

Rapid identification of chloroplast haplotypes using High Resolution Melting analysis

XIAO-DONG DANG,* COLIN T. KELLEHER,† EMMA HOWARD-WILLIAMS,*† and CONOR V. MEADE*

*Molecular Ecology Laboratory, Department of Biology, National University of Ireland Maynooth, Co. Kildare, Ireland, †DBN Plant Molecular Laboratory, National Botanic Gardens, Glasnevin, Dublin 9, Ireland

Abstract

We have evaluated High Resolution Melting (HRM) analysis as a method for one-step haplotype identification in phylogeographic analysis. Using two adjoined internal amplicons (c. 360 and 390 bp) at the chloroplast rps16 intron (c. 750 bp) we applied HRM to identify haplotypes in 21 populations of two European arctic-alpine herb species *Arenaria ciliata* and *Arenaria norvegica* (Caryophyllaceae). From 446 accessions studied, 20 composite rps16 haplotypes were identified by the melting-curve protocol, 18 of which could be identified uniquely. In a comparative sensitivity analysis with *in silico* PCR-RFLP, only seven of these 20 haplotypes could be identified uniquely. Observed *in vitro* experimental HRM profiles were corroborated by *in silico* HRM analysis generated on uMeltSM. *In silico* mutation analysis carried out on a 360 bp wild-type rps16I amplicon determined that the expected rate of missed single-nucleotide polymorphisms (SNP) detection *in vitro* was similar to existing evaluations of HRM sensitivity, with transversion SNPs being more likely to go undetected compared to transition SNPs. *In vitro* HRM successfully discriminated between all amplicon templates differing by two or more base changes (352 cases) and between 11 pairs of amplicons where the only difference was a single transition or transversion SNP. Only one pairwise comparison yielded no discernable HRM curve difference between haplotypes, and these samples differed by one transversion (C/G) SNP. HRM analysis represents an untapped resource in phylogeographic analysis, and with appropriate primer design any polymorphic locus is potentially amenable to this single-reaction method for haplotype identification.

Keywords: angiosperms, conservation genetics, phylogeography, population genetics—empirical, speciation

Received 16 February 2012; revision received 27 April 2012; accepted 4 May 2012

Introduction

An issue of ongoing concern in haplotype-based phylogeographic analysis is that polymorphism levels in targeted field populations may be underestimated due to insufficient sampling, as rare haplotypes can potentially go unrepresented where less than ten individuals per population are analysed (Bettin *et al.* 2007; Teacher *et al.* 2009), and biased or incorrect conclusions can thus emerge due to insufficiency of data (Petit *et al.* 2005). While the commonly used low-cost technique PCR-RFLP (Taberlet *et al.* 1998) can be applied to large quantities of samples, it cannot detect point mutations that are not covered by restriction enzyme cut sites. Alternatively, exhaustive sequencing of all accession samples in a given analysis is in most cases cost-prohibitive. Thus a more

efficient low-cost method with increased sensitivity to small polymorphic differences would assist greatly in detecting both common and under-represented haplotypes from the field.

High Resolution Melting (HRM) analysis has emerged as a powerful method for genotype identification in short DNA amplicons and currently is applied most often in biomedical analyses (Wittwer 2009). HRM is based on real-time PCR (polymerase chain reaction) techniques, where duplex DNA-binding fluorescent dyes, e.g. LC Green and SYBR Green I, are incorporated into PCR reactions to monitor the progress of DNA amplification (Wittwer *et al.* 1997). The melting process involves the programmed increase of temperature to dissociate the amplified double-strand DNA amplicons, leading to a decrease in the strength of detected fluorescent signals. The melting curve is thus obtained by plotting the decline of fluorescence against real-time increase in temperature (Fig. 1). As the shape of the melting curve and the

Correspondence: Conor Meade, Fax: +353-1-7083845;
E-mail: conor.v.meade@nuim.ie

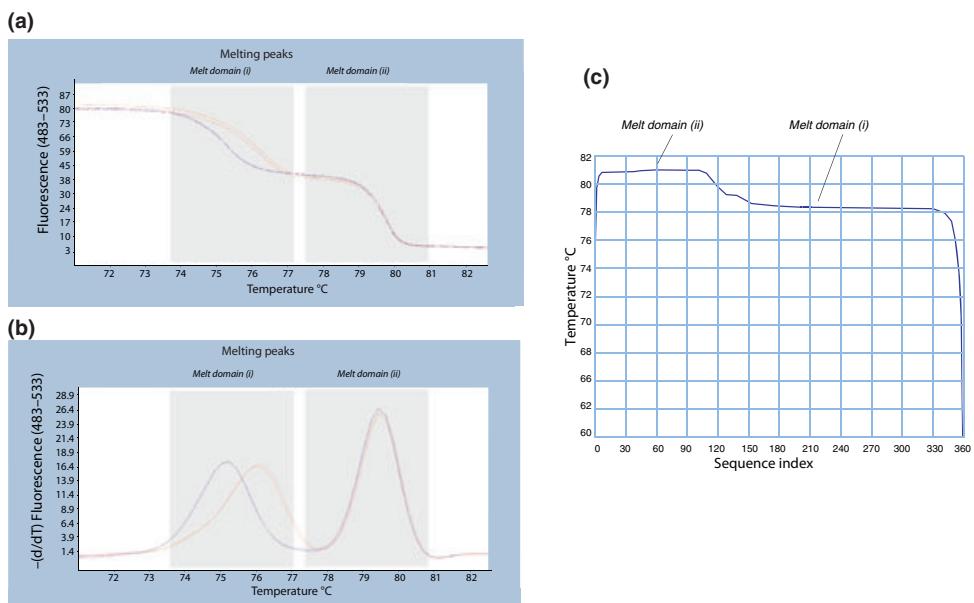


Fig. 1 Derivation of melting peak data. (a) *In vitro* melt-curve profile for two *rps16I* amplicons recording sequential decline in fluorescence caused by dissociation of double-stranded PCR amplicons with increasing temperature. Maximum declines are evident in melting domains (i) and (ii). (b) Calculated melting peaks for these profiles where peak amplitudes indicate temperature point of maximum amplicon dissociation. In this example for *rps16I*, the two principal melting domains give rise to two distinct High Resolution Melting peaks (i and ii). [generated on Roche LC480 system]. (c) Modelled *in silico* dissociation behaviour of *rps16I* amplicon by sequence position and temperature, indicating the template location of two principal melt domains (i and ii). [generated via HRM simulation on uMeltSM].

temperature of maximum dissociation rate (the melting peak, T_m , Fig. 1) are determined by the sequence content of the DNA amplicon involved, melting curve analysis is increasingly used in the identification of variation in sequence content of PCR amplicons (Ririe *et al.* 1997), including studies in population biology (Mader *et al.* 2010; Smith *et al.* 2010), where the exact sequences and expected identities of the studied template alleles have been predefined before HRM analysis is applied.

High Resolution Melting analysis has recently been used to detect unexpected mutations in for example lineage screening in perennial ryegrass breeding (Studer *et al.* 2009), and this advantage points to the potential utility of HRM in phylogeographic analysis where neither the number nor sequence identities of alleles at a particular DNA locus are known.

A key consideration however is the technical limits that apply to HRM, as the size and specific nucleotide content of amplicons influences the discriminating ability of HRM analysis, due to the complex physical-chemical mechanisms of DNA duplex denaturation (Ririe *et al.* 1997; Wittwer *et al.* 2003). In general, for a single point mutation between two genotypes the divergence in T_m values becomes greater when less flanking DNA is included in the targeted amplicon, thus small fragments are preferred for many HRM applications because HRM analysis has less sensitivity in discriminating between larger amplicons (Liew *et al.* 2004). In an experimental

evaluation of HRM sensitivity across a range of amplicon size classes, identification of one single-nucleotide polymorphisms (SNP) between otherwise homologous amplicons was achievable in 100% of cases for product <300 bp in length, but less reliable between 400 and 1000 bp (Reed & Wittwer 2004). In phylogeographic analysis, however, larger DNA regions (300–1000 bp) are routinely sequenced to identify multiple nucleotide polymorphisms (Shaw *et al.* 2005, 2007), one advantage of which is obtaining extensive sequence information with only one pair of primers for each locus. If HRM analysis is to be performed with these loci, internal primers are thus needed to amplify two or more shorter amplicons covering the whole length of the given locus. The challenge for amplicon design is to seek a balance between HRM resolution and sequence coverage.

High Resolution Melting is known to have sensitivity limits for different classes of point mutation between amplicons; A/T substitutions are generally regarded as the most difficult to identify (Reed & Wittwer 2004) though G/C substitutions can also be problematic (Liew *et al.* 2004), and these difficulties persist for larger amplicons. Alternately one useful attribute of larger amplicons is the possibility of more than one T_m melting peak being generated per amplicon, for example, fragments >c. 200 bp can have two or more melting domains, and these additional T_m peaks provide elevated sensitivity for template polymorphisms (Reed & Wittwer 2004;

Vossen *et al.* 2009; Wittwer 2009). The number of melting domains within any given amplicon is determined by its sequence composition, thus some capacity to run an a-priori evaluation of amplicon design and performance would be a key tool for HRM-based haplotype detection in phylogeographic studies.

This tool has recently appeared with the development *in silico* HRM modelling. *In silico* simulation methods have been compiled into the web-based software, uMeltSM (Dwight *et al.* 2011), and with this software it is possible to use a limited number of known sequences as a basis for evaluating the likely sensitivity of HRM for a given range of haplotypes, reviewing the performance of *in vitro* reactions compared to empirical expectations, and proofing the likely HRM sensitivity for certain neutral substitutions within a given amplicon. Taken together these *in silico* applications can assist greatly in helping to optimize stand-alone *in vitro* analyses.

In a general sense, and bearing in mind the need to clarify the various analysis limits associated with HRM, the low unit cost and high resolution of the method point to the potential benefit of a usable HRM protocol for phylogeographic analysis. In the present work we have set out to establish the utility of HRM analysis as a haplotype screening method in plant phylogeography; in this case by analysing chloroplast diversity in populations of two closely related cold-tolerant herbaceous species, *Arenaria ciliata* L. and *Arenaria norvegica* Gunnerus (Caryophyllaceae) which have disjunct distributions across Europe. In this context our objectives have been three-fold: (i) to evaluate the limitations of HRM in terms of sensitivity, repeatability and consistency with regard to the identification of haplotypes at a polymorphic chloroplast DNA locus; (ii) to validate the efficacy of HRM in identifying the range of chloroplast haplotypes that are present in a large biogeographically diverse sample set and (iii) to determine the utility of *in silico* HRM modelling as an aid to *in vitro* HRM experiments and as a tool for visualizing the sensitivity limits of HRM with respect to single point mutations between haplotypes.

Materials and methods

Materials

The two target species, *Arenaria ciliata* L. and *Arenaria norvegica* Gunnerus, are perennial herbs of the family Caryophyllaceae (Angiosperm Phylogeny Group III 2009). *A. ciliata* is an arctic-alpine calcicole herb with several subspecies widely distributed in the high mountains of Europe, with isolated populations in Scandinavia and northwest Ireland. *A. norvegica* is a quasi-arctic species occurring in Norway, Iceland, the Shetland and Faroe Islands, Britain and Ireland (Jalas & Suominen 1983;

Tutin *et al.* 1993). The two species are closely related taxonomically (Wyse Jackson & Parnell 1987), and while both are associated with shallow poorly formed soil on exposed limestone habitats, they are not known to co-occur at any single location. For this study we used 15 populations of *A. ciliata* from throughout Europe and six populations of *A. norvegica* from Ireland, Britain and Iceland, totalling 446 individual accessions and a median population sample size of 26 (Table S1, Supporting information).

DNA preparation and initial assays

Total genomic DNA was extracted from silica-dried leaf tissue using a modified CTAB protocol (Doyle & Doyle 1987) and dissolved in de-ionized H₂O. The target locus for haplotype identification was selected based on comparative sequence analysis of selected *A. ciliata* and *A. norvegica* accessions at five non coding chloroplast loci prompted by Shaw *et al.* (2005, 2007); rpl16, rps16, trnS-trnG, trnT-trnL and rpl32-trnL. Primer sequences and corresponding PCR programmes for all the tested loci were as per Shaw *et al.* (2005, 2007) and primers were synthesized by Applied Biosystems BV. PCR was performed on a PTC-200 Thermal Cycler (MJ Research) with GoTaq[®] Flexi DNA polymerase/buffer system (Promega). Reaction volume was 25 µL containing 1× buffer, 2 mM MgCl₂, 0.2 mM each of the four dNTPs, 0.4 µM each of the forward and reverse primers, 1 µL of DNA template (ranging from 20 to 200 ng DNA), 0.5 U Taq polymerase and de-ionized water. PCR product was screened on a 1.5% agarose gel using SYBR Safe DNA gel stain (Invitrogen). Confirmed PCR products were purified using the mi-PCR Purification Kit (Metabion) and sequenced by Eurofins MWG Operon. Sequencing primer sets were as used for initial PCR. Completed sequences were aligned using BioEdit 7.0.9 (Hall 1999) and compared between samples from different populations. Selection of optimal loci for HRM analysis was based on evident inter population variation and outline size of primed PCR amplicons.

Primer design and locus selection for High-Resolution Melting analysis

The five tested loci varied in size between 640 and 1350 bp in length, thus internal primers were necessary to divide each of the loci into c. 3–400 bp amplicons more suitable for HRM. Primer design rested on two criteria, (i) internal primers must bind to conserved regions of the DNA locus to ensure broad intra-specific sensitivity and (ii) as few internal primers as possible were sought for each locus to maximize the efficiency of HRM analysis. Internal primers were designed with AlleleID 7 (Premier

Biosoft International) using its SYBR® Green Design function. *Cross Species Design* was conducted with aligned locus sequences from different populations of *A. ciliata* and *A. norvegica*. For primer searching, the target amplicon length was set between 250 and 400 bp, the length of primers was set between 18 and 30 bp and the primer *T_a* was set at 55.0 ± 5.0 °C. Considering the requirements for appropriate amplicon lengths, optimal coverage of polymorphic sites and conservation of sequence at primer sites, rps16 was finally selected for this analysis with two pairs of internal primers (rps16I: forward 5'-ATG CTC TTG ACT CGA CAT CTT-3', reverse 5'-GGG TTT AGA CAT TAC TTC GTT GAT T-3', amplicon size c. 360 bp; rps16II: forward 5'-AAG TAA TGT CTA AAC CCA ATG ATT CAA-3', reverse 5'-CGT ATA GGA AGT TTT CTC CTC GTA-3', c. 390 bp; the two internal amplicons are adjacent with 17 bp overlap, and cover c. 750 bp of the rps16 intron).

High-Resolution Melting analysis: in vitro protocols

Real-time PCR protocols were conducted in Roche Light-Cycler® 480 (LC480) Multiwell plates (96-well white) on the LC480 instrument (Hoffman-La Roche, Basel, Switzerland) using the LC480 HRM Master Mix reagent kit (Roche). Real-time PCR programmes and reaction mixes were optimized to generate the required quantities of amplicon necessary for full melting analysis (Table 1). Following DNA quantification of stock sample aliquots using a Nanodrop 1000 spectrophotometer (Thermo Scientific) DNA template samples were diluted to 20–30 ng/μL. HRM analysis was carried out in single-population batches, with the aim of managing inter individual and inter plate errors arising from variation of template concentration and quality across samples from different population extractions. Only population samples IT2, IT3 and FR1 were assayed with other populations because of their small sampling sizes. Each

individual accession from each population (total 446 samples) was run in duplicate or triplicate for HRM analysis, depending on the total sample size in the run. Each real-time PCR reaction was conducted in a 15 μL final volume containing 7.5 μL of the Taq enzyme mix (No. 1 reagent from the Roche HRM Master kit containing Taq polymerase, dNTPs, buffer system and saturating dsDNA binding dye), 1.5 μL of 25 mM MgCl₂ solution (No. 2 reagent from the Roche HRM Master kit, final Mg concentration 2.5 mM), 1.5 μL of primer mix (final concentration 1 μM each), 3.5 μL of de-ionized water (No. 3 reagent from the Roche HRM Master kit) and 1 μL of DNA template (20–30 ng). Fluorescence values for each sample replicate in each PCR cycle were recorded through the SYBR Green (483–533 nm) channel using the default LC480 data acquisition settings.

Interpretation and analysis of High Resolution Melting curve profiles

Peak *T_m* Calling Analysis was performed after each real-time PCR reaction run using the LC480 Software (release 1.5.0 SP3, Version 1.5.0.39) based on which two clear melting peaks were revealed for amplicon rps16I (Fig. 1), while one prominent clear peak, named *T_{m3}*, and a less prominent lower temperature shoulder (*T_{mX}*) were revealed for amplicon rps16II (Fig. 2D). *T_m* values for the obtained melting peaks were given initially by automatic *T_m* calculation under 'SYBR Green I Format' where the number of 'Maximal Peaks' was set as '2 or less' by default, and these values were manually adjusted to mark the temperatures where the melting peaks were at their maximum heights.

Two or three replicates were used to validate the consistency of the recorded rps16I and II amplicon melting profiles for each analyzed accession, depending on the sample analysis run. Each new amplicon type recorded for rps16I and II was assigned an identity (which we have

Table 1 Cycle settings for real-time PCR high resolution melt analysis of amplicons rps16I and II on Roche LightCycler® 480 system

Stage of High Resolution Melting programme	Target temperature (°C)	Duration (mm:ss)	Ramp rate (°C/s)	Number of cycles	Acquisitions/°C	Analysis mode
Pre incubation	95	10:00	4.4	1	—	Quantification
	95	00:15	4.4	—	—	
Amplification	60	00:15	2.2	30	—	Quantification
	72	00:25	1	—	—	
	95	00:05	4.4	—	—	
Melting	65	01:00	2.2	1	—	Melt curve generation
	97	—	0.01	—	50	
Cooling	40	30	1.5	1	—	—

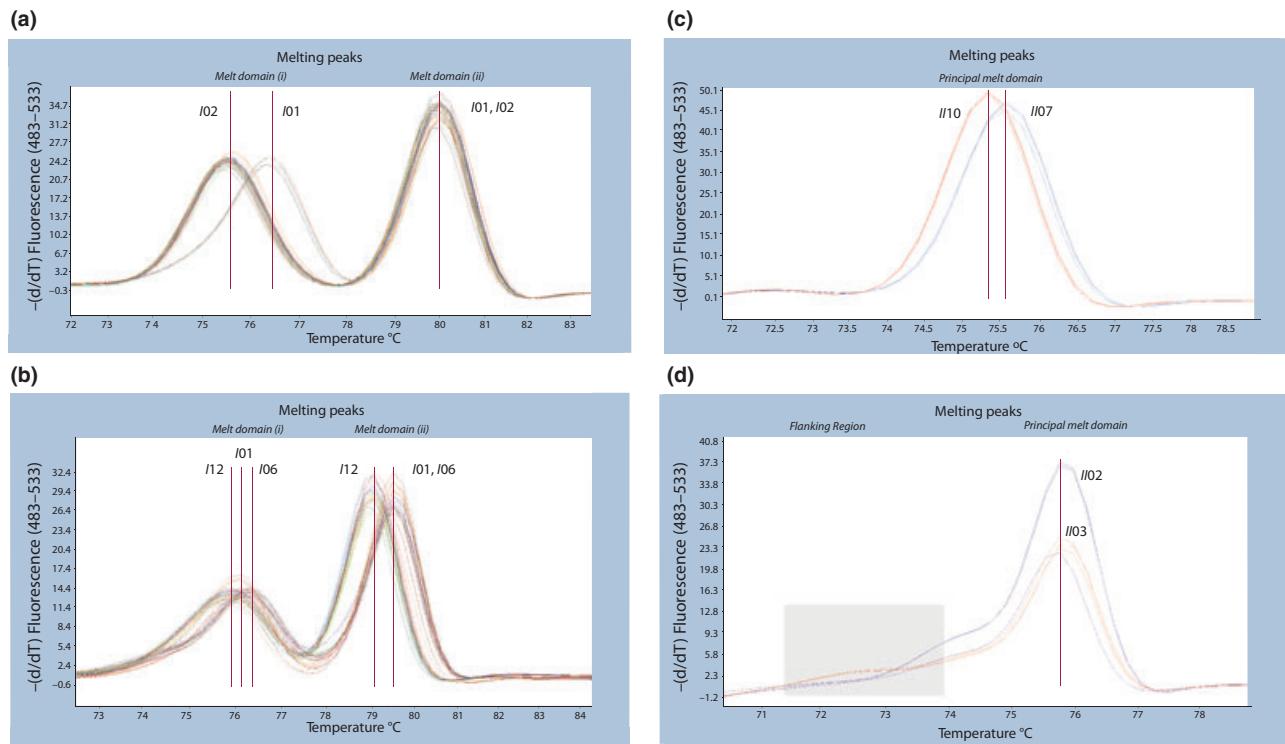


Fig. 2 Sensitivity of *in vitro* High Resolution Melting to haplotype variation. (a) Melting peak data for amplicon rps16I (c. 360 bp) in an Irish population of 30 *Arenaria ciliata* individuals showing amplootypes I01 and I02 with distinct T_{m1} values, arising from one single-nucleotide polymorphisms (SNP), one 6 bp indel and one 7 bp indel difference between templates in melt domain (i) (detailed in Table 2). (b) melting peak data for amplicon rps16I (c. 360 bp) in an Austrian population of 25 *A. ciliata* individuals revealing 3 amplootypes I01, I06 and I12 with distinct T_{m1} and T_{m2} peak values, arising from 6 distinct SNP differences in melt domains (i) and (ii) (detailed in Table 2). (c) Melting peak data for amplicon rps16II (c. 390 bp) showing amplootypes II07 (in blue) and II10 (in red) that deviate by a single C/T SNP in the principal melt domain (detailed in Table 3). (d) The T_{m3} flanking region ' T_{mX} ' in the rps16II amplicon provides additional discriminatory value, in this case distinguishing amploypes II02 [purple, sample Ir1.25] from II03 [orange, sample Ir2.14] which differ by a 13 bp indel that flanks the principal melt domain (detailed in Table 3).

termed ‘amplopotype’), e.g. I01, I02 or II01, II02 etc. Final T_m values were manually validated for each accession based on the standardized curves of the replicates for each amplicon. Two additional analyses were carried out to validate the consistency of the method; (i) single-plate analysis of multiple accessions of one amplopotype drawn from several different populations, to evaluate the relative variation in HRM profiles associated with population-specific factors and (ii), full-plate replicates comprising 60 individual HRM analyses of a single sample accession, to evaluate well-to-well variation across the plate.

Amplopotype confirmation for each sample was carried out using a standard protocol. Within each run batch, samples were first grouped into different hypothesized amplopotypes based on their validated melting curves (T_m values). After initial analysis runs (Table 1), variation of ≥ 0.2 °C in T_m was considered the likely level of divergence which would be evident between variant amplopotypes, with the proviso that ongoing analysis might alter this working assumption (no guidance was available from the literature on haplotype T_m divergence in

amplicons larger than 150 bp). Sample profiles that differed by ≥ 0.2 °C at any of the three T_m peaks or the shoulder region (T_{mX}) beside rps16II T_{m3} were regarded as having potentially distinct amplopotypes, the difference between putative amplopotypes termed ‘ ΔT_m ’.

For each discrete T_m group within each population the amplopotype sequence identity was obtained. Where within-group T_m variance was evident but at < 0.1 °C, two individuals were randomly selected from this group to be sequenced over the whole rps16 region. When within-group variance was between 0.1 and 0.2 °C, the individuals with the lowest and the highest T_m values were chosen to be sequenced to confirm rps16 sequence composition. If different amplopotypes were detected within a pre classified group in the latter manner, one or two additional samples from each new subgroup were chosen for sequencing to confirm the revised amplopotype groupings. To control inter population error, the sampling protocol required that in each population a minimum of two replicates of each putative amplopotype were sequenced, even where the same melting curve profile

had been encountered in another population. Each putative amplicon for both *rps16I* and *II* was sequenced over the full c. 750 bp length of *rps16*, and following comparative alignment of these sequences, the final compilation of composite *rps16* haplotypes was completed.

Analysis runs were carried out on the central 60 wells of the 96-well plate to minimize the possibility of edge-effect variation. In total some 2000 individual HRM reactions for the two *rps16* loci were carried out over 38 runs, covering all 446 individual sample accessions and validation checks/re-runs, where necessary.

In addition to collation of recorded T_{m1} , T_{m2} , T_{m3} and T_{mx} shoulder values for each composite haplotype, a pairwise matrix of inter haplotype ΔT_m values was also created for each of these peaks to provide an overview of expected degree Celsius divergence between all haplotypes that could potentially occur in a single population sample. To further facilitate evaluation of the correlation (if any) between ΔT_m and evolutionary distance between haplotypes (and in particular the extent to which closely related haplotypes can be discriminated using their pairwise ΔT_m) the ΔT_m matrix was supplemented by an identically composed pairwise matrix of Tajima-Nei genetic distance D (Tajima & Nei 1984) between haplotypes, based on sequence composition. A Mantel test (Mantel 1967) was used for correlation analysis between these matrices, with the *ade4* package (Thioulouse *et al.* 1997) within the R software 2.12.2 (R Development Core Team 2011).

High-Resolution Melting analysis: in silico simulation

Evaluation of the utility of in silico HRM modelling as a support for in vitro HRM analysis. For all the amplicons identified in the *in vitro* analysis, obtained sequences were uploaded to uMeltSM (Dwight *et al.* 2011), and subjected to *in silico* simulation of the melting process. uMeltSM simulations applied the thermodynamic settings of Huguet *et al.* (2010), with the temperature range of melting set between 65 and 95 °C and resolution set to maximal resolution at 0.1 °C. Concentration of free Mg²⁺ was set at 2.5 mM to be consistent with *in vitro* analysis, and DMSO was set at 10% to minimize the difference between the absolute *in silico* and *in vitro* T_m values. Derived values for T_{m1} , T_{m2} , T_{m3} and the T_{m3} shoulder (T_{mx}) were recorded for each unique amplicon, and as carried out for the *in vitro* data, pairwise inter amplicon matrices of ΔT_m values at each T_m peak were created. As all *in vitro* data, especially the estimated optimal T_m values for individual amplicons, is subject to experiment-specific environmental factors, some divergence between observed absolute *in vitro* and *in silico* T_m values was expected for each individual amplicon (Fig. 1B,C). Thus to establish whether comparative

analysis of *in silico* HRM data could in this case be used to support *in vitro* work, the underlying pattern of T_m estimation in the two methods was analysed. A Mantel test-based correlation analysis was carried out between *in vitro* and *in silico* data in two ways: (i) direct comparison of the calculated amplicon T_m values from *in vitro* and *in silico* analysis respectively, and (ii) comparison of pairwise inter amplicon ΔT_m matrices derived from *in vitro* and *in silico* analysis (Mantel 1967).

Sensitivity of HRM to single-substitution mutations between amplicons: in silico evaluation of single substitutions in a wild-type template. In order to appraise inter amplicon SNP variation in recently diverged lineages, it is necessary to understand the sensitivity limitations that may apply to HRM in the detection of single base differences between amplicons. Using locus *rps16I*, the sequence of haplotype *rps16I01* (amplicon 352 bp in length) was selected as a wild type template to generate single substitutions on all possible nucleotide sites within the amplicon outside the primer-binding sites (sites 1–21 and 328–352). All of these mutant amplicons were inputted to uMeltSM to generate their melting peaks and corresponding T_m values. As the highest resolution of uMeltSM is 0.1 °C, if the ΔT_m between two amplicons is <0.1 °C, the two amplicons were considered to be indistinguishable by HRM analysis. This approach facilitated an *in silico* evaluation of the theoretical missing-detection rate for A/T and C/G mutations in *in vitro* experiments.

Comparison of HRM performance against in silico RFLP

To evaluate the performance of HRM against PCR-RFLP, an *in silico* RFLP analysis was undertaken on the composite haplotypes identified using HRM in this study. DNA sequences from each haplotype were subjected to *in silico* digestions using the NEBCutter software (Vincze *et al.* 2003). Following, this enzyme combinations were selected to evaluate the *in silico* RFLP profiles using the Bioperl tool DistinctiEnz (http://www.bioinformatics.org/~docreza/rest_html/help_DistinctiEnz.htm). Among the trialled set of enzyme combinations the most informative was a combination of ApoI (6-base cutter) with AluI (4-base cutter).

Results

Haplotype detection in the sampled populations of Arenaria ciliata and Arenaria norvegica

Within the 446 individual samples from 21 populations of *A. ciliata* and *A. norvegica*, 14 amplicons of *rps16I* and 14 amplicons of *rps16II* were detected. When all 28

putative amplootypes were aligned over the full length of *rps16I* and *II*, a total of 20 composite haplotypes were identified, 18 of them for *A. ciliata* and two for *A. norvegica* (Tables 2–4). The T_m values shown in Table 4 are adjusted mean values that incorporate plate-to-plate T_m variations in the inter plate reference samples. These are validated by additional HRM analysis of selected haplotype individuals, either single accessions analysed in 60 replicate HRM reactions on the same plate, to determine well-to-well consistency, or multiple individuals of the same amplootypes drawn from different populations and re-run on the same single plate, to determine inter population and inter plate consistency. Within the replicate analyses of single accessions, total variation in observed T_m ranged between 0.04 and 0.16 °C, while total variation between replicates of the same amplootypes drawn from different populations ranged between 0.12 and 0.23 °C, based on three replicates of each amplopotype accession from each population. Recorded haplotypes differed at multiple nucleotide sites both by point mutations and insertion-deletion events (Table 2 and 3), and in a maximum likelihood phylogenetic analysis of *A. ciliata* and several sister species, the observed *rps16* haplotypes all emerged within an *A. ciliata* clade with 100% bootstrap support (Fig. S1, Supporting information). In every case

it was possible to assign individual samples from each population to a validated composite haplotype identity (Table 4).

Rps16I data

All but two amplootypes of *rps16I* were readily distinguishable from one another by their combined T_{m1} and T_{m2} values (Fig. 2A,B), including seven cases where these T_m differences were based on a single SNP between amplopotype templates. Amplootypes *rps16I04* and *I11* differed greatly in sequence composition (by five substitutions and one 6-bp indel) but displayed the same T_{m1} and T_{m2} values. These amplootypes are seen to be endemic to populations in Ireland and Spain, respectively, however by sequencing at least two replicates of each putative amplopotype in each population, incorrect assignment of amplopotype identity in the initial population analysis was avoided. In addition the composite *rps16* sequence of both amplootypes differed in *rps16II* and respective T_{m3} values were clearly distinct, providing a valid HRM diagnostic. Table 2 details the template differences between amplootypes *I01* and *I02* and between *I01*, *I06* and *I12* that give rise to the distinct HRM profiles and T_{m1}/T_{m2} values for each as shown in Fig. 2A,B respectively.

Table 2 Polymorphic sites and principal melting domains among 14 *rps16I* amplootypes identified by High Resolution Melting (HRM) analysis

Amplotype	Base position																	
	Melt domain (ii)							Melt domain (i)										
	29	49	93	113	124–129	160	166	217	231–237	241	245	246	252	290–295	296	306	314	321–325
<i>I01</i>	G	T	C	A	AAAGAA	G	G	G	<u>ATATA<u>TC</u></u>	T	A	C	C	—	C	T	G	—
<i>I02</i>	G	T	C	A	AAAGAA	G	G	G	—	T	A	A	C	TTATAA	C	T	G	—
<i>I03</i>	G	T	C	A	AAAGAA	G	G	G	ATATA <u>TC</u>	T	A	C	C	—	C	T	G	TTTT
<i>I04</i>	G	T	C	A	—	G	G	G	ATATA <u>TC</u>	T	A	C	C	—	C	T	G	—
<i>I05</i>	G	T	C	C	AAAGAA	G	T	G	—	T	A	A	C	—	C	T	G	—
<i>I06</i>	G	T	C	A	AAAGAA	G	G	G	ATATA <u>GC</u>	C	A	A	C	—	C	T	C	—
<i>I07</i>	G	G	C	A	AAAGAA	G	G	G	ATATA <u>GC</u>	C	A	A	C	—	C	T	C	—
<i>I08</i>	A	T	C	A	AAAGAA	G	G	G	ATATA <u>TC</u>	C	C	A	G	—	C	T	C	—
<i>I09</i>	G	T	C	A	AAAGAA	G	T	G	—	T	A	A	C	—	C	T	G	—
<i>I10</i>	G	T	T	A	AAAGAA	G	T	G	—	T	A	A	C	—	C	T	G	—
<i>I11</i>	G	T	C	A	AAAGAA	T	G	G	ATATA <u>GC</u>	C	A	A	C	—	C	T	C	—
<i>I12</i>	A	T	C	A	AAAGAA	G	G	G	ATATA <u>TC</u>	C	C	A	G	—	T	T	C	—
<i>I13</i>	G	T	C	A	AAAGAA	G	T	T	—	T	A	A	C	—	C	G	G	—
<i>I14</i>	G	T	C	A	AAAGAA	G	G	G	—	T	A	A	T	TTATAA	C	T	G	—

Observed composite haplotype identities covering both *rps16I* and *II* are listed in Table 4. Base positions 160–325 and 29–129 comprise melt domains (i) and (ii) respectively. Template differences between amplootypes give rise to discriminating melting curve T_m values in HRM. For example, *I01* and *I02* differ only in domain (i) [underlined] giving rise to divergent T_{m1} peak values for these amplootypes in Fig. 2A; *I01* and *I06* also differ in domain (i) [shaded], however *I06* and *I12* differ in both domains (i) and (ii) [shaded] giving rise to divergent T_{m1} and T_{m2} peak values in Fig. 2B.

Table 3 Polymorphic sites and position of principal melting domain among 14 rps16 δ /II amplicons identified by High Resolution Melting (HRM) analysis

	Base position	Melt domain flanking region ($T_{m\Delta}$)	Principal melt domain
	Amplotype 38 79 94 111 119 129–155		179 183 186 194 200 213 221 223–227 281 291 303 322 336 338 354–359 372–384
<i>II</i> 01	— C G T C	T C A G T T	T T C C T T
<i>II</i> 02	— C G T G	T T A G T T	T T C C T T
<i>II</i> 03	— C G T G	T T A G T T	T T C C T T
<i>II</i> 04	— C G T G	T C A G T T	T G T C C T
<i>II</i> 05	— C G T G	T C A G T T	T T C C T T
<i>II</i> 06	— C G T G	T C A G T T	T C C T C T
<i>II</i> 07	T T G T G T G A T T C T AAAATGAGACACAACT	T C A G T G — T	T T C C T T
<i>II</i> 08	T T G T G T G	T C A G T G — T	T T C C T T
<i>II</i> 09	— G A T G	T C T G T T	G T C T G
<i>II</i> 10	T T G T G T G A T T C T AAAATGAGACACAACT	T C A G T G — T	T T C C T T
<i>II</i> 11	— C G T C	T C A G T T	T T C C T T
<i>II</i> 12	— C G T G	T C A G T T	T T C C T T
<i>II</i> 13	— C G T G	T C A G T T	T T C C T T
<i>II</i> 14	— C G G G	T C A G T T	T T C C T T

Observed composite haplotype identities covering both rps16 δ and *II* are listed in Table 4. Base positions 38–213 and 221–384 comprise the principal melt domain and melt domain flanking region ($T_{m\Delta}$) respectively. Template differences between amplicons give rise to discriminating melting curve profiles in HRM, for example *II*02 and *II*03 differ by a 13 bp indel in the flanking region (shaded) resulting in divergent melt curves in the HRM profile (Fig. 2C), while *II*07 and *II*10 differ by one single-nucleotide polymorphism in the principal melt domain (shaded) giving rise to discrete peak T_{m3} values in Fig. 2D

Table 4 Summary of all rps16 haplotypes identified in *Arenaria ciliata* (rpsC-) and *Arenaria norvegica* (rpsN-) using High Resolution Melting *in vitro* and *in silico* analysis of rps16I and rps16II amplicons

Composite rps16 haplotype	Genbank accession no.	rps16I amplo type	<i>In vitro</i> T _{m1} (°C)	<i>In silico</i> T _{m1} (°C)	<i>In vitro</i> T _{m2} (°C)	<i>In silico</i> T _{m2} (°C)	rps16II amplo type	<i>In vitro</i> T _{m3} (°C)	<i>In silico</i> T _{m3} (°C)	<i>In silico</i> T _{mX} (°C)	Frequency of haplotype by population†
rpsC01	JQ945739	I01	76.25	77.7	79.65	80.6	II01	76.15	77.7	75.8	Ir1 (8/30) Ir2 (28/30) Ir3 (13/30) Ir4 (1/30) Pi2 (3/15)
rpsC02	JQ945740	I02	75.35	77.3	79.65	80.6	II02	75.70	77.4	76.3	Ir1 (22/30) Ir3 (10/30) Ir4 (29/30); Au1 (13/27)
rpsC03	JQ945741	I02	75.35	77.3	79.65	80.6	II03	75.70	77.4	75.7	Ir2 (2/30)
rpsC04	JQ945742	I03	75.45	77.5	79.65	80.6	II04	76.05	77.7	76.5	Ir3 (6/30)
rpsC05	JQ945743	I04	76.35	77.7	79.65	80.7	II01	76.15	77.7	75.8	Ir3 (1/30)
rpsC06	JQ945744	I01	76.25	77.7	79.65	80.6	II05	76.15	77.7	76.3	Py1 (7/20) Py2 (1/16)
rpsC07	JQ945745	I01	76.25	77.7	79.65	80.6	II06	76.00	77.7	76.3	Py1 (2/20)
rpsC08	JQ945746	I05	75.95	77.5	79.85	80.8	II05	76.15	77.7	76.3	Py1 (8/20) Pi2 (2/15)
rpsC09	JQ945747	I06	76.45	78.1	79.65	80.6	II07	75.80	77.3	76.1	It1 (26/26) It2 (1/1) It3 (3/4) Au2 (8/25)
rpsC10	JQ945748	I07	76.40	78.1	79.95	80.9	II08	75.60	77.2	76.1	It3 (1/4)
rpsC11	JQ945749	I08	76.50	78.0	79.10	80.3	II09	75.70	77.4	76.3	Fr1 (2/5)
rpsC12	JQ945750	I09	75.95	77.5	79.65	80.6	II05	76.15	77.7	76.3	Py1 (3/20) Py2 (1/16)
rpsC13	JQ945751	I06	76.45	78.1	79.65	80.6	II10	75.60	77.0	76.1	Fr1 (3/5)
rpsC14	JQ945752	I01	76.25	77.7	79.65	80.6	II11	76.15	77.7	76.3	Pi1 (16/18) Py2 (14/16) Au2 (4/25)
rpsC15	JQ945753	I10	75.60	77.5	79.00	80.0	II12	76.55	77.9	76.3	Pi1 (2/18) Pi2 (9/15) Pi3 (19/19)
rpsC16	JQ945754	I11	76.35	77.8	79.65	80.6	II07	75.80	77.3	76.1	Pi2 (1/15)
rpsC17	JQ945755	I01	76.25	77.7	79.65	80.6	II13	76.10	77.7	76.6	Au1 (14/27)
rpsC18	JQ945756	I12	75.85	77.9	79.10	80.3	II09	75.70	77.4	76.3	Au2 (13/25)
rpsN01	JQ945757	I13	76.10	77.4	79.65	80.6	II05	76.15	77.7	76.3	NB (30/30) NE (20/20) NiC (2/2) Nin (30/30) NR (11/29) NS (29/29)
rpsN02	JQ945758	I14	75.05	77.0	79.65	80.6	II14	76.50	77.9	76.3	NR (18/29)

†Key to Population identity. *A. ciliata*: Ir1, Ir2, Ir3, Ir4, Ben Bulben Plateau, NW Ireland; Pi1, Pi2, Pi3, Picos de Europa, Spain; Py1, Py2, Valle de Benasque, Pyrenees, Spain; It1, It2, It3, Piemonte, Alps, Italy; Fr1, Col D'Agnel, Alps, France; Au1, Au2, Steiermark, Alps, Austria. *A. norvegica*: NB, Burren, Ireland; NE, Yorkshire, England; Nic, Eldgja gorge, Iceland; Nin, Inchnadamph, Scotland; NR, Isle of Rum, Scotland; NS, Shetland Islands, Scotland.

Rps16II data

Amplicon melting curves for rps16II showed lower resolution than for rps16I. Three different amplootypes, II01, II05 and II11 that differ by one G/C substitution and/or a 6 bp indel have the same T_{m3} at 76.15 °C. II01 can be distinguished from both of the latter at T_{mX}, however II05 and II11 (which co-occur in population Py2 as part of composite haplotypes C06 and C14, respectively, Fig. 3) cannot be distinguished from one another based on the rps16II melting curve. Amplopotype II09 differs markedly (minimum seven point mutations) from both II02 and II03, but has the same T_{m3} value as both (75.70 °C). The latter two amplootypes differ by a single indel of 13 bp, which is associated with divergent T_{mX} shoulder curves between the two (Fig. 2D), however none of these rps16II amplootypes co-occur within a single population (Fig. 3).

The averaged T_{m3} values of two geographically distant rps16II amplootypes II12 (*A. ciliata* Spain) and II14 (*A. norvegica* Scotland) varied by only 0.05 °C, which is below the limit of HRM resolution, however these amplootypes were successfully identified by following the protocol to sequence all putative amplootypes in each population.

In vitro sensitivity to SNPs

Among all possible pairwise comparisons between amplootypes, single nucleotide polymorphisms constituted the sole difference between templates in 12 cases, 11 of which were identifiable by HRM (e.g. II07 and II10, Fig. 2C). In eight cases the pairwise ΔT_m was >0.2 °C. In two cases discrimination at <0.2 °C was achieved using the T_{mX} shoulder in rps16II, and in one case a shift of <0.2 °C was evident in T_{m1} in rps16I. Only one case, a class III transversion SNP in

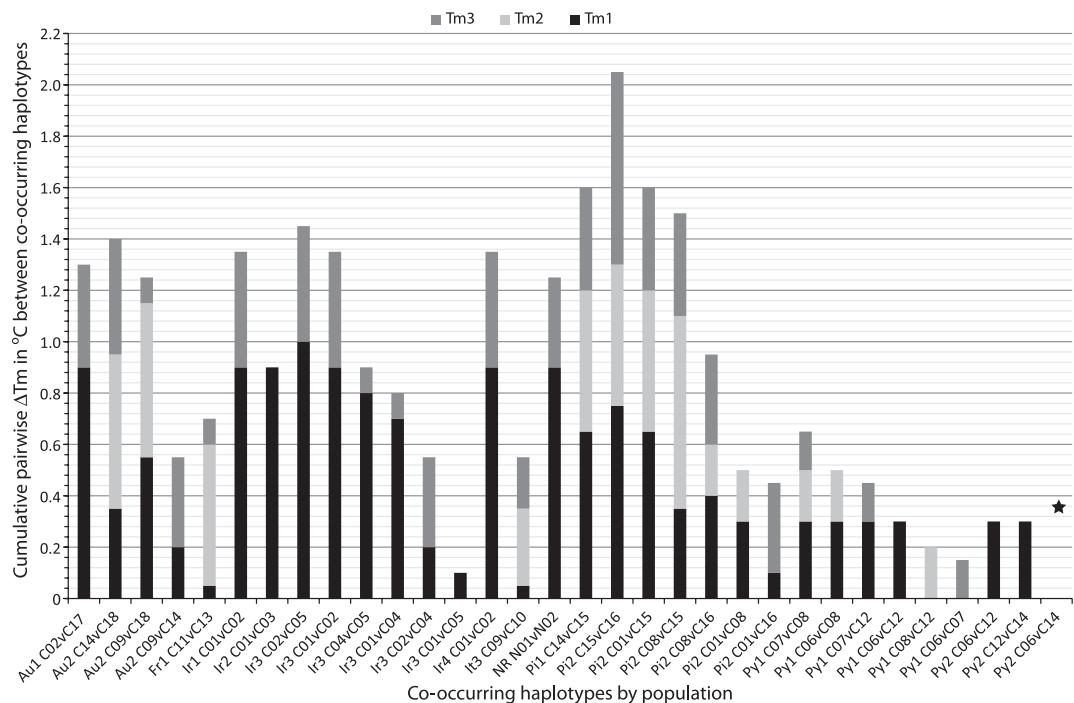


Fig. 3 Observed total pairwise ΔT_m between haplotypes that co-occur in the same sampled population based on combined differences in T_{m1} , T_{m2} and T_{m3} values. Only Py2 CO6 vs. C14 failed to yield any discrete ΔT_m value between the haplotypes, indicated by a star.

rps16II (G–C), was not discernable via initial HRM analysis. Overall class I transition SNPs (G–T or A–C) were most easily detected (five of six $\Delta T_m > 0.2$ °C), followed by class II transversion SNPs (A–C or G–T) (three of five $\Delta T_m > 0.2$ °C). No class IV transversion SNPs were recorded.

No differences were discernable in the melting curve profile between *II05* and *II11*, and *II05* was initially detected by sequencing at least 2 samples that showed the putative *II11* amplicon curve for rps16II in population Py2. Sequence analysis of all 14 putative *II11* amplicons in Py2 confirmed just one individual of *II05*.

Sensitivity to differences between co-occurring haplotypes

Among co-occurring haplotypes within the sampled populations, only three of 32 inter haplotype comparisons failed to yield at least one ΔT_m value that exceeded the nominal discriminating threshold of 0.2 °C for T_{m1} , T_{m2} or T_{m3} (Fig. 3). Two of these inter haplotype comparisons were reliably distinct below the 0.2 °C threshold; C06/C07 in Py1 (ΔT_{m3} 0.15) and C01/C05 in Ir3 (ΔT_{m1} 0.1). Only one inter haplotype comparison, C06/C14 in Py2 (ΔT_{m1} , 2, 3 = 0), as discussed above, was not distinguishable by HRM. The combined array of T_{m1} , T_{m2} , T_{m3} and T_{mX} values thus provided a unique identifier for 18 of 20 composite haplotypes identified in the analysis and

validated by sequencing (Table 4, Fig. 3). While the overall sequence composition for each of these composite haplotypes was unique, seventeen of the twenty shared at least one rps16I or II sequence identity with another haplotype, only three composite haplotypes included entirely unique sequences in both regions (C04, C10 and C15).

Correlation between ΔT_m and genetic distance

A Mantel test analysis of compounded T_{m1} , T_{m2} and T_{m3} differences between haplotypes and the corresponding pairwise inter haplotype Tajima-Nei genetic distance D showed that the two measures of difference were significantly correlated at $P < 0.05$ (Table 5), however when analysed separately no significant correlation was evident for either T_{m1} , T_{m2} vs. D or for T_{m3} vs. D .

In silico simulation (i): evaluating the utility of in silico HRM modelling as a support for in vitro HRM analysis

In silico simulation of rps16I melting using uMeltSM generates two clear T_m peaks, as for the *in vitro* results, while *in silico* rps16II generates two peaks, one clear peak corresponding to T_{m3} , and a second less-pronounced peak corresponding to the shoulder T_{mX} region in the *in vitro* melting analysis. The corresponding values of T_{m1} , 2, 3

Table 5 Results of Mantel correlation tests between observed *in vitro* inter haplotype ΔT_m , modelled *in silico* inter haplotype ΔT_m , and calculated inter haplotype genetic distance (Tajima-Nei's D)

		ρ	P
Inter amplopotype	ΔT_{m1} <i>in vitro</i> vs. ΔT_{m1} <i>in silico</i>	0.6247	0.0001***
ΔT_m <i>in vitro</i> vs. <i>in silico</i>	ΔT_{m2} <i>in vitro</i> vs. ΔT_{m2} <i>in silico</i>	0.9257	0.0001***
	ΔT_{m3} <i>in vitro</i> vs. ΔT_{m3} <i>in silico</i>	0.7964	0.0001***
Inter haplotype	$ \Delta T_{m1} + \Delta T_{m2} $ vs. D_{rps16I}	<i>In vitro</i> 0.2168	0.0858
ΔT_m vs. Tajima-Nei genetic distance D between haplotypes	ΔT_{m3} vs. $D_{rps16II}$	<i>In silico</i> 0.4468	0.0021**
	$ \Delta T_{m1} + \Delta T_{m2} + \Delta T_{m3} $ vs. D_{rps16}	<i>In vitro</i> 0.2118	0.0886
	$ \Delta T_{m1} + \Delta T_{m2} + \Delta T_{m3} + \Delta T_{mX} $ vs. D_{rps16}	<i>In silico</i> 0.5088 <i>In vitro</i> 0.3593 <i>In silico</i> 0.4189	0.0015** 0.0122* 0.0012**

*Correlation is significant at $P < 0.05$.

**Correlation is significant at $P < 0.01$.

***Correlation is significant at $P < 0.001$.

and T_{mX} (for the shoulder with *rps16II*) from *in silico* simulation are listed in Table 4.

A linear regression was performed to evaluate the consistency between the *in vitro* and *in silico* T_{m1} , T_{m2} and T_{m3} values for each haplotype (estimated $R^2 = 0.9902$, Fig. S2, Supporting information). The similarity between *in vitro* and *in silico* pairwise inter haplotype ΔT_m matrices was evaluated by Mantel test for T_{m1} ($r = 0.6247$, $P = 0.0001$), T_{m2} ($r = 0.9257$, $P = 0.0001$) and T_{m3} ($r = 0.7964$, $P = 0.0001$), with detailed results shown in Table 5. Both regression and Mantel tests suggest that *in vitro* and *in silico* T_m values are significantly correlated.

All the amplopotypes that are distinguishable by *in silico* HRM analysis (varying by ≥ 0.1 °C) can also be distinguished by *in vitro* analysis, and amplopotypes that share the same *in silico* T_m values also share the same *in vitro* T_m values, e.g. II05 and II11 share *in silico* $T_{m3} = 77.7$ °C and *in vitro* $T_{m3} = 76.15$ °C. However, one exception emerged between amplopotypes II05 and II06 which share the same *in silico* T_{m3} value (77.7 °C) but have distinct T_{m3} values *in vitro* (76.15 and 76.00 °C, respectively).

In silico simulation (ii): sensitivity of HRM to single-substitution mutations between haplotypes -*in silico* evaluation of single substitutions in a wild-type template

Within the 352 bp region of *rps16I*, class I (transition A to G or C to T) and class II (transversion A-C or G-T) mutations were generated on all the nucleotide sites between site 22 and 327 (i.e. 306 mutant amplopotypes obtained for each of class I and II mutations), while class III (transversion C-G) and IV (transversion A-T) mutations were generated on all possible sites within the same region (95 class III mutants by mutation on C and G sites and 211 class IV mutants on A and T sites) (Fig. S3, Supporting information).

In total 303 out of the 306 class I mutant amplopotypes were given T_{m1} or T_{m2} values more than 0.1 °C different from the wild type ($T_{m1} = 77.7$ °C and $T_{m2} = 80.6$ °C), indicating a detection rate of 303/306 = 99.02% for class I single mutations (Fig. S3, Supporting information). In the same way, the detection rate for class II single mutation is also 303/306 = 99.02%, while it is 46/95 = 48.42% for class III and 81/211 = 38.39% for class IV single mutations, respectively.

Within the 306 bp detectable region, there are in total 918 possible single substitution mutations, of which there are 95 class III, 211 class IV and 612 class I and II mutations. The estimated overall missed SNP detection rate for the *rps16I* *A. ciliata* wild-type amplicon is thus estimated at 20.15%. This indicates that if there is a single substitution mutation in the target region between haplotypes, the possibility we cannot detect it by HRM analysis is around 20%.

Performance of *in silico* RFLP

The *in silico* RFLP undertaken on the 20 distinct haplotype sequences obtained through HRM showed that 12 putative haplotypes were distinguishable by this method. Only seven of these corresponded exactly to one of the HRM-identified haplotypes (Table S2, Supporting information), the other five putative haplotypes corresponded to either two or three different haplotypes identified through HRM. In terms of discriminating haplotype identity the overall resolution of the *in silico* RFLP is 35% (7/20) in comparison to the 90% (18/20) for HRM.

Discussion

The protocol presented here was designed to evaluate whether HRM can provide a practical and effective

method for haplotype identification. The size of the target amplicons was determined based on joint consideration of DNA coverage, sensitivity and cost, an approach that was not guaranteed a priori to provide optimal performance. The threshold of T_m difference for amplopype discrimination was also questionable at 0.2 °C before our study, so we applied posterior DNA sequencing after HRM analysis to evaluate the efficacy of this threshold. The method as tested has generated highly detailed haplotype identity and frequency data for *Arenaria ciliata* and *Arenaria norvegica*. Working with the two targeted amplicons rps16 I and II, HRM analysis revealed all but one of the haplotypes that were finally confirmed by DNA sequencing, allowing discrete haplotype identification in 379 of 380 possible pairwise haplotype comparisons, including 31 of 32 observed co-occurrences of haplotypes within single populations (Fig. 3). Initially there was concern that the chosen amplicons (350–400 bp) might be too long compared to established HRM norms; however, the accuracy of the analysis was sustained over these amplicon size intervals, in particular due to the presence of multiple melting domains in the double-stranded amplified DNA.

From both *in vitro* and *in silico* analyses, the melting peak corresponding to T_{m1} in melting domain (i) of the rps16I amplicon was the most discriminative indicator (Fig. 2A,B). The average inter amplopype ΔT_{m1} values were 0.46 °C *in vitro* and 0.31 °C *in silico*, compared to 0.24/0.20 °C (for ΔT_{m2}), and 0.31/0.26 °C (for ΔT_{m3}). While amplopypes of rps16I were identified distinctly with T_{m1} and T_{m2} , amplopypes of rps16II were not so easily distinguished with T_{m3} alone, and this may be related to the slightly larger amplicon size compared to rps16I. However the T_{mx} shoulder that flanks the principal melting domain T_{m3} did discriminate between a number of otherwise similar rps16II amplopypes, e.g. between II02 and II03 (Fig. 2B). While the T_{mx} region lacked a discernable peak in our *in vitro* HRM profiles and it was not possible to assign precise T_{mx} values to amplopypes, discrete T_{mx} curve patterns were evident between many amplopypes.

In only three cases did haplotypes that co-occurred within a single population have no pairwise ΔT_m value of ≥0.2 °C at any of T_{m1} , T_{m2} or T_{m3} (Fig. 3). In two of these cases a smaller but consistent ΔT_m of 0.10–0.15 °C was evident, and once sequence content and observed T_m variation were confirmed, discrimination of these haplotype samples by HRM was possible. In one single case all ΔT_m values between a haplotype pair were 0 °C, (C06 and C14 in population Py2 deviated by a single class III C/G mutation in rps16II), which is the only case where HRM analysis failed to distinguish between two haplotypes in our study.

Sensitivity of the analysis may have been impacted by the quality of the DNA template. This work utilized a

DNA sample set that had been extracted using a modified CTAB method, which doesn't guarantee a uniform chemical composition for the extracted DNA. The potentially lower quality of our DNA template may thus have had an impact on subsequent melting analysis, generating system errors between T_m readings of different batches of HRM assays. However despite these errors within *in vitro* analysis, the Mantel-test correlation with modelled *in silico* ΔT_m values for the equivalent amplopype templates was significant at $P < 0.001$, indicating the obtained *in vitro* ΔT_m values do reflect DNA variation among the haplotypes as predicted by *in silico* simulation.

Also it is clear that where two amplopypes can be distinguished by *in silico* simulation (by 0.1 °C) using uMeltSM, they can also be distinguished by HRM analysis on the LightCycler 480 system. Occasionally *in vitro* HRM analysis produces even higher resolution than *in silico* simulation, e.g. rps16I amplopypes I09 and I10 shared the same *in silico* $T_{m1} = 77.5$ °C but have different *in vitro* T_{m1} values (0.35 °C difference). While estimates of precise °C values of T_m peaks differ between the two methods as we have applied them, the underlying pattern of inter haplotype ΔT_m identification is equivalent between the two (mantel test of correlation between pairwise ΔT_m matrices is significant at $P < 0.001$ Table 5). Overall this data affirms both that *in silico* modelling is a valid support for *in vitro* HRM work, and that the model parameters applied in uMeltSM generate slightly conservative estimates of HRM curve difference between amplopypes.

As mentioned above, the limited resolution of HRM means that it is not an equivalent to template sequencing, however a more general correlation between inter haplotype ΔT_m values and inter haplotype genetic distance is a possibility. Mantel tests on the *in vitro* and *in silico* data here (Table 5) showed a significant positive correlation between the compounded differences in ΔT_{m1} , two and three between haplotypes and the corresponding pairwise Tajima-Nei genetic distance D between haplotypes. Thus while certain haplotype pairs in the study with large pairwise D values did have very similar melting-curve profiles, these correlation tests on our data support the case that in general a greater ΔT_m value between haplotypes indicates a greater level of evolutionary divergence.

In silico modelling of HRM sensitivity to different SNP mutation classes on the wild-type rps16I amplicon confirmed that transition SNPs (class I mutations) and class II transversion SNPs are easier to detect than Class III and IV transversions (Liew *et al.* 2004). Our *in vitro* results show the former classes can be detected by HRM analysis when occurring in an amplicon up to 360 bp (rps16I, one case of class II SNP) or 390 bp (rps16II, one case of class I SNP) at or above the 0.2 °C threshold we selected for identifying between-amplicon deviation. *In silico* HRM model-

ling estimated the likely incidence of silent SNPs in *rps16*I to be 51.58% and 61.61% for class III and IV mutations respectively, compared to just 0.98% for Class I and II mutations. In our haplotype dataset four Class III and two Class IV SNPs were observed, however only in one case (a Class III G/C mutation between *II05* and *II11*) did this mutation constitute the sole difference between composite haplotypes—and this was the only instance where HRM failed to distinguish between haplotypes.

By assuming the likelihood of some silent mutations in the analysis, our protocol required that two or three individual representatives from each putative amplopotype were sequenced from each population. In total over 80 sequences were obtained, and in only one case was an amplopotype sequence returned that was not discernable by HRM analysis. This contrasts with the results of the *in silico* RFLP analysis (Table S2, Supporting information), where only 35% of the recorded haplotypes could be detected uniquely. HRM sensitivity to silent SNPs in larger amplicons can be improved by reducing amplicon size at the specific SNP locus using new internal primers. In addition, it is common in diagnostic HRM applications to identify the presence of specific SNPs by categorizing nuclear genotypic (dual template) samples as either homozygous or heterozygous for the SNP in question (Wittwer *et al.* 2003). While we have not evaluated this approach for discriminating between dual templates in the current work, pooling amplopotype samples may provide additional sensitivity for otherwise silent SNPs among samples that share the same HRM curve. However the only assurance of total haplotype coverage is to sequence every sample (Liew *et al.* 2004; Wittwer 2009).

Compared to the alternative approach of sequencing every single accession (446) to generate a total count of haplotype frequency data across the population samples, the resources required for the HRM method were significantly reduced. Although sequence analysis was not exhaustive, we are confident that the method effectively identified haplotypes in the sampled populations. In particular, while the low haplotype diversity of *A. norvegica* suggests a characteristic post pleistocene south-to-north expansion similar to many European plant species (Taberlet *et al.* 1998), *A. ciliata* appears to be much more diverse, and the multiple endemic haplotypes identified in populations in Ireland are suggestive of an origin that predates the last glacial maximum (LGM) c. 20 000 years BP (Table 4, Fig. S1, Supporting information), akin to other patterns recently emerging from analysis of the arctic-alpine flora (Schneeweiss & Schonswetter 2011). This may be a significant finding and offers tentative phylogeographic support for a postulated nunatak on the Ben Bulben plateau in Northwest Ireland throughout the LGM (Synge & Wright 1969). A more detailed phylogeo-

graphic analysis of this and other data, including taxonomic inferences, will be presented in an upcoming paper (*In prep*).

Based on the results we have detailed, HRM haplotype analysis is suitable for a wide range of phylogeographic studies where polymorphism is evident in commonly used 400 + bp spacer regions in chloroplast and mitochondrial DNA (e.g. Shaw *et al.* 2005). The limitations of the method are similar to direct sequencing in that individual loci may not contain informative polymorphisms below the species level. However, in any studied organism where discrete organelle DNA differentiation is evident between population sub-groups, or where cryptic speciation has occurred, HRM has the capacity to greatly increase the scope and sensitivity of haplotype analysis.

These advantages may be relevant for future work both on regional meta-populations with significant intermixing of diverse biogeographic stock (e.g. European tree species., Petit *et al.* 2003), and in species populations with disjunct distributions, (e.g. *Arenaria* and *Sagina* spp in the Holarctic, Westergaard *et al.* 2011). Population sample groups with similar levels of haplotype diversity to what we have recorded for *A. ciliata* are already common in the literature; recent examples include the Mediterranean tree frog *Hyla sarda* (37 mitochondrial *cyt b* gene haplotypes, Bisconti *et al.* 2011) and the Mexican mountain shrub *Palicourea padifolia* (14 chloroplast *rpl32-trnL* spacer haplotypes, Gutiérrez-Rodríguez *et al.* 2011). Phylogeographic data compiled by Petit *et al.* (2005) indicated an average of 12.3 haplotypes per species across 183 angiosperm and gymnosperm taxa, with 47 displaying more than 15 discrete chloroplast or mitochondrial haplotypes across their distribution. Equally, many newly detected cryptic species have also come to light via organelle haplotype variation (e.g. the Neotropical tree *Cedrela fissilis*, Garcia *et al.* 2011; and the north Atlantic *Littorina saxatilis*, Doellman *et al.* 2011).

Bearing in mind the sensitivity limitations and error-minimization requirements documented in this current work, all of the above studies are amenable to HRM haplotype analysis. There is considerable scope for improvement in this technique, for example in the optimization of amplicon design and selection protocols, and evaluation of (potentially heterozygous) large nuclear loci such as ITS. However, for any organism where comparative sequence data has already been generated, *in silico* HRM analysis can be carried out to evaluate the utility of deploying the method alongside existing protocols for direct sequencing or PCR-RFLP. With wider availability of real-time PCR equipment, and the superior sensitivity of HRM to mutational differences compared to PCR-RFLP, the method likely represents an untapped resource in phylogeographic analysis.

Acknowledgements

Xiaodong Dang is funded by a John and Pat Hume Scholarship from the National University of Ireland Maynooth. This project and the graduate scholarship for Emma Howard-Williams were funded by Science Foundation Ireland (Grant No.: SFI/08/RFP/EOB1545). Quantitative PCR instrumentation was funded by Science Foundation Ireland (Grant No.: SFI/07/RFP/GEN/-F571/ECO7). Andreas Tribsch (Salzburg), Kevin Walker (BSBI), Pablo Vargas (Madrid) and Patrick Kuss (Bern) assisted with field collections.

References

- Angiosperm Phylogeny Group III (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society*, **161**, 105–121.
- Bettin O, Cornejo C, Edwards PJ, Holderegger R (2007) Phylogeography of the high alpine plant *Senecio halleri* (Asteraceae) in the European Alps: in situ glacial survival with postglacial stepwise dispersal into peripheral areas. *Molecular Ecology*, **16**, 2517–2524.
- Bisconti R, Canestrelli D, Colangelo P, Naselli G (2011) Multiple lines of evidence for demographic and range expansion of a temperate species (*Hyla sarda*) during the last glaciation. *Molecular Ecology*, **20**, 5313–5327.
- Doellman MM, Trussell GC, Grahame JW, Vollmer SV (2011) Phylogeographic analysis reveals a deep lineage split within North Atlantic *Littorina saxatilis*. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character. Royal Society (Great Britain)*, **278**, 3175–3183.
- Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, **19**, 11–15.
- Dwight Z, Palais R, Wittwer CT (2011) uMELT: prediction of high-resolution melting curves and dynamic melting profiles of PCR products in a rich web application. *Bioinformatics*, **27**, 1019–1020.
- Garcia MG, Silva RS, Carnielo MA et al. (2011) Molecular evidence of cryptic speciation, historical range expansion, and recent intraspecific hybridization in the Neotropical seasonal forest tree *Cedrela fissilis* (Meliaceae). *Molecular Phylogenetics and Evolution*, **61**, 639–649.
- Gutiérrez-Rodríguez C, Ornelas JF, Rodríguez-Gómez F (2011) Chloroplast DNA phylogeography of a distylous shrub (*Palicourea padifolia*, Rubiaceae) reveals past fragmentation and demographic expansion in Mexican cloud forests. *Molecular Phylogenetics and Evolution*, **61**, 603–615.
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.
- Huguet JM, Bizarro CV, Forns N et al. (2010) Single-molecule derivation of salt dependent base-pair free energies in DNA. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 15431–15436.
- Jalas J, Suominen J (1983) *Atlas Flora Europaea: Distribution of Vascular Plants in Europe, Volume 6: Caryophyllaceae (Alsinoideae and Paronychioideae)*. Committee for Mapping the Flora of Europe, Helsinki.
- Liew M, Pryor R, Palais R et al. (2004) Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons. *Clinical Chemistry*, **50**, 1156–1164.
- Mader E, Lohwasser U, Börner A, Novak J (2010) Population structures of genebank accessions of *Salvia officinalis* L. (*Lamiaceae*) revealed by high resolution melting analysis. *Biochemical Systematics and Ecology*, **38**, 178–186.
- Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–220.
- Petit RJ, Aguinagalde I, de Beaulieu J-L et al. (2003) Glacial refugia: hot-spots but not melting pots of genetic diversity. *Science*, **300**, 1563–1565.
- Petit RJ, Duminil J, Fineschi S et al. (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology*, **14**, 689–701.
- R Development Core Team (2011) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.
- Reed GH, Wittwer CT (2004) Sensitivity and specificity of single-nucleotide polymorphism scanning by high-resolution melting analysis. *Clinical Chemistry*, **50**, 1748–1754.
- Ririe KM, Rasmussen RP, Wittwer CT (1997) Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. *Analytical Biochemistry*, **245**, 154–160.
- Schneeweiss GM, Schonswetter P (2011) A re-appraisal of nunatak survival in arctic-alpine phylogeography. *Molecular Ecology*, **20**, 190–192.
- Shaw J, Lickey EB, Beck JT et al. (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany*, **92**, 142–166.
- Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *American Journal of Botany*, **94**, 275–288.
- Smith BL, Lu C-P, Alvarado Bremer JR (2010) High-resolution melting analysis (HRMA): a highly sensitive inexpensive genotyping alternative for population studies. *Molecular Ecology Resources*, **10**, 193–196.
- Studer B, Jensen LB, Fiil A, Asp T (2009) “Blind” mapping of genic DNA sequence polymorphisms in *Lolium perenne* L. by high resolution melting curve analysis. *Molecular Breeding*, **24**, 191–199.
- Synge FM, Wright HEJ (1969) The Würm ice limit in the West of Ireland. In: *Quaternary Geology and Climate* (ed. Wright HE), pp. 89–92. National Academy of Sciences, Washington, District of Columbia.
- Taberlet P, Fumagalli L, Wust-Saucy A-G, Cosson J-F (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453–464.
- Tajima F, Nei M (1984) Estimation of evolutionary distance between nucleotide sequences. *Molecular Biology and Evolution*, **1**, 269–285.
- Teacher AGF, Garner TWJ, Nichols RA (2009) European phylogeography of the common frog (*Rana temporaria*): routes of postglacial colonization into the British Isles, and evidence for an Irish glacial refugium. *Heredity*, **102**, 490–496.
- Thioulouse J, Chessel D, Dolédec S, Olivier JM (1997) ADE-4: a multivariate analysis and graphical display software. *Statistics and Computing*, **7**, 75–83.
- Tutin TG, Burges NA, Chater AO et al. (1993) *Flora Europaea, Volume 1: Psilotaceae to Platanaceae*. Cambridge University Press, Cambridge.
- Vincze T, Posfai J, Roberts RJ (2003) NEBCutter: a program to cleave DNA with restriction enzymes. *Nucleic Acids Research*, **31**, 3688–3691.
- Vossen RHAM, Aten E, Roos A, den Dunnen JT (2009) High-resolution melting analysis (HRMA): more than just sequence variant screening. *Human Mutation*, **30**, 860–866.
- Westergaard KB, Alsos IG, Popp M et al. (2011) Glacial survival may matter after all: nunatak signatures in the rare European populations of two west-arctic species. *Molecular Ecology*, **20**, 376–393.
- Wittwer CT (2009) High-resolution DNA melting analysis: advancements and limitations. *Human Mutation*, **30**, 857–859.
- Wittwer CT, Herrmann MG, Moss AA, Rasmussen RP (1997) Continuous fluorescence monitoring of rapid cycle DNA amplification. *BioTechniques*, **22**, 130–131, 134–138.
- Wittwer CT, Reed GH, Gundry CN, Vandersteen JG, Pryor RJ (2003) High-resolution genotyping by amplicon melting analysis using LCGreen. *Clinical Chemistry*, **49**, 853–860.
- Wyse Jackson MB, Parnell JAN (1987) A biometric study of the *Arenaria ciliata* L. complex Caryophyllaceae. *Watsonia*, **16**, 373–382.

C.M., X.-D. D. and C.T.K. conceived the ideas for this work; All the contributing authors collected samples; X.-D. D. led the data analysis; C.M. and X.-D. D. led the writing.

Data Accessibility

DNA sequences for composite haplotypes C01–N02 in Table 4 are archived on Genbank accession nos

JQ945739–JQ945758 inclusive. The final DNA sequence assembly used for genetic distance and phylogenetic analysis is deposited in FASTA format on DRYAD, doi: 10.5061/dryad.s904k.

Supporting Information

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

Figure S1. Maximum likelihood phylogenetic tree generated from recorded rps16 haplotype sequences and sister-group taxa using MEGA 5.0 (Tamura *et al.*, 2011)^{*1}.

Figure S2. Correlation between HRM T_m calculation *in-vitro* and *in-silico*.

Figure S3. *In-silico* simulation of SNP mutation sensitivity for HRM analysis using *Arenaria ciliata* rps16I haplotype J01 as a wild-type template (352 bp length).

Table S1. Location and size of sample populations of *Arenaria ciliata* (*Ac*) and *A. norvegica* (*An*) used in High Resolution Melt Analysis.

Table S2. Results of *in silico* RFLP analysis carried out using NEBcutter (Vincze *et al.* 2003).

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.