



OPEN

# Chloroplast DNA analysis of the invasive weed, Himalayan balsam (*Impatiens glandulifera*), in the British Isles

Daisuke Kurose<sup>1</sup>✉, Kathryn M. Pollard<sup>1</sup> & Carol A. Ellison<sup>1,2</sup>

*Impatiens glandulifera* or Himalayan balsam (HB), is an invasive alien weed throughout the British Isles (BI). Classical biological control of HB in the BI using a rust fungus from the Himalayan native range was implemented in 2014. However, not all HB populations are susceptible to the two rust strains currently released. Additional strains are needed that infect resistant populations in order to achieve successful control. These are best sourced from the historical collecting sites. A molecular analysis was conducted using six chloroplast DNA sequences from leaf material from across the BI and the native range. Herbarium samples collected in the Himalayas between 1881 and 1956 were also included. Phylogenetic analyses resulted in the separation of two distinct groups, one containing samples from the BI and the native range, and the other from the BI only; suggesting that HB was introduced into the BI on at least two occasions. The former group is composed of two subgroups, indicating a third introduction. Ten and 15 haplotypes were found in the introduced and native range respectively, and with two of these found in both regions. Results show where to focus future surveys in the native range to find more compatible rust strains.

*Impatiens glandulifera* Royle (Balsaminaceae), commonly known as Himalayan balsam (HB), is an annual plant native to the foothills of the Himalayas. Introduced as an ornamental due to its showy flowers, HB has become a troublesome invader across Europe and North America<sup>1,2</sup>. The introduction of an invasive species often occurs over large temporal and spatial scales and, as a result introductions are often poorly documented and difficult to trace<sup>3</sup>. For HB, however, there is much consensus that it was first introduced into the Royal Botanic Gardens at Kew (UK) in 1839 from Kashmir<sup>4</sup>. Nevertheless, there is some dispute over the date of introduction with Tanner<sup>5</sup> proposing that HB was introduced prior to 1839, since the collector, John Forbes Royle, then curator of the East India Company's botanical garden in Saharanpur (Northern India), took up a post at King's College London in 1837. However, Morgan<sup>6</sup> states that the species was first introduced into England from Nepal. Regardless, since its introduction, HB has spread rapidly (both naturally and human-mediated) throughout the British Isles (BI), where it is now one of the most prevalent alien plant species. HB typically invades riparian habitats, but is also found in damp woodlands, on waste lands and along railway lines where it can have negative ecological and environmental impacts. Its prolific growth means that this species can rapidly form dense monospecific stands which can outcompete the native flora<sup>7</sup> and decrease the diversity of above and below ground invertebrates<sup>8</sup>. Plants can lower the abundance of arbuscular mycorrhizal communities in the soil<sup>9</sup> and, due to their high sugar nectar production, flowers can draw pollinators away from native species<sup>10</sup>. Being an annual species, regarded as the tallest in the UK, HB dies back in the autumn leaving dead plant material to be incorporated in the water body, as well as bare ground, increasing flood potential and bank erosion<sup>11</sup>.

In 2006, the UK Government initiated a classical biological control (CBC) programme against HB and a rust fungus, *Puccinia komarovii* var. *glanduliferae*, was prioritised due to its prevalence and high level of damage in the native range<sup>12</sup>. The culmination of eight years of research into the life-cycle and safety of the rust formed the basis of a Pest Risk Assessment, and a strain of the rust fungus from India was approved for release by the UK Government in 2014<sup>13</sup>. Field observations and inoculation studies revealed significant variation in the susceptibility of selected HB populations to the rust in England and Wales; many populations were either weakly susceptible or resistant to the Indian strain of the rust<sup>13</sup>. Fortunately, a viable strain of the rust from Pakistan was retrieved

<sup>1</sup>CABI-UK, Bakeham Lane, Egham, Surrey TW20 9TY, UK. <sup>2</sup>Carol A. Ellison is deceased. ✉email: D.Kurose@cabi.org

from storage in liquid nitrogen and inoculation studies conducted under controlled conditions found that this strain was able to infect some of the populations resistant to the Indian strain. Following ministerial approval, the strain from Pakistan was released into the UK in 2017; the Pakistani strain performed well in the field at many of the previously resistant sites, with high levels of leaf infection recorded<sup>14</sup>. Whilst the release of the Pakistani rust strain significantly improved field infection across the UK, during pre-release screening, conducted under controlled glasshouse conditions, HB populations at 30–50% of sites selected for potential new releases between 2018 and 2019 were found to be either immune or resistant to infection by both rust strains<sup>14</sup>. It became necessary, therefore, to identify additional strains of the rust that can infect these resistant HB populations in order to better manage the weed invasions. Biotrophic plant pathogens, such as rusts, have co-evolved with their host plants, and often can only infect one species, and may even be specific to certain genotypes within that species<sup>15</sup>. An incompatible interaction between some weed populations in the invasive range and the introduced pathogen has hindered the success of a number of CBC programmes<sup>16,17</sup>. In order to achieve the best host–pathogen match, and to successfully control HB in the BI, it is important that plant genotypes are matched to the most compatible rust strains. Logically, such strains should be sourced from areas where the original plant collections were made. It is crucial, therefore, to understand the genetic diversity of HB both in the BI and the native range in order to identify dominant genotypes present in the BI and to pinpoint those areas where these match with populations in the native range.

Molecular techniques offer a unique opportunity to enhance the success of CBC and, as such, they are now a prerequisite of many programmes<sup>3,18,19</sup>. The ability to gain insight into the introduction of a species—including information on where introductions originated from, how many times the species was introduced, and the distribution and occurrence of specific genotypes (genetic diversity)—is an invaluable tool which can reduce both the time and costs associated with identifying, matching and therefore prioritising natural enemies for such projects<sup>20,21</sup>. These molecular techniques are now widely used as a tool in CBC programmes against invasive, non-native or alien weeds, where both herbivores and pathogens have been exploited as biological control agents<sup>22–25</sup>. Chloroplast DNA (cpDNA) has frequently been used as a phylogeographical marker for analysing weed invasions<sup>22,26,27</sup>. The typical maternal inheritance of cpDNA in angiosperms means that any genetic structure that was present in the original introductions may be retained, in comparison to nuclear markers that are subject to gene flow via both pollen and seed<sup>22</sup>.

There is much phenotypic variation observed between plants of HB, both within and among populations, which, according to Beerling and Perrins<sup>4</sup>, is determined by abiotic factors. Most obvious is the variation in flower colour, from pink to purple to white, which is observed throughout the BI. To date, a number of molecular studies of HB have been undertaken which all concluded that, despite being introduced into the UK multiple times, HB populations here and in mainland Europe have a relatively low level of genetic diversity, when compared to populations from the native range<sup>28–30</sup>. The level of genetic variation within and between populations of an invasive species has been shown to influence the potential success of biological control<sup>21,31–33</sup>. In this instance, it is important, therefore, to not only include HB leaf samples from UK rust-release sites but also leaf samples from where the two rust strains were collected in the native range.

Our aims in this study are to: (1) clarify how many HB genotypes are present in the BI and understand the distribution of the most dominant genotypes in this region; (2) compare these results with plant samples from India and Pakistan in order to identify where these introductions originated; (3) identify genetically matched haplotypes, particularly for the most dominant genotypes, between the BI and the native range so that future surveys for additional strains of *P. komarovii* var. *glanduliferae* can be better targeted. The results of this study should improve the chances of a more rapid success of the CBC programme against this prolific and ecologically-damaging weed in the BI.

## Results

Our preliminary work showed that variability in the internal transcribed spacer regions, including the 5.8S rDNA (rDNA-ITS) was very low (data not shown), which is consistent with the previous research<sup>30</sup>. The six cpDNA regions (*trnL-trnF*, *atpB-rbcL*, *rps16* Intron, *trnG* Intron, *psbA-trnH*<sup>(GUG)</sup> and *rpl32-trnL*<sup>(UAG)</sup>) were selected due to high variability. The aligned sequences of *trnL-trnF*, *atpB-rbcL*, *rps16* Intron, *trnG* Intron, *psbA-trnH*<sup>(GUG)</sup> and *rpl32-trnL*<sup>(UAG)</sup> were 906, 804, 813, 632, 397 and 1,047 bp in length, respectively. Out of the total of 4,599 bp, there were 42 variable sites including 30 substitutions and 12 indels (Table 1). Variable sites in each region were as follows: *trnL-trnF*, one substitution; *atpB-rbcL*, one substitution; *rps16* Intron, five substitutions and three indels; *trnG* Intron, three substitutions and three indels; *psbA-trnH*<sup>(GUG)</sup>, nine substitutions and two indels; *rpl32-trnL*<sup>(UAG)</sup>, 11 substitutions and four indels. Based on the variable sites as mentioned above, 23 haplotypes (A–W) (eight and 13 haplotypes unique to the introduced and native range, respectively, and two haplotypes which occur in both regions) were found across the 86 individuals (52 populations) of HB sampled. From the 24 populations where samples were analysed from two or more individuals, 22 were fixed for a single haplotype and only two populations in the introduced range were polymorphic with two haplotypes.

A phylogenetic tree for HB was reconstructed by using the Maximum-likelihood (ML) and Bayesian inference (BI) methods with the 23 different haplotype sequences of cpDNA, along with *Impatiens parviflora* and *Cornus controversa* as outgroups (Fig. 1). The ML tree was identical to that reconstructed by the BI method. All the individuals of HB formed a monophyletic group with high bootstrap values and posterior probabilities. Based on the phylogenetic analyses, the 23 haplotypes could be divided into two groups, containing haplotypes from the introduced and native range (Group 1) and from the introduced range of the BI only (Group 2). Although Group 1 forms a monophyletic group, it could be divided into two subgroups (Subgroups 1A and 1B). The majority of samples from Pakistan are included in Subgroup 1A and those from India in Subgroup 1B (Table 2, Fig. 1). Haplotypes J and K belonging to Group 2 diverged earlier than Group 1 with a bootstrap value of 55%.

Haplotype	trnL-trnF										trnG Intron										psbA-trnH										rpl32-trnL <sup>(UAG)</sup>											
	260	74	104	123	195	425	532	533	699	703	9	256	266	278	280	431	22	134	141	146	175	179	180	181	182	186	253	6	128	169	220	288	353	354	366	380	381	382	383	384	522	530
A	A	A	G	G	T	-	-	T	T	C	A	-	-	-	T	A	T	-	-	A	C	T	G	G	T	A	T	A	G	T	T	C	-	-	T	T	A	T	C	C	T	C
B	A	A	G	G	T	-	-	T	T	C	A	-	-	-	T	A	T	-	-	A	C	T	G	G	T	A	T	A	G	T	T	C	A	-	T	T	A	T	C	C	T	C
C	A	A	G	G	T	-	-	T	T	C	A	-	-	-	T	A	T	A	-	A	C	T	G	G	T	A	T	A	G	T	T	C	-	-	T	T	A	T	C	C	T	C
D	A	A	G	G	T	-	-	T	T	C	T	-	T	-	T	C	T	-	-	A	C	T	G	G	T	A	T	G	G	A	-	C	A	-	-	T	A	T	C	C	T	C
E	A	C	G	G	T	-	-	T	T	C	T	-	T	-	T	A	T	A	-	A	C	T	G	G	T	A	T	A	G	A	T	C	A	-	-	T	A	T	C	C	T	C
F	A	C	G	G	T	-	-	T	T	C	T	-	T	-	T	A	T	A	-	A	C	T	G	G	T	A	T	A	G	A	T	C	A	-	-	T	A	T	C	C	T	C
G	A	C	G	G	T	-	-	T	T	C	T	-	T	-	T	A	T	A	-	A	C	T	G	G	T	A	T	A	G	A	T	C	A	-	-	T	A	T	C	C	T	G
H	A	C	G	G	T	-	-	T	T	C	T	-	T	-	T	A	T	A	-	C	C	T	G	G	T	A	T	A	G	A	T	C	A	-	-	T	A	T	C	C	T	C
I	A	C	G	G	T	-	-	T	T	C	T	-	T	-	T	A	T	A	-	A	C	T	G	G	T	A	T	A	T	A	T	C	-	-	T	T	A	T	C	C	T	C
J	A	C	G	G	T	-	-	T	T	C	T	-	-	-	T	A	G	A	-	A	C	T	G	G	T	A	T	A	G	A	T	C	-	-	-	T	A	T	C	C	T	C
K	A	C	G	G	C	-	-	T	A	T	T	-	-	-	T	A	G	A	-	A	C	T	G	G	T	A	T	A	G	A	T	C	-	-	-	T	A	T	C	C	T	C
M	A	C	G	G	T	-	-	T	T	C	T	-	-	-	T	A	T	A	-	A	C	T	G	G	T	A	T	A	G	A	T	C	A	-	-	T	A	T	C	C	T	C
N	A	C	G	G	T	5 bp indel	-	-	T	C	T	-	-	-	T	A	T	A	-	A	C	T	G	G	T	A	T	A	G	A	T	C	A	-	-	T	A	T	C	C	T	C
L	A	C	G	G	T	-	-	T	T	C	A	-	-	-	T	A	T	-	5 bp indel	A	C	T	G	G	T	A	T	A	G	T	T	C	-	-	T	T	A	T	C	C	T	C
O	A	C	G	G	T	-	-	T	T	C	A	-	-	-	T	A	T	A	-	A	C	T	G	G	T	A	T	A	G	A	T	C	A	-	-	T	A	T	C	C	T	C
P	A	C	G	G	T	-	-	T	T	C	A	-	-	-	T	A	T	-	-	A	C	T	G	G	T	A	T	A	G	T	T	C	-	-	T	T	A	T	C	C	T	C
Q	A	C	G	G	T	-	-	T	T	C	T	-	-	-	G	A	T	A	-	A	C	T	G	G	T	A	T	A	G	A	T	C	-	-	-	T	A	T	C	C	G	C
R	A	C	G	A	T	-	-	T	T	C	T	C	-	-	T	A	T	-	-	A	C	T	G	G	T	A	C	A	T	A	T	C	A	A	T	G	G	A	T	A	T	C
S	A	C	A	G	T	-	-	T	T	C	T	-	T	-	T	A	T	-	-	A	C	T	G	G	T	A	T	A	G	A	T	G	-	-	T	T	A	T	C	C	T	C
T	A	C	A	G	T	-	-	T	T	C	T	-	T	A	T	A	T	-	-	A	C	T	G	G	T	A	T	A	G	A	T	G	-	-	T	T	A	T	C	C	T	C
U	A	C	G	G	T	-	-	T	T	C	A	-	-	-	T	A	T	A	-	A	C	T	G	G	T	A	T	A	G	T	T	C	A	A	T	T	A	T	C	C	T	C
V	A	C	G	G	T	-	-	T	T	C	T	-	-	-	T	A	T	A	-	A	C	T	G	G	T	A	T	A	G	A	T	C	-	-	-	T	A	T	C	C	T	C
W	C	C	G	G	T	5 bp indel	-	-	T	C	T	-	T	-	T	A	T	A	-	A	T	A	C	C	A	G	T	A	G	A	T	C	-	-	-	T	A	T	C	C	T	C

**Table 1.** Polymorphic sites in the aligned sequences of the six chloroplast DNA regions in the haplotypes from 52 populations of *Impatiens glandulifera* in the native and introduced range. <sup>a</sup>Indel; ACTAT. <sup>b</sup>Indel; AGTAA.

There are examples from the seven samples from the Natural History Museum, collected from across England more than 100 years ago, in both of the groups and each of the subgroups.

The unrooted network analysis of cpDNA using all haplotypes, in which indels were treated as missing data, is similar with the trees created from the ML and BI methods (Fig. 2). DNA sequences of ACCA in the positions 179–182 at *psbA-trnH* in haplotype W and those of GGATA in the positions 380–384 at *rpl32-trnL*<sup>(UAG)</sup> in haplotype R were reverse complement of TGGT and TATCC, respectively, found in the other haplotypes (Table 1), and therefore these were treated as a single inversion event in the analysis. In this analysis, haplotypes A and L in Subgroup 1A are same as haplotypes B and C, and P and U, respectively. In Subgroup 1B, haplotype E is identical to haplotypes F, M, N, and V. Haplotype S is same as haplotype T. When each indel was treated as a fifth state in the analysis, homoplasy could be increased due to indels (see Supplementary Fig. S1). Based on both the network and phylogenetic analyses, Group 2, containing haplotypes J and K, is located in the outer branch and is likely to be the ancestral lineage of HB. In addition, the phylogenetic analyses suggest that the centre of diversification of HB in the native range is likely to be in the eastern part of the India Himalayas/Western Nepal, with the plant spreading and potentially evolving new genotypes in a north-westerly direction along the Himalayas through India and into Pakistan. Among the 23 cpDNA haplotypes, haplotype E was found to be the most widespread (Figs. 3, 4). Twenty of the 34 populations sampled in the BI region of the introduced range (18 of the 31 populations in the UK and two of the three populations in Ireland) and one of the 18 populations in the native range, Wangat Valley, India, were categorised as haplotype E. The geographical distributions for each haplotype, based on the cpDNA sequences, in the UK showed that haplotype E was most widespread in the south (Fig. 3). Haplotype J, the second most common haplotype, was found in seven populations, mainly concentrated in the north and east of the UK. Haplotype A was distributed in three populations in the introduced range (UK and Ireland) and in two populations in the native range including an unknown site in Kashmir, India. As would be expected in the native range, the haplotypes were found to be more diverse when compared with those in the introduced range (Fig. 4).

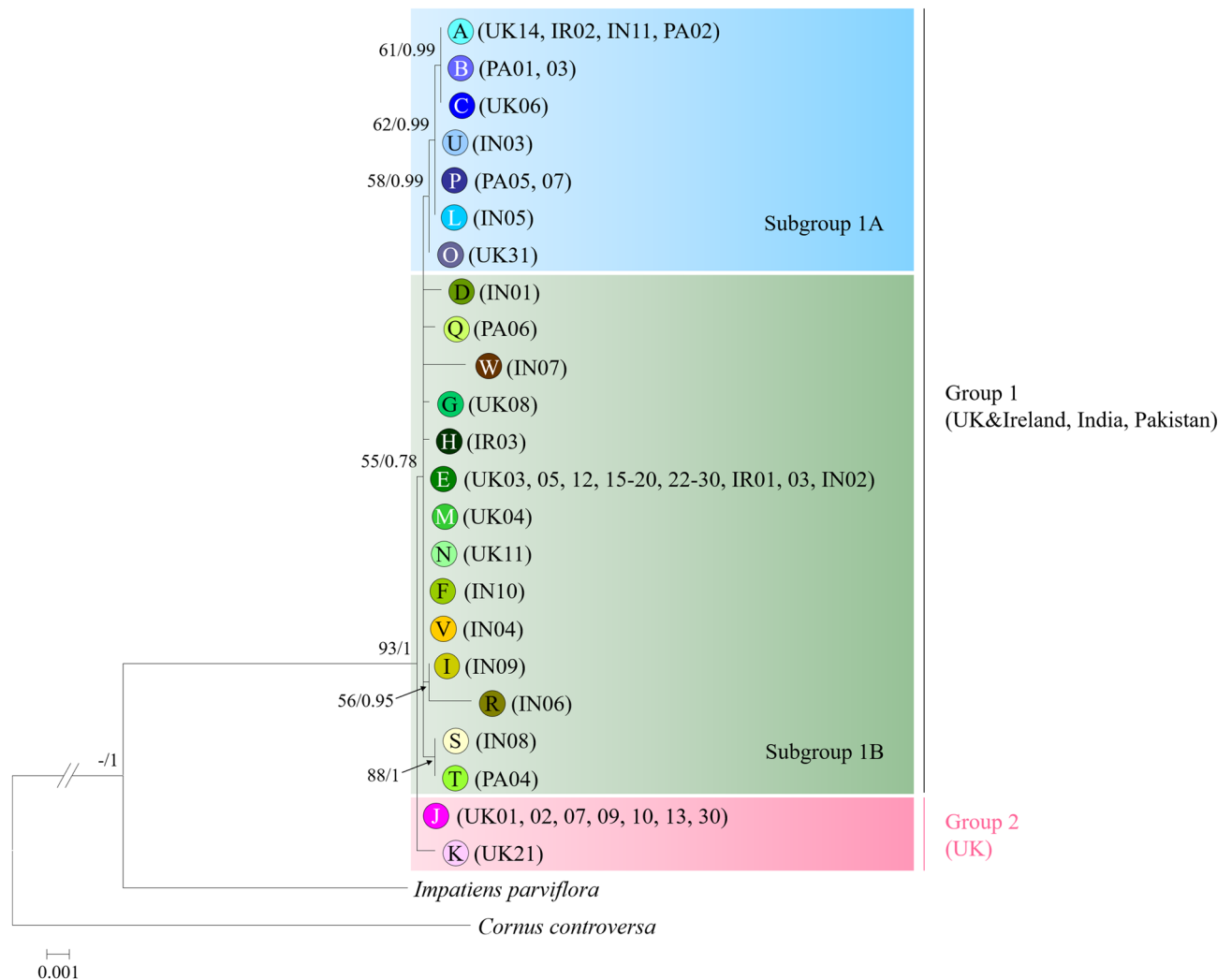
## Discussion

It is important to identify the native range of an invasive weed prior to the start of a CBC programme. However, the centre of origin of the invasive population within the native range can prove to be equally important, particularly if there are high levels of genetic diversity in the target plant in its invasive range. Molecular methods are one of the tools that can be employed to determine the origin of introduced weeds, particularly when they have a wide native range. The success rate of CBC programmes has been shown to significantly improve when natural enemies are collected from populations in the native range which genetically match those in the introduced range<sup>41</sup>.

CpDNA is inherited maternally<sup>42</sup> and lacks recombination<sup>43</sup>. Therefore, data from cpDNA can provide important insights into evolutionary processes in plants such as hybridisation, population structure and phylogeography<sup>44</sup>. A total of 23 haplotypes of HB were found during this study across the introduced and native

Population no.	Location	Country	Latitude	Longitude	Altitude (m)	Collection year	N	Haplotypes	Group/subgroup <sup>e</sup>
<b>Introduced range</b>									
UK01	Killlearn, Stirling	Scotland, UK	56.00297	- 4.35907	-	2016	2	J	2
UK02	Forth, South Lanarkshire	Scotland, UK	56.13263	- 3.93942	-	2016	2	J	2
UK03	River Till, Northumberland <sup>a</sup>	England, UK	55.67944	- 2.18283	-	2016	1	E	1B
UK04	Norham, Northumberland <sup>a</sup>	England, UK	55.73161	- 2.14806	-	2016	2	M	1B
UK05	River Tweed, Northumberland <sup>a</sup>	England, UK	55.67851	- 2.21306	-	2016	1	E	1B
UK06	Harrogate, North Yorkshire <sup>b</sup>	England, UK	54.36659	- 1.92535	-	1914	1	C	1A
UK07	Newbus Grange, Durham <sup>a</sup>	England, UK	54.48018	- 1.50706	-	2016	2	J	2
UK08	Colden Clough, West Yorkshire <sup>a</sup>	England, UK	53.74747	- 2.03029	-	2016	2	G	1B
UK09	Gorpley Clough, West Yorkshire <sup>a</sup>	England, UK	53.70696	- 2.12721	-	2016	2	J	2
UK10	Stanley Marsh, West Yorkshire <sup>a</sup>	England, UK	53.70769	- 1.47883	-	2016	2	J	2
UK11	Meanwood Grove, West Yorkshire <sup>b</sup>	England, UK	53.61275	- 1.51271	-	1908	1	N	1B
UK12	Boston, Lincolnshire <sup>b</sup>	England, UK	53.36102	- 0.23627	-	1945	1	E	1B
UK13	Framingham Pigot, Norfolk	England, UK	52.585	1.36972	-	2016	2	J	2
UK14	Harmondsworth Moor, Middlesex <sup>a</sup>	England, UK	51.493	- 0.48321	-	2016	8	A	1A
UK15	Silwood Park, Berkshire <sup>a</sup>	England, UK	51.40756	- 0.64374	-	2016	4	E	1B
UK16	Sunningdale, Berkshire <sup>a</sup>	England, UK	51.39426	- 0.63567	-	2016	2	E	1B
UK17	Bexley, Kent <sup>a</sup>	England, UK	51.44735	0.16217	-	2016	2	E	1B
UK18	Gatwick, West Sussex <sup>a</sup>	England, UK	51.15257	- 0.19838	-	2016	2	E	1B
UK19	Andover, Hampshire <sup>b</sup>	England, UK	51.20461	- 1.17105	-	1898	1	E	1B
UK20	Loughwood, Devon <sup>b</sup>	England, UK	50.92382	- 4.01043	-	1916	1	E	1B
UK21	Bovey Tracey, Devon <sup>b</sup>	England, UK	50.60359	- 3.64852	-	1918	1	K	1B
UK22	Looe, Cornwall <sup>b</sup>	England, UK	50.50383	- 4.60633	-	1901	1	E	1B
UK23	Nanstallon, Cornwall <sup>a</sup>	England, UK	50.47344	- 4.7692	-	2016	1	E	1B
UK24	Lanivet, Cornwall <sup>a</sup>	England, UK	50.45195	- 4.76513	-	2016	4	E	1B
UK25	Helligan and Grogley Woods, Cornwall <sup>a</sup>	England, UK	50.47995	- 4.79778	-	2016	2	E	1B
UK26	Polmorla, Cornwall <sup>a</sup>	England, UK	50.501	- 4.85685	-	2016	2	E	1B
UK27	Lampeter, Ceredigion <sup>a</sup>	Wales, UK	52.11126	- 4.07258	-	2016	2	E	1B
UK28	Rhosmaen, Carmarthenshire <sup>a</sup>	Wales, UK	51.92316	- 3.97646	-	2016	2	E	1B
UK29	Swansea Vale, Swansea <sup>a</sup>	Wales, UK	51.66526	- 3.9036	-	2016	2	E	1B
UK30	Clyne Valley, Swansea <sup>a</sup>	Wales, UK	51.59975	- 3.99761	-	2016	2	E, J	1B,2
UK31	Unknown <sup>b</sup>	UK	-	-	-	-	1	O	1A
IR01	Carragh, Kildare	Ireland	53.23044	- 6.72179	-	2016	2	E	1B
IR02	Bunclody, Wexford	Ireland	52.65607	- 6.65031	-	2016	2	A	1A
IR03	Kilkenny, Kilkenny	Ireland	52.65206	- 7.24265	-	2016	2	E, H	1B
<b>Native range</b>									
IN01	Tangmarg, Jammu and Kashmir <sup>b</sup>	India	34.0 <sup>d</sup>	74.4 <sup>d</sup>	1,829	1956	1	D	1B
IN02	Wangat Valley, Jammu and Kashmir <sup>b</sup>	India	34.3 <sup>d</sup>	74.9 <sup>d</sup>	-	1940	1	E	1B
IN03	Liddar Valley, Jammu and Kashmir <sup>b</sup>	India	33.7 <sup>d</sup>	75.1 <sup>d</sup>	3,048-3,353	1893	1	U	1A
IN04	Palgam, Jammu and Kashmir <sup>b</sup>	India	34.0 <sup>d</sup>	75.3 <sup>d</sup>	-	1913	1	V	1B
IN05	Sumadah Chun, Ladakh, Jammu and Kashmir <sup>b</sup>	India	33.8 <sup>d</sup>	77.0 <sup>d</sup>	4,115	1947	1	L	1A
IN06	Solang, Solang Valley, Himachal Pradesh	India	32.32293	77.15746	2,450	2009	1	R	1B
IN07	Rohtang, Kullu Valley, Himachal Pradesh <sup>c</sup>	India	32.32983	77.21162	3,067	2009	2	W	1B
IN08	Jibhi, Himachal Pradesh	India	31.57635	77.36196	1,938	2009	1	S	1B
IN09	Dimanjan Dogri, Baspa Valley, Himachal Pradesh <sup>b</sup>	India	31.3 <sup>d</sup>	78.2 <sup>d</sup>	3,810	1939	1	I	1B
IN10	Gauges Valley, Tihri-Gathwal <sup>b</sup> Himachal Pradesh	India	30.7 <sup>d</sup>	78.9 <sup>d</sup>	1,829-2,134	1881	1	F	1B
IN11	Jammu and Kashmir <sup>b</sup>	India	-	-	-	1945	1	A	1A
PA01	Jalkhand, Khyber Pakhtunkhwa	Pakistan	34.97462	73.92942	3,098	2009	1	B	1A
PA02	Batakundi, Khyber Pakhtunkhwa	Pakistan	34.9337	73.75732	2,651	2009	1	A	1A
PA03	Lalazar, Khyber Pakhtunkhwa <sup>c</sup>	Pakistan	34.91687	73.76405	3,130	2009	1	B	1A
PA04	Naran Lake, Khyber Pakhtunkhwa	Pakistan	34.90726	73.67806	2,829	2009	1	T	1B
PA05	Naran, Khyber Pakhtunkhwa	Pakistan	34.86858	73.6088	2,300	2009	1	P	1A
PA06	Kaghan Valley <sup>b</sup>	Pakistan	34.5 <sup>d</sup>	73.3 <sup>d</sup>	-	1954	1	Q	1B
PA07	Thandiani, Hazara <sup>b</sup>	Pakistan	34.2 <sup>d</sup>	73.3 <sup>d</sup>	2,591	1956	1	P	1A

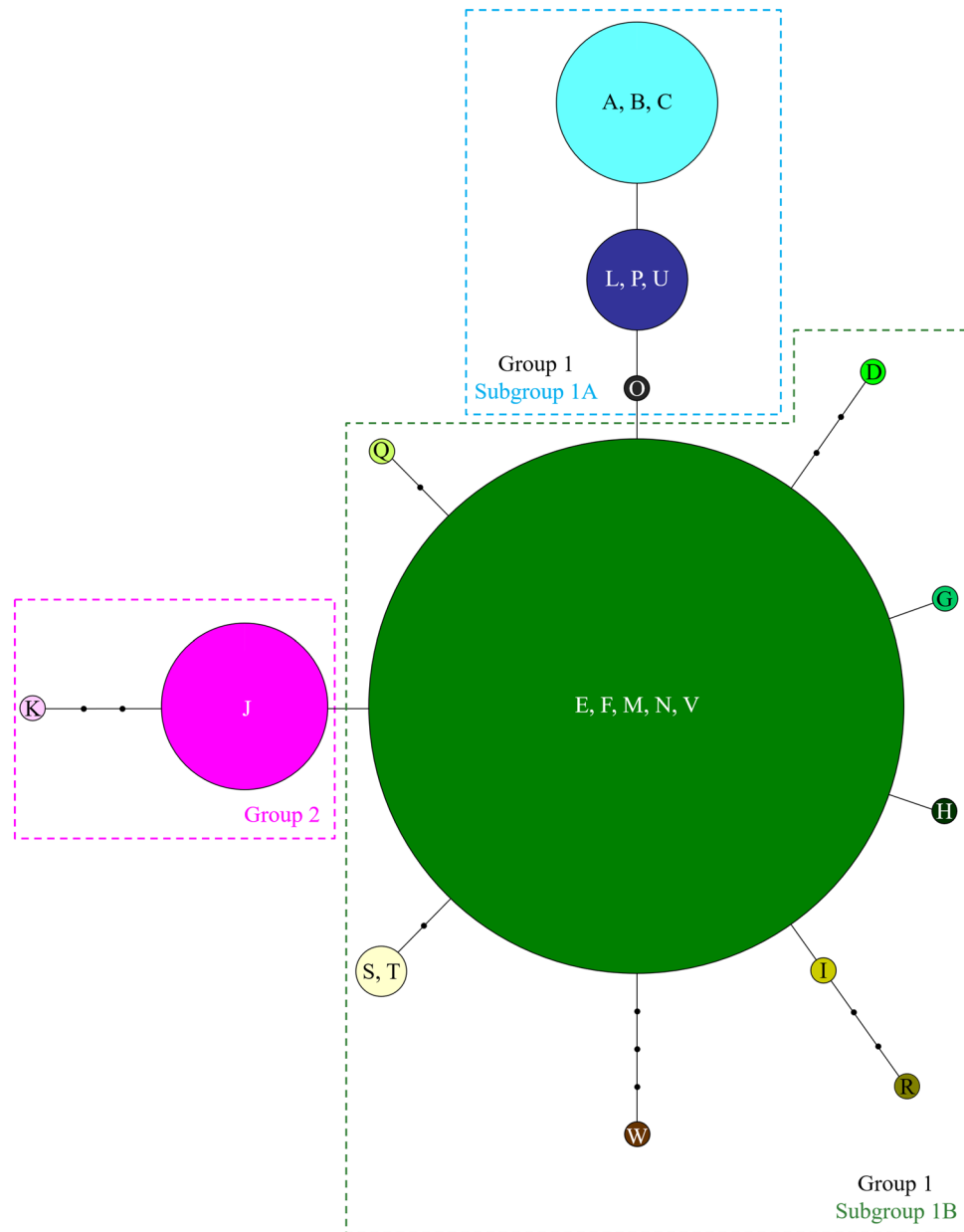
**Table 2.** Site details of the Himalayan balsam populations from the British Isles, India and Pakistan used in the study. <sup>a</sup>Populations where the rust was released<sup>14</sup>. <sup>b</sup>Samples were obtained from the Natural History Museum, UK. <sup>c</sup>Populations where the rust was collected<sup>14</sup>. <sup>d</sup>Data estimated based on location. <sup>e</sup>Groups and subgroups were determined based on the phylogenetic analysis (Fig. 1).



**Figure 1.** Phylogenetic tree based on sequence data of the six chloroplast DNA regions for *Impatiens glandulifera* haplotypes, as constructed by maximum-likelihood (ML) and Bayesian inference (BI) methods. The sequences of *Impatiens parviflora* and *Cornus controversa* was used as the outgroup species. Numbers above branches represent the bootstrap values (> 50% are shown) for ML (Left, 1,000 replicates) and BI analyses (Right).

ranges. Ten of these HB haplotypes were distributed across the 34 populations in the BI and 15 in the 18 native populations. These results corroborate those of Hagenblad et al.<sup>29</sup> and Nagy and Korpelainen<sup>30</sup>, demonstrating that there is more variation in HB in the native range than in the introduced range. HB is native to the foothills of the Himalayas, with populations of the weed occurring in valleys separated by mountain ranges. As a result, populations have evolved in isolation and as shown here, represent different haplotypes.

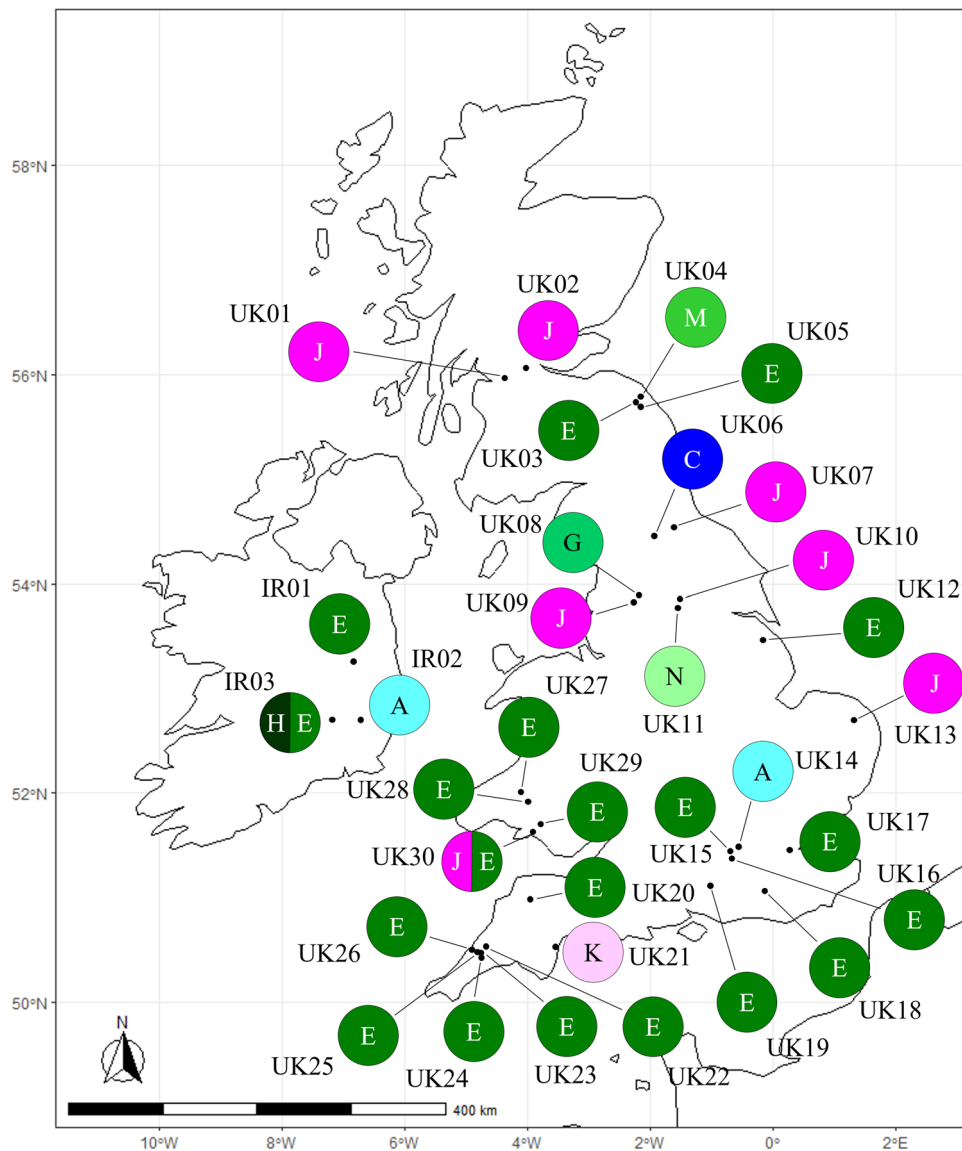
This study demonstrates that haplotypes of HB in the BI and the native range form two distinct genetic groups: Group 1, which can be divided into two subgroups; Subgroup 1A contains two haplotypes which are present in England and Ireland (A and C); Subgroup 1B contains seven haplotypes which are present in the BI (E, G, H, J, K, M, N). Haplotype E is the most common genotype found in the BI and is particularly prevalent in southern England and Wales. Haplotypes from the native range could be divided between the two subgroups, with the majority from Pakistan belonging to Subgroup 1A and those from India to Subgroup 1B. Group 2 consists of two haplotypes which are present in the UK (J and K), but as yet have not been matched to haplotypes in the native range. Haplotype J is predominately distributed in the northern and eastern parts of England and in Scotland. The earliest herbarium samples provided by the Natural History Museum, dating back to before the First World War, gave particular insight into the introduction of HB into the BI. Stately homes and country residences of the aristocracy were significant growers of new plant species brought to the BI by plant hunters and botanists<sup>45</sup> such as John Forbes Royle who described *I. glandulifera*. It is likely that HB would have first been grown on these estates, and indeed the plant can be found in many of them today. These historical samples dating back more than 100 years are likely to represent genotypes with very little genetic divergence from the initial founder populations. Interestingly, our analyses placed these oldest samples into haplotypes from both genetic groups, and the two subgroups suggesting that HB was introduced from the native range to the BI on at least two, but probably three



**Figure 2.** Parsimony network of chloroplast DNA haplotypes of *Impatiens glandulifera*. Each indel was treated as missing data in the analysis. Each link between haplotypes represents one mutational difference. Unlabelled nodes indicate inferred steps not found in the sampled populations. The size of each circle is roughly proportional to the haplotype frequency. The Groups indicated in the dashed boxes correspond to those in Fig. 1.

separate occasions. These herbarium specimens, in addition to those from the native range, can provide significant evolutionary insights into the species, as they represent a genetic snapshot at a particular time and place<sup>46</sup>.

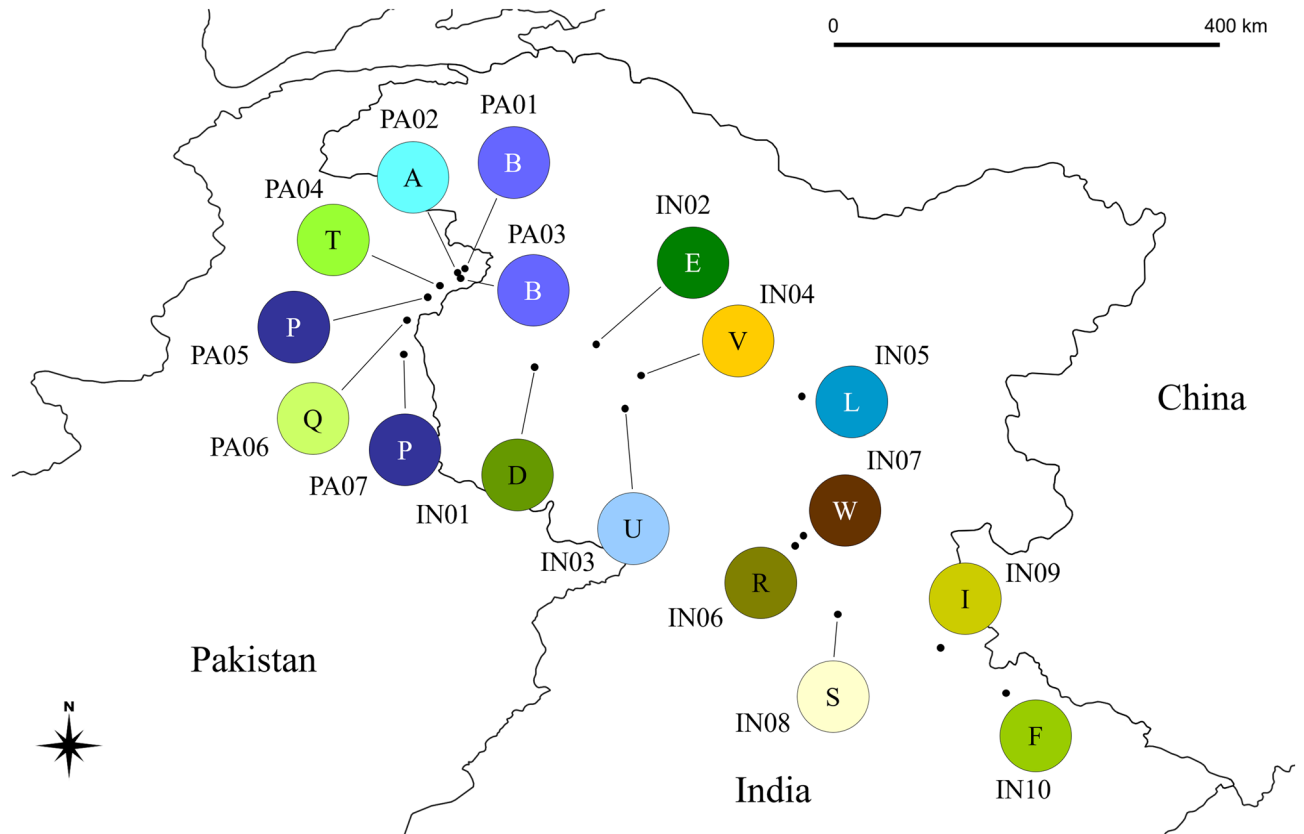
In Group 1, it was possible to match both subgroups to populations in the native range. Subgroup 1A had an identical cpDNA profile (haplotype A) to two populations in the Batakundi, Kaghan Valley of Pakistan (PA02) and in an unknown location in Jammu and Kashmir of India (IN11); whilst Subgroup 1B had an identical cpDNA profile (haplotype E) to the population analysed from the Wangat Valley in Kashmir (IN02). As for those sites with full information, these two regions between the Kaghan Valley and the Wangat Valley are geographically distinct from each other, separated by the Pir Panjal range and Nanga Parbat region, which are high mountain ranges that divide the Vale of Kashmir in India and the North-West Province of Pakistan. However, Group 2 which consists of the haplotypes J and K was not associated with any of the populations included in the study from the native range. The position of Group 2 containing the haplotypes J and K, in the phylogenetic tree (closest to the outgroup) indicates that these haplotypes may be an ancestral lineage and it is possible that Groups 1 and 2 have diversified from north-eastern India and potentially north-western Nepal (the most easterly part of the native range of Himalayan balsam). The theory of the diversification of *I. glandulifera* along the mountain



**Figure 3.** Geographical distribution of the chloroplast DNA haplotypes of *Impatiens glandulifera* in the introduced range. Numbers with the letter UK or IR correspond to the population numbers in Table 2. The letters in the circles correspond to the haplotypes in Figs. 1 and 2 and Tables 1 and 2. Haplotype O (UK31) is not included as its location is unknown. The map was generated using the `ggplot2`<sup>34</sup>, `ggspatial`<sup>35</sup>, `sf`<sup>36</sup>, `rnaturalearth`<sup>37</sup> and `rnaturalearthdata`<sup>38</sup> packages in R version 3.5.1<sup>39</sup>.

slopes of Nepal and India towards Pakistan is supported by a study by Janssens et al.<sup>47</sup>; who concluded that the centre of origin for *Impatiens* species is south-west China. In order to confirm this, additional samples collected more widely from the native range, especially from Nepal, should be included in future studies. In addition, to confirm whether the haplotypes J and K represent an ancestral lineage, due to the low bootstrap value in the phylogenetic analysis shown in this study, other methods such as inter-simple sequence repeats PCR or next-generation sequencing should be used.

The results of this study provide an insight into the genetic diversity of HB in the BI and yield crucial data concerning where to focus future searches for additional compatible strains of the rust *P. komarovii* var. *glanduliferae* in the native range of the species. In a mountainous region such as the Himalayas, it is probable that the rust has evolved with distinct plant biotypes in isolation and that, as such, distinct strains or pathotypes of the rust exist<sup>48, 49</sup>. Consequently, the potential for intraspecific specificity of biological control agents, particularly co-evolved, biotrophic pathogens such as *P. komarovii* var. *glanduliferae*, is significant. The results of the releases of two strains of this rust in England and Wales from the Kullu Valley of India (IN07) and the Kaghan Valley of Pakistan (PA03), have demonstrated the need to consider the introduction of more strains of the rust, since there are weed populations resistant to both strains in the BI<sup>14</sup>. There are numerous examples where matching the genotypes of a target invasive alien weed with the isolate of a fungal biological control agent has proven to be



**Figure 4.** Geographical distribution of the chloroplast DNA haplotypes of *Impatiens glandulifera* in the native range. Numbers with the letter IN or PA correspond to the population numbers in Table 2. The letters in the circles correspond to the haplotypes in Figs. 1 and 2 and Tables 1 and 2. Haplotype A (IN11) is not included as its location is unknown. The map was created using QGIS 3.12.0<sup>40</sup>.

critical: most notably, control of rush skeleton weed, *Chondrilla juncea* in Australia<sup>41</sup>. In this example, a strain of the rust *Puccinia chondrillina* collected from Italy, was released and had a severe impact on one of three forms of the weed in the field; with populations levels being decreased by up to 99%<sup>50,51</sup>. Unfortunately, as not all forms were targeted, due to them not being recognised initially as distinct morphotypes, the distribution of the two other forms increased significantly. This necessitated the need to introduce additional rust strains, more virulent towards the resistant forms of the plants<sup>52</sup> to achieve successful control<sup>41</sup>. This case illustrates clearly the possibility to counter the presence of natural resistance in weed populations by the introduction of new pathogen strains<sup>53</sup>. There are, however, other instances where this appears not to be a factor for successful control, for example, mistflower, *Ageratina riparia*, in Australia, New Zealand and Hawaii<sup>54</sup>. In some cases, an isolate of a pathogen has broad intraspecific specificity, independent of where it was collected from<sup>55</sup>. The susceptibility of HB genotypes to strains of the rust, however, is not clear cut; the rust strain originally collected from India was from the centre location of the native range (IN07), an area where the molecular evidence reported here indicates that none of the BI haplotypes originated. However, this rust strain is able to infect some of the BI populations in England and Wales and has established in the field at a site where haplotype E is the dominant type (UK18)<sup>14</sup>. The release of the rust strain from Pakistan has resulted in the infection of a separate cohort of HB populations<sup>14</sup>, enhancing the success of the CBC programme. In the case of HB, it will be advisable to collect a range of rust strains from throughout the native range, in addition to focusing on the areas matching the likely origin of a plant genotype. Potentially, this will also create more genetic diversity within the introduced populations of the rust making local adaptations to evolving HB populations more likely and rapid<sup>56</sup>.

The sample number used in this study was limited and the number of plants sampled at each site in the BI was low, and may have masked a more significant number of mixed haplotypes at the sites. Only at two sites were the two haplotypes found (UK30 and IR03) and at the four sites where 4–8 plants were sampled only one haplotype was revealed. The results of this research have confirmed previous findings that populations of *I. glandulifera* in the BI have been introduced from both India and Pakistan<sup>30</sup>. The samples from the BI can be organised into two groups, one being composed of two subgroups, based on the sequences of the six cpDNA regions. Although caution should be taken when interpreting these results, they do provide compelling evidence where searches to find additional strains of the rust that are fully compatible with the dominant haplotypes in the BI should be targeted. Of the two native range haplotypes present in the introduced range, one is from the Butakundi, Kaghan Valley (Pakistan) and positioned within Subgroup 1A, whilst the other, from the Wangat Valley, Kashmir (India) is positioned in Subgroup 1B. No haplotypes were found amongst the native range samples that matched UK haplotypes J and K from Group 2. However, the position of the haplotypes J and K on the phylogenetic tree suggests



that these haplotypes may be found in the most easterly part of the native range (where India borders Nepal). Thus, this region may yield rust strains compatible with the HB haplotypes J and K, and should, therefore, also be targeted in future surveys. These additional rust strains would thereby enhance the likelihood of successful biological control of HB in the BI.

## Methods

**Plant material.** *British Isles (BI).* A total of 34 distinct HB populations, including some rust-release sites, were used in this study from across the BI (Table 2). The majority of the leaf samples (26 populations) were collected in 2016 from living plants, dried immediately in silica gel and stored at 4 °C, using the techniques described by Gaskin et al.<sup>18</sup>. The number of plants sampled from each population (*N*) varied from one to eight; although only for three populations were single samples available (UK03, River Till, Northumberland; UK05, River Tweed, Northumberland; and UK23, Nanstallon, Cornwall). For 23 populations, leaf samples were collected from a minimum of two individual plants. In order to check the validity of using limited sample numbers at each site, eight plants were sampled from four separate parts of the HB infestation at Harmondsworth Moor, Middlesex (UK14); and four plants from Silwood Park, Berkshire (UK15) and also Lanivet, Cornwall (UK24).

Leaf samples from seven sites in England were provided by the Natural History Museum (London, UK) from historic herbarium material dating back to before the end of the First World War; collection dates are from 1898 to 1945 (and one sample from the UK from an unknown date and location). The aim was to maximise the chance of obtaining samples from the original introductions, before HB became widespread.

*Himalayan native range.* Eighteen HB populations from the native range (11 from India and seven from Pakistan) were also included in the study (Table 2). Leaf samples from eight populations were taken from CABI's fungal herbarium material that had been collected during surveys to look for natural enemies of HB from 2006–2010. Consequently, leaf material was only available from one plant, apart from at Rohtang, Himachal Pradesh, India (IN07) where two plants were available. The herbarium samples had been dried in a plant press and stored in wax packets at room temperature. The remaining 10 herbarium samples were obtained from the Natural History Museum. These samples included eight Indian samples and two from Pakistan collected 50–100 years ago. In addition, in order to use *I. parviflora* as an outgroup species for phylogenetic analyses, the leaf samples of this species were collected from one plant in Egham (Surrey, UK).

**DNA extraction, PCR amplification and sequencing.** Total genomic DNA was extracted from 10–20 mg of dried leaf tissue, depending on the availability, particularly for the herbarium samples, using DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) or Power Plant Pro DNA Isolation Kit (MO BIO Laboratories Inc., West Carlsbad, USA). Preliminary screening of the rDNA-ITS region as well as 15 non-coding regions of cpDNA using three individuals from geographically distant HB populations (UK, India and Pakistan) observed polymorphisms in six regions of cpDNA, *trnL-trnF* (*trnL* intron, and *trnL* [UAA] 3' exon-*trnF* [GAA] intergenic spacer), *atpB-rbcL* intergenic spacer, *rps16* Intron, *trnG* Intron, *psbA-trnH*<sup>(GUG)</sup> intergenic spacer and *rpl32-trnL*<sup>(UAG)</sup> intergenic spacer (see Supplementary Table S1). Consequently, these six cpDNA fragments were amplified and sequenced for all samples from the introduced and native range. PCR amplifications were undertaken in reaction volumes of 20 µl containing 10 ng of genomic DNA templates, 10 µl of MegaMix-Royal (Microzone, Haywards, UK) and 0.5 µM of each primer or the same reaction volumes containing of 10 ng of genomic DNA templates, 0.5 U KOD Hot Start DNA polymerase (Novagen/Toyobo, Darmstadt, Germany), 1 × PCR Buffer, 0.2 mM of dNTPs, 1.0 mM MgSO<sub>4</sub> and 0.3 µM of each primer. DNA amplification was performed in a Mastercycler (Eppendorf AG, Hamburg, Germany). The PCR conditions were as follows: denaturation at 96 °C for 5 min, followed by 36 cycles of 94 °C for 45 s, 50–60 °C (depending on the primers) for 30 s, and 72 °C for 90 s. PCR products were fractionated in 1.5% (w/v) agarose gels, and visualised using SafeView Nucleic Acid Stain (NBS Biologicals, Huntingdon, UK) and UV illumination. The PCR products were purified with MicroCLEAN (Microzone, Haywards, UK) and directly sequenced bidirectionally using an ABI 3100 Genetic Analyzer (Applied Biosystems, Tokyo, Japan) with a Big Dye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems, Austin, USA) with the same primers used for PCR. All haplotype sequences were deposited in DDBJ/EMBL/GenBank under the accession number LC379653-LC379796.

**Phylogenetic analysis.** Sequence alignments were generated using the MUSCLE algorithm in the program MEGA 7.0.14<sup>57</sup> and manually optimised. In order to confirm that *I. parviflora* became the outgroup species, the sequence of *C. controversa* (Cornaceae), designated as one of the outgroup species, was obtained from the GenBank database. ML and Bif analyses were performed using RAxML version 8.2.9<sup>58</sup> and MrBayes version 3.2.6<sup>59</sup>, respectively, for both the individual data partitions as well as the combined aligned dataset. The best fit nucleotide substitution model was determined using Kakusan 4<sup>60</sup>, which also generates input files for ML and Bif. Best fit models were evaluated using the Akaike Information Criterion (AIC)<sup>61</sup> for ML and the Bayesian Information Criterion (BIC) with significant determined by Chi-square analysis. Equalrate model among regions with GTR + G in AIC and non-partitioned model among regions in BIC were selected. In the ML analyses, the resultant tree was evaluated by bootstrap analysis with 1,000 replicates. In the Bif analyses, two independent runs, each with four Markov chain Monte Carlo were run for 3,000,000 generations. Samples were taken every 100 generations. Burn-in values were estimated using Tracer v1.6<sup>62</sup>. The first 25% of generations were discarded as burn-in and then a 50% majority rule consensus tree was generated from the sampled trees. Phylogenetic trees outputted from the ML and Bif analyses were processed via FigTree v1.4.3<sup>63</sup>. A network tree of the combined cpDNA regions was constructed with a statistical parsimony network using TCS v. 1.21<sup>64</sup>. Insertions of 5 bp observed at

*rps16* Intron and *psbA-trnH* were considered as single mutation events. The analyses were conducted twice, in which each indel was treated as missing data (Fig. 2) and a fifth state (see Supplementary Fig. S1).

## Data availability

Sequence data were deposited in DDBJ/EMBL/GenBank under the accession number LC379653-LC379796.

Received: 28 October 2019; Accepted: 26 May 2020

Published online: 03 July 2020

## References

- Clements, D. R., Feenstra, K. R., Jones, K. & Staniforth, R. The biology of invasive alien plants in Canada. 9. *Impatiens glandulifera* Royle. *Can. J. Plant Sci.* **88**, 403–417. <https://doi.org/10.4141/CJPS06040> (2008).
- Cockell, C. P. & Tanner, R. A. *Impatiens glandulifera* Royle (Himalayan balsam) in *A Handbook of Global Freshwater Invasive Species* (ed Francis, R. A.) 67–77 (Earthscan, London, 2011).
- Prentis, P. J. *et al.* Understanding invasion history: genetic structure and diversity of two globally invasive plants and implications for their management. *Divers. Distrib.* **15**, 822–830. <https://doi.org/10.1111/j.1472-4642.2009.00592.x> (2009).
- Beerling, D. J. & Perrins, J. M. *Impatiens glandulifera* Royle (*Impatiens roylei* Walp.). *J. Ecol.* **81**, 367–382. <https://doi.org/10.2307/2261507> (1993).
- Tanner, R. A. *An Ecological Assessment of Impatiens glandulifera in its Introduced and Native Range and the Potential for its Classical Biological Control*. Ph.D thesis thesis, Royal Holloway, University of London, London (2011).
- Morgan, R. J. *Impatiens: The Vibrant World of Busy Lizzies, Balsams, and Touch-Me-Nots* (Timber Press, Oregon, 2007).
- Hulme, P. E. & Bremner, E. T. Assessing the impact of *Impatiens glandulifera* on riparian habitats: partitioning diversity components following species removal. *J. Appl. Ecol.* **43**, 43–50. <https://doi.org/10.1111/j.1365-2664.2005.01102.x> (2006).
- Tanner, R. A. *et al.* Impacts of an invasive non-native annual weed, *Impatiens glandulifera*, on above- and below-ground invertebrate communities in the United Kingdom. *PLoS ONE* **8**, e67271. <https://doi.org/10.1371/journal.pone.0067271> (2013).
- Pattison, Z. *et al.* Positive plant–soil feedbacks of the invasive *Impatiens glandulifera* and their effects on above-ground microbial communities. *Weed Res.* **56**, 198–207. <https://doi.org/10.1111/wre.12200> (2016).
- Chittka, L. & Schürkens, S. Successful invasion of a floral market. *Nature* **411**, 653. <https://doi.org/10.1038/35079676> (2001).
- Greenwood, P. & Kuhn, N. J. Does the invasive plant, *Impatiens glandulifera*, promote soil erosion along the riparian zone? An investigation on a small watercourse in northwest Switzerland. *J. Soils Sed.* **14**, 637–650. <https://doi.org/10.1007/s11368-013-0825-9> (2014).
- Tanner, R. A. *et al.* *Puccinia komarovii* var. *glanduliferae* var. nov.: a fungal agent for the biological control of Himalayan balsam (*Impatiens glandulifera*). *Eur. J. Plant Pathol.* **141**, 247–266. <https://doi.org/10.1007/s10658-014-0539-x> (2015).
- Varia, S., Pollard, K. & Ellison, C. Implementing a novel weed management approach for Himalayan balsam: progress on biological control in the UK. *Outlooks Pest Manage.* **27**, 198–203. [https://doi.org/10.1564/v27\\_oct\\_02](https://doi.org/10.1564/v27_oct_02) (2016).
- Ellison, C. A., Pollard, K. M. & Varia, S. Potential of a coevolved rust fungus for the management of Himalayan balsam in the British Isles: first field releases. *Weed Res.* **60**, 37–49. <https://doi.org/10.1111/wre.12403> (2020).
- Ellison, C. A., Evans, H. C. & Ineson, J. The significance of intraspecific pathogenicity in the selection of a rust pathotype for the classical biological control of *Mikania micrantha* (mile-a-minute weed) in Southeast Asia. In *Proceedings of the XI International Symposium on Biological Control of Weeds* (eds Cullen, J. M. *et al.*) 102–107 (CSIRO Entomology, Canberra, 2004).
- Smith, L. *et al.* The importance of cryptic species and subspecific populations in classic biological control of weeds: a North American perspective. *Biocontrol* **63**, 417–425. <https://doi.org/10.1007/s10526-017-9859-z> (2018).
- Thrall, P. H. & Burdon, J. J. Host–pathogen life history interactions affect biological control success. *Weed Technol.* **18**, 1269–1274. [https://doi.org/10.1614/0890-037X\(2004\)018\[1269:HLHIAB\]2.0.CO;2](https://doi.org/10.1614/0890-037X(2004)018[1269:HLHIAB]2.0.CO;2) (2004).
- Gaskin, J. F. *et al.* Applying molecular-based approaches to classical biological control of weeds. *Biol. Control* **58**, 1–21. <https://doi.org/10.1016/j.biocontrol.2011.03.015> (2011).
- St. Quinton, J. M., Fay, M., Ingrouille, M. & Faull, J. Characterisation of *Rubus niveus*: a prerequisite to its biological control in oceanic islands. *Biocontrol Sci. Technol.* **21**, 733–752. <https://doi.org/10.1080/09583157.2011.570429> (2011).
- Goolsby, J. A., Van Klinken, R. D. & Palmer, W. A. Maximising the contribution of native-range studies towards the identification and prioritisation of weed biocontrol agents. *Aust. J. Entomol.* **45**, 276–286. <https://doi.org/10.1111/j.1440-6055.2006.00551.x> (2006).
- Nissen, S. J., Masters, R. A., Lee, D. J. & Rowe, M. L. DNA-based marker systems to determine genetic diversity of weedy species and their application to biocontrol. *Weed Sci.* **43**, 504–513. <https://doi.org/10.1017/S0043174500081546> (1995).
- Gaskin, J. F., Zhang, D. Y. & Bon, M. C. Invasion of *Lepidium draba* (Brassicaceae) in the western United States: distributions and origins of chloroplast DNA haplotypes. *Mol. Ecol.* **14**, 2331–2341. <https://doi.org/10.1111/j.1365-294X.2005.02589.x> (2005).
- Morin, L. & Hartley, D. Routine use of molecular tools in Australian weed biological control programmes involving pathogens. In *Proceeding of the XIIIth International Symposium on Biological Control of Weeds* (eds Julien, M. H. *et al.*) (CAB International, Wallingford, 2008).
- Paterson, I. D., Downie, D. A. & Hill, M. P. Using molecular methods to determine the origin of weed populations of *Pereskia aculeata* in South Africa and its relevance to biological control. *Biol. Control* **48**, 84–91. <https://doi.org/10.1016/j.biocontrol.2008.09.012> (2009).
- Sutton, G., Paterson, I. & Paynter, Q. Genetic matching of invasive populations of the African tulip tree, *Spathodea campanulata* Beauv. (Bignoniaceae), to their native distribution: Maximising the likelihood of selecting host-compatible biological control agents. *Biol. Control* **114**, 167–175. <https://doi.org/10.1016/j.biocontrol.2017.08.015> (2017).
- Saltonstall, K. Cryptic invasion by a non-native genotype of the common reed, *Phragmites australis*, into North America. *Proc. Natl. Acad. Sci.* **99**, 2445–2449 (2002).
- Treweek, S. A., Morgan-Richards, M. & Chapman, H. M. Chloroplast DNA diversity of *Hieracium pilosella* (Asteraceae) introduced to New Zealand: reticulation, hybridization, and invasion. *Am. J. Bot.* **91**, 73–85 (2004).
- Love, H. M., Maggs, C. A., Murray, T. E. & Provan, J. Genetic evidence for predominantly hydrochoric gene flow in the invasive riparian plant *Impatiens glandulifera* (Himalayan balsam). *Ann. Bot.* **112**, 1743–1750. <https://doi.org/10.1093/aob/mct227> (2013).
- Hagenblad, J. *et al.* Low genetic diversity despite multiple introductions of the invasive plant species *Impatiens glandulifera* in Europe. *BMC Genet.* **16**, 103. <https://doi.org/10.1186/s12863-015-0242-8> (2015).
- Nagy, A.-M. & Korpelainen, H. Population genetics of Himalayan balsam (*Impatiens glandulifera*): comparison of native and introduced populations. *Plant Ecol. Divers.* **8**, 317–321. <https://doi.org/10.1080/17550874.2013.863407> (2015).
- Barrett, S. Genetic variation in weeds. In *Biological control of weeds with plant pathogens* (eds Charudattan, R. & Walker, H.) 73–98 (1982).
- Burdon, J. & Marshall, D. Biological control and the reproductive mode of weeds. *J. Appl. Ecol.* <https://doi.org/10.2307/2402423> (1981).

33. Charudattan, R. Ecological, practical, and political inputs into selection of weed targets: what makes a good biological control target?. *Biol. Control* **35**, 183–196. <https://doi.org/10.1016/j.biocontrol.2005.07.009> (2005).
34. Wickham, H. *ggplot2: elegant graphics for data analysis* (Springer, New York, 2016).
35. Dunnington, D. *ggspatial: spatial data framework for ggplot2*. <https://cran.r-project.org/package=ggspatial> (2018).
36. Pebesma, E. Simple features for R: standardized support for spatial vector data. *R J.* **10**, 439–446. <https://doi.org/10.32614/RJ-2018-009> (2018).
37. South, A. R natural earth: world map data from natural earth. <https://CRAN.R-project.org/package=rnaturalearth> (2017).
38. South, A. *rnaturalearthdata: world vector map data from Natural Earth used in 'rnaturalearth'*. <https://CRAN.R-project.org/package=rnaturalearthdata> (2017).
39. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org/> (2018).
40. QGIS Development Team. QGIS Geographic Information System. Open Source Geospatial Foundation Project. <https://qgis.osgeo.org/> (2020).
41. Hasan, S. Search in Greece and Turkey for *Puccinia chondrillina* strains suitable to Australian forms of skeleton weed. In *Proceeding of the VI International Symposium on Biological Control of Weeds* (ed. Delfosse, E. S.) 625–632 (Canada Agriculture, Ottawa, 1985).
42. Kress, W. J. *et al.* Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. USA* **102**, 8369–8374. <https://doi.org/10.1073/pnas.0503123102> (2005).
43. Birky, C. W. Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proc. Natl. Acad. Sci. USA* **92**, 11331–11338. <https://doi.org/10.1073/pnas.92.25.11331> (1995).
44. Provan, J., Powell, W. & Hollingsworth, P. M. Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends Ecol. Evol.* **16**, 142–147. [https://doi.org/10.1016/S0169-5347\(00\)02097-8](https://doi.org/10.1016/S0169-5347(00)02097-8) (2001).
45. Fry, C. *The Plant Hunters* (Andre Deutsch, London, 2017).
46. Jukoniene, I., Subkaitė, M. & Ričkienė, A. Herbarium data on bryophytes from the eastern part of Lithuania (1934–1940) in the context of science history and landscape changes. *Botanica* **25**, 41–53. <https://doi.org/10.2478/botlit-2019-0005> (2019).
47. Janssens, S. B. *et al.* Rapid radiation of *Impatiens* (Balsaminaceae) during Pliocene and Pleistocene: result of a global climate change. *Mol. Phylog. Evol.* **52**, 806–824. <https://doi.org/10.1016/j.ympev.2009.04.013> (2009).
48. Barreto, R. W., Ellison, C. A., Seier, M. K. & Evans, H. C. Biological control of weeds with plant pathogens: four decades on. In *Integrated Pest Management: Principles and Practice* (eds Abrol, D. P. & Shankar, U.) 299–350 (CAB International, Wallingford, 2012).
49. Kniskern, J. & Rausher, M. D. Two modes of host–enemy coevolution. *Popul. Ecol.* **43**, 3–14. <https://doi.org/10.1007/PL00012012> (2001).
50. Evans, H. C., Greaves, M. P. & Watson, A. K. Fungal biocontrol agents of weeds. In *Fungi as Biocontrol Agents: Progress, Problems and Potential* (eds Butt, T. M. *et al.*) 169–192 (CABI Publishing, Wallingford, 2001).
51. Cullen, J. M., Kable, P. F. & Catt, M. Epidemic spread of a rust imported for biological control. *Nature* **244**, 463–464 (1973).
52. Hasan, S. A new strain of the rust fungus *Puccinia chondrillina* for biological control of skeleton weed in Australia. *Ann. Appl. Biol.* **99**, 119–124. <https://doi.org/10.1111/j.1744-7348.1981.tb05137.x> (1981).
53. Charudattan, R. & Dinoor, A. Biological control of weeds using plant pathogens: accomplishments and limitations. *Crop Protect.* **19**, 691–695. [https://doi.org/10.1016/S0261-2194\(00\)00092-2](https://doi.org/10.1016/S0261-2194(00)00092-2) (2000).
54. Fröhlich, J. *et al.* Biological control of mist flower (*Ageratina riparia*, Asteraceae): transferring a successful program from Hawai'i to New Zealand. In *Proceeding of the X International Symposium on Biological Control of Weeds* (ed. Spencer, N.) 51–57 (Montana State University, Montana, 2000).
55. Ellison, C. A. & Cock, M. J. W. Classical biological control of *Mikania micrantha*: the sustainable solution. In *Invasive Alien Plants: Impacts on Development and Options for Management* (eds Ellison, C. A. *et al.*) 162–190 (CABI, Wallingford, 2017).
56. Roderick, G. K. & Navajas, M. Genes in new environments: genetics and evolution in biological control. *Nat. Rev. Genet.* **4**, 889–899. <https://doi.org/10.1038/nrg1201> (2003).
57. Kumar, S., Stecher, G. & Tamura, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874. <https://doi.org/10.1093/molbev/msw054> (2016).
58. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033> (2014).
59. Ronquist, F. *et al.* MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542. <https://doi.org/10.1093/sysbio/sys029> (2012).
60. Tanabe, A. S. Kakusan4 and Aminomas: two programs for comparing nonpartitioned, proportional and separate models for combined molecular phylogenetic analyses of multilocus sequence data. *Mol. Ecol. Resour.* **11**, 914–921. <https://doi.org/10.1111/j.1755-0998.2011.03021.x> (2011).
61. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Automat. Control* **19**, 716–723. <https://doi.org/10.1109/TAC.1974.1100705> (1974).
62. Rambaut, A., Suchard, M., Xie, D. & Drummond, A. Tracer v1.6. <https://tree.bio.ed.ac.uk/software/tracer/> (2016).
63. Rambaut, A. FigTree ver. 1.4.3. <https://tree.bio.ed.ac.uk/software/figtree> (2016).
64. Clement, M., Posada, D. & Crandall, K. TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* **9**, 1657–1660. <https://doi.org/10.1046/j.1365-294x.2000.01020.x> (2000).

## Acknowledgements

We are very thankful to H. C. Evans for providing us leaf samples in addition to revising and providing valuable comments on the manuscript. We express our gratitude to R. Tanner, K. Orman, Z. Pattinson, L. O'Neill, J. Callis and J.-R. Baars for collecting the leaf materials and J. Yesilyurt from the Natural History Museum for providing the herbarium material used in this study. We would also like to acknowledge the Indian Government Germplasm Exchange Scheme (Reference FS No. 11/2010), for allowing the inclusion of selected field collected leaf samples in this study. We thank M. Rabiye for her help with laboratory work, H. Hayakawa for his technical and fruitful advice on earlier versions of this manuscript, and M. K. Seier for critical review of the manuscript. This work was financially supported by Natural England and Defra. CABI is an international intergovernmental organisation and we gratefully acknowledge the core financial support from our Member Countries (and lead agencies) including the United Kingdom (Department for International Development), China (Chinese Ministry of Agriculture), Australia (Australian Centre for International Agricultural Research), Canada (Agriculture and Agri-Food Canada), Netherlands (Directorate-General for International Cooperation), and Switzerland (Swiss Agency for Development and Cooperation). See <https://www.cabi.org/about-cabi/who-we-work-with/key-donors/> for full details.

### Author contributions

D.K., K.M.P., and C.A.E. designed the research. D.K. performed the molecular analyses and analysed the data. All authors contributed to the writing and reviewing of the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-67871-0>.

**Correspondence** and requests for materials should be addressed to D.K.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020