus, we discarded these loci from further analyses. Similarly, three samples were discarded from the further analysis due to poor chromatograms. PGDSPIDER version 2.0.5.0 (Lischer & Excoffier 2012), MICROSATELLITE TOOLS (Park 2001) and GENALEX version 6.5 (Peakall & Smouse 2012) were used as data conversion tools, and the latter two were also used to check errors in genotypic data. Genotypic error rate was estimated by taking the ratio of mistyped genotypes to the total observed genotypes during the replication (the per-genotype error rate) whereas the ratio of miss-called allele to the total number of observed allele in the replication was considered as the per-allele error rate (Morin *et al.* 2009).

Hardy–Weinberg equilibrium and linkage disequilibrium. The test of Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) was performed in GENEPOP version 4.3 (Raymond & Rousset 1995; Rousset 2008) with 10 000 dememorization and in 1000 batches with 10 000 iterations per batch. We also performed a HWE jackknife test (Morin *et al.* 2009) using package 'strataG' (Archer 2014) in R version 3.1.2 (Team 2014) to detect the influential samples in populations. We reran the HWE test to evaluate the impact of influential samples on HWE by omitting samples with unusually large odds ratio (>99% of the rest of the distribution) as suggested by Morin *et al.* (2009).

Molecular diversity and genetic differentiation. The percentage of polymorphic loci (P%), Shannon's information index (I), unbiased expected heterozygosity (UH_E), average (N_A) and effective (N_E) number of alleles, observed (H_O) and expected heterozygosity (H_E), inbreeding coefficient (F_{IS}), and frequencies of private alleles were calculated for populations with ≥4 samples, that is 38 populations and 25 loci. All the analyses were performed in GENALEX version 6.5 (Peakall & Smouse 2012).

Allelic richness (A_R) was calculated to account for the possible bias due to difference in population size. The pairwise population genetic differentiation (F_{ST}) was calculated and tested for significance based on 1000 permutation without assuming HWE. Both analyses were performed in FSTAT version 2.9.3.2 (Goudet 1995). FSTAT is sensitive to missing loci and produces error while calculating A_R and does not provide *P*-values for F_{ST}. The locus Hp_30 was not present in Danish populations; loci Hp_07, Hp_10 and Hp_24 were missing in Latvia, and in Gryllefjord locus Hp_23 was present in two individuals. Thus, we included populations with nine or more samples (30 populations) and excluded the aforementioned loci, that is, 20 loci included, while calculating A_R and F_{ST}. Null alleles overestimate population differentiation by reducing within-population genetic diversity. The frequency of null allele was estimated following expectation maximization (EM) algorithm (Dempster *et al.* 1977) as implemented in FREENA (Chapuis & Estoup 2007). The global F_{ST} was calculated with and without correction for null allele, using FREENA with 1000 bootstrap resampling over loci, to evaluate the impact of null alleles in estimation of genetic differentiation.

Native and introduced populations were not equally represented in this study due to unequal sampling. Thus, when comparing diversity estimates between native and introduced ranges we used Welch two sample *t*-test, which corrects the problem of unequal sampling by incorporating variance in the analysis and adjusting the degrees of freedom (Ruxton 2006). The tests were performed in R version 3.1.2 (Team 2014).

Genetic bottleneck. To assess the effects of population bottlenecks in *H. persicum,* tests of heterozygosity excess and deficiency, were performed in BOTTLENECK version 1.2.02 (Piry *et al.* 1999), using all available mutation models, with 1000 iterations. Infinite allele model (IAM) overestimates, whereas stepwise mutation model (SMM) underestimates the bottleneck signature (Cornuet & Luikart 1996). Two-phase mutation model (TPM) is one of the complex but realistic mutational models that also includes the possibility of non-stepwise mutations to SMM (Selkoe & Toonen 2006). Thus, a TPM was used with 70% proportion of SMM along with 30% variance for TPM. To get an overview, results based on all mutation models were evaluated by applying Wilcoxon's test as it is the most powerful method when <20 polymorphic loci are considered (Cornuet & Luikart 1996; Piry *et al.* 1999). We also used mode shift test available in BOTTLENECK version 1.2.02 (Piry *et al.* 1999) to explore the recent bottleneck-induced distortion in the allele frequency (Luikart *et al.* 1998; Awad *et al.* 2014). The signature of subsequent population expansion after the bottleneck was tested with k and g tests (Reich & Goldstein 1998) using an excel macro KGTESTS (Bilgin 2007). Populations with ≥4 samples, that is 38 populations and 20 loci, were included in both of the analyses.

Population genetic structure. All 25 loci and 575 samples from 50 populations (Table 1) were assessed for genetic relationship by principal coordinate analysis (PCoA) in GENALEX version 6.5 (Peakall & Smouse 2012). The number of genetic clusters in *H. persicum* was estimated in STRUCTURE version 2.3.4 (Pritchard *et al.* 2000). The genetic structures of native and introduced populations were first evaluated separately. Altogether, 25 loci and 548 samples from 38 populations (with ≥4 samples) from native and introduced ranges were included in a global analysis. To detect the most likely native sources of the introduced populations, Denmark, England and Finland were analysed separately as well as jointly with native populations. To identify the likely sources of Norwegian populations, they were analysed separately with English and Finnish populations as well as in combination with all others. The analysis was performed on the Lifeportal computing platform (https://lifeportal.uio.no/) with initial burnin period of 200 000 followed by 250 000 Markov Chain Monte Carlo steps. The independent allele frequency and admixture model was assumed when performing Bayesian clustering analyses. The expected number of clusters (K) was set to 1–10 with 10 iterations for each K. The structure output was further processed in STRUCTURE HARVESTER (Earl & vonHoldt 2012). The best K was selected based on the Evanno *et al.* (2005) as implemented in STRUCTURE HARVESTER (Earl & vonHoldt 2012). Finally, summation of the individual file for different runs from STRUCTURE was performed in CLUMPAK (Kopelman *et al.* 2015).

Colonization routes. To trace the most likely introduction route of *H. persicum* in Europe, we tested four competing hypotheses by implementing approximate Bayesian computation (ABC) approach in DIY-ABC version 2.0.4 (Cornuet *et al.* 2014). Sweden and Latvia consisted of only 8 and 6 multilocus genotypes without missing loci, respectively, and their genetic structures were similar to England and Finland. The addition of less informative populations not only increases the number and complexity of the ABC scenarios, but also poses challenges in the result interpretation (Estoup *et al.* 2012). Thus, Latvia and Sweden were excluded from the ABC analysis, and 20 random multilocus genotypes without missing genotypes were selected each from England, Finland, Iran and Norway, and 19 from Denmark. The theoretical rationale for such regional sampling is provided in Stenøien *et al.* (2011).

Testing historical scenarios within the ABC framework is inherently a post hoc analysis, and the hypotheses (historical scenarios) are generally based on the available historical information and genetic population structures (Estoup *et al.* 2012; Lombaert *et al.* 2014). Our hypotheses were also based on historical records and we used genetic evidence to test those hypotheses. Most of the introduced alleles (nearly 78%) were in a subset of Iranian alleles, and private alleles of the introduced range were seemingly recently mutated from alleles introduced from Iran (Table S3). Thus, we tested the following scenarios (Fig. 2) by considering Iranian populations as the native source of the introduced populations: (i) scenario 1 was based on the historical account which assumes that *H. persicum* was first introduced from Iran to England and then to Norway, and finally to Denmark and Finland from Norway; (ii) scenario 2 assumed serial introductions from Iran to Denmark to

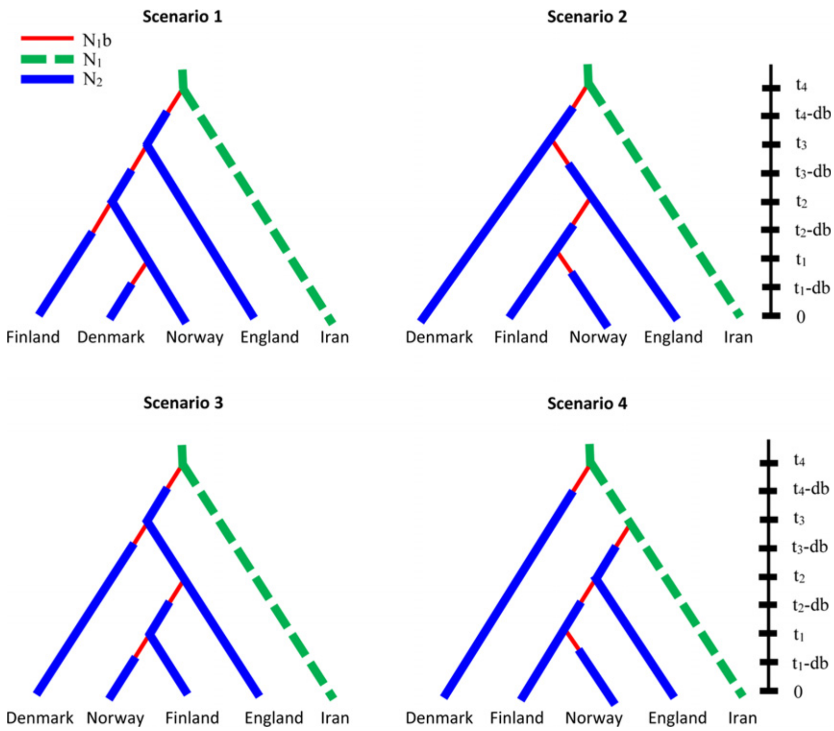


Fig. 2 Illustrations of four historical scenarios for introduction route of *Heracleum persicum* into Europe.

England to Finland to Norway; (iii) scenario 3 assumed two independent introductions from England to Denmark and from Denmark to Finland, while Finland acted as source for Norway; and (iv) scenario 4 hypothesized two independent introductions from Iran to Denmark and England. The Finnish population was assumed to have originated in England and acted as source for Norwegian populations.

The priors in the ABC analysis were defined based on the available information and later adjusted according to the results of initial runs. The effective population size of the native range (Iran) and introduced ranges were considered as N_1: 10–2000 and N_2: 10–200, respectively. Due to high abundance of *H. persicum* in Norway, but low genetic diversity, different ABC runs were performed assuming effective size of Norwegian population equal to Iran as well as less than or equal to other introduced populations. Invasive species suffer through an initial bottleneck as only few individuals invade the new area (Sakai *et al.* 2001). *Heracleum persicum* produces hermaphrodite flowers and like most of the members of Apiaceae, the species is considered to be self-compatible (Perglová *et al.* 2007). On this basis, we assume that even a single plant of *H. persicum* can produce seeds. Thus, we arbitrarily specified population size during bottleneck (N_1b) as 1–100. A variation of 30–100 years in the lag phase of invasive weeds has been reported (Aikio *et al.* 2010). If we assume the upper limit as the lag phase for *H. persicum* and a generation time of 3–6 years, then bottleneck duration may also vary from 17 to 33 generations. In general, defining narrow bottleneck duration prior reduces the accuracy of scenario identification (Guillemaud *et al.* 2010). Thus, we defined a wide period, that is 2–100 generations as the bottleneck duration (*db*). The species was present in Europe as early as 1819, which gives an estimate of 32–65 generations if we assume 3–6 years as the generation time of *H. persicum*. To cover the uncertainties in the divergence time we chose to use widely divergent time priors. Thus, the time since divergence of the recent to the oldest clades was considered as 2–100, 2–200, 2–300 and 2–400 generations ago and defined as t_1, t_2, t_3 and t_4, respectively. All the microsatellite loci were included in a single group and assumed to follow the identical mutation model with minimum mutation rate of 10^{-6} to maximum 10^{-2} per generation as reported for plant microsatellites (Udupa & Baum 2001; McConnell *et al.* 2007). The reference table was generated by 8×10^6 randomizations, twice the number considered optimal by the program (Cornuet *et al.* 2014). We compared the posterior probabilities of competing scenarios based on the logistic regression of the raw and the linear discriminant analysis (LDA) transformed summary statistics (Estoup *et al.* 2012; Lombaert *et al.* 2014). We used 4×10^6 simulated data sets while performing logistic regression on LDA transformed summary statistics. The type I and II error rates were used to discriminate the most plausible scenario. Type I error was the proportion of the number of times other scenarios have the highest posterior probability than the scenario under consideration. Type II error rate was based on the scenario II which had the largest type II error rate (as suggested by Estoup *et al.* 2012) and calculated as the proportion of the number of times the scenario under consideration has the highest posterior probability in scenario II.

Migration rates. To quantify the demographic parameters, especially migration rate between Norway and Finland, we used isolation with migration analysis in IMa software, which allows subsequent migration between two lineages being split from an ancestral population (Nielsen & Wakeley 2001; Hey & Nielsen 2004, 2007). The isolation with migration analysis was performed setting the upper limit of the prior distribution of population mutation parameter as 1 for both Norway and Finland and 10 for the ancestral population. The upper migration priors for both lineages were set to 250. The divergence time prior for two lineages was set to 0.5. Burn-in period was set as 10 000 and genealogy was saved each hour. Metropolis coupling was implemented with 20 chains and two geometric heating terms, that is 0.8 and 0.9. Average mutation rate of microsatellite loci was considered as 10^{-5} (Udupa & Baum 2001; McConnell *et al.* 2007). Three replicates of isolation with migration analyses were performed with identical settings until 50 million MCMC steps had been generated after burn-in.

Results

Genotypic error

Four samples had a replicate with poor chromatograms and were removed from downstream analyses. The absolute difference between loci varied from 0.07 to 1.03 base pairs (bp) with mean (±SE) of 0.26 (±0.06) bp based on two replicates of 92 samples. We observed a per-genotype error of 2.2%, which was slightly higher than the per-allele error rate of 1.5%.

Hardy–Weinberg equilibrium and linkage disequilibrium

Out of 950 population–locus combinations, 37 departed from HWE after Bonferroni correction (about 4%, Table S1). Most of the combinations (29) deviating from HWE were confined to three loci – Hp_13, Hp_14 and Hp_20 – and the remaining eight deviations were distributed among populations, occurring no more than twice per population and locus (Table S1). Jackknife analysis produced odd ratios for loci Hp_14 and Hp_20, indicating that these two loci had a comparatively large impact on tests for deviations from HWE (result not shown). Removal of 18 samples with ≥1.2 odd ratio did not change the overall HWE result (result not shown). The test of genotypic disequilibrium was significant for two loci pairs (Hp_27 × Hp_30 and HMNSSR_132B × HMNSSR_206) after Bonferroni correction (Table S2).

Molecular diversity and genetic differentiation

The average percentage of polymorphic loci was lowest for Norway (52.1%) and highest for Sweden (86.4%) (Table 1). Out of 205 alleles recorded, 163 were common and 25 and 17 were private to the native and the introduced populations, respectively. There were 48 and 35 alleles private to native and introduced ranges, respectively (Table S3). The Latvian population did not contain any private alleles. The Shannon's information index, allelic richness, expected and unbiased expected heterozygosities were lowest in Norway and highest in Iran (Table 1). The average number of alleles ranged from 1.72 (Latvia) to 3.34 (Iran). Minimum and maximum values of the observed heterozygosity were found for Norway and England, respectively. Similarly, the inbreeding coefficient ranged from −0.24 (England) to 0.11 (Iran). Locus-wise diversity statistics for native and invaded ranges are provided in Table S4.

Out of 435 comparisons, F_{ST} values of 295 population pairs were significant after Bonferroni correction (Table S5). One population from Iran (Mazandara) was not significantly differentiated from any native or introduced populations (nonsignificant pairwise F_{ST}). Three populations from Norway (Kvaløyvegen of Tromsø, Hammerfest and Nesna) were not significantly differentiated from most of the native and introduced populations. The mean (±SE) country-wise F_{ST} (averaged over population) was lowest between England and Sweden, that is 0.267 (±0.006), and highest between Norway and Denmark, that is 0.552 (±0.005) (Table 2). The average (±SE) frequency of null allele per locus varied from 0 ± 0 to 0.140 ± 0 (Table S6). There was a strong positive correlation between number of alleles and frequency of null allele, and only five loci had >0.05 null allele frequency (Fig. S1, Supporting information). The average (±SE) frequency of null alleles per population ranged from 0.001 ± 0 to 0.137 ± 0.023 (Table S6). The genetic differentiation between native and introduced ranges remained nonsignificant, when F_{ST} was estimated by including and excluding null alleles (result not shown).

The percentage of polymorphic loci, Shannon's information index, average numbers of alleles, effective number of alleles, private alleles, allelic richness; observed, expected (gene diversity) and unbiased expected heterozygosities, as well as inbreeding coefficients were significantly higher in the native range than

Table 2 The country-wise F_{ST} values averaged over populations of *Heracleum persicum*. Standard errors are given in the parentheses

	Iran	Denmark	England	Finland	Latvia	Norway	Sweden
Iran	0.253 (0.023)						
Denmark	0.388 (0.015)	0.037 (0.000)					
England	0.385 (0.014)	0.336 (0.010)	0.082 (0.000)				
Finland	0.409 (0.019)	0.392 (0.028)	0.272 (0.016)	0.286 (0.023)			
Latvia	0.407 (0.019)	0.452 (0.009)	0.306 (0.003)	0.354 (0.025)	0.000 (0.000)		
Norway	0.503 (0.006)	0.552 (0.005)	0.421 (0.008)	0.396 (0.009)	0.480 (0.008)	0.109 (0.005)	
Sweden	0.405 (0.021)	0.465 (0.005)	0.267 (0.006)	0.327 (0.028)	0.304 (0.000)	0.432 (0.014)	0.000 (0.000)

in the introduced range (Table 3). The loss of genetic diversity ranged from 16 to 49% in the introduced range, and on average nearly 42% of the gene diversity (H_E, Table 3) was lost by the introduced populations compared to the native populations. The average frequency of null alleles was significantly higher in native compared to introduced range. The fixation index, F_{ST}, was lower in the native compared to the introduced range, but the difference was marginal and nonsignificant (Table 3).

Genetic bottleneck

The tests of heterozygosity excess was significant after Bonferroni correction for one native and seven introduced populations when infinite allele model was considered (Table S7). However, the numbers were reduced to four and three introduced populations when two-phase and stepwise mutation models were assumed, respectively. Neither heterozygosity excess nor deficiency was observed in one native and twelve introduced populations. Similarly, mode of the allele frequency was shifted in 79% of the populations. About 67% native and 81% of the introduced populations showed mode shifts in the allele frequency distributions indicating recent bottlenecks (Table S7).

The within-locus k tests were significant for five introduced populations, indicating a signal of population expansion (Table S7). The interlocus g test was not very informative, as there were no clear trends between g ratios and significant k values (Table S7).

Table 3 Comparison of overall genetic diversity statistics between native and introduced populations of *Heracleum persicum*

Estimates	Native	Introduced	t	d.f.	P-value
P (%)	85.50	59.81	4.82	15.78	0.000
I	0.80	0.40	5.66	6.60	0.001
N_A	3.40	1.88	4.62	5.50	0.005
N_E	2.21	1.50	5.04	5.74	0.003
A_R	2.16	1.61	5.00	6.88	0.002
H_O	0.38	0.30	2.43	8.58	0.039
H_E	0.43	0.25	5.86	8.20	0.000
uH_E	0.45	0.27	5.88	8.43	0.000
F_{IS}	0.11	−0.14	3.95	10.88	0.002
P_A	4.17	1.89	3.07	7.89	0.016
F_{ST}	0.25	0.30	−1.94	19.77	**0.066**
Null allele	0.07	0.03	3.11	5.72	0.022

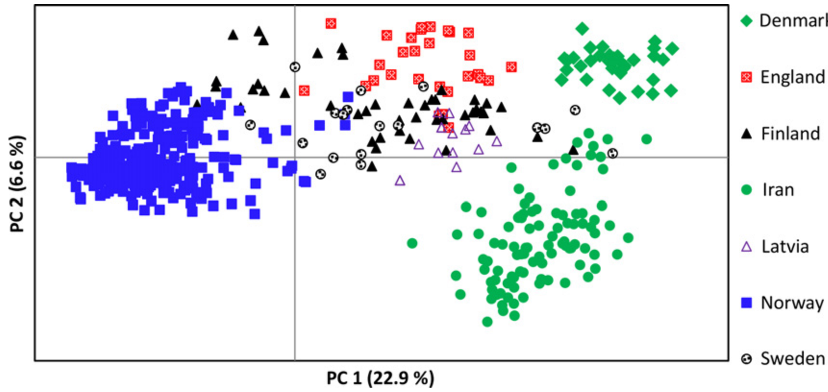
Nonsignificant P-value is in bold.
P (%), percentage of polymorphic loci; I, Shannon's information index; N_A, average number of alleles over loci; N_E, effective number of alleles; A_R, allelic richness based on three samples; H_O, observed heterozygosity; H_E, expected heterozygosity; uH_E, unbiased expected heterozygosity; F_{IS}, inbreeding coefficient; P_A, number of private alleles; F_{ST}, fixation index.

Population genetic structure

Ordination of microsatellites revealed that the Iranian, Danish and Norwegian populations of *Heracleum persicum* were distinct from each other. Populations from England, Finland, Latvia and Sweden appeared in between the former populations in the ordination plot (Fig. 3). Most of the variation (22.9%) in ordination plot was explained by the first axis while the second axis explained 6.6% of the variation. Finland consisted of highly variable samples scattering across most of the length of the first axis (Fig. 3).

There were three and two distinct genetic clusters in the native and the introduced ranges of *H. persicum*, respectively (Fig. 4). The two genetic clusters remained consistent when native populations were analysed with introduced populations from each country or in combinations (Fig. S2, Supporting information). Based on the rate of change of the likelihood distribution and the delta K value (Fig. 4C), two genetic clusters were detected for *H. persicum* in a global analysis (Figs 1 and 4D). More than 90% of the genomes of Norwegian samples were assigned to cluster I (hatched cluster in Figs 1 and 4D,F). However, more than 90% of the genomes of Iranian and Danish samples were assigned to cluster II (plain cluster in Figs 1 and 4D,F). Samples from England, Finland, Latvia and Sweden shared a higher

Fig. 3 Principal coordinate analysis of *Heracleum persicum* showing genetic relationship among samples originating both from native (Iran, 99 samples) and introduced ranges (476 samples).



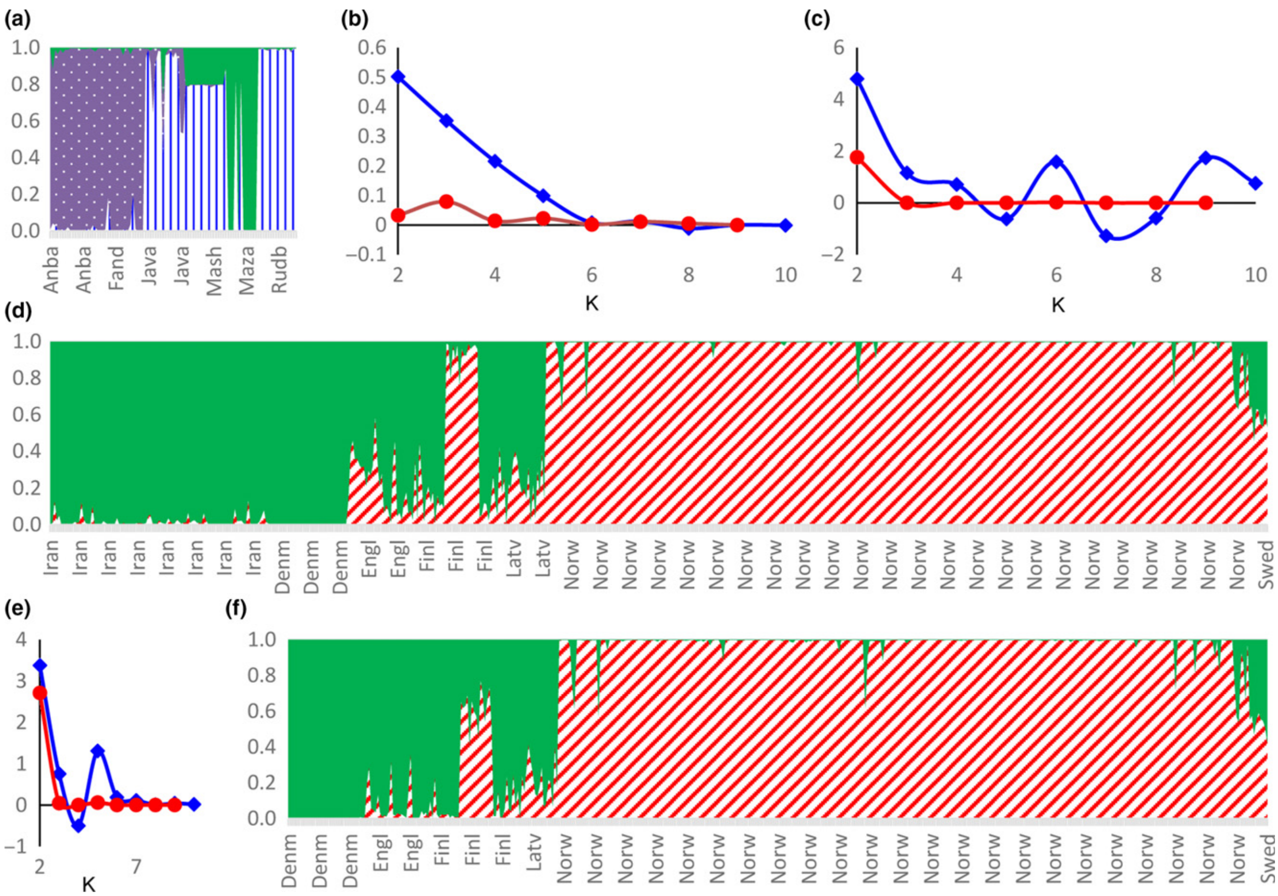


Fig. 4 Genetic structure of *Heracleum persicum* in Iran based on $K = 3$ (A), and global analysis (D) and introduced populations based on $K = 2$ (F). The transformed values (1/1000) of the rate of change of the likelihood distributions (diamond) and delta K (circle) for Iran (B), global analysis (C) and introduced populations (E). Delta K value of Iran was 1/100 transformed. Vertical bar represents proportion of individual genome assigned to each cluster. The abbreviated names consist of the first four characters of populations and countries from table 1.

proportion of both clusters. Assignment graphs of higher K values (2–4) for native, introduced, native-Denmark, native-England, global analyses and Norway are provided as supporting information (Fig. S2, Supporting information).

Colonization routes

The pre-evaluation of the scenarios suggested that priors were satisfactory delimited as the simulated data surrounded observed data in the ordination plot. There were no differences in the overall scenario discrimination patterns when the effective population size of Norway varied. The third scenario, which assumed two independent introductions from England to Denmark and Finland as well as another introduction to Norway from Finland, appeared more plausible than other scenarios when raw summary statistics were used. The posterior probability of scenario III was slightly higher in both the direct and logistic methods (average posterior probabilities 0.390 ± 0.010 and 0.648 ± 0.014, respectively) (Fig. S3 and Table S8, Supporting information). However, the highest posterior probability (0.651 ± 0.004) was observed for the fourth scenario, which assumed multiple introductions to Denmark and England from Iran, when LDA transformed summary statistics were used. The type I and II error rates were 3.0 and 1.9 times higher for the scenario III compared to the scenario IV, respectively, when using raw summary statistics (Table 4). The LDA transformed summary statistics produced 5.8 and 0.9 times higher type I and II error rates, respectively, for the scenario III compared to the scenario IV. The observed data of the scenario IV was more properly surrounded by the posteriors than the scenario III (Fig. S4, Supporting information), which further indicated that the fourth scenario was more likely than others.

The effective population sizes of Iran and Denmark/England/Finland/Norway under scenario IV were estimated to 1250 and 132, respectively (median of N_1 and N_2, Table 5). The result indicated that the Danish and the English lineages of *H. persicum* were derived from Iran about 218 and 139 generations ago respectively (median of t_4 and t_3, Table 5). However, the Finnish and the Norwegian lineages were split from their respective common ancestors about 75 and 57 generations ago, respectively (median of t_2 and t_1, Table 5). The medians of the biases were found within the range of -0.046 to 0.839 for t_1 and db, respectively (Table S9).

Table 4 Type I and II error rates for scenarios 3 and 4 (see Fig. 2 for the details) based on the logistic regression with raw (from 8×10^6 simulated data) and LDA transformed (from 4×10^6 simulated data) summary statistics

		Scenarios		Magnitude of error difference compared to scenario 4
Errors	Summary statistics	3	4	
Type I	Raw	0.43	0.11	3.0
	LDA transformed	0.44	0.06	5.8
Type II	Raw	0.26	0.09	1.9
	LDA transformed	0.25	0.14	0.9

LDA, linear discriminant analysis.

Migration rate

Exact mutation rates of *Heracleum* microsatellites have not been reported. When minimum (4.4×10^{-4}) and maximum (1.4×10^{-3}) mutation rate estimates from ABC analysis (Table 5, 25 and 97.5% quintiles) were used, population divergence time (τ/μ) varied from 24 to 75 generations for highest to lowest mutation rates. Average divergence time of Norwegian and Finnish lineages estimated by isolation with migration model was nearly 50 generations, which was approximately similar to the ABC estimates. The IM model suggested a higher rate of migration from Finland to Norway than vice versa (Table 6 and Fig. 5).

Discussion

We found significantly lower percentages of polymorphic loci, allelic richness and private alleles in the introduced range of *Heracleum persicum* compared to its native range. In addition, a significant loss of genetic diversity, as revealed by reduced expected heterozygosity and effective number of alleles, was also observed in the introduced range. Heterozygosity excess, an indicator of a genetic bottleneck, was observed in a few introduced populations.

Genetic diversity, population differentiation and bottleneck

Several monomorphic loci, lower genetic diversity, shifts in allele frequency and bottleneck signatures detected in the introduced range indicate that the introduced populations were established by few founders (Cornuet & Luikart 1996; Luikart *et al.* 1998; Piry *et al.* 1999; Sakai *et al.* 2001). Meanwhile, tests of recent population expansion was significant for five Norwegian populations growing south of Tromsø. Spread of *H. persicum* south of Tromsø is considered as a more recent event in Norway (Alm 2013). Successful invaders are expected to experience frequent bottlenecks without dramatic changes in genetic variation (Dlugosch *et al.* 2015). Thus, detection of bottleneck signature and population expansion characterizes a general process of initial establishment and colonization of *H. persicum* as it is spreading to new locations (Alm 2013; Wasowicz *et al.* 2013). Some of the earliest records of *H. persicum* in Norway come from Hammerfest, Honningsvåg, Talvik and Tromsø (see Fig. 1) (Alm 2013 & references therein), and none of them showed signatures of bottlenecks. Thus, evidence of bottlenecks is more common in the most recent populations, which agrees with general principles of the currently employed test that expect detection of bottleneck signatures for relatively

Table 5 ABC results of historical parameters estimated from 2008 pseudo-observed data sets simulated under scenario III (see Fig. 2) for *Heracleum persicum*. Mean, median, mode as well as 2.5, 5, 95 and 97.5% quintiles of estimated values are provided. N_1 and N_2, current effective population size of Iran and Norway, and Denmark, England and Finland, respectively; db, duration of bottleneck; N_1b, population size during bottleneck; t_1, t_2, t_3 and t_4 time since divergence of the youngest to the oldest lineages (see Fig. 2 and text for details)

Parameter	Mean	Median	Mode	q25	q50	q95	q97.5
N_1	1250	1250	1250	528	637	1880	1940
N_2	130	132	136	58	70	186	193
db	29	22	2	2	3	81	89
N_1b	62	64	69	17	24	95	97
t_1	56	57	54	19	25	87	92
t_2	79	75	66	28	34	139	157
t_3	144	139	142	54	65	242	261
t_4	222	218	215	83	99	362	379
Mutation rate	0.00075	0.0007	0.0006	0.00044	0.00047	0.00122	0.0014

Table 6 Maximum-likelihood estimates (MLE) along with the 95% highest posterior density (HPD) intervals for divergence time ($\tau = t\mu$ where t is the generation since divergence) of Norwegian (*N*) and Finnish (*F*) lineages of *Heracleum persicum*. Estimates of ancestral (θ_A), Norwegian (θ_N) and Finnish (θ_F) population size as well as migration rate to Norway ($m_{F>N}$) and to Finland ($m_{N>F}$) are provided

Parameter	95% hpd low	MLE high point	95% HPD high
τ	0.015	0.033	0.474
θ_N	0.003	0.003	0.037
θ_F	6.038	0.836	498.340
θ_A	203.423	331.607	965.098
$m_{F>N}$	18.708	48.542	237.292
$m_{N>F}$	10.208	10.458	156.625

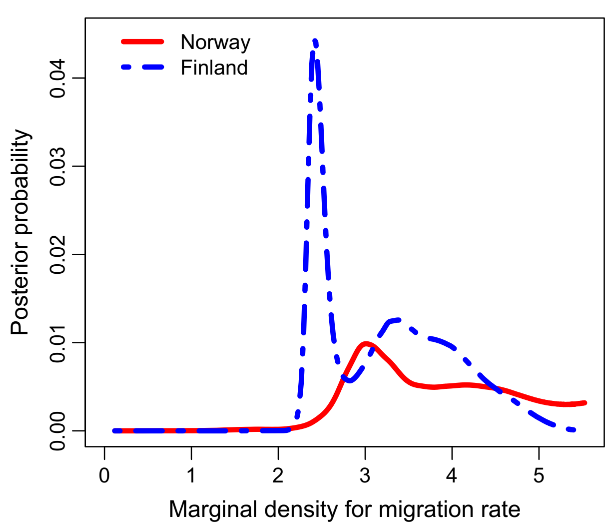


Fig. 5 Log-scaled marginal densities of migration rate of *Heracleum persicum* from Norway to Finland (dashed line) and Finland to Norway (solid line) estimated by IM analysis.

recently bottlenecked populations ($2N_e$-$4N_e$ generations in the past) (Cornuet & Luikart 1996; Piry *et al.* 1999).

The inbreeding coefficients were significantly lower for introduced populations, indicating a genetic bottleneck. Inbreeding depression depends on several factors including life history stages and population history (Husband & Schemske 1996). In general, due to fewer individuals, mating between close relatives (biparental inbreeding) is nearly unavoidable in smaller populations, which could force species towards the verge of extinction as a consequence of inbreeding depression and loss of alleles (Newman & Pilson 1997; Frankham & Ralls 1998). Thus, one would expect severe inbreeding in introduced species, as they are generally founded by few individuals, which in turn may reduce fitness. Surprisingly, inbreeding coefficients were either close to zero (an indication of perfect outcrossing) or negative (an indication of heterozygote excess) for introduced populations of *H. persicum*. Inbreeding can be avoided and outcrossing promoted through protandry in Apiaceae, a feature that has been reported for *Heracleum mantegazzianum* (Perglová *et al.* 2007). Inbreeding coefficients close to zero for several native and introduced populations indicate that the phenomenon is pervasive in both ranges. Negative inbreeding coefficients, on the other hand, have been frequently reported for the introduced populations of invasive species (Walker *et al.* 2003; Henry *et al.* 2009; Hagenblad *et al.* 2015). Thus, it could perhaps be viewed as a phenomenon linked with reduction in population size during expansion of the invasive species. Populations which showed relatively more negative inbreeding coefficients were those that predominantly had bottleneck signatures under IAM (Table 1 and Table S7). Thus, populations exhibiting a significant heterozygosity excess or negative inbreeding coefficient might have experienced a recent genetic bottleneck (Cornuet & Luikart 1996).

In general, introduced populations are genetically less diverse than native populations (Barrett & Kohn 1991; Sakai *et al.* 2001; Lavergne & Molofsky 2007) and this is also the case for introduced and native populations of *H. persicum*. This pattern is expected when only a fraction of the genetic diversity of the native population is introduced during initial colonization (Barrett & Kohn 1991). In addition, introduced populations generally suffer from population bottlenecks often for a longer period of time, which also reduces the genetic diversity (Allendorf & Lundquist 2003). However, Dlugosch *et al.* (2015) argue that invaders often retain significant amount of genetic variation if the founding populations are large enough to overcome the demographic constraints. In a closely related species, *H. mantegazzianum*, Walker *et al.* (2003) found a large genetic differentiation among populations at different river catchments in the introduced range and credited the observed variation to several independent introductions and relatively large initial founder populations. Niinikoski & Korpelainen (2015) found high genetic differentiation and a modest level of genetic variation in the introduced Finnish populations of *H. mantegazzianum*. It should be noted that both studies had no comparison with the native range and thus the differentiation is relative. Similarly, while comparing genetics of giant hogweeds, Jahodová *et al.* (2007) found high overall genetic variability in the invaded ranges and concluded that the invasive populations were not affected by genetic bottlenecks. In contrast, by comparing native and introduced populations of *H. mantegazzianum*, Henry *et al.* (2009) found a significant reduction in the genetic diversity in the introduced range and concluded that a founder event might have occurred. In extreme cases, some of the Norwegian invasive populations of *H. persicum* have lost >65% of the genetic diversity compared to native populations (Nesna & Gryllefjord, Table 1); otherwise, on average 16–35% of the genetic diversity was lost in the other introduced regions. Although nearly 50% of the genetic diversity is lost by the Norwegian populations compared to native populations (average H_E, Table 1), *H. persicum* is most abundant and vigorous in Norway compared to other introduced areas. Although neutral genetic markers may be poorly correlated with quantitative traits (Merilä & Crnokrak 2001; Reed & Frankham 2001; McKay & Latta 2002), a low level of genetic diversity does not seem to limit the invasiveness in giant hogweeds. Genetic diversity *per se* appears less important in determining the invasiveness of *H. persicum* in the introduced range. Genetics of invasive species, thus, represents a paradox in terms of the role of genetic diversity in adaptability (Simberloff 2013; Edelaar *et al.* 2015).

Route of introduction

We found higher population structuring within the native range as indicated by three distinct genetic clusters. However, two genetic clusters were consistent when some of the initially established introduced populations (Danish and English) were analysed separately or in combination with native populations, and populations from north-central Iran appeared more likely to be the sources of these introduced populations (Fig. S2C and D, Supporting information). A global Bayesian cluster analysis and ordination plot revealed two pure and one admixed genetic structures for introduced populations of *H. persicum* (Figs 1 and 4D,F). Denmark and Norway were clustered separately with distinct genetic structures, whereas England, Finland, Latvia and Sweden showed admixed genetic structure. Based on this result, we inferred that the Danish and all introduced genotypes (except the Norwegian) originated from two independent introductions from the native range, and the Norwegian genotypes originated from one of the introduced populations composed of mixed genotypes.

Although we could not include samples from Turkey and Iraq, genetic diversity, structure analyses and the *post hoc* ABC analysis indicated Iran as the source area for the European *H. persicum*. Nearly 78% of the introduced alleles were subset of the Iranian alleles and the remaining 22% private alleles were seemingly recent deviants of the Iranian alleles (1–4 mutational steps, Table S3). Although our six populations covered the major geographical distribution of the species in Iran (see Fig. 1), relatively higher genetic differentiation among Iranian populations (Figs 4A and S2, Supporting information) indicates that inclusion of more populations from Iran would have encompassed most of the introduced private alleles. Nevertheless, the apparent similarity in the allelic composition between Iran and the introduced range of *H. persicum* is unlikely to be a chance effect alone. The narrow distribution of *H. persicum* in Turkey, as well as its morphological mismatch with the Scandinavian specimens (Øvstedal 1987), makes it less likely to assume Turkey (and even more so, Iraq, with only a single, recently discovered station 400 m from the Iranian border) as sources of the European *H. persicum,* although we cannot exclude this as those populations were not sampled. The wide distribution of *H. persicum* in Iran as well as its morphological and genetic similarity with the European specimens (Jahodová *et al.* 2007; Fröberg 2010) indicate Iran as the more likely source of the European *H. persicum*.

Our findings do not corroborate the contemporary hypothesis that assumes an English population of *H. persicum* as the source of Norwegian population and all other European populations as descendant of the latter (Nielsen *et al.* 2005; Jahodová *et al.* 2007; EPPO 2009). In an earlier study, Jahodová *et al.* (2007) concluded that, as the Danish population appeared completely different from other introduced populations but more similar to Iran, multiple introductions from Iran might be responsible for invasion of *H. persicum* in Nordic countries. Structure analysis revealed that the Danish populations are more genetically similar to the Iranian than to the other introduced populations. As introduced populations tend to be more genetically similar to the source population(s) than to each other (Bond *et al.* 2002), our data indicate that the introduced populations were founded by more than one independent introduction from Iran.

In the ABC analyses, the LDA transformed summary statistics provided the highest support for the scenario IV that assumed two independent introductions to Denmark and England from the native source, and the subsequent spread in other parts from England. Although direct summary statistics provided the highest support for the scenario III, we considered scenario IV as the most likely scenario based on LDA transformed summary statistics. LDA reduces the number of dimensions, which decreases the number of explanatory variables and maximizes the differences among the scenarios, thereby improving the accuracy of the ABC approximation by avoiding correlations among explanatory variables (Estoup *et al.* 2012). In addition, scenario IV had lower type I and II error rates compared to scenario III. The ABC result was also supported by Bayesian cluster analysis showing shared clustering between English, Norwegian and Finnish but not Danish populations (Figs 1 and 4D,F). The genetic variation of introduced populations depends on the genetic diversity of the source population, and a relative decrease (due to bottleneck) or increase (due to multiple introductions and admixture) in the diversity of the introduced population is likely to happen (Edelaar *et al.* 2015). However, neither structure analysis nor genetic diversity patterns indicate any genetic admixture in the introduced range. Multiple introductions do not seem to have increased genetic variation. Instead, the pattern of loss of the genetic diversity in the introduced range closely resembled the introduction events indicated by the ABC analyses. For instance, Danish and English populations most likely originated from the similar native source from Mazandaran of central Iran close to the capital city Tehran (see Fig. S2, Supporting information) and have lost nearly 16% and 19% of the genetic diversity of the source; Finnish populations lost 6% of the English genetic variation; and Norway lost nearly 33% of the Finnish genetic variation. Thus, genetic diversity patterns of *H. persicum* appear to have been shaped largely by diversity of the source and the introduction history.

Although ABC appears as a promising methodology for inferring invasion scenarios, incorporating too many populations exponentially decreases the probability of accepting a simulation, a phenomenon known as the 'curse of dimensionality'. It also increases the number of scenarios and parameters to be tested (Beaumont *et al.* 2002; Cornuet *et al.* 2010). We traced the invasion history of *H. persicum* by ABC analysis and expected managers to utilize this information to avoid further introduction by isolating or eliminating small, introduced populations from the important source populations. We still suggest caution while interpreting ABC outcomes as our results were based on only four competing scenarios (out of 120 possible introduction scenarios).

Nevertheless, IM analysis provided new insights into the spread of *H. persicum* into Europe. As migration rate was higher from Finland to Norway than the reverse, it is quite likely that Norwegian populations were founded by Finnish propagules. Though the first seed record for *H. persicum* comes from Royal Botanic Gardens, Kew, the first verified Nordic escape record comes from Finland from 1871 (see Fig. 1) (Fröberg 2010). The first verified record of species in Denmark dates back to 1888 and the first Norwegian record to 1899 (Fröberg 2010). In contrast, the Norwegian records of *H. persicum* cultivation date back to the 1830s (Christy 1837; Fröberg 2010; Alm 2013). One probable explanation for this discrepancy may be the lack of historical records of *H. persicum* in Finland. In Denmark, past authors failed to realize that the introduced plants could belong to several species, generally interpreting both extant stands and the historical records as relating to *H. mantegazzianum* (e.g. Brøndegaard 1990). Brøndegaard (1979: p.307) cites anecdotal evidence of introduction of (presumed) *H. mantegazzianum* to Denmark in the 1830s. The timing is probably more reliable than the mode (as packing material for statues) and route (from Italy) of transport. In the light of our molecular data, early cultivations in Denmark are likely to have included *H. persicum*.

In addition, historical records of workers' movement from Finland to Norway, especially in the area where *H. persicum* was first recorded, further link Finnish and Norwegian populations of *H. persicum*. The earliest documented introduction of a large *Heracleum* species to northern Norway was made by a British traveller, W. Christy, in 1836. He visited Kåfjord at Alta and Hammerfest, and distributed seeds from England at both stations (Christy 1837). In 1835, Kåfjord was the largest single settlement in the otherwise sparsely populated county of Finnmark, due to the English-owned and run copper mines. In 1840, the mines employed 651 workers, with Finns constituting the largest ethnic group, outnumbering Norwegians (Moberg 1968; Nielsen 1995). It is probable that seeds from northern Norway may have been transferred to Finland and vice versa. Thus, while genetic data confirms the historical record of link between Finland and Norway, the inferred direction of spread is opposite.

Extensive populations of *H. persicum* in Norway suggest that it might be one of the oldest European populations. However, if Norwegian populations were older than Finnish and Danish populations, and founded the latter two, we should expect to observe higher level of polymorphisms in Norway than in other places. Norwegian populations are composed of quite distinct genotypes (Figs 1, 3 and 4; Fig. S2, Supporting information) and genetically highly structured compared to other regions (highest average regional F_{ST}, Table 2), indicating limited dispersal. Reduced gene flow is a prerequisite for local adaptation (Lenormand 2002). Thus, despite the lowest genetic diversity, spatially extensive populations in Norway may be due to local adaptations or success of pre-adapted genotypes from Iranian temperate mountains. These genotypes may be favoured in cool northern Norwegian climate compared to other countries. From its present distribution in Norway, it is evident that *H. persicum* thrives in the humid coastal areas with mild winters and avoids the drier inland areas with their cold winters, which may also explain the general scarcity of records of naturalized plants in Sweden and Finland. Also, fewer ornamental plants are able to thrive in northern Norway than England and Denmark may have increased its popularity. The current genetic (dis)similarity among regional populations might be due to discrepancy in regional climate and local adaptation.

Management implications

The genetic diversity of *H. persicum* is comparatively lower in the introduced than in the native range. *Heracleum persicum*, however, is vigorous and highly invasive in the introduced range despite lower genetic diversity.

As it is now generally regarded as an obnoxious weed in Norway, we assume that the historical vector (i.e. frequent cultivation in gardens) responsible for the original introduction and dispersal of *H. persicum* is now obsolete, indicating no further risk of intentional introductions from the native sources (unless Iranian immigrants are tempted to cultivate it from fruits imported for culinary use). However, a successfully established invasive population may pose greater risk of spread than the native source as the former needs a single evolutionary shift to acquire invasiveness while the latter needs multiple changes along with independent evolution of traits to be invasive (Estoup & Guillemaud 2010; Lombaert *et al.* 2010). Further introduction and expansion of *H. persicum* are quite likely in Europe due to high frequency of cross-border travels and transportations. While tracing the route of the introduction of *H. persicum*, the English and the Finnish populations appeared as the important sources for founding introduced populations. We urge managers to pay special attention while formulating management interventions to avoid the possible second introduction from the respective sources. Otherwise, successive waves of introduction from similar sources may augment further invasions (Benazzo *et al.* 2015). In addition, population admixture due to multiple introductions is considered a stimulus for rapid evolutionary changes (Kolbe *et al.* 2004; Lavergne & Molofsky 2007; Facon *et al.* 2008; Dlugosch *et al.* 2015). Thus, it is important to emphasize that some populations in the introduced range of *H. persicum* (i.e. Denmark, England, Finland, and Sweden) still have higher genetic diversity and may contribute to increase genetic diversity of neighbouring populations, for example Norwegian populations, by multiple introductions.

In general, biological control agents are chosen from the native (source) range of the invasive species (Roderick & Navajas 2003). *Heterodera persica*, a cyst-forming nematode, has been reported to parasitize on *H. persicum* in Iran (Maafi *et al.* 2006). *Heterodera persica* may be considered as a candidate biocontrol agent in the introduced range of *H. persicum*; however, so far, there has been no effort to test the effectiveness of *H. persica* as biological control agent against *H. persicum*. Meanwhile, we suggest to carefully assess the pitfalls of biological control agents as it has received both negative and positive responses (Messing & Wright 2006; Seastedt 2015). Moreover, it is important to note that single agent from the native range adapted against certain genotypes of *H. persicum* may not be sufficient for biological control (Marrs *et al.* 2008) as there are two distinct and one admixed groups of *H. persicum* in Europe.

Most microsatellite markers used in this study are also polymorphic for other giant hogweeds, that is *H. mantegazzianum* and *H. sosnowskyi*, the native *H. sphondylium* which has been reported to hybridize with giant hogweeds (EPPO 2009), their invasive hybrids and some also for *Anthriscus sylvestris* (Rijal *et al.* 2015). Hybridization can impede management interventions through creation of unique characteristics, for example production of novel chemicals, which in turn makes hybrids unrecognizable or unpalatable to specific herbivores or biological control agents (Schoonhoven *et al.* 2005; Williams *et al.* 2014). In general, hybridization appears a common phenomenon within the genus *Heracleum* (EPPO 2009). In particular, *H. persicum* commonly hybridizes with *H. sphondylium*, producing fertile and vigorous hybrids. They have already shown their presence and effect in Scandinavia (Fröberg 2010; Alm 2013; Rijal *et al.* 2015), and may further pose management challenges due to enhanced invasive abilities in hybrids as a consequence of interspecies hybridization (Ellstrand & Schierenbeck 2000; Schierenbeck & Ellstrand 2009). Thus, population genetics of *H. persicum* may shed light on the genetic attributes of other giant hogweeds as well as their invasive hybrids.

Conclusions

Even though the genetic data indicated at least two independent introductions of *H. persicum* to Europe, a clear genetic bottleneck was inferred, increasing with the stepwise introduction to more northern ranges within Europe. In contrast to the contemporary hypothesis of English origin of Norwegian populations, Finland appears as a more likely source for Norwegian populations of *H. persicum*. Despite the lowest level of genetic diversity, Norwegian populations are the most vigorous in the introduced range, suggesting no effect of bottlenecks on the invasiveness of *H. persicum*. Thus, genetic diversity *per se* does not seem to be an important determinant of invasiveness in *H. persicum*. Our result indicates that, due to either pre-adaptations or rapid local adaptation, introduced populations may acquire invasiveness after subsequent introductions when a suitable environment is encountered.

Acknowledgements

We are thankful to Eli Fremstad for providing geographical coordinates of the most recent Norwegian records of *H. persicum*, Mohsen-Falahati Anbaran for contributing valuable samples from Iran, Rämä Teppo for contributing important reagents, and all the sample contributors listed in the Table 1. DPR would like to appreciate Georgy Semenev for introducing cost-effective microsatellite genotyping methods, Subash Basnet and Anup Gupta for assisting in the field work, and Madan K. Suwal for sharing his high-speed computer to run ABC analyses. The first draft of this manuscript was prepared when DPR was in the Bergelson Lab (Department of Ecology and Evolution, University of Chicago) as a visiting PhD scholar. DPR is grateful to Joy Bergelson for providing working space in her laboratory. DPR appreciates Benjamin Brachi, Matthew Perisin and Timothy C. Morton, all from the University of Chicago, for fruitful discussion during data analysis. Authors are thankful to Matthew Perisin for English editing. We are thankful to Philippe Henry and three anonymous reviewers for their constructive comments on the earlier versions of the manuscript. The germination experiment was performed in the Climate Laboratorium, Holt, Tromsø. This project was funded by Tromsø Museum, University of Tromsø-The Arctic University of Norway.

References

Ahmad SA (2014) Eighteen species new to the flora of Iraq. *Feddes Repertorium*, **124**, 65–68.

Aikio S, Duncan RP, Hulme PE (2010) Lag-phases in alien plant invasions: separating the facts from the artefacts. *Oikos*, **119**, 370–378.

Allendorf FW, Lundquist LL (2003) Introduction: population biology, evolution, and control of invasive species. *Conservation Biology*, **17**, 24–30.

Alm T (2013) Ethnobotany of *Heracleum persicum* Desf. ex Fisch., an invasive species in Norway, or how plant names, uses, and other traditions evolve. *Journal of Ethnobiology and Ethnomedicine*, **9**, 42.

Archer E (2014) strataG: summaries and population structure analyses of haplotypic and genotypic data. R package version 0.9.2. http://CRAN.R-project.org/package=strataG.

Arslan ZF, Uludag A, Uremis I (2015) Status of invasive alien plants included in EPPO Lists in Turkey. *EPPO Bulletin*, **45**, 66–72.

Awad L, Fady B, Khater C *et al.* (2014) Genetic structure and diversity of the endangered fir tree of Lebanon (*Abies cilicica* Carr.): implications for conservation. *PLoS ONE*, **9**, e90086.

Baker HG (1965) Characteristics and modes of origin of weeds. In: *The Genetics of Colonizing Species* (eds Baker HG, Stebbins GL), pp. 147–172. Academic Press, New York City, New York.

Barrett SCH, Kohn JR (1991) Genetic and evolutionary consequences of small population size in plants: implication for conservation. In: *Genetics and Conservation of Rare Plants* (eds Falk DA, Holsinger KE), pp. 3–30. Oxford University Press, New York.

Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025–2035.

Bedano JC, Sacchi L, Natale E *et al.* (2014) Saltcedar (*Tamarix ramosissima*) invasion alters decomposer fauna and plant litter decomposition in a temperate xerophytic deciduous forest. *Advances in Ecology*, **2014**, 8.

Benazzo A, Ghirotto S, Vilaca ST *et al.* (2015) Using ABC and microsatellite data to detect multiple introductions of invasive species from a single source. *Heredity*, **115**, 262–272.

Besnard G, Dupuy J, Larter M *et al.* (2014) History of the invasive African olive tree in Australia and Hawaii: evidence for sequential bottlenecks and hybridization with the Mediterranean olive. *Evolutionary Applications*, **7**, 195–211.

Bilgin R (2007) Kgtests: a simple Excel Macro program to detect signatures of population expansion using microsatellites. *Molecular Ecology Notes*, **7**, 416–417.

Bock DG, Caseys C, Cousens RD *et al.* (2015) What we still don't know about invasion genetics. *Molecular Ecology*, **24**, 2277–2297.

Bond JM, Veenendaal EM, Hornby DD *et al.* (2002) Looking for progenitors: a molecular approach to finding the origins of an invasive weed. *Biological Invasions*, **4**, 349–357.

Bonin A, Bellemain E, Bronken Eidesen P *et al.* (2004) How to track and assess genotyping errors in population genetics studies. *Molecular Ecology*, **13**, 3261–3273.

Brøndegaard VJ (1979) *Folk og Flora: Dansk Etnobotanik*. Rosenkilde og Bagger, København.

Brøndegaard VJ (1990) Massenausbreitung des Bärenklaus. *Naturwissenschaftliche Rundschau*, **43**, 438–439.

Brooks ML, D'Antonio CM, Richardson DM *et al.* (2004) Effects of invasive alien plants on fire regimes. *BioScience*, **54**, 677–688.

Brouat C, Tollenaere C, Estoup A *et al.* (2014) Invasion genetics of a human commensal rodent: the black rat *Rattus rattus* in Madagascar. *Molecular Ecology*, **23**, 4153–4167.

CBD (2001) *Invasive Alien Species: Status, Impacts and Trends of Alien Species That Threaten Ecosystems, Habitats and Species*. United Nations Environment Programme and Convention on Biological Diversity, Montreal.

Chapuis M-P, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution*, **24**, 621–631.

Christy W (1837) Notes of a voyage to Alten, Hammerfest, etc.. *Entomological Magazine*, **4**, 462–483.

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, **144**, 2001–2014.

Cornuet J-M, Ravigne V, Estoup A (2010) Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics*, **11**, 401.

Cornuet J-M, Pudlo P, Veyssier J *et al.* (2014) DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism. DNA sequence and microsatellite data. *Bioinformatics*, **30**, 1187–1189.

Cristescu ME (2015) Genetic reconstructions of invasion history. *Molecular Ecology*, **24**, 2212–2225.

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B (Methodological)*, **39**, 1–38.

Dlugosch KM, Anderson SR, Braasch J *et al.* (2015) The devil is in the details: genetic variation in introduced populations and its contributions to invasion. *Molecular Ecology*, **24**, 2095–2111.

Earl D, vonHoldt B (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Edelaar P, Roques S, Hobson EA *et al.* (2015) Shared genetic diversity across the global invasive range of the monk parakeet suggests a common restricted geographic origin and the possibility of convergent selection. *Molecular Ecology*, **24**, 2164–2176.

Ellstrand NC, Schierenbeck KA (2000) Hybridization as a stimulus for the evolution of invasiveness in plants? *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 7043–7050.

EPPO (2009) *Heracleum mantegazzianum*, *Heracleum sosnowskyi* and *Heracleum persicum*. *EPPO Bulletin*, **39**, 489–499.

Estoup A, Guillemaud T (2010) Reconstructing routes of invasion using genetic data: why, how and so what? *Molecular Ecology*, **19**, 4113–4130.

Estoup A, Lombaert E, Marin J-M *et al.* (2012) Estimation of demo-genetic model probabilities with Approximate Bayesian Computation using linear discriminant analysis on summary statistics. *Molecular Ecology Resources*, **12**, 846–855.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Facon B, Pointier J-P, Jarne P *et al.* (2008) High genetic variance in life-history strategies within invasive populations by way of multiple introductions. *Current Biology*, **18**, 363–367.

Forsman A (2014) Effects of genotypic and phenotypic variation on establishment are important for conservation, invasion, and infection biology. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 302–307.

Frankham R, Ralls K (1998) Conservation biology: inbreeding leads to extinction. *Nature*, **392**, 441–442.

Fremstad E, Elven R (2006) *De Store Bjørnekjeksartene Heracleum i Norge*. Norges teknisk-naturvitendkapelige universitet, Trondheim, Norge.

Fröberg L (2010) *Heracleum* L.. In: *Flora Nordica (Thymelaeaceae to Apiaceae)* (eds Jonsell B, Karlsson T), pp. 224–234. The Swedish Museum of Natural History, Stockholm.

Gederaas L, Moen TL, Skjelseth S *et al.* (eds) (2012) *Fremmede Arer i Norge-med Norsk Svarteliste 2012*. Artsdatabanken, Trondheim, Norway.

Genovesi P, Butchart SHM, McGeoch MA *et al.* (2013) Monitoring trends in biological invasion, its impact and policy responses. In: *Biodiversity Monitoring and Conservation: Bridging the gap Between Global Commitment and Local Action* (eds Collen B, Pettorelli N, Baillie JEM *et al.*), pp. 138–158. John Wiley & Sons, Ltd., Hoboken, New Jersey.

Goudet J (1995) FSTAT (Version 1.2): a computer program to calculate F-statistics. *Journal of Heredity*, **86**, 485–486.

Guillemaud T, Beaumont MA, Ciosi M *et al.* (2010) Inferring introduction routes of invasive species using approximate Bayesian computation on microsatellite data. *Heredity*, **104**, 88–99.

Hagenblad J, Hülskötter J, Acharya KP *et al.* (2015) Low genetic diversity despite multiple introductions of the invasive plant species *Impatiens glandulifera* in Europe. *BMC Genetics*, **16**, 103.

Henry P, Provan J, Goudet J *et al.* (2008) A set of primers for plastid indels and nuclear microsatellites in the invasive plant *Heracleum mantegazzianum* (Apiaceae) and their transferability to *Heracleum sphondylium*. *Molecular Ecology Resources*, **8**, 161–163.

Henry P, Le Lay G, Goudet J *et al.* (2009) Reduced genetic diversity, increased isolation and multiple introductions of invasive giant hogweed in the western Swiss Alps. *Molecular Ecology*, **18**, 2819–2831.

Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*, **167**, 747–760.

Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences*, **104**, 2785–2790.

Husband BC, Schemske DW (1996) Evolution of the magnitude and timing of inbreeding depression in plants. *Evolution*, **50**, 54–70.

Jahodová Š, Trybush S, Pyšek P *et al.* (2007) Invasive species of *Heracleum* in Europe: an insight into genetic relationships and invasion history. *Diversity and Distributions*, **13**, 99–114.

van Kleunen M, Dawson W, Maurel N (2015) Characteristics of successful alien plants. *Molecular Ecology*, **24**, 1954–1968.

Kolbe JJ, Glor RE, Rodriguez Schettino L *et al.* (2004) Genetic variation increases during biological invasion by a Cuban lizard. *Nature*, **431**, 177–181.

Kopelman NM, Mayzel J, Jakobsson M *et al.* (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, **15**, 1179–1191.

Lambdon PW, Pyšek P, Basnou C *et al.* (2008) Alien flora of Europe: species diversity, temporal trends, geographical patterns and research needs. *Preslia*, **80**, 101–149.

Lavergne S, Molofsky J (2007) Increased genetic variation and evolutionary potential drive the success of an invasive grass. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 3883–3888.

Lawton-Rauh A (2008) Demographic processes shaping genetic variation. *Current Opinion in Plant Biology*, **11**, 103–109.

Lee CE (2002) Evolutionary genetics of invasive species. *Trends in Ecology & Evolution*, **17**, 386–391.

Lenormand T (2002) Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, **17**, 183–189.

Lischer HEL, Excoffier L (2012) PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, **28**, 298–299.

Lockwood JL, Hoopes MF, Marchetti MP (2013) *Invasion Ecology*, 2nd edn. Willey-Blackwell, Chichester, West Sussex, UK.

Lombaert E, Guillemaud T, Cornuet J-M *et al.* (2010) Bridgehead effect in the worldwide invasion of the biocontrol Harlequin Ladybird. *PLoS ONE*, **5**, e9743.

Lombaert E, Guillemaud T, Lundgren J *et al.* (2014) Complementarity of statistical treatments to reconstruct worldwide routes of invasion: the case of the Asian ladybird Harmonia axyridis. *Molecular Ecology*, **23**, 5979–5997.

Luikart G, Allendorf F, Cornuet J-M *et al.* (1998) Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity*, **89**, 238–247.

Maafi ZT, Sturhan D, Subbotin SA *et al.* (2006) *Heterodera persica* sp. n. (Tylenchida: Heteroderidae) parasitizing Persian hogweed *Heracleum persicum* (Desf. ex Fisch.) in Iran. *Russian Journal of Nematology*, **14**, 171–178.

Marini L, Battisti A, Bona E *et al.* (2012) Alien and native plant life-forms respond differently to human and climate pressures. *Global Ecology and Biogeography*, **21**, 534–544.

Marrs RA, Sforza R, Hufbauer RA (2008) Evidence for multiple introductions of *Centaurea stoebe micranthos* (spotted knapweed, Asteraceae) to North America. *Molecular Ecology*, **17**, 4197–4208.

McConnell R, Middlemist S, Scala C *et al.* (2007) An unusually low microsatellite mutation rate in *Dictyostelium discoideum*, an organism with unusually abundant microsatellites. *Genetics*, **177**, 1499–1507.

McKay JK, Latta RG (2002) Adaptive population divergence: markers, QTL and traits. *Trends in Ecology & Evolution*, **17**, 285–291.

Merilä J, Crnokrak P (2001) Comparison of genetic differentiation at marker loci and quantitative traits. *Journal of Evolutionary Biology*, **14**, 892–903.

Messing RH, Wright MG (2006) Biological control of invasive species: solution or pollution? *Frontiers in Ecology and the Environment*, **4**, 132–140.

Moberg A (1968) *Kopparverket i Kåfjord. Ett Bidrag til Nordkalottens Historia*. Norrbottens Museum, Luleå.

Morin PA, Leduc RG, Archer FI *et al.* (2009) Significant deviations from Hardy-Weinberg equilibrium caused by low levels of microsatellite genotyping errors. *Molecular Ecology Resources*, **9**, 498–504.

Newman D, Pilson D (1997) Increased probability of extinction due to decreased genetic effective population size: experimental populations of *Clarkia pulchella*. *Evolution*, **51**, 354–362.

Nielsen JP (1995) *Altas Historie. Bind 2. Det Arktiske Italia (1826-1920)*. Alta kommune, Alta.

Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov Chain Monte Carlo approach. *Genetics*, **158**, 885–896.

Nielsen C, Ravn HP, Nentwig W *et al.* (eds) (2005) *The Giant Hogweed Best Practice Manual. Guidelines for the Management and Control of an Invasive Weed in Europe*. Forest and Landscape Denmark, Hoersholm.

Niinikoski P, Korpelainen H (2015) Population genetics of the invasive giant hogweed (*Heracleum* sp.) in a northern European region. *Plant Ecology*, **216**, 1155–1162.

Often A, Graff G (1994) Skillekarakterer for kjempebjørnekjeks *Heracleum mantegazzianum* og tromsøpalme *H. laciniatum*. *Blyttia*, **52**, 129–133.

Øvstedal DO (1987) Er tromsøpalma sit namn *Heracleum persicum* Desf.? *Polarflokken*, **11**, 25–26.

Paetkau D, Slade R, Burden M *et al.* (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology*, **13**, 55–65.

Park SDE (2001) *Microsatellite Tools*. Deaprtment of Genetics, Trinity College, Dublin, Ireland.

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research —an update. *Bioinformatics*, **28**, 2537–2539.

Pemberton RW, Ferriter AP (1998) Old world climbing fern (*Lygodium microphyllum*), a dangerous invasive weed in Florida. *American Fern Journal*, **88**, 165–175.

Perglová I, Pergl J, Pyšek P (2007) Reproductive ecology of *Heracleum mantegazzianum*. In: *Ecology and Management of Giant Hogweed (Heracleum Mantegazzianum)* (eds Pyšek P, Cock MJW, Nentwig W *et al.*), pp. 55–73. CAB International, Wallingford, UK.

Pimentel D (2011) Introduction: non-native species in the world. In: *Biological Invasions: Economic and Environmental Costs of Allien Plant, Animal, and Microbe Species* (ed. Pimentel D), pp. 3–8. CRC Press, Boka Raton, USA.

Piry S, Luikart G, Cornuet JM (1999) BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity*, **90**, 502–503.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Pyšek P, Lambdon PW, Arianoutsou M *et al.* (2009) Alien vascular plants of Europe. In: *Handbook of Alien Species in Europe* (ed. Drake JA), pp. 43–61. Springer, Dordrecht, Netherlands.

Pyšek P, Pergl J, Jahodová Š *et al.* (2010) The hogweed story: invasion of Europe by large *Heracleum* species. In: *Atlas of Biodiversity Risk* (eds Settele J, Peven L, Georgiev T, *et al.*), pp. 150–151. Pensoft Publishers, Sofia, Bulgaria.

R Core Team (2014) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences*, **94**, 9197–9201.

Raymond M, Rousset F (1995) GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Rechinger KH (ed.) (1987) *Flora Iranica*. Akademische Druck- u, Verlagsanstalt, Graz, Austria.

Reed DH, Frankham R (2001) How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. *Evolution*, **55**, 1095–1103.

Reich DE, Goldstein DB (1998) Genetic evidence for a Paleolithic human population expansion in Africa. *Proceedings of the National Academy of Sciences*, **95**, 8119–8123.

Rhymer JM, Simberloff D (1996) Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics*, **27**, 83–109.

Rijal D, Falahati-Anbaran M, Alm T *et al.* (2015) Microsatellite markers for *Heracleum persicum* (apiaceae) and allied taxa: Application of next-generation sequencing to develop genetic resources for invasive species management. *Plant Molecular Biology Reporter*, **33**, 1381–1390.

Roderick GK, Navajas M (2003) Genes in new environments: genetics and evolution in biological control. *Nature Reviews Genetics*, **4**, 889–899.

Roman J, Darling JA (2007) Paradox lost: genetic diversity and the success of aquatic invasions. *Trends in Ecology & Evolution*, **22**, 454–464.

Rousset F (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.

Ruxton GD (2006) The unequal variance t-test is an underused alternative to Student's t-test and the Mann-Whitney U test. *Behavioral Ecology*, **17**, 688–690.

Sakai AK, Allendorf FW, Holt JS *et al.* (2001) The population biology of invasive specie. *Annual Review of Ecology and Systematics*, **32**, 305–332.

Schierenbeck K, Ellstrand N (2009) Hybridization and the evolution of invasiveness in plants and other organisms. *Biological Invasions*, **11**, 1093–1105.

Schoonhoven LM, Loon J, Dicke M (2005) *Insect-Plant Biology*. Oxford University Press, Oxford.

Seastedt TR (2015) Biological control of invasive plant species: a reassessment for the Anthropocene. *New Phytologist*, **205**, 490–502.

Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters*, **9**, 615–629.

Sell P, Murrell G (2009) *Flora of Great Britain and Ireland. Volume 3 (Mimosaceae-Lentibulariaceae)*. Cambridge University Press, New York.

Simberloff D (2009) The Role of Propagule Pressure in Biological Invasions. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 81–102.

Simberloff D (2013) *Invasive Species: What Everyone Needs to Know*. Oxford University Press, New York.

Stace C (2010) *New Flora of the British Isles*, 3rd edn. Cambridge University Press, Cambridge.

Stenøien HK, Shaw AJ, Shaw B *et al.* (2011) North American origin and recent European establishments of the amphi-Atlantic peat moss *Sphagnum angermanicum*. *Evolution*, **65**, 1181–1194.

Udupa S, Baum M (2001) High mutation rate and mutational bias at (TAA)n microsatellite loci in chickpea (*Cicer arietinum* L.). *Molecular Genetics and Genomics*, **265**, 1097–1103.

Vilà M, Espinar JL, Hejda M *et al.* (2011) Ecological impacts of invasive alien plants: a meta-analysis of their effects on species, communities and ecosystems. *Ecology Letters*, **14**, 702–708.

Vitousek PM, Walker LR (1989) Biological invasion by *Myrica faya* in Hawai'i: plant demography, nitrogen fixation, ecosystem effects. *Ecological Monographs*, **59**, 247–265.

Vitousek PM, Walker LR, Whiteaker LD *et al.* (1987) Biological invasion by *Myrica faya* alters ecosystem development in Hawaii. *Science*, **238**, 802–804.

Walker NF, Hulme PE, Hoelzel AR (2003) Population genetics of an invasive species, *Heracleum mantegazzianum*: implications for the role of life history, demographics and independent introductions. *Molecular Ecology*, **12**, 1747–1756.

Wang C, Xiao H, Liu J *et al.* (2015) Insights into ecological effects of invasive plants on soil nitrogen cycles. *American Journal of Plant Sciences*, **6**, 34–46.

Ward SM, Gaskin JF, Wilson LM (2008) Ecological genetics of plant invasion: what do we know? *Invasive Plant Science and Management*, **1**, 98–109.

Wasowicz P, Przedpelska-Wasowicz EM, Kristinsson H (2013) Alien vascular plants in Iceland: diversity, spatial patterns, temporal trends, and the impact of climate change. *Flora - Morphology, Distribution, Functional Ecology of Plants*, **208**, 648–673.

Watt MS, Kriticos DJ, Manning LK (2009) The current and future potential distribution of Melaleuca quinquenervia. *Weed Research*, **49**, 381–390.

Williams WI, Friedman JM, Gaskin JF *et al.* (2014) Hybridization of an invasive shrub affects tolerance and resistance to defoliation by a biological control agent. *Evolutionary Applications*, **7**, 381–393.

Wright S (1943) Isolation by distance. *Genetics*, **28**, 114–138.

Yu X, He T, Zhao J *et al.* (2014) Invasion genetics of *Chromolaena odorata* (Asteraceae): extremely low diversity across Asia. *Biological Invasions*, **16**, 2351–2366.

Zavaleta E (2000) The economic value of controlling an invasive shrub. *AMBIO: A Journal of the Human. Environment*, **29**, 462–467.

D.P.R., T.A. and I.G.A. designed the project, obtained funding and participated in the fieldworks. H.K.S. was also involved in sampling design and provided appropriate training on statistical analyses to D.P.R.. S.J. has contributed important samples from England, Finland, Denmark and Iran. D.P.R. performed the laboratory work, analysed the data and wrote the manuscript. All co-authors commented on the manuscript.

Data accessibility

Geographical coordinates and sampling locations are provided in Table 1. DNA and primer sequences of 25 microsatellite markers used in this study are available as supplementary material in Rijal *et al*. (2015) at http://link.springer.com/article/10.1007/s11105-014-0841-y. Final microsatellite genotypes for 25 loci and 575 samples, input files and important analysis scripts are available on Dryad doi: http://dx.doi.org/10.5061/dryad.kg66r

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Relationship between number of alleles per locus and null allele frequency in *Heracleum persicum*.

Fig. S2 The graphical output of structure analysis based on K = 2–4 for (A) Iran, (B) introduced range, (C) Iran-Denmark, (D) Iran-England, (E) global analysis, and (F) Norway. Based on the delta K value, the best K = 3 for Iran and 2 for all other analyses.

Fig. S3 Comparison of four scenarios for introduction history of *Heracleum persicum* based on (A) direct, and logistic regression with (B) raw and (C) LDA transformed summary statistics as implemented in DIYABC (see Fig. 2 and the text for details). X-axes, the number of simulated data closest to the observed; and y-axes, posterior probabilities.

Fig. S4 Principal component analysis of priors (open circle), posteriors (solid circle) and observed data (yellow solid circle) for (A) scenario III and (B) scenario IV (see Fig. 2 and the text for details).

Table S1 Exact test of Hardy-Weinberg equilibrium using a Markov chain with 10 000 demorization steps (1000 batches and 10 000 iterations per batch).

Table S2 Test of linkage disequilibrium using a Markov chain with 10 000 demorization steps (1000 batches and 10 000 iterations per batch).

Table S3 Number of private alleles in the native and the introduced ranges of *Heracleum persicum* along with their frequencies and nearest alleles.

Table S4 Locus wise genetic diversity statistics between native and introduced populations of *Heracleum persicum*.

Table S5 Pairwise population F_{ST} values significantly different from each other after Bonferroni correction are indicated in bold.

Table S6 Estimation of frequency of null alleles in each population and locus by expectation maximization algorithm.

Table S7 Test of the signature of bottleneck following infinite allele model (IAM), two-phase model (TPM) and step-wise mutation model (SMM).

Table S8 Comparison of four historical scenarios of route of introduction of *Heracleum persicum* based on direct and logistic regression with raw and LDA transformed summary statistics as implemented in DIYABC.

Table S9 ABC results for mean relative biases of historical parameters for *Heracleum persicum* based on present data.