MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2012) doi: 10.1111/j.1755-0998.2012.03139.x

Two highly informative dinucleotide SSR multiplexes for the conifer *Larix decidua* (European larch)

STEFANIE WAGNER,*†‡ SOPHIE GERBER*† and RÉMY J. PETIT*†
*INRA, UMR1202 BIOGECO, F-33610 Cestas, France, †Univ. Bordeaux, UMR1202 BIOGECO, F-33400 Talence, France, ‡University of Bonn, Steinmann Institute of Geology, Mineralogy and Paleontology, D-53115 Bonn, Germany

Abstract

We have designed two highly polymorphic microsatellite multiplexes for *Larix decidua* Mill (European larch), a coniferous tree species with a fragmented distribution across Europe. The multiplexes combine microsatellites previously designed for the sister species *L. kaempferi* and newly identified microsatellites obtained by pyrosequencing of an enriched microsatellite library and subsequent marker candidate selection. As we wanted to target highly polymorphic markers, only microsatellite motifs with a high number of repeats (≥12) were selected. An important proportion of the marker candidates presented multiple bands, bad amplification or insufficient polymorphism. Such difficulties were expected owing to the large genome size of the studied species. Our strategy for marker validation followed most recent recommendations for microsatellite development, for example verifying marker quality in terms of polymorphism and accurate allele binning before multiplexing. The most promising loci were combined in two multiplexes, a 7-plex and a 6-plex. These were tested on a sample of 413 individuals from 18 populations distributed across the natural range. The 13 loci had from 9 to 36 alleles. Markers were successfully tested in another laboratory, confirming robustness of the marker protocols. We also tested transferability on six other larch species from Asia and North America. Overall, this study shows that, even in species with large genome size and relatively low overall polymorphism, microsatellites can be successfully developed using next-generation sequencing technologies, provided that some additional precautions are taken compared to species lacking these characteristics.

Keywords: large genome size, *Larix decidua*, microsatellite, next-generation sequencing, polymorphism

Received 27 October 2011; revision received 27 January 2012; accepted 6 February 2012

Introduction

Larix decidua is an endemic European conifer with a highly fragmented montane to subalpine distribution in the Alps, the Sudety, the Tatra and the Carpathians as well as some exceptional lowland occurrences in Poland (Rubner 1953; McComb 1955). As these regions, especially the mountainous ones (e.g. Tinner & Kaltenrieder 2005; Colombaroli *et al.* 2010), are strongly exposed to climate change, there is a need for detailed range-wide genetic studies that can provide information for preservation of valuable genetic resources and sustainable forest management. So far, range-wide genetic studies based on nuclear markers in *L. decidua* have exclusively relied on allozyme markers. Such studies have yielded useful information about genetic relationships among populations (Lewandowski & Mejnartowicz 1991a,b; Lewandowski *et al.* 1991; Maier 1992). However, there is a need for increased resolution to better understand the past population dynamics of the species. We have therefore endeavoured to develop highly polymorphic nuclear microsatellites. The development was based on the transfer of existing markers from sister species, and the selection and design of new markers based on 454 pyrosequencing. First, we tested existing markers from the Asian sister species *L. kaempferi* (Isoda & Watanabe 2006) and the North American species *L. occidentalis* (Khasa *et al.* 2000; Chen *et al.* 2009). As this did not lead to a satisfactory number of markers matching recommended quality criteria, for example sufficient polymorphism and clear binning of alleles (Guichoux *et al.* 2011a), we developed new markers based on pyrosequencing of an enriched microsatellite library. This approach can be substantially more cost-effective as the conventional method based on the screening of cloned libraries by Sanger sequencing (e.g. Santana *et al.* 2009). However, in our study, we had to face two difficulties. First, initial tests had shown that markers with <12 repeats (i.e. for the individual that was initially sequenced) were characterized by low variation. This low variation might in part be attributed to the fact that

Correspondence: Rémy J. Petit, Fax: +33557122881;
E-mail: petit@pierroton.inra.fr

trees and shrubs, which have relatively long generation times, generally show lower rates of molecular evolution than related herbaceous plants (Smith & Donoghue 2008). As microsatellite polymorphism increases exponentially with the number of microsatellite repeats (Ellegren 2000; Kelkar *et al.* 2008), we focused on microsatellites with a high number of repeats (≥12) to increase the chances to find variable markers. Second, in coniferous species with large genomes such as *L. decidua* (11 198 Mb, Greibhuber 1986), a large proportion of microsatellites can be expected to be located within repetitive DNA sequences, leading to marker candidates that are difficult to amplify (e.g. Pfeiffer *et al.* 1997). We therefore targeted a large number of marker candidates to select those showing both the clear amplification profiles and sufficient polymorphism. Finally, we decided to combine the markers in multiplex reactions, that is, PCR amplifying several markers simultaneously, because once such multiplexes are established, they enormously reduce laboratory costs and labour time and thus enable high-throughput analyses and promote accuracy and precision of the genetic result (Guichoux *et al.* 2011a,b; Lefèvre *et al.* 2011). By following all these steps, we obtained two highly variable multiplex for *Larix decidua* (one 7-plex and one 6-plex).

Materials and methods

Plant material and DNA isolation

In 2010, we collected phloem and needle samples from 18 populations forming a gradient over the natural distribution range (Table 1). Eight had been collected *in situ* and 10 *ex situ* in four German provenance trials. In the latter case, each population sample (consisting in 24 individuals) originated from one single trial. Further samples for marker validation as well as for transferability tests were provided by colleagues. These included six progenies (each comprising one female parent and seven offspring) that were collected in a progeny trial (Planches, France). Note that progeny tests in this study only give a rough insight into Mendelian segregation but should help detect null alleles. Originally, we planned to start with 12 progenies (12 female parents and 7 offspring) based on seeds. Due to problems with material, we were had to work with the six progenies described above. Furthermore, we obtained DNA samples of another six *Larix* species. These were *L. sibirica* (21 individuals from one population), *L. kaempferi* (12 individuals from 12 populations) and *L. gmelinii* var. *japonica* (12 individuals from 12 populations), all from Eurasia. From North America, we obtained samples from *L. laricina* (10 individuals from 10 populations), *L. lyallii* (four individuals from four populations) and *L. occidentalis* (four individuals from four populations). For the samples that we collected ourselves, we mostly relied on phloem as tree height (up to 40 m in the trials) made it difficult to collect needles. Phloem was sampled by using a hammer and small leather punch (Ø = 1 cm, length = 10 cm). The sampling technique we developed was rapid and easy. The leather punch was positioned between the bark scales (which can be very thick), and with one to three slight hammer strokes, a small but sufficient sample (Ø = 1 cm, depth 1.5 cm) was recovered. Damage to the tree was minimal,

Table 1 Geographical origin of the 18 *Larix decidua* populations used in this study

Pop ID	Provenance	Country	Lat	Long	Altitude	Trial name	No. of individuals analysed
4	Fernpass	Austria	47.37	10.90	1150	Münden	24
8	Semmering	Austria	47.63	15.77	1200	Riedesel	20
9	Lammerau	Austria	48.11	15.93	610	Münden	21
10	Neulengbach	Austria	48.05	15.93	560	Winnefeld	21
21	Pragelato	Italy	45.02	4.93	1900	Riedesel	22
24	Briançon/Montgenèvre	France	44.93	6.72	1730	Riedesel	22
39	Zabřeh-Dubicko	Czech Republic	49.83	16.97	400	Riedesel	24
40	Ruda	Czech Republic	49.98	16.90	480	Riedesel	24
42	Góra Chełmova	Poland	50.80	21.10	347	Sellhorn	24
43	Bliźyn	Poland	51.07	20.73	330	Sellhorn	24
72	Rekowo	Poland	54.08	17.46	184	*In situ*	20
73	Ruciane Nida	Poland	53.64	21.54	132	*In situ*	24
79	Brusturjany	Ukraine	48.42	24.02	1100	*In situ*	24
81	Vallée de la Tinée	France	44.09	7.09	1070	*In situ*	24
82	Sils Maria	Switzerland	46.43	9.77	2000	*In situ*	24
83	Zinal	Switzerland	46.11	7.64	2000	*In situ*	23
84	Sinaia Forest	Romania	45.33	25.50	1500	*In situ*	24
85	Voineasa Forest	Romania	45.37	23.93	1000	*In situ*	24

and fast regeneration was ensured by sampling during the growing season. Samples were then put into tea bags that were stored in sealed plastic bags with 10 g of silica gel. We isolated DNA from all individuals using 96-well plates. Starting material was mostly phloem (1-cm disc, 0.5 mm thick), but in some cases, needles were used (1–3 needles, cut into 2-mm pieces). For material disruption, we added two 4 mm-tungsten beads to the wells with the starting material. The plates were frozen during 1 min in liquid nitrogen before a 1:30 min disruption by a Mixer Mill (Retsch, Germany). This step was repeated once. An Invisorb DNA 96 plant HTS kit (Invitek, Germany) was then used for DNA isolation following the manufacturer protocol. After isolation, DNA quality was evaluated on a 1% (w/v) agarose gel stained with GelRed (Biotium, USA). DNA concentration was determined by an eight-channel NanoDrop spectrometer and adjusted to 10 ng/μL on a STARTlet 8-channel robot (Hamilton, USA).

Multiplex optimization

Multiplex optimization followed the recommendations given by Guichoux *et al.* (2011a) to guarantee high marker quality, sufficient polymorphism and unambiguous allele binning. Screening for marker candidates started by a literature search of available markers from the closest sister species *Larix kaempferi* (Isoda & Watanabe 2006) and the North American species *L. occidentalis* (Khasa *et al.* 2000; Chen *et al.* 2009). These were tested in simplex reactions. As this did not lead to a sufficient number of suitable candidates, we designed new markers based on 454 sequencing of an enriched microsatellite library. Sequencing was performed by the Swiss company *ecogenics GmbH*. In this sequencing approach, fragments were selected according to their size from genomic DNA enriched with simple sequence repeat (SSR) motifs by using magnetic streptavidin beads and biotin-labelled CT and GT repeat oligonucleotides. The SSR-enriched library was analysed on a Roche 454 platform using the GS FLX Titanium reagents. We worked on 1/16th of a 454 run. For the selection of candidate markers, we used as main criterion the number of dinucleotide repeats (≥12). Note that this decision was taken after a failed attempt based on 454 sequencing of another enriched library produced by another company (data not shown). The problem was that most of the marker candidates had <12 repeats and were monomorphic or displayed limited variation (≤4 alleles). Those having 12 or more repeats (15 out of 100) showed bad amplification or multiple bands. Hence, this experiment was abandoned. The main reason for the failure of the first experiment seemed to be poor sequencing quality and insufficient read length, making it difficult to design primers in the flanking regions of microsatellites with the targeted repeat unit number. In the second attempt, more candidates were obtained. Polymorphism and profile quality were first checked in simplex reactions using the M13 technique (Schuelke 2000). The PCR products were separated on a capillary sequencer (ABI-3730; Applied Biosystems, USA). Each marker was tested on seven individuals from across the range and on an additional DNA pool composed of 12 individuals from different geographical origins. Markers with low polymorphism (<5 alleles/seven individuals, unless additional alleles could be found in the pooled DNA sample) and profiles of low quality (multiple bands, bad amplification) were discarded. Second, the remaining candidates were tested for the presence of null alleles and for large allele dropout (e.g. nonamplification of the longer allele due to differential amplification success) by using segregation analyses. For this purpose, six families composed of the female parent and seven offspring were screened. Third, based on the observations of optimal annealing temperature for the markers and by taking into account the size range of each locus, we composed the two multiplexes. For the multiplex reaction, we used the Qiagen Multiplex PCR kit (Qiagen, Germany). Final volume and final concentration of Mastermix were optimized to reduce final costs (Guichoux *et al.* 2011a). PCR mix for both multiplexes was composed of 4.75 μL of sterile water, 4 μL of Qiagen Multiplex Buffer (2×), 1.25 μL of primer premix and 3 μL of DNA (10 ng/μL). Concentrations of the primer pairs in the primer premix are given in Table 2. The cycling conditions for the two multiplexes differed only in the number of PCR cycles and were as follows: an initial step at 95 °C for 15 min; then 35 (multiplex 1)/30 (multiplex 2) cycles at 94 °C for 30 s, 56 °C for 1 min and 72 °C for 60 s; and a final incubation at 60 °C for 30 min. PCR products were separated on 3% agarose gels stained with GelRed (Biotium, USA) and diluted 18 times in pure water for genotyping with the internal lane size standard LIZ600 on an ABI 3730 (Applied Biosystems, USA). This optimized protocol was communicated to a collaborating laboratory (INRA Orléans, France) to test its robustness. We also tested transferability on the six other larch species from Asia and North America (Table 3).

Genotype scoring and analyses

For subsequent genotype scoring and accurate allele binning based on raw sizes (Guichoux *et al.* 2011a,b), we used STRand (http://www.vgl.ucdavis.edu/informatics/strand.php) and Autobin (http://www4.bordeaux-aquitaine.inra.fr/biogeco/layout/set/print/Ressources/Logiciels/Autobin). For each marker, clear reading rules were defined and illustrated in a handout provided to all readers to increase reading consistency. A first

Table 2 Characteristics of multiplexes 1 and 2 based on 413 *Larix decidua* samples from across its distribution range

Locus*	Reference	Primer sequences (5′–3′)	Motif	Dye	[C]	Size (bp)	N_A	H_O	H_E	F_{IS}	F_{ST}	Mism.†
Multiplex 1												
Ld31	This study	F: TTGAACTAGGGAGATCCGGC R: AATAAAATAGCATTCCATGTGTAGC	$(AC)_{18}$	FAM	2.0	104–147	23	0.751	0.776	0.053	0.074	0
bcLK211	Isoda & Watanabe 2006	F: CCATTCTCCATAGGTTCATTG R: ATGCTCCTTACTAAGTCAGATACAC	$(CT)_{16}$	FAM	2.5	174–242	28	0.746	0.745	0.037	0.052	0
Ld30	This study	F: TTGTAGGTGTGTATGAAAGTTCTG R: TGCCACTCTATTTCCTTAATGCC	$(AC)_{18}$	VIC	1.68	100–138	17	0.674	0.730	0.100	0.133	2
bcLK228	Isoda & Watanabe 2006	F: CCCTAACCCTAGAATCCAATAA R: GAGGAAGGCGACAAGTCATT	$(AG)_{18}$	VIC	0.6	165–215	15	0.869	0.837	−0.013	0.064	0
Ld50	This study	F: GAAGGCGACTTTACATGCCC R: TCCATCTTTATGTCTCTTCCATGC	$(CA)_{18}$	PET	2.2	157–205	19	0.729	0.761	0.066	0.069	2
bcLK189	Isoda & Watanabe 2006	F: ACCATACGCATACCCAATAGA R: AGTTTTCCTTTCCCACACAAT	$(AG)_{17}AT(AG)_6$	NED	1.2	142–172	15	0.733	0.789	0.096	0.074	2
bcLK253	Isoda & Watanabe 2006	F: AACACCATAGTGCAATGTGC R: TCCTCTTGTTGATGCCACTT	$(AG)_{17}$	NED	1.1	195–227	17	0.821	0.806	0.006	0.071	0
Multiplex 2												
Ld58	This study	F: AATGGCAAGAGCAGCAATCC R: TCCAGGAATGATTTATCGAGAGC	$(AC)_{15}$	FAM	2.1	131–183	25	0.775	0.815	0.077	0.073	0
Ld45	This study	F: TGTGGGAGGTATAGCTTGGC R: AGTAGGATGGAATGATGGAAACAC	$(CA)_{13}$	FAM	2.25	198–216	12	0.653	0.723	0.121	0.092	2
Ld42	This study	F: TCGTATGCATTGTCCAAATTTCC R: TCCAAGTGAGGTCACACGAG	$(TG)_{14}$	VIC	1.8	167–191	9	0.551	0.590	0.084	0.134	1
bcLK263	Isoda & Watanabe 2006	F: CGATTGGTATAGTGGTCATTGT R: CCATCATACCTTCTTGAAGAG	$(TC)_{20}$	PET	3.2	185–259	36	0.801	0.868	0.099	0.056	2
Ld101	This study	F: ACACCAAGGACTCTCTGACTAC R: GGTGATTCCAGAAGCAGGTG	$(AC)_{12}$	NED	1.3	179–215	15	0.382	0.395	0.057	0.080	0
Ld56	This study	F: AGCCATCGTGGTTCTTCTTTG R: CTTGTAACTGTGCACCCACC	$(AC)_{16}$	NED	1.6	219–247	14	0.749	0.769	0.049	0.099	0

[C], primer concentration in the primer premix [μM]; N_A, number of alleles; H_O, observed heterozygosity; H_E, expected heterozygosity; F_{IS}, inbreeding index; F_{ST}, fixation index; *Sequence codes for primers: Ld31: lardec012611, Ld30: lardec001529, Ld50: lardec022835, Ld58: lardec022359, Ld45: lardec024823, Ld42: lardec023929, Ld101: lardec025807, Ld56: lardec023228. †Number of Mendelian inconsistencies identified by comparing the genotype of the six half-sib families and their mothers.

estimate of error rate was obtained by counting mismatches on the basis of six positive controls repeated five times and of 10 randomly repeated controls, leading to 34 repetitions (8% of the complete data set). It should be emphasized that positive controls are important to include right from the start as they allow verifying differences across sequencing runs of a given laboratory but also, importantly, across different laboratories. This enables accurate and easy data exchange and combined data analyses. A second estimate of error rate was obtained by comparing scoring across readers at a subset of 264 individuals. Two types of errors were distinguished. Type A corresponds to cases where reader 1 calls a genotype as heterozygous and reader 2 as homozygous or vice versa. Type B corresponds to cases where a wrong allele is called by one of the readers. These error types can be corrected after being identified. There can be other kinds of incoherencies between readers, for instance if reader 1 calls a genotype and reader 2 sets missing data or if readers are calling different genotypes due to profile ambiguities which, in contrast to type A and B errors, cannot be corrected and eventually have to be set as missing data. Total allele number (n_A), observed (H_O) and expected (H_E) heterozygosity, inbreeding index (F_{IS}) and fixation index (F_{ST}) were estimated based on 413 individuals from 18 populations from across the range of 20–24 individuals per population (Table 2) using GENEPOP (Raymond & Rousset 1995; Rousset 2008) and GENEALEX (Peakall & Smouse 2006). Presence of null alleles was checked on a subset of 12 populations that had a minimum of 19 individuals without missing data using MICRO-CHECKER v.2.2.0.3 (Van Oosterhout *et al.* 2004).

Results and discussion

Multiplex optimization

We tested the nine most promising candidate loci originating from the literature in simplex reactions. Most promising means that these loci had already been proven to amplify well and to be polymorphic (Pluess 2011). The 454 sequencing approach resulted in 27 041 reads with an average length of 184 bp. Of these, 3311 contained a microsatellite insert with a tetra- or a trinucleotide of at least six repeat units or a dinucleotide of at least 10 repeat units. Primer design was possible in 312 reads, of which 100 were tested. Simplex PCR conditions had to be adjusted for each marker. This was because some candidates with long motifs needed more PCR cycles than others to be successfully amplified, whereas such an increase resulted in nonspecific products for other candidates. Testing marker profile quality and polymorphism (with a target of ≥5 alleles) over seven individuals and an additional DNA pool resulted in the selection of seven out of nine loci, originating from the literature and 26 out of the 100 newly designed candidates. For identifying the latter, we started by excluding 34 markers with stutter bands or bad amplification, as identified on agarose gels. After genotyping the 66 remaining candidates with ABI capillary sequencer, we removed loci with multiple bands (19 markers), bad amplification (nine markers) or insufficient polymorphism (12 markers), leaving only 26 suitable markers. Low success rates have also been reported in other plant species with large genomes (e.g. Röder *et al.* 1995; Pfeiffer *et al.* 1997). In particular, a detailed investigation of the causes of amplification failures of microsatellite markers in *Picea abies* using Southern blot experiments showed that, in six out of seven cases tested, loci showing high-quality profiles corresponded to single- or low-copy sequences (Pfeiffer *et al.* 1997). This suggests that to identify sufficient candidate loci in species with large genomes, careful evaluation of the sequencing depth is necessary. Indeed, many markers will fail to produce clear PCR if the flanking regions (used to design the primers) are repeated elsewhere in the genome.

The other evaluation steps included allele binning as well as tests of Mendelian segregation to search for large allele dropout or null alleles. This led to the selection of a subset of 22 candidates for multiplexing for which binning was clear. There was no indication for nonamplification corresponding to large allele dropout. Null alleles were apparent at some markers (Ld50, bcLK189, bcLK263, Ld42, Ld45, Ld30, Tables 2 and S1, Supporting Information). They were identified by comparing the female parent genotype with the genotypes of the offsprings. As there were too few loci combining sufficient polymorphism and zero mismatches in the progeny test, we retained those loci with no more than two Mendelian discrepancies and sufficient polymorphism. The number of possible multiplex combinations was limited as allele size ranges of the markers were large, and as we had to take into account hetero-

Table 3 Results of transferability tests: Number of alleles per locus for the six other *Larix* species tested

Locus	*L. kaempferi** (12)‡	*L. gmelinii** (12)‡	*L. sibirica** (21)‡	*L. laricina*† (10)‡	*L. lyallii*† (4)‡	*L. occidentalis*† (4)‡
Multiplex 1						
Ld31	8	8	5¶	5	2	2
bcLK211	4	4	4	5	2	4
Ld30	–§	–§	–§	–§	3	–§
bcLK228	3	7	9	8	4	–††
Ld50	12	8	3¶	–§	–§	–§
bcLK189	6	8	6¶	–§	2	–§
bcLK253	4	3	3	8	4	3
Multiplex 2						
Ld58	–**	–§	6	5	3	3
Ld45	–**	11	2¶	3	1	4
Ld42	6	7	–§	5	3	6
bcLK263	5	3	15	12	7	6
Ld101	7	12	9	2	2	3
Ld56	9	9	6	2	2	8

*Eurasian species.
†American species.
‡Number of tested individuals in brackets.
§No amplification.
¶Amplification in only a subset of individuals.
**Only one locus (either Ld58 or Ld45) was amplified but it was not clear which one.
††Profile ambiguities.

geneities in PCR conditions. We were finally able to design one 6-plex and one 7-plex that we validated over the 413 individuals coming from across the range. Laboratory protocols were successfully tested in another laboratory (INRA Orléans, UR AGPF, Vanina Guérin, personal communication), confirming their robustness.

Genotype scoring and analyses

Clear binning of alleles was confirmed over the whole sample. The marker allelograms, that is, diagrams showing all detected allele raw sizes ranked in increasing order, are shown in Figs 1 and 2. They indicate that most markers are characterized by a clear succession of dinucleotide repeats over large size ranges (between 28 and 74 bp; mean, 42 bp) and a remarkably high number of alleles (9–36, mean of 20). Off-ladder microvariants, that is, allele sizes that are in-between the sizes expected from the repetition of dinucleotide units, were only observed for one of the 13 loci (Ld45, see Fig. 2), and as these intermediate size variants were clearly separated from the neighbouring size classes, they did not cause problems for binning. In fact, such clearly defined off-ladder microvariants can improve the precision of the analyses when they are correctly identified (Guichoux *et al.* 2011b). There was no incoherency across repetitions (positive and random controls). Type A and type B error rates ranged between 0% and 0.19% (mean of 0.05%) for multiplex 1, and between 0 and 1.3% (mean, 0.32%) for multiplex 2. The frequencies of cases that could not be scored were slightly higher. They ranged from 0% to 2.1% (mean of 0.32%) for multiplex 1 (mainly caused by an unspecific product being amplified at locus bcLK211) and from 0% to 0.94% (mean of 0.16%) for multiplex 2 (mainly caused by excessive stuttering of some long alleles at locus bcLK263). All markers were highly polymorphic. Number of allele (N_A), observed (H_O) and expected (H_E) heterozygosity, inbreeding coefficient (F_{IS}) and coefficient of differentiation (F_{ST}) are shown in Table 2. F_{IS} values were slightly positive except at locus bcLK228, which is an indication of the presence of null alleles. To better estimate the frequency of null alleles, we relied on the software MICRO-CHECKER v.2.2.0.3 (Van Oosterhout *et al.* 2004). The method searches for loci with a significant homozygote excess evenly distributed across all homozygote classes. The analysis of a subsample of 12 populations pointed to 13 cases out of 156 population/marker combinations suggesting the presence of null alleles, involving 7 of the 13 loci (Table S2, Supporting Information). The average null allele frequency across all loci per population ranged between 2.4% and 6.6%. As shown in simulation studies of Chapuis & Estoup (2007), null alleles with frequencies between 5% and 8% should have only reduced effects on classical estimates of population differentiation. Hence, the genotyping results could be used directly with no need of data correction for null allele genotypes. However, as recommended by Oddou-Muratorio *et al.* (2009), correction of null allele genotypes remains of interest for some other analyses. This can be carried out by using the correction options implemented in different software (e.g. MICRO-CHECKER).

Transferability of the markers was tested in six other *Larix* species and led to heterogeneous results, although new loci have been identified in all species (Table 3). Note that allelic richness estimates for the different species are not directly comparable as we used different sample sizes. In cases of transfer problems, we differentiated four different kinds of problems that are given in the table (a–d) and explained in its footnote. While our data sets did not allow detailed conclusions about success of transferability according to phylogenetic relationships, transfer rate was lower for *L. sibirica*. According to nuclear phylogenies based on AFLP and ITS data (Gernandt & Liston 1999; Semerikov *et al.* 2003), this species forms a clade with *L. decidua* and the other Eurasian species. However, other phylogenies based on chloroplast data separate *L. sibirica* from all other *Larix* species, in agreement with our observations (Quian *et al.* 1995; Semerikov *et al.* 2003).

Conclusions and perspectives

This study shows that even for species with very large genome size and low overall polymorphism, new highly informative microsatellites can be developed based on 454 sequencing, provided that some precautions are taken. These precautions involve (i) verifying whether the initial sequencing of the enriched microsatellite library can guarantee a sufficient read length and sequencing depth to identify enough candidates and (ii) focusing from the beginning on microsatellites with high number of repeats. The two *Larix decidua* multiplexes developed in this study are currently applied on a range-wide sample of 43 populations to resolve in detail past population dynamics, which should help identify and protect valuable genetic resources.

Acknowledgements

Genotyping was performed in the Genome-Transcriptome facility of the Functional Genomic Centre of Bordeaux. We thank Franck Salin, Sarah Monllor, Christophe Boury, Philippe Chaumeil and Camille Lepoittevin as well as Andres Buser (*ecogenics GmbH*) for support and assistance during genotyping and primer design. For assistance in the various fields of the project, we thank Vanina Guérin, Patrick Léger, Guillaume Lalanne-Tisné, Anne Laure Bouillot, Erwan Guichoux, Lélia Lagache, Olivier Lepais and Maxime Nardin. For support with plant material and assistance during sampling, we thank Luc Pâques, Jutta Mäsgen, Laurent Bouffier, Frédéric Lagane, Christoph Sperisen, Yuri

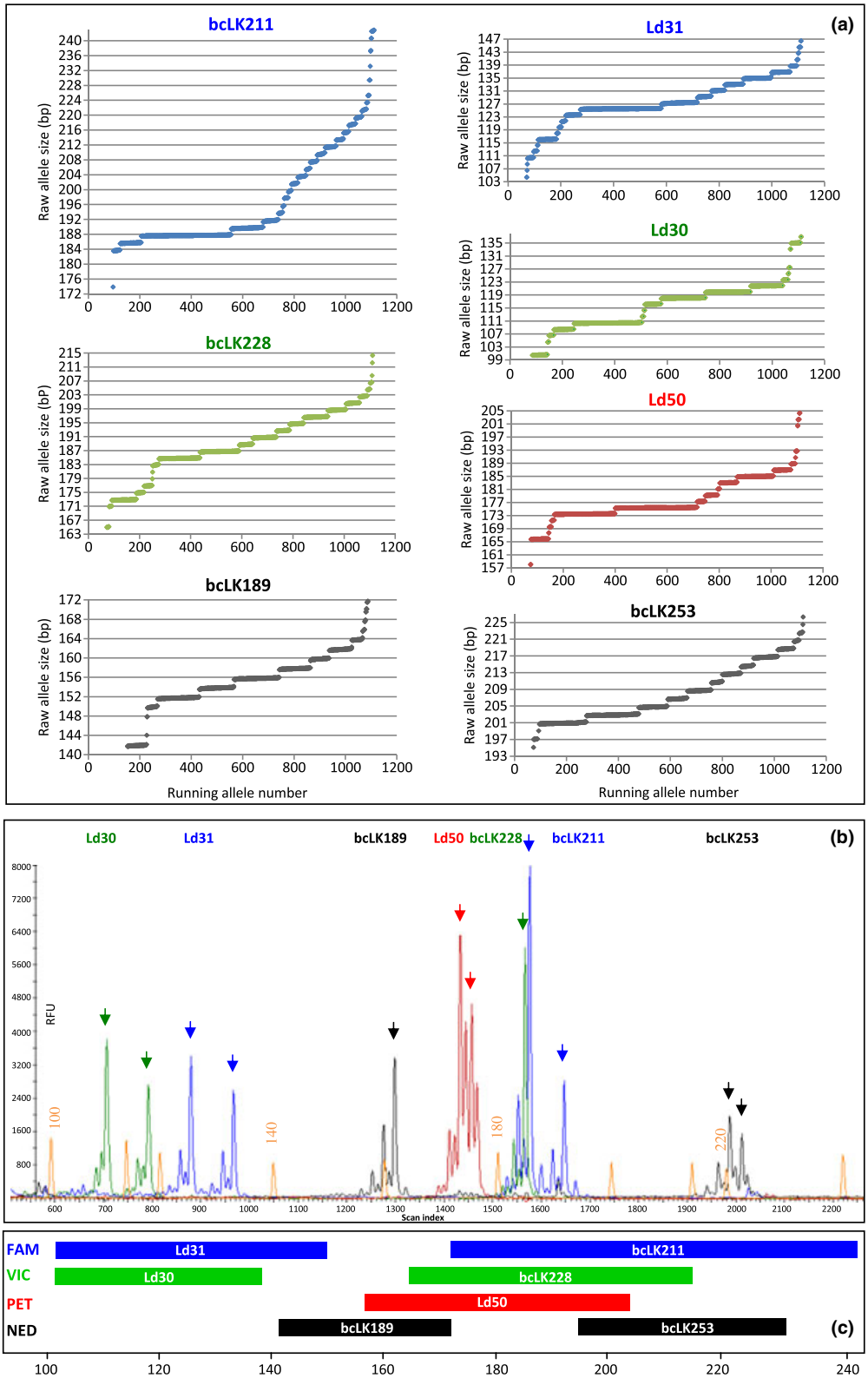

Fig. 1 Binning, profiles and marker ranges for multiplex 1. (a) Allelograms based on 413 individuals from across the range. The running allele number corresponds to the ranking number of the allele in the list of allele raw sizes, ranked in increasing order. (b) Example of an individual electropherograms. (c) Range sizes based on the same sample as in (a). In (b), arrows point to alleles at each locus, and small peaks with numbers (base pairs) correspond to fragments of the internal size standard LIZ600.

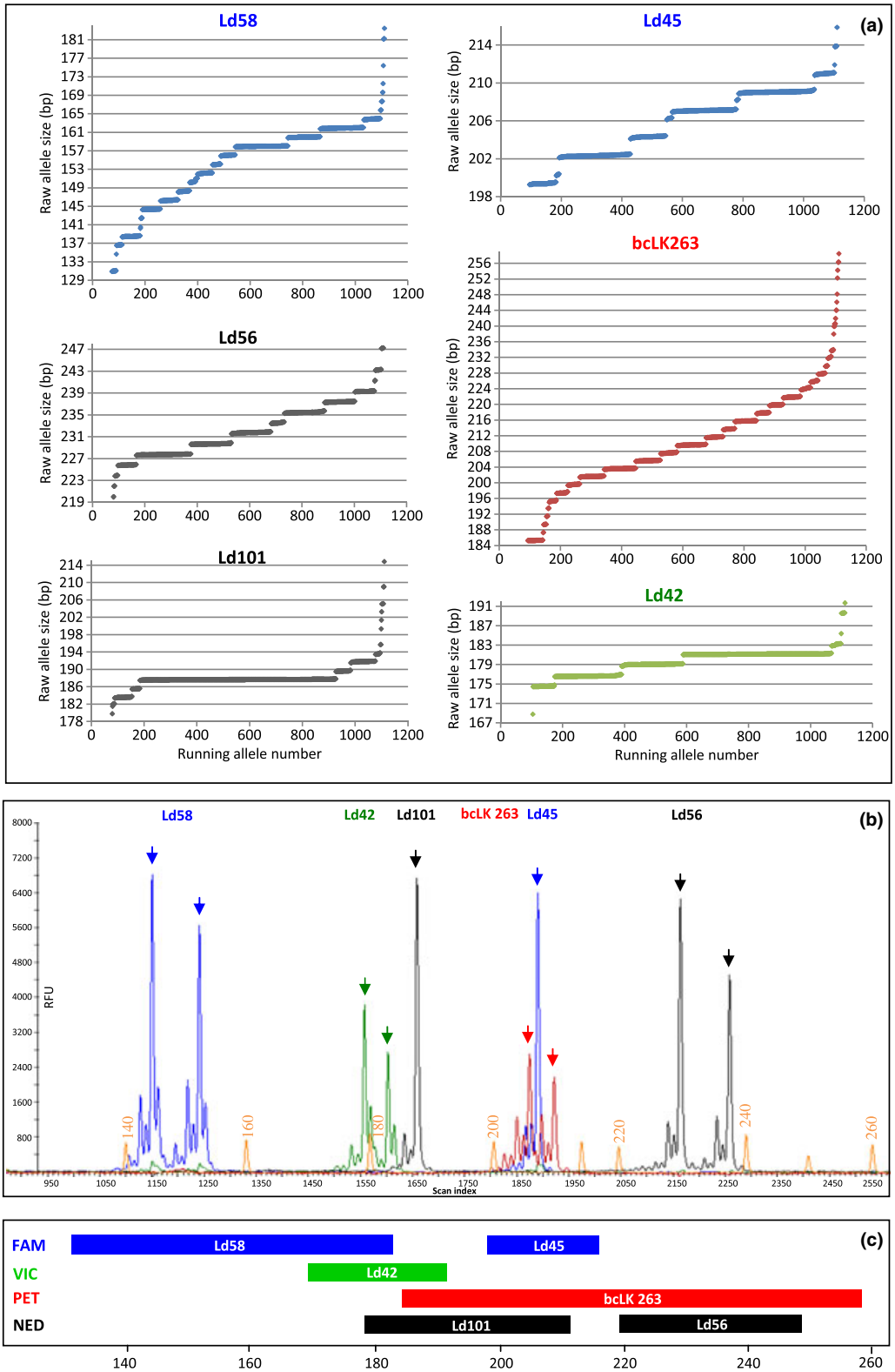

Fig. 2 Binning, profiles and marker ranges for multiplex 2. (a) Allelograms based on 413 individuals from across the range. The running allele number corresponds to the ranking number of the allele in the list of allele raw sizes, ranked in increasing order. (b) Example of an individual electropherograms. (c) Range sizes based on the same sample as in (a). In (b), arrows point to alleles at each locus, and small peaks with numbers (base pairs) correspond to fragments of the internal size standard LIZ600.

Hayda, Ihor Neyko, Guy Bour, Czeslaw Koziol, Philippe Rozenberg, Leopoldo Sanchez, Frédéric Millier, Jean-Paul Charpentier, Owe Martinsson, Lars Karlman, Hermann Spellmann (NW-FVA), Keiya Isoda, Nathalie Isabel, Marie-Claude Gros-Louis, the Ukrainian Research Institute for Mountain forestry, Ivano-Frankivsk, the Polish Osusznica Forest Institute (Nadleśnictwo Osusznica), and the Polish Maskulińskie Forest Institute (Nadleśnictwo Maskulińskie).

References

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

Chen C, Liewlaksaneeyanawin C, Funda T, Kenawy A, Newton CH, El-Kassaby YA (2009) Development and characterization of microsatellite loci in western larch (*Larix occidentalis* Nutt.). *Molecular Ecology Resources*, **9**, 843–845.

Colombaroli D, Henne PD, Kaltenrieder P, Gobet E, Tinner W (2010) Species responses to fire, climate and human impact at the tree line in the Alps as evidenced by palaeo-environmental records and a dynamic simulation model. *Journal of Ecology*, **98**, 1346–1357.

Ellegren H (2000) Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in Genetics*, **5**, 435–445.

Gernandt DS, Liston A (1999) Internal transcribed spacer region evolution in *Larix* and *Pseudotsuga* (Pinaceae). *American Journal of Botany*, **86**, 711–723.

Greibhuber J (1986) Severely distorted Feulgen DNA amounts in *Pinus* (Coniferophytina) after non-additive fixations as a result of meristematic self- tanning with vacuole contents. *Canadian Journal of Genetics and Cytology*, **28**, 409–415.

Guichoux E, Lagache L, Wagner S *et al*. (2011a) Current trends in microsatellite genotyping. *Molecular Ecology Resources*, **11**, 591–611.

Guichoux E, Lagache L, Wagner S, Léger P, Petit RJ (2011b) Two highly validated multiplexes (12-plex and 8-plex) for species delimitation and parentage analysis in oaks (*Quercus* spp.). *Molecular Ecology Resources*, **11**, 578–585.

Isoda K, Watanabe A (2006) Isolation and characterization of microsatellite loci from *Larix kaempferi*. *Molecular Ecology*, **6**, 664–666.

Kelkar YD, Tyekucheva S, Chiaromonte F, Makova KD (2008) The genome-wide determinants of human and chimpanzee microsatellite evolution. *Genome Research*, **18**, 30–38.

Khasa DP, Newton CH, Rahman MH, Jaquish B, Dancik BP (2000) Isolation, characterization, and inheritance of microsatellite loci in alpine larch and western larch. *Genome*, **43**, 439–448.

Lefèvre S, Wagner S, Petit RJ, De Lafontaine G (2011) Multiplex kits of microsatellite markers for genetic studies of beech. *Molecular Ecology Resources*, doi: 10.1111/j.1755-0998.2011.03094.x.

Lewandowski A, Mejnartowicz L (1991a) Levels and patterns of allozyme variation in some European Larch (*Larix decidua*) populations. *Hereditas*, **115**, 221–226.

Lewandowski A, Mejnartowicz L (1991b) Linkage analysis of allozyme loci in Polish larch (*Larix decidua* subsp. *polonica* (Racib.) (Domin). *Hereditas*, **114**, 107–109.

Lewandowski A, Burczyk J, Mejnartowicz L (1991) Genetic structure and the mating system in an old stand of Polish Larch. *Silvae Genetica*, **40**, 75–79.

Maier J (1992) Genetic variation in European larch. *Annals Forest Science*, **49**, 39–47.

McComb AL (1955) The European Larch: its races, site requirements and characteristics. *Forest Science*, **1**, 298–318.

Oddou-Muratorio S, Vendramin G, Buiteveld J, Fady B (2009) Population estimators or progeny tests: what is the best method to assess null allele frequencies at SSR loci? *Conservation Genetics*, **10**, 1343–1347.

Peakall ROD, Smouse PE (2006) GENEALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.

Pfeiffer A, Olivieri AM, Morgante M (1997) Identification and characterization of microsatellites in Norway spruce (*Picea abies* K.). *Genome*, **40**, 411–419.

Pluess AR (2011) Pursuing glacier retreat: genetic structure of a rapidly expanding *Larix decidua* population. *Molecular Ecology*, **20**, 473–485.

Quian T, Ennos RA, Helgason T (1995) Genetic relationships among larch species based on analysis of restriction fragment variation for chloroplast DNA. *Canadian Journal of Forest Research*, **25**, 1197–1202.

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

Röder MS, Plaschke J, König SU *et al*. (1995) Abundance, variability and chromosomal location of microsatellites in wheat. *Molecular and General Genetics MGG*, **246**, 327.

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

Rubner K (1953) *Die pflanzengeographischen Grundlagen des Waldbaus*. Neumann, Radebeul, Berlin.

Santana QC, Coetzee MPA, Steenkamp ET *et al*. (2009) Microsatellite discovery by deep sequencing of enriched genomic libraries. *BioTechniques*, **46**, 217–223.

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology*, **18**, 233–234.

Semerikov VL, Zhang H, Sun M, Lascoux M (2003) Conflicting phylogenies of *Larix* (Pinaceae) based on cytoplasmic and nuclear DNA. *Molecular Phylogenetics and Evolution*, **27**, 173–184.

Smith SA, Donoghue MJ (2008) Rates of molecular evolution are linked to life history in flowering plants. *Science*, **322**, 86–89.

Tinner W, Kaltenrieder P (2005) Rapid responses of high-mountain vegetation to early Holocene environmental changes in the Swiss Alps. *Journal of Ecology*, **93**, 936.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

S.W. conceived the study with R.J.P., performed the experiments, and produced and analysed the data. S.W. wrote the paper with the help of R.J.P. and S.G. All authors have checked and approved the final version of the manuscript.

Data Accessibility

DNA sequences: Genbank accessions JQ340312-JQ340319.

Sample locations and microsatellite data: DRYAD entry doi: 10.5061/dryad.08507r35.

Supporting Information

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

Table S1 Simplex test results final markers

Table S2 MICRO-CHECKER results

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.