

Microsatellite markers for *Heracleum persicum* (Apiaceae) and allied taxa: application of next generation sequencing to develop genetic resources for invasive species management

Dilli P. Rijal^{1,*}, Mohsen Falahati-Anbaran^{2,3}, Torbjørn Alm¹ and Inger G. Alsos¹

¹ Department of Natural Resources, Tromsø Museum, UiT The Arctic University of Norway, 9037 Tromsø, Norway

² School of Biology and Center of Excellence in Phylogeny of Living Organisms, University of Tehran, 14155-6455, Tehran, Iran.

³ Department of Biology, Norwegian University of Science and Technology, 7491 Trondheim, Norway

* Corresponding author: dilli.p.rijal@uit.no; Tel.: +47-776-20786; Fax: +47-776-45520

Abstract: Invasive giant hogweeds are infamous in Europe for causing ecological and economic damage, but genetic resources for their study are limited. We used next generation sequencing to develop a microsatellite library for *Heracleum persicum*, a widespread invasive in Nordic countries. In addition, these markers were cross amplified with the closely related *H. mantegazzianum*, *H. sosnowskyi*, *H. sphondylium*, and the putative hybrid *H. persicum* × *H. sphondylium*, as well as the more distantly related *Anthriscus sylvestris*. We designed and validated 164 primer pairs. A cost-effective PCR approach with modified forward primer, reverse primer and fluorescently labelled universal tail was used to test the functionality of each marker. Twenty-five of thirty markers screened on eight geographically distant samples of *H. persicum* were polymorphic. The number of alleles was 2-4 whereas the expected and observed heterozygosity varied from 0.06-0.84 and 0.0-1.0 respectively. The cross-species amplification efficiency was 84-100%, in which 60-76% of the cross-species amplified markers were polymorphic for *Heracleum* taxa including *H. persicum* × *sphondylium*. Three out of eight of the cross-amplified markers were polymorphic in *Anthriscus sylvestris*. Ordination revealed a clear genetic structure of *Heracleum* taxa. Thus, these markers can serve as important genetic resources for understanding taxonomy, population genetics and phylogeny of giant hogweeds and their hybrids, which in turn, is expected to contribute to invasive species management.

Keywords: invasive species; giant hogweeds; hybridization; microsatellites; universal tails; biodiversity conservation

Introduction

Invasive giant hogweeds and their hybrids can pose serious threats to the native biodiversity and need to be managed at local and regional scales. However, morphological similarity, taxonomic uncertainty, and hybridization impede the management process. In such cases, understanding population genetics of the species is helpful. Studies of the population genetics of giant hogweeds are limited due to paucity of genetic resources

(Henry et al. 2009; Jahodova et al. 2007; Walker et al. 2003). Among different genetic markers, microsatellites are one of the most popular markers for the study of genetic structure and mixture analysis due to their high allelic diversity and possibility of amplification in other closely related species (Guichoux et al. 2011). On the other hand, lack of sequence information for non-model organisms has stymied microsatellite development, but the emergence of next-generation sequencing has now opened the door for genetic exploration of organisms (Guichoux et al. 2011) like

giant hogweeds which have broad impact on biodiversity, ecology and socio-economy of the invaded country (Hejda et al. 2009; Pimentel et al. 2005; Thiele and Otte 2007).

The invasive giant hogweeds include *Heracleum persicum* Desf. ex Fisch., *H. mantegazzianum* Sommier & Levier and *H. sosnowskyi* Mandenova (Nielsen et al. 2005). The taxonomy of giant hogweeds is disputed, for example, the northern Norwegian giant hogweed has been classified as *H. panaces*, *H. giganteum*, *H. laciniatum*, and *H. tromsøensis* (Alm 2013). Recently, Jahodova et al. (2007) suggested *H. persicum* as the name for the northern Norwegian giant hogweed based on amplified fragment length polymorphism (AFLP). *H. persicum* was further used by Fröberg (2010) in the 'Flora Nordica', however, taxonomy of the northern Norwegian giant hogweed is still unclear (Alm 2013). Similarly, the taxonomy of English giant hogweed is controversial. In a recent flora, Sell and Murrell (2009) described *H. trachyloma* Fisch. & C. A. Mey., *H. lehmanianum* Bunge and *H. grossheimii* Manden. ex. Grossh. from the United Kingdom and have treated *H. mantegazzianum* as synonym. However, Stace (2010) retained *H. mantegazzianum* as the valid name for giant hogweed of the British Isles. These uncertainties clearly show that *Heracleum* is a taxonomically complex genus.

Giant hogweeds were famous in the European gardens during the 19th century due to their spectacular appearance. However, within less than two centuries of their introduction, they have become prominent and problematic invasive species (Alm 2013; Elvebakk 1992; EPPO 2009). *Heracleum persicum*, commonly known in Norway as tromsøpalme, is a native of Persia and invasive in Nordic countries. This herbaceous polycarpic species is generally 2 m tall and sometimes reaches up to 3 m with large leaves of up to 2.5 m long. The giant hogweed, *H. mantegazzianum*, derives from western Caucasus and has aggressively colonized most of the Europe (EPPO 2009). *H. mantegazzianum* is monocarpic and can grow up to 5 m with leaves reaching up to 3 m. Similarly, *H. sosnowskyi*, commonly known as, Sosnowsky's hogweed, is a native of

Caucasus and Transcaucasia which has colonized most of the Baltic region (EPPO 2009; Nielsen et al. 2005). This species, which grows up to 3 m tall, is also monocarpic and morphologically closer to *H. mantegazzianum*. All hogweed species are perennial, seed-propagated plants. These three giant hogweeds have damaged natural systems of geographically distinct areas (EPPO 2009), and there is still risk of further expansion and invasion into adjacent areas.

In addition, giant hogweeds often hybridize with *H. sphondylium* L. (common hogweed), a smaller plant which reaches up to 1.4 m in height (Fröberg 2010). *Heracleum sphondylium* is considered indigenous to most European countries. Scattered natural hybrids of giant hogweeds with native *H. sphondylium* have been reported from the British Isles (Sell and Murrell 2009; Stace 2010) as well as in Scandinavia (Fröberg 2010). Several well established populations of natural hybrids have been observed in northern Norway which are sometimes morphologically very similar to *H. persicum*, but without its characteristic anise smell (Alm 2013). Norwegian hybrids are as invasively vigorous as *H. persicum* (Alm 2013), and biological systems face even stronger challenges due to the combined threat of giant hogweeds and hybrids.

As a consequence of their large size and toxic sap, giant hogweeds have rarely received priority during plant collection, which in turn, has always been hindering the plant identification process (Fröberg 2010). In such cases, microsatellite markers can serve as a useful tool to resolve taxonomy. However, microsatellite markers have not been reported so far for *H. persicum*, *H. sosnowskyi*, and the putative hybrid *H. persicum* × *sphondylium*. Our aims were to design and test a suite of novel microsatellite markers for identification, population genetic studies, and hybrid characterization of giant hogweeds. Specifically, we consider *H. persicum* as a model species for designing and testing microsatellite markers that also cross-amplify *H. sosnowskyi*, *H. mantegazzianum*, *H. sphondylium* and hybrids. To extend the application of this approach, we also test cross-generic transferability to *Anthriscus sylvestris*

from Apiaceae. Finally, we evaluate taxonomic applications of the validated markers.

Materials and Methods

Plant material

Leaf samples were collected from up to eight individuals per population, if possible at 5–10 m intervals, and dried in silica gel (Table 1 and Table 2). The leaf samples, DNA extracts, and voucher specimens are deposited at the Tromsø Museum (TROM).

DNA extraction and standardization

DNA was extracted using a DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) following manufacturer's protocol. The quality of DNA was checked by running 2% agarose gels. DNA concentrations were measured by NanoDrop 2000 (Thermo Scientific, Waltham, USA), and all the samples were normalized to 10 ng/μl for PCR.

Microsatellite library preparation and primer design

The facility available at GENOSCREEN (Lille, France) was used to prepare libraries and design primers. High quality DNA of ten individuals from Norwegian, Danish and English populations of *H. persicum* were pooled for microsatellite library preparation. The pooled DNA was fragmented and hybridized with eight probes namely: TG, TC, AAC, AAG, AGG, ACG, ACAT and ACTC to enrich the DNA fragments containing microsatellite repeats. Following the protocol described in Malausa et al. (2011), the enriched microsatellite DNA libraries were further amplified by PCR using high fidelity Taq (Roche, Indiana, USA) to increase their concentrations. The concentrations of the libraries were measured for quality assurance prior to sequencing by 454-GS-FLX Titanium chemistry (Malausa et al. 2011). QDD software (Megléczy et al. 2010) with default settings was used to design the microsatellite primers.

Biological validation and multiplex optimization

We selected 30 pairs of primers from the validated set with wide ranges of expected PCR products to screen *Heracleum* samples. We chose primers targeting di, tri and tetra nucleotide microsatellite repeats with less than 15 total repeats. Multiplex manager (Holleley and Geerts 2009) was used to assess the minimum number of primers for multiplexed reactions. Eight individuals of *H. persicum* from different populations covering its entire range were used for the initial screening of the markers (Table 1). Four universal primers suggested by Blacket et al. (2012) were fluorescently labelled with PET, NED, HEX and 6-FAM, and all forward primers were modified in their 5' end with universal tails (Table 3). The universal primer remains inactive during 1st and 2nd cycle of the PCR; however, it acts as a locus specific primer from the 3rd PCR cycle onwards (Blacket et al. 2012). Each reaction contained 1 pmol of forward primer and 10 pmol each of reverse and universal primers. The total volume of each reaction was 6 μl including 1 μl of standardized template DNA, 1 μl of primer mix, 3 μl of 2x Type-it master mix and 1 μl of RNA free water (Type-it Microsatellite PCR Kit, Qiagen). The cycling conditions for singleplex PCR were as follows: initial denaturation at 95 °C for 15 min followed by 15 cycles of 95 °C for 30s, 64-56.5 °C of touch down PCR for 1 min with 0.5 °C decrease per cycle, and 72 °C for 40s; 17 cycles of 95 °C for 30s, 57 °C for 45s, 72 °C for 40s; 8 cycles of 95 °C for 30s, 53 °C for 45s, 72 °C for 40s; and a final extension of 60 °C for 30 min. The overnight holding temperature was set as 4 °C. To find the optimal dilution level, undiluted, 1:10, 1:20 and 1:30 diluted PCR product were used in capillary electrophoresis. Finally, 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 3130xL genetic analyzer (Applied Biosystems).

Table 1 Country of origin and number of samples used for primer testing and multiplex optimization for *Heracleum persicum*

Location	# samples	Country	City	Collection date	Latitude	Longitude	Collector
Bognæsvej	8 ^{ab}	Denmark	Roskilde	2003-07-26	55.6833	12.0333	S. Jahodova
Buckingham Palace	1	England	London	2004-08-31	51.4983	-0.1456	O. Booy
Kensington Garden	4 ^a	England	London	2012-07-04	51.5102	-0.1751	D.P. Rijal
Weckstrominpolku	2	Finland	Helsinki	2004-08-11	60.2558	24.9711	S. Jahodova
Strandvägen	3 ^a	Finland	Tammisaari	2004-07-21	59.9836	23.4152	S. Jahodova
Anbaran	1	Iran	Anbaran	2013-09-27	38.5244	48.4625	M. Falahati-Anbaran
Fandoughlu	2	Iran	Fandoughlu	2013-09-25	38.4159	48.5719	M. Falahati-Anbaran
Kang Road	1 ^a	Iran	Mashad	2005-07-20	36.3611	59.3500	R. Sobhian
Pol-e Zanguleh	1 ^a	Iran	Mazandaran	2013-06-20	36.1918	51.3384	A. Pirany
Hamnegata	1	Norway	Andenes	2012-07-18	69.3220	16.1281	D.P. Rijal, I.G. Alsos, T. Alm
Sørgårdselva	1	Norway	Hårvik	2012-07-20	68.5675	16.3509	D.P. Rijal, I.G. Alsos, T. Alm
Strandlandsvegen	1 ^a	Norway	Nesna	2012-09-22	66.1951	13.0298	L.U. Tveraabak
Talvikbukta	1 ^a	Norway	Talvik	2012-08-26	70.0468	22.9610	D.P. Rijal, I.G. Alsos, T. Alm
Breiviklia	1	Norway	Tromsø	2012-08-31	69.6781	18.9773	D.P. Rijal
Näsvägen	3 ^a	Sweden	Lit	2013-08-21	63.3170	14.8387	D.P. Rijal
Vännäs	1	Sweden	Tväräbäck	2013-08-22	63.9978	19.7241	D.P. Rijal
Pilgatan	1	Sweden	Umeå	2013-08-22	63.8237	20.2783	D.P. Rijal

^a samples used in initial screening and ^b samples used for molecular diversity analysis

Table 2 Country of origin and number of samples used in cross-species amplification and ordination analyses

Taxa	# samples	Country	City	Location	Latitude	Longitude	Collector
<i>H. mantegazzianum</i>	2	England	Woking	Surrey	51.2532	-0.5608	D.P. Rijal
	2	England	Cambridge	Bin Brook	59.6464	10.8362	D.P. Rijal
	2	Norway	Stavanger	Nato	58.9079	5.7238	M.K. Føreid
	8 ^a	Norway	Hedmark	Stavsjø	60.8137	10.8393	S. Lundemo
<i>H. persicum</i>	2	Norway	Tromsø	Breiviklia	69.6780	18.9766	D.P. Rijal
	2	Norway	Honningsvåg	Elvebakken	70.9945	25.9732	D.P. Rijal
	2	Norway	Gryllefjord	Gryllefjord	69.3626	17.0561	D.P. Rijal
	2	Norway	Sandnessjøen	Novikveien	66.0220	12.6355	D.P. Rijal
<i>H. sosnowskyi</i>	8 ^a	Georgia	Khone	Khevsureti	42.5745	45.2365	S. Shetekauri
	1	Russia	Kasplya	Smolenskaya	54.9638	31.6468	G. Konechnaya
<i>H. sphondylium</i>	1	Norway	Bergen	Lagune	60.2935	5.3010	K.B. Shrestha
	1	Norway	Andøya	Gavlen	68.8761	15.4759	D.P. Rijal, I.G. Alsos, T. Alm
	3	Norway	Bodo	Near Train station	67.2871	14.4011	D.P. Rijal
	8 ^a	Norway	Sandtorg	Sandtorg	68.5649	16.5030	D.P. Rijal, I.G. Alsos, T. Alm
Hybrid	8 ^a	Norway	Andenes	Andøya Flyplass	69.3114	16.1131	D.P. Rijal, I.G. Alsos
	2	Norway	Mo i Rana	Båsmoveien	66.3377	14.1142	D.P. Rijal
	1	Norway	Andøya	Gavlen	68.8761	15.4759	D.P. Rijal, I.G. Alsos, T. Alm
	2	Norway	Trondheim	Othilienborgvegen	63.4073	10.4455	D.P. Rijal
	1	Norway	Tromsø	Åsgard-Giværbukta	69.6679	18.9123	D.P. Rijal
<i>Anthriscus sylvestris</i>	4 ^a	Norway	Tromsø	Telegrafbukta	69.6332	18.9073	D.P. Rijal
	4	Norway	Tromsø	Hansine Hansens veg	69.6871	18.9904	D.P. Rijal

^a samples used for molecular diversity analysis

Table 3 List of the functional markers along with primer concentration, annealing temperature and multiplex assignment tested on *H. persicum* (see Table 1). Forward primers were modified by respective universal tails as shown in the table

Markers	Universal tail	Fluorophore	Concentration of forward primer (μM)	Annealing temperature ($^{\circ}\text{C}$)	Multiplex
Hp_06	GCCTCCCTCGCGCCA	6-FAM	0.2	62	I
Hp_08	CGGAGAGCCGAGAGGTG	PET	0.2	62	I
Hp_14	CAGGACCAGGCTACCGTG	NED	0.1	64	I
Hp_16	CAGGACCAGGCTACCGTG	NED	0.1	63	I
Hp_18	CGGAGAGCCGAGAGGTG	PET	0.2	61	I
Hp_20	CGGAGAGCCGAGAGGTG	PET	0.15	59	I
Hp_22	GCCTTGCCAGCCCGC	HEX	0.2	59	I
Hp_23	GCCTTGCCAGCCCGC	HEX	0.25	64	I
Hp_26	GCCTCCCTCGCGCCA	6-FAM	0.25	61	I
Hp_29	GCCTCCCTCGCGCCA	6-FAM	0.2	60	I
Hp_05	GCCTTGCCAGCCCGC	HEX	0.25	62	II
Hp_07	CGGAGAGCCGAGAGGTG	PET	0.2	62	II
Hp_10	CAGGACCAGGCTACCGTG	NED	0.25	62	II
Hp_17	GCCTCCCTCGCGCCA	6-FAM	0.15	62	II
Hp_24	CGGAGAGCCGAGAGGTG	PET	0.1	63	II
Hp_30	CAGGACCAGGCTACCGTG	NED	0.2	60	II
Hp_02	CGGAGAGCCGAGAGGTG	PET	0.2	64	III
Hp_09	GCCTCCCTCGCGCCA	6-FAM	0.2	65	III
Hp_11	CGGAGAGCCGAGAGGTG	PET	0.2	63	III
Hp_12	GCCTCCCTCGCGCCA	6-FAM	0.2	62	III
Hp_13	CAGGACCAGGCTACCGTG	NED	0.2	61	III
Hp_19	GCCTTGCCAGCCCGC	HEX	0.1	58	III
Hp_21	GCCTTGCCAGCCCGC	HEX	0.2	62	III
Hp_25	GCCTTGCCAGCCCGC	HEX	0.2	62	III
Hp_27	CAGGACCAGGCTACCGTG	NED	0.15	58	III

All markers which were successfully amplified in single reactions (see examples in Fig. S1) were further accommodated in three multiplex reactions by using multiplex manager (Holleley and Geerts 2009). The modified forward primer, reverse primer, and fluorescently labelled universal tail were mixed in 1:2:1 ratio (Blacket et al. 2012). The concentrations of forward primers were 0.1 μM for the first screening. Each multiplex PCR reaction consisted of 3 μl of master mix and 0.5 μl of RNA free water (Type-it Microsatellite PCR Kit, Qiagen); 1 μl of primer mix; and 1.5 μl of template DNA. The initial multiplex PCR amplification was performed with identical thermal

conditions listed above. We observed high variation in amplification efficiency between markers. Thus, concentrations of the weaker markers were increased in 0.05 μM increments until a satisfactory amplification was observed. The final concentration of the primers is given in Table 3. This step increased the peak height of weakly amplified loci; however, the amplification was still not uniform across all loci within each multiplex reaction. Several trouble shooting strategies mentioned in the Type-it microsatellite PCR handbook (Anonymous 2009), for example, increasing the annealing time, decreasing the number of PCR cycles, increasing the time for final

extension, etc. were attempted. Finally, PCR conditions of the multiplex were as follows: initial denaturation at 95 °C for 10 min followed by 10 cycles of 95 °C for 30s, 60-50 °C of touch down PCR for 1 min with 1°C decrease per cycle, and 72 °C for 45s; 25 cycles of 95 °C for 30s, 50 °C for 1 min, 72 °C for 45s; and a final extension of 60 °C for 15 min. Five samples of *H. persicum* each from Norway, Iran, England, Finland, Sweden, and Denmark (Table 1) were screened for functional markers in three multiplex reactions as indicated in Table 3 (also see Fig. S2-S4).

Cross-species amplification

Cross-species amplification was performed on eight distant samples for each species (Table 2). Initial screening revealed that the alleles were very close to each other among species. Thus, final cross-species amplification was performed with the conditions as described for *H. persicum*. Cross-generic amplification tests were also performed on eight samples of *Anthriscus sylvestris* collected from Tromsø, Norway. We selected 25 markers, which were well amplified in *Heracleum*, for this test keeping all other PCR conditions identical with *H. persicum*. The details on the number of samples for cross-amplification are given in the Table 2.

Data analysis

The fragments were further analyzed in Geneious (v. 6.1.6, Biomatters) for allele calling. The data were checked for genotyping errors due to null alleles, large allele drop outs, and stutter peaks in Micro-Checker v. 2.2.3 (Van Oosterhout et al. 2004). Allele size range was calculated based on the distant samples of each taxa used in the initial screening. GeneAIEx v. 6.5 (Peakall and Smouse 2012) was used for molecular diversity analysis, i.e. the number of allele (N_A), expected (H_E) and observed heterozygosity (H_O) as well as Shannon's information index (I), of eight samples from a single population of each taxon, except *A. sylvestris* which was based on four samples. Arlequin v. 3.5.1.3 (Excoffier and Lischer 2010) was used to test Hardy-Weinberg equilibrium. The potential role of validated

primers for resolving *Heracleum* taxonomy was assessed by principal coordinate analysis (PCoA) in GenAIEx v. 6.5 (Peakall and Smouse 2012).

Results and Discussion

Microsatellite library and primer design

The concentration of the library was 1.03×10^{10} molecule/ μ l, thus much higher than the threshold value of 1.46×10^8 molecule/ μ l (GenoScreen, France). A bead recovery value of 88% was also higher than the minimum required value of 65% (Anonymous 2011). The enrichment value was 10%, well within the expected range (Anonymous 2011). Out of the 25951 raw sequences, 3904 sequences contained microsatellite motifs (Online Resource 1). A total of 164 primer pairs (Online Resource 2) were designed for *H. persicum* from sequences which varied from 118-580 bp with an average sequence length of 345.78 bp. The expected product length for the designed primers ranged from 90-320 bp. About 80% of the microsatellite motifs were dinucleotide repeats with 5-15 repeats followed by trinucleotide repeats (17%) with repeat numbers ranging from 5-9. Only four of the loci were tetra-nucleotide with 5-7 repeats. There was a single motif with six nucleotides repeated 5 times.

The differences in salt unadjusted melting temperature (T_m) of 15 primer pairs were between 5 to 8 °C (Online Resource 2); however, 91% of the primer pairs had a melting temperature difference below 5 °C, an important multiplexing criterion (Butler 2005). Similarly, high variation in the range of expected PCR products creates possibilities of accommodating several markers in a single reaction, allowing substantial reduction of project costs (Butler 2005; Guichoux et al. 2011).

Biological validation of primers

Out of the 30 markers selected for biological validation on *H. persicum*, 26 (87%) amplified in singleplex PCR yielding 83% polymorphic markers. One of the markers did not amplify in the multiplex reaction and was thus discarded

Table 4 Characteristics of 25 novel microsatellite markers tested on *Heracleum persicum* (see Table 1). The locus range also includes the length of the respective universal tails (see Table 3). Note: N_A = number of alleles; As= Allele size range; H_E = expected & H_O = observed heterozygosity; and I = Shannon's information index

Loci	Forward primer sequence (5'-3')	Reverse primer sequence (5'-3')	Motif	As (bp)	N _A	H _O	H _E	I
Hp_02	TGCAACAACATAAGACAAAGCC	GCAATCAGCGAGAAGGTGAT	(ATAG) ₆	133-140	2	0.25	0.22	0.38
Hp_05	GGCATCAAAGCAAGTGTC	CCAATGGGTGTTACTTGAGG	(CAA) ₅	118-124	2	0.13	0.12	0.23
Hp_06	ATCTTAATGAAGGCATGAAGCATCT	TCCTCTACAGATTTGGATTTAGTAGG	(GAT) ₅	202-211	2	0.50	0.38	0.56
Hp_07	AATTATAACGCGGTGCTGGC	TGATTACTTGTCTGGTTAGTTCAACA	(TTC) ₅	211-217	2	0.63	0.49	0.69
Hp_08	CAACAAAGCATTCTTAGGTGGA	TGGTTGTGGAAGAAGATGAGG	(TCA) ₅	155-176	1	-	-	-
Hp_09	AGGACGAACGTGTAAGCCAC	ATGCCACCAACTAGACGAC	(GAA) ₆	159-165	2	0.75	0.47	0.66
Hp_10	GATGATCGTGATCGGAGGAC	TTCCATGTTACCAGGGTTGTT	(TGT) ₉	179-200	2	0.38	0.30	0.48
Hp_11	TCTGTAATCTCATTTCTGAACTCATT	AACAGATCGAACCTGCTCTGA	(CTT) ₆	153-174	2	0.13	0.12	0.23
Hp_12	TCATCACATCATCAAGAGGATCA	AACAATTTCTCCATGGCTGC	(TCT) ₅	203-212	4	1.00	0.73	1.35
Hp_13	CAGTAGCAGCATTCCACATGA	TTGGTTTTGGAGCACAGTTG	(CCT) ₅	164-167	1	-	-	-
Hp_14	TCATGTCTACACATTCAATTCCC	GACATCACTCCTACACCTTCCA	(ATG) ₅	209-227	2	1.00	0.50	0.69
Hp_16	AGGAAGAAGAGCAGGATAGTTTTG	ATCAGTATCACTATCATCACCAGC	(ATG) ₅	133-139	1	-	-	-
Hp_17	ATGTGAATGGGCTGGGAAT	CCACTTCCAATCATCTCCTCA	(GGA) ₅	218-227	2	0.63	0.43	0.62
Hp_18	GGATGATATGGGGATGGGTA	TGCATTGTCAAGGTTTCCTG	(TGA) ₆	249-258	3	0.63	0.54	0.92
Hp_19	ACACAAGCAAAGAAAAGAAATTCA	ACACAATCAAAGTTGAAAATTGA	(TCT) ₅	175-181	1	-	-	-
Hp_20	TACTGTGGGATTGGCAATGT	TGGTTAAATTACACAAACGAAGAAA	(GTAT) ₆	185-197	2	1.00	0.50	0.69
Hp_21	TCATTGCTCTTCTTCTCCCC	TATGGATGACTGCTGCAAGG	(TCC) ₅	120-123	2	0.38	0.43	0.62
Hp_22	ATGTTACAAGGTCAAGCCCG	TTACCCACAATCCTGATGA	(GTAT) ₅	118-126	3	0.63	0.62	1.02
Hp_23	GGATGGGAGAGTGAATGAGG	CAGCCACGATACTTGAGCG	(GA) ₁₄	189-207	3	0.25	0.46	0.78
Hp_24	GCTTCAAGGCGATATCTTCG	ATCGATATAACAACCGCCGC	(GA) ₁₃	126-170	4	0.63	0.48	0.92
Hp_25	TGAACACAAACTTCTGCTCAT	TGCGATTGCCTATTGTCTGT	(AC) ₁₃	273-285	2	1.00	0.50	0.69
Hp_26	CGAATTGAGATTATGCAGATCG	TCGCTCTTCTCGATCCTTTC	(GA) ₁₂	160-174	4	0.75	0.54	0.99
Hp_27	TGTAATTTTGTTCCTTCCCTTTG	TCCATTTCTGCATTGGATCA	(TC) ₁₀	190-240	2	0.25	0.22	0.38
Hp_29	AGAAGATGCAACCGGATGAG	ACAACAATACCCTTATTTACATTTGA	(AC) ₁₀	113-121	3	0.75	0.51	0.83
Hp_30	TTGTTTTCCCTTTGTCACC	AATCAACGTGATAAGAAATAAACGA	(TC) ₁₀	109-143	3	0.38	0.32	0.60

Table 5 Cross-species amplification of 25 novel microsatellite markers in *Heracleum* including the putative hybrid *H. persicum* × *spondylium* and *Anthriscus sylvestris* (see Table 2). The locus range also includes the length of the respective universal tails (see Table 3). Note: N_A = number of alleles; As= Allele size range; H_E = expected and H_O = observed heterozygosity

Locus	<i>H. mantegazzianum</i>				<i>H. spondylium</i>				<i>H. sosnowskyi</i>				<i>H. persicum</i> × <i>spondylium</i>				<i>A. sylvestris</i>			
	As (bp)	N _A	H _O	H _E	As (bp)	N _A	H _O	H _E	As (bp)	N _A	H _O	H _E	As (bp)	N _A	H _O	H _E	As (bp)	N _A	H _O	H _E
Hp_02	131-135	1	-	-	-	-	-	-	135-143	3	0.88	0.65	139	1	-	-	-	-	-	-
Hp_05	123	1	-	-	120-123	1	-	-	120-123	2	0.13	0.12	123	1	-	-	118	1	-	-
Hp_06	202-211	2	1.00	0.50	202-208	1	-	-	205	1	-	-	202	1	-	-	-	-	-	-
Hp_07	213	1	-	-	213	1	-	-	213	1	-	-	213-216	2	0.13	0.12	-	-	-	-
Hp_08	171-174	1	-	-	168-177	2	0.25	0.22	171	1	-	-	168-177	3	0.63	0.63	156-171	3	0.50	0.37
Hp_09	161-164	2	0.50	0.47	161-182	1	-	-	161-164	2	0.63	0.49	161	1	-	-	156	1	-	-
Hp_10	185	1	-	-	179-182	2	0.50	0.38	179-188	3	0.50	0.41	179-200	4	0.63	0.70	-	-	-	-
Hp_11	157-160	1	-	-	163-175	2	0.29	0.41	157-160	2	0.13	0.30	160-175	3	0.50	0.40	-	-	-	-
Hp_12	203-209	2	1.00	0.50	203-206	2	1.00	0.50	203-212	3	0.57	0.50	203-209	2	1.00	0.50	212	1	-	-
Hp_13	162	1	-	-	159	1	-	-	162-165	2	0.29	0.24	159-162	2	1.00	0.50	-	-	-	-
Hp_14	206-221	2	1.00	0.50	221-224	2	0.63	0.43	206-221	2	1.00	0.50	206-221	2	0.88	0.49	-	-	-	-
Hp_16	136-139	2	0.75	0.50	133-139	2	0.75	0.47	136	1	-	-	133-139	3	0.50	0.40	-	-	-	-
Hp_17	221	-	-	-	218-230	2	0.43	0.46	221-224	2	0.00	0.48	218-227	3	0.88	0.59	-	-	-	-
Hp_18	256	1	-	-	247-259	2	0.50	0.38	256	1	-	-	247-256	2	0.25	0.47	-	-	-	-
Hp_19	178-181	1	-	-	178	1	-	-	178	1	-	-	178-181	2	0.88	0.49	-	-	-	-
Hp_20	185-193	2	1.00	0.50	185	1	-	-	185-213	2	0.88	0.49	185-193	2	0.38	0.30	-	-	-	-
Hp_21	123	1	-	-	-	-	-	-	120-123	2	0.25	0.22	123	1	-	-	120	1	-	-
Hp_22	124-132	-	-	-	120-124	1	-	-	120-132	2	0.25	0.47	116-124	-	-	-	-	-	-	-
Hp_23	193	-	-	-	199	-	-	-	189-205	5	0.43	0.67	-	-	-	-	207	1	-	-
Hp_24	139-171	3	0.63	0.51	125-133	1	-	-	135-163	4	0.63	0.66	143-159	2	0.00	0.50	-	-	-	-
Hp_25	277-287	2	1.00	0.50	277-289	3	0.75	0.53	277-283	2	1.00	0.50	269-283	2	0.88	0.49	-	-	-	-
Hp_26	163-165	2	0.25	0.22	-	-	-	-	165-171	3	0.25	0.59	163	-	-	-	171-179	3	0.50	0.40
Hp_27	193	1	-	-	-	-	-	-	203-239	6	1.00	0.76	209-227	1	-	-	-	-	-	-

Hp_29	115-119	2	0.13	0.30	107-121	2	0.00	0.32	115-117	2	0.50	0.38	117-123	2	0.00	0.22	116-144	7	1.00	0.75
Hp_30	125	1	-	-	125	1	-	-	125-145	3	0.63	0.46	122-140	1	-	-	-	-	-	-

from further analysis (Hp_04, Online Resource 2). We successfully amplified 10 markers in a single reaction. However, the number of loci per multiplex reaction could be increased if the forward primer is carefully modified. For example, multiplex manager accommodated 17-20 markers in a singleplex reaction. The *in silico* estimation is tractable as designed primers target PCR products of varying size and have similar T_m values. In addition, use of fluorescently labelled universal tails could be highly flexible in terms of re-modification of forward primers and cost reduction (Blacket et al. 2012). If some markers do not amplify with the given modification, one can re-modify the forward primer using another universal tail. Sometimes this strategy improves the primer amplification. In such cases, one can re-order the newly modified forward primer and continue to use the previously ordered reverse primer, thus reducing the extra cost for reverse primers.

Polymorphism in *Heracleum persicum*

Out of the 25 functional markers, 21 were polymorphic for the Danish population of *H. persicum*. However, all markers were polymorphic for more than one population. The Shannon's information index ranged from 0.23-1.35 among polymorphic loci (Table 4). The number of alleles varied from 2-4. The expected and observed heterozygosities ranged from 0.12-0.73 and 0.13-1.0 respectively. All the loci were found to be in Hardy-Weinberg equilibrium (HWE). There was no evidence of null alleles, large allele drop-outs and scoring errors due to stutter bands. Observed and expected heterozygosities for *H. persicum* are comparable with the closely related *H. mantegazzianum* (Henry et al. 2008; Walker et al. 2003) except for a few loci which are completely heterozygous in this study. The number of alleles that we found for Danish *H. persicum* is remarkably lower than in the English population of *H. mantegazzianum* (Walker et al. 2003).

Cross-species amplification

Ten loci were polymorphic for *H. mantegazzianum* collected from a single Norwegian population (Table 5).

Additionally, 15 markers were polymorphic for more than one population (see allele size range, Table 5). The number of alleles varied from 2-3 depending on the polymorphic marker. The expected and observed heterozygosity ranged from 0.22-0.51 and 0.13-1.0, respectively. The number of alleles for *H. sphondylium* ranged from 2-3 among 10 polymorphic loci. Observed heterozygosity ranged from 0.25-1.0 whereas expected heterozygosity varied from 0.22-0.53. Similarly, the number of alleles in *H. sosnowskyi* ranged from 2-6 for 19 polymorphic loci. The expected and observed heterozygosity ranged from 0.12-0.76 and 0.13-1.0, respectively (Table 5). The number of alleles was 2-4 for the putative hybrid *H. persicum* × *sphondylium* for 15 polymorphic loci with observed heterozygosity ranging from 0.13-1.0. The expected heterozygosity ranged from 0.12-0.7. Similarly, 2-7 alleles were present in *A. sylvestris*. Expected and observed heterozygosity varied from 0.36-0.76 and 0.29-0.86, respectively. All the markers are bi- or tri-allelic for *H. mantegazzianum* and *H. sphondylium*, which contrasts, previous studies that have reported 2-10 alleles for *H. mantegazzianum* and 2-8 alleles for *H. sphondylium* for single populations (Henry et al. 2008; Walker et al. 2003). The differences may be due to the higher number of samples used in those studies.

The cross-species amplification efficiencies for multiple populations varied from 84-100% for functional markers, and the polymorphism ranged from 60% (*H. sphondylium* and *H. mantegazzianum*) to 76% (*H. sosnowskyi* and hybrid). These results indicate that the set of designed primers could be used for morphologically distant species of *Heracleum*. There are about 70 species of *Heracleum* distributed primarily in Asia and Europe. Some species are important ingredients of traditional medicines in China (Fading and Watson 2005) and Persia (Alm 2013). Thus, the given set of primers could possibly be used to evaluate the genetic status e.g. of medicinally important plants.

Out of the 25 functional markers, eight loci were amplified in *A. sylvestris* (Table 5). However, only three of the loci were polymorphic. This low level of polymorphism was expected as we used samples from a range of only few

kilometers. Nevertheless, the proportion of polymorphic loci was 37% of the amplified markers. *Heracleum* belongs to the family Apiaceae which includes between 250 and 440 genera and numerous economically important plants (Fading and Watson 2005). Microsatellite markers have also been used for cross-genus amplification (Fan et al. 2013). In our case, 32% of the functional markers amplified in *A. sylvestris*. As *Heracleum* is genetically closer to several other genera than *Anthriscus* (Downie et al. 2000), we assume there is a high possibility to amplify several markers in other genera of Apiaceae.

Taxonomic application

Out of the 25 markers, six were highly polymorphic with the number of alleles ranging from 3-6 (Table 6). These polymorphic markers explain most of the genetic variation in the data. The tested markers provided clear taxonomic resolution as all five taxa were noticeably separated in the PCoA analysis (Fig. 1). The genetic structure of the taxa revealed by PCoA also mimics the general morphological similarity demonstrated by the close position of the morphologically similar *H. mantegazzianum* and *H. sosnowskyi*. The controversial samples from England, which have been named either *H. trachyloma* (Sell and Murrell 2009) or *H. mantegazzianum* (Stace 2010), indeed clustered with other *H. mantegazzianum*. Thus, the tested markers can readily be used to resolve the *Heracleum* taxonomy.

In addition, the tested markers could be used for studying hybridization among different species of *Heracleum*. The putative hybrid *H. persicum* × *sphondylium* was clustered between the two parental species (Fig. 1), which corroborates other studies (Alm 2013). The quick replacement of local flora by a stable hybrid has been observed previously (Alm 2013). This could happen anywhere if different species of *Heracleum* grow in the vicinity, as has e.g. been observed for *H. mantegazzianum* (Fröberg 2010). Our designed primers provide additional applications in tracking invasive hybrids. We observed that

72% of the tested markers were polymorphic for hybrids which implies that a good number of markers could be developed from the given set of untested markers (Online Resource 2).

Table 6 List of the most polymorphic markers with their respective allelic forms

Locu s	Hp_0 6	Hp_0 8	Hp_1 2	Hp_1 7	Hp_2 0	Hp_2 9
107						1
115						1
117						1
119						1
121						1
123						1
168		1				
171		1				
174		1				
177		1				
178						
181						
185					1	
193					1	
202	1					
203			1			
205	1					
206			1			
208	1					
209			1			
211	1					
212			1			
213						1
218				1		
221				1		
224				1		
227				1		
230				1		
Tota						
1	4	4	4	5	3	6

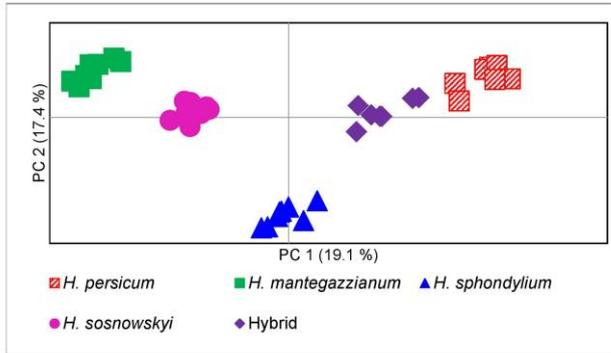


Fig. 1 The first and second axes of PCoA showing genetic structure among *Heracleum* taxa

Conclusions

Identification and population genetics of invasive species are crucial for formulating management plans. This study provides a suite of 25 biologically validated markers as well as 134 newly designed primer pairs which could be used to discriminate *Heracleum* taxa, which has the potential to illuminate population genetics of giant hogweeds and their putative hybrids. This suite of markers could also be used for QTL analysis or genetic mapping to explore new adaptive traits in giant hogweeds as well as in their hybrids. Thus, these markers can contribute to biodiversity conservation through invasive species management. In addition, cross-generic amplification of a few markers indicates the possibility of wider application of the designed markers within the Apiaceae, for example, for phylogenetic studies. Future studies on several

References

- Alm T (2013) Ethnobotany of *Heracleum persicum* Desf. ex Fisch., an invasive species in Norway, or how plant names, uses, and other traditions evolve. *J Ethnobiol Ethnomed* 9 (42):1-12. doi:10.1186/1746-4269-9-42
- Anonymous (2009) Type-it microsatellite PCR hand book: for reliable multiplex PCR-based analysis of microsatellites without the need for optimization. Qiagen, Hilden, Germany. <http://www.qiagen.com/resources/resourcedetail?id=d6135896-0466-4d9d-aeef-1c4f23f8964e&lang=en>. Accessed 15 December 2012

economically important species could benefit from these markers and primers.

Author Contributions

DPR, TA and IGA designed the project, obtained funding, and performed the fieldwork. MFA instructed DPR during primer optimization and laboratory work. DPR performed the laboratory work and wrote the manuscript. All co-authors commented on the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgments

We would like to thank Atefeh Pirany for collecting samples from Iran. We are grateful to Jens-Petter Kjemprud, Embassy of Norway in Iran, for delivering samples from Iran. We would also like to thank Sarka Jahodova for sample contribution. Authors are thankful to all the sample contributors listed in the Table 1 and 2. We would like to thank two anonymous reviewers for their constructive comments which were invaluable while preparing this manuscript. Finally, we would like to thank Matthew Perisin from Department of Ecology and Evolution, University of Chicago for language correction. This project was funded by Tromsø Museum, UiT- the Arctic University of Norway.

- Anonymous (2011) emPCR amplification method manual - Lib L LV. 454 Sequencing, Branford, Connecticut, USA. http://454.com/downloads/my454/documentation/gs-flx-plus/emPCR-Amp-Lib-L-LV-Method-Manual_XLPlus_May2011.pdf. Accessed 07 November 2014
- Blacket MJ, Robin C, Good RT, Lee SF, Miller AD (2012) Universal primers for fluorescent labelling of PCR fragments—an efficient and cost-effective approach to genotyping by fluorescence. *Mol Ecol Resour* 12:456-463. doi:10.1111/j.1755-0998.2011.03104.x

- Butler J (2005) Constructing STR multiplex assays. In: Carracedo A (ed) *Forensic DNA Typing Protocols*, vol 297. *Methods in Molecular Biology*. Humana Press, Totowa, New Jersey, pp 53-65. doi:10.1385/1-59259-867-6:053
- Downie SR, Katz-Downie DS, Watson MF (2000) A phylogeny of the flowering plant family Apiaceae based on chloroplast DNA rpl16 and rpoC1 intron sequences: towards a suprageneric classification of subfamily Apioideae. *Am J Bot* 87:273-292
- Elvebakk A (1992) Has *Heracleum laciniatum* auct. scand. been introduced to North Norway through England? *Polarflokken* 16:313-316
- EPPO (2009) *Heracleum mantegazzianum*, *Heracleum sosnowskyi* and *Heracleum persicum*. EPPO Bulletin 39:489-499. doi:10.1111/j.1365-2338.2009.02313.x
- Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10:564-567. doi:10.1111/j.1755-0998.2010.02847.x
- Fading P, Watson MF (2005) *Heracleum* Linnaeus. In: Zhengyi W, Raven P (eds) *Flora of China (Apiaceae through Ericaceae)*, vol 14. Missouri Botanical Garden Press, St. Louis, pp 194-202
- Fan L et al. (2013) Transferability of newly developed pear SSR markers to other Rosaceae species. *Plant Mol Biol Rep* 31:1271-1282. doi:10.1007/s11105-013-0586-z
- Fröberg L (2010) *Heracleum* L. In: Jonsell B, Karlsson T (eds) *Flora Nordica (Thymelaeaceae to Apiaceae)*, vol 6. The Swedish Museum of Natural History, Stockholm, pp 224-234
- Guichoux E et al. (2011) Current trends in microsatellite genotyping. *Mol Ecol Resour* 11:591-611. doi:10.1111/j.1755-0998.2011.03014.x
- Hejda M, Pyšek P, Jarošík V (2009) Impact of invasive plants on the species richness, diversity and composition of invaded communities. *J Ecol* 97:393-403. doi:10.1111/j.1365-2745.2009.01480.x
- Henry P, Le Lay G, Goudet J, Guisan A, Jahodová Š, Besnard G (2009) Reduced genetic diversity, increased isolation and multiple introductions of invasive giant hogweed in the western Swiss Alps. *Mol Ecol* 18:2819-2831. doi:10.1111/j.1365-294X.2009.04237.x
- Henry P, Provan J, Goudet J, Guisan A, Jahodová Š, Besnard G (2008) A set of primers for plastid indels and nuclear microsatellites in the invasive plant *Heracleum mantegazzianum* (Apiaceae) and their transferability to *Heracleum sphondylium*. *Mol Ecol Resour* 8:161-163. doi:10.1111/j.1471-8286.2007.01911.x
- Holleley CE, Geerts PG (2009) Multiplex Manager 1.0: a cross-platform computer program that plans and optimizes multiplex PCR. *BioTechniques* 46:511-517. doi:10.2144/000113156
- Jahodova S, Trybush S, Pysek P, Wade M, Karp A (2007) Invasive species of *Heracleum* in Europe: an insight into genetic relationships and invasion history. *Divers Distrib* 13:99-114. doi:10.1111/j.1472-4642.2006.00305.x
- Malausa T et al. (2011) High-throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries. *Mol Ecol Resour* 11:638-644
- Megléczy E, Costedoat C, Dubut V, Gilles A, Malausa T, Pech N, Martin J-F (2010) QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics* 26:403-404. doi:10.1093/bioinformatics/btp670
- Nielsen C, Ravn HP, Nentwig W, Wade M (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, Hoersholm, Denmark, pp.44
- 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. doi:10.1093/bioinformatics/bts460
- Pimentel D, Zuniga R, Morrison D (2005) Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecol Econ* 52:273-288. doi:10.1016/j.ecolecon.2004.10.002
- Sell P, Murrell G (2009) *Flora of Great Britain and Ireland. Volume 3 (Mimosaceae-Lentibulariaceae)* vol 3. Cambridge University Press, New York
- Stace C (2010) *New flora of the British Isles*. Third edn. Cambridge University Press, Cambridge
- Thiele J, Otte A (2007) Impact of *Heracleum mantegazzianum* on invaded vegetation and

human activities. In: Pysek P, Cock MJW, Nentwig W, Ravn HP (eds) Ecology and management of giant hogweed (*Heracleum mantegazzianum*). CABI, Wallingford, UK, pp 144-156

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) micro-checker: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4:535-538. doi:10.1111/j.1471-8286.2004.00684.x

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. Mol Ecol 12:1747-1756. doi:10.1046/j.1365-294X.2003.01866.x

Supplementary materials

Online resource 1 Microsatellite library of *H. persicum* with 3904 sequences containing microsatellite motifs

Online resource 2 List of 30 biologically and 134 bioinformatically validated microsatellite markers

List of supplementary figures

Fig. S1 Screen shot of some of the markers amplified in *H. persicum* in a singleplex. PET labelled marker Hp_02 with two alleles at 131.5 and 135.6, Hp_08 with two alleles at 170.6 and 173.5, and Hp_18 with single allele at 256.8 bp (base pair); HEX labelled marker Hp_19 with two alleles at 170 and 180.1, and Hp_05 with one allele at 125 bp; 6-FAM labelled marker Hp_09 with single allele at 159 bp; and NED labelled marker Hp_10 and Hp_16 with single allele each at 187.6 and 135.6 bp respectively (see Table 3 and 4 of the paper for details)

Fig. S2 Screen shot of 10 co-amplified markers of multiplex I in *H. persicum* (see Table 3 and 4 of the paper for details)

Fig. S3 Screen shot of 6 co-amplified markers of multiplex II in *H. persicum* (see Table 3 and 4 of the paper for details)

Fig. S4 Screen shot of 9 co-amplified markers of multiplex III in *H. persicum* (see Table 3 and 4 of the paper for details)