

## Lessons learned from microsatellite development for nonmodel organisms using 454 pyrosequencing

C. N. SCHOEBEL\*, S. BRODBECK\*, D. BUEHLER\*, C. CORNEJO\*, J. GAJUREL\*†, H. HARTIKAINEN‡, D. KELLER\*, M. LEYS§, Š. ŘÍČANOVÁ¶, G. SEGELBACHER\*\*, S. WERTH\*†† & D. CSENCICS\*

\*Biodiversity & Conservation Biology, WSL Swiss Federal Research Institute, Birmensdorf, Switzerland

†Central Department of Botany, Tribhuvan University, Kathmandu, Nepal

‡Department of Zoology, Natural History Museum, London, UK

§Department of Aquatic Ecology, eawag Swiss Federal Research Institute, Dübendorf, Switzerland

¶Department of Population Biology, Institute of Vertebrate Biology, ASCR, Brno, Czech Republic

\*\*Department Wildlife Ecology and Management, University of Freiburg, Freiburg, Germany

††Faculty of Life and Environmental Sciences, University of Iceland, Sturlugata 7, 101, Reykjavik, Iceland

### Keywords:

comparative studies;  
 conservation genetics;  
 massively parallel sequencing;  
 next generation sequencing technology;  
 population genetics;  
 shotgun sequencing.

### Abstract

Microsatellites, also known as simple sequence repeats (SSRs), are among the most commonly used marker types in evolutionary and ecological studies. Next Generation Sequencing techniques such as 454 pyrosequencing allow the rapid development of microsatellite markers in nonmodel organisms. 454 pyrosequencing is a straightforward approach to develop a high number of microsatellite markers. Therefore, developing microsatellites using 454 pyrosequencing has become the method of choice for marker development. Here, we describe a user friendly way of microsatellite development from 454 pyrosequencing data and analyse data sets of 17 nonmodel species (plants, fungi, invertebrates, birds and a mammal) for microsatellite repeats and flanking regions suitable for primer development. We then compare the numbers of successfully lab-tested microsatellite markers for the various species and furthermore describe diverse challenges that might arise in different study species, for example, large genome size or nonpure extraction of genomic DNA. Successful primer identification was feasible for all species. We found that in species for which large repeat numbers are uncommon, such as fungi, polymorphic markers can nevertheless be developed from 454 pyrosequencing reads containing small repeat numbers (five to six repeats). Furthermore, the development of microsatellite markers for species with large genomes was also with Next Generation Sequencing techniques more cost and time-consuming than for species with smaller genomes. In this study, we showed that depending on the species, a different amount of 454 pyrosequencing data might be required for successful identification of a sufficient number of microsatellite markers for ecological genetic studies.

### Introduction

Microsatellites, that is, short tandemly repeated motifs of 2–6 bases, also known as simple sequence repeats (SSRs), are still widely used as genetic markers in evolutionary and ecological studies (e.g. Dolgener *et al.*, 2012; Hurtado *et al.*, 2012; Kopatz *et al.*, 2012; Manoel *et al.*, 2012). Their ease of use, their co-dominant inheritance and their supposed neutrality are the key benefits

Correspondence: Dr. Corine N. Schoebel, Biodiversity & Conservation Biology, WSL Swiss Federal Research Institute, Birmensdorf, Switzerland.

Tel.: +41 44 739 25 28; fax: +41 44 739 22 54;  
 e-mail: corine.schoebel@wsl.ch

of microsatellites (Goldstein & Schlötterer, 1999; Selkoe & Toonen, 2006). Traditionally, a major obstacle for their use in evolutionary, ecological and conservation studies of nonmodel species was the need to isolate a reasonable number of microsatellite markers by repeat enrichment, cloning and Sanger sequencing, a cost and labour intensive process. Mostly, microsatellites had to be developed de novo for every species under study, as cross-amplification from congeneric species is not generally feasible (Barbara *et al.*, 2007). Cross-amplifying markers are often monomorphic and may contain null alleles (Selkoe & Toonen, 2006).

With the advent of Next Generation Sequencing technologies, the identification and characterization of microsatellites in nonmodel species has become more feasible. Among the different Next Generation Sequencing platforms, the 454 pyrosequencing technology by Roche is currently the most used for microsatellite development (Gardner *et al.*, 2011; Guichoux *et al.*, 2011; Zalapa *et al.*, 2012), as it produces relatively long reads (up to 600 bp; <http://454.com/products/gs-flx-system/index.asp>). Longer reads are more likely to contain microsatellites including their flanking regions, which is required for successful primer development. As the availability of 454 pyrosequencing platforms has expanded, traditional repeat enrichment and cloning is no longer the method of choice for microsatellite development (Abdelkrim *et al.*, 2009; Castoe *et al.*, 2010). Nonetheless, repeat enrichment steps can still be combined with 454 pyrosequencing (Santana *et al.*, 2009; Lepais & Bacles, 2011, Malausa *et al.*, 2011).

In this article, we highlight the general ease of marker development from 454 pyrosequencing data for nonmodel species, but also discuss potential problems as well as likely solutions. We review a broad range of nonmodel species, some of which had proved to be challenging in microsatellite development, such as species with large genome sizes, fungi, whose genomes are known to contain few microsatellites relative to other taxa, or cases where genomic DNA extracted of the focal species is always intermixed with DNA from other species, such as food or parasites. More precisely, we address and discuss the following questions. How much 454 pyrosequencing data is necessary to be able to identify a sufficient number of microsatellite markers for a typical population genetic study? What are the benefits and problems of a pooled run with several individually labelled species? What can be done if the extracted genomic DNA of an organism of interest is not pure but intermixed with DNA of other organisms (e.g. in metagenomic studies or symbiotic organisms)? What are the prospects of success for microsatellite identification in fungi, regarding the diverse difficulties of microsatellite development for this organismic group discussed in literature (e.g. Dutech *et al.*, 2007)?

## Materials and methods

### Study species

We studied nonmodel organisms belonging to a range of taxonomic groups, i.e. fungi (including Oomycetes), plants, bryozoans, insects, birds and mammals.

#### Fungi

The lichen *Lobaria pulmonaria*, commonly known as tree lungwort, consists of an ascomycete fungus living in a symbiotic relationship with a green alga and a cyanobacterium. *Lobaria pulmonaria* is sensitive to air pollution and is also affected by habitat loss and changes in forest management. Its populations have declined across Europe, and *L. pulmonaria* is considered threatened in many lowland areas (Hallingbäck & Martinsson, 1987). Here, we investigate the fungal partner of this lichen symbiosis, obtained from an axenic culture.

Species of the genus *Phytophthora* (Straminipila, Oomycetes) are well known to cause devastating diseases on a great number of crops, ornamentals and native plants worldwide (Erwin & Ribeiro, 1996). The most famous example is presumably *Phytophthora infestans*, the causal agent of the so called 'Irish potato famine' (Oxford, 1996). Many significant recent declines and dieback phenomena in forest ecosystems worldwide have been associated with *Phytophthora* species. Recently, it was shown based on DNA sequence data that the *P. citricola* species complex comprises four species, namely *P. multivora*, *P. plurivora*, *P. pini* (also called *P. citricola* 1) and *P. citricola sensu strictu* (s.s.), which show different geographical distributions and host ranges (Jung & Burgess, 2009). *Phytophthora plurivora* is frequent in forests, semi-natural ecosystems and nurseries in Europe where it is involved in widespread beech and oak declines (Jung & Burgess, 2009). Moreover, it is also found in plantations and nurseries in North America. *Phytophthora gonapodyides* is considered a saprophytic species commonly present in streams but its ecological role is still unclear (Kroon *et al.*, 2011). Population genetic analyses are being conducted to better understand the population structure and the main pathways of spread of these cosmopolitan plant pathogens.

#### Plants

The alpine rock-cress, *Arabis alpina*, is a diploid alpine rosette plant with an extensive arctic-alpine distribution. Its growth conditions reflect varying low-competition habitats between 400 and 3200 m a.s.l. (Schultze-Motel, 1986). *Arabis alpina* has recently become important in eco-genomic research (Ansell *et al.*, 2008; Wang *et al.*, 2009; Manel *et al.*, 2010).

The dwarf bulrush (*Typha minima*) is an endangered pioneer plant species of riparian flood plains. *T. minima* was once widespread along braided rivers of Europe

and central Asia, but is now threatened with extinction: its European distribution is limited to a few rivers in the Alps (Müller, 1991; Prunier *et al.*, 2010).

The Swiss stone pine (*Pinus cembra*) is a key species of the alpine tree line ecotone. Its natural range is in the European Alps and the Carpathians, mainly near or at the timber line. Occurrences are locally abundant, in particular in the Alpine core range. In peripheral areas, populations tend to be small and fragmented. *Pinus cembra* is not vulnerable in the Alps, but it is of conservation concern in the Carpathians. In addition, *P. cembra* is anticipated to be sensitive to climate change (Casalegno *et al.*, 2010).

The Himalayan Yew (*Taxus wallichiana*) is a medium-sized evergreen coniferous tree, endemic to the Himalaya and occurring at altitudes from 2000 to 3500 m a.s.l. (Rushforth, 1987; Farjon, 1998).

#### Invertebrates

The alpine mayfly *Baetis alpinus* is a widespread and abundant macroinvertebrate of the Palearctic region (Humpesch, 1979). Larvae are eurythermal and are found in swift stony streams up to 2600 m a.s.l. (Sartori & Landolt, 1999).

The large marsh grasshopper (*Stethophyma grossum*) is specialized to wet meadows often found in agricultural landscapes and is frequently used as an indicator species for extensively managed wet meadows. In Switzerland, it is red-listed as vulnerable (Monnerat *et al.*, 2007). Its distribution ranges throughout Europe and Siberia.

The freshwater bryozoan *Fredericella sultana* is a benthic, colonial suspension feeder, commonly found growing on submerged roots, stones and other substrates in a wide range of freshwater habitats (Wood & Backus, 1992). *Fredericella sultana* colonies grow rapidly and reach high densities during summer, regressing in the autumn and largely overwintering as asexual, dormant resting stages (statoblasts). *Fredericella sultana* acts as the most common host for the causative parasite of the Proliferative Kidney Disease (PKD) of salmonid fish. PKD is an economically important disease, causing losses on trout farms and impacting populations of wild salmonids (Okamura *et al.*, 2011).

#### Vertebrates

The rock ptarmigan (*Lagopus muta*) is a grouse species inhabiting alpine areas. Although still hunted in parts of its circumpolar range, it is regarded as susceptible to climate change (BirdLife International: <http://www.birdlife.org/datazone/species/index.html>).

The El Oro parakeet (*Pyrrhura orcesi*) is an endangered parrot species, which is endemic to south-western Ecuador. Because of habitat fragmentation and habitat loss due to logging and cattle ranging, population size has been declining and is now estimated to be less than 1000 individuals (BirdLife International: <http://www.birdlife.org/datazone/species/index.html>).

The blackcap (*Sylvia atricapilla*) is a widespread and common songbird across most of Europe. As a species with different migratory strategies, it serves as study organism for evolutionary studies and is one of the few demonstrated cases of fast microevolutionary processes (Rolshausen *et al.*, 2009).

The garden dormouse (*Eliomys quercinus*) is a small mammal of West-Mediterranean origin, which has spread to Central and Eastern Europe as a Holocene immigrant (Horáček, 1986). It occurs in broadleaved and coniferous forests, shrub land, vineyards, orchards, in treeless landscapes and regularly enters buildings. In recent decades, the species has declined (Mitchell-Jones *et al.*, 1999) and is now regarded as an endangered species by IUCN (Bertolino *et al.*, 2008).

#### Sample preparation

Genomic DNA was extracted with different methods. All DNA fulfilled the criteria of the commercial companies providing the 454 pyrosequencing service, that is, a DNA amount of > 2 µg and a DNA concentration of > 50 ng/µL. Apart from the DNA of *F. sultana*, only pure, single species DNA was subject to 454 pyrosequencing.

#### 454 pyrosequencing

DNA of all species was subject to 454 pyrosequencing using GS-FLX titanium chemistry. The detailed technique is described in Margulies *et al.* (2005). With the exception of *E. quercinus* (Genoscreen, F), *F. sultana* (University of Liverpool, GB) and *B. alpinus* (Ecogenics, CH), all samples were sequenced at Microsynth (Balgach, CH). Most species were sequenced in single runs, without prior repeat enrichment. As exceptions, all 5 *Phytophthora* species were sequenced as a pooled sample with individual MID tags, and, for *B. alpinus* and *E. quercinus* (pooled with seven other species in 1/4 of a full run) a repeat enrichment step was included prior to sequencing. 454 pyrosequencing runs can generally be ordered as a full plate or as a fraction of a plate, for the exact run characteristics per species, see Table 1.

#### Data analysis

We used the software MSATCOMMANDER 1.0.8 for Mac (Faircloth, 2008) to screen raw data for microsatellite motifs, as it is straightforward, user friendly and was often successfully used for microsatellite development (e.g. Csencsics *et al.*, 2010; Reid *et al.*, 2012). We separately searched for perfect di-, tri- and tetranucleotide motifs with five or more repeats. We decided to neglect penta- and hexanucleotide motifs as they are less frequently used in the literature. In addition, it is less likely that loci comprising penta- and hexanucleotide

**Table 1** Species-specific information on genome size (c-value) and 454 run.

| Species                            | Family          | c-value (pg) and reference   | Run size (proportion of full run) | Total number of reads | Average read length (bp) | Repeat enrichment | Pooled sample |
|------------------------------------|-----------------|--|-----------------------------------|-----------------------|--------------------------|-------------------|---------------|
| <i>Lobaria pulmonaria</i>          | Lobariaceae     | 0.03* \$; Armaleo & May, 2009  | 0.25                              | 181 541               | 313.8                    | no                | no            |
| <i>Phytophthora gonapodyides</i>   | Pythiaceae      | 0.082* \$; <i>P. ramorum</i> & <i>P. sojae</i> ; Tyler <i>et al.</i> , 2006  | 0.025                             | 14 857                | 376.8                    | no                | yes           |
| <i>Phytophthora multivora</i>      | Pythiaceae      | 0.082* \$; <i>P. ramorum</i> & <i>P. sojae</i> ; Tyler <i>et al.</i> , 2006  | 0.025                             | 20 716                | 376.8                    | no                | yes           |
| <i>Phytophthora plurivora</i>      | Pythiaceae      | 0.082* \$; <i>P. ramorum</i> & <i>P. sojae</i> ; Tyler <i>et al.</i> , 2006  | 0.025                             | 16 162                | 376.8                    | no                | yes           |
| <i>Phytophthora citricola</i> s.s. | Pythiaceae      | 0.082* \$; <i>P. ramorum</i> & <i>P. sojae</i> ; Tyler <i>et al.</i> , 2006  | 0.025                             | 15 009                | 376.8                    | no                | yes           |
| <i>Phytophthora pini</i>           | Pythiaceae      | 0.082* \$; <i>P. ramorum</i> & <i>P. sojae</i> ; Tyler <i>et al.</i> , 2006  | 0.025                             | 18 465                | 376.8                    | no                | yes           |
| <i>Arabis alpina</i>               | Brassicaceae    | 0.65 \$; Greilhuber <i>et al.</i> , 2006   | 0.0625                            | 40 367                | 354                      | no                | no            |
| <i>Typha minima</i>                | Typhaceae       | 0.33; <a href="http://data.kew.org/cvalues">http://data.kew.org/cvalues</a>  | 0.125                             | 76 692                | 341.3                    | no                | no            |
| <i>Pinus cembra</i>                | Pinaceae        | 24.1; Greilhuber, 1986   | 0.0625                            | 37 496                | 336                      | no                | no            |
| <i>Taxus wallichiana</i>           | Taxaceae        | 10.22* \$; Leitch <i>et al.</i> , 2001   | 0.1875                            | 119 508               | 337                      | no                | no            |
| <i>Baetis alpinus</i>              | Baetidae        | 0.547*; <i>Baetis bicaudatus</i> ; Vance, 1996   | 0.0625                            | 7075                  | 139.8                    | yes               | no            |
| <i>Stethophyma grossum</i>         | Acrididae       | 8.77*; <i>Locusta migratoria</i> ; <i>Acrida cononica</i> ; <i>Caledia captiva</i> ; <a href="http://www.genomesize.com">http://www.genomesize.com</a> | 0.0625                            | 52 693                | 355                      | no                | no            |
| <i>Fredericella sultana</i>        | Fredericellidae | 0.2*; <i>Watersipora</i> sp.; Schopf, 1983   | 0.25                              | 196 025               | 388                      | no                | no            |
| <i>Lagopus muta</i>                | Phasianidae     | 1.07 \$; <i>Gallus gallus</i> ; <a href="http://www.ncbi.nlm.nih.gov/genome">http://www.ncbi.nlm.nih.gov/genome</a>                                    | 0.0625                            | 33 362                | 360.6                    | no                | no            |
| <i>Pyrrhura orcesi</i>             | Psittacidae     | 1.26 \$; Zebra finch; <a href="http://www.ncbi.nlm.nih.gov/genome">http://www.ncbi.nlm.nih.gov/genome</a>  | 0.125                             | 80 093 #              | 324.2                    | no                | no            |
| <i>Sylvia atricapilla</i>          | Silviidae       | 1.09; <a href="http://www.genomesize.com">http://www.genomesize.com</a>  | 0.0625                            | 43 939                | 286.5                    | no                | no            |
| <i>Eliomys quercinus</i>           | Gilridae        | 4.17*; <a href="http://www.genomesize.com">http://www.genomesize.com</a>   | 0.0313                            | 33 461                | 240.9                    | yes               | yes           |

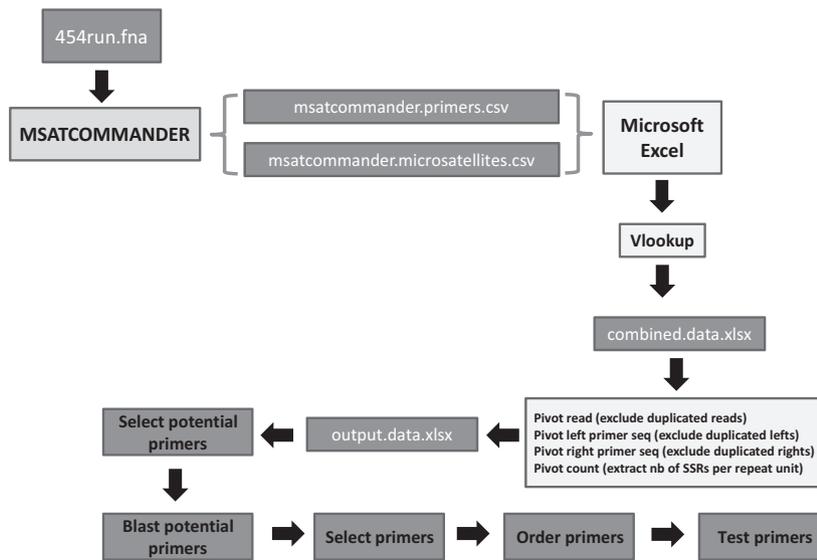
\*Mean of several closely related species

\$calculated after Doležel *et al.*, 2003#For *Pyrrhura orcesi*, the run was repeated by the commercial supplier due to technical issues; hence the data set is twice the regular size.

motifs with sufficient flanking sequence to allow primer development would be identified, regarding the restricted read length obtained from a 454 pyrosequencing run. We searched for microsatellites  $\geq 5$  repeats, as we were interested to detect the relationship of microsatellite repeat length (di-, tri- and tetranucleotides) and repeat number. Further, we chose the option 'design primers', in which case the software searches for microsatellite repeats and identifies possible primer annealing sites in a single step. We did not use any primer tagging option. Primers were developed with PRIMER3 (Rozen & Skaletsky, 2000) implemented in MSATCOMMANDER using default parameters: amplification of products within a size range of 100–500 bp, optimal melting temperature of 60.0 °C (range 57.0–62.0 °C), optimal GC content of 50%, possession of at least one 1 bp GC clamp, low levels of self- or pair-complementarity and maximum end-stability ( $\Delta G$ ) of 8.0 (Faircloth, 2008; <http://frodo.wi.mit.edu/primer3/input.htm>).

For each search, MSATCOMMANDER generated two output files. One contained all reads with microsatellite motifs of five or more repeats and the other file listed

all reads for which PRIMER3 could successfully develop potential up- and downstream primers in the flanking region. We therefore received six output files per species, two for each di-, tri- and tetranucleotide motif respectively. The two output files obtained from each search for each repeat length were then combined in one MICROSOFT EXCEL (v. 14.2.2) spread sheet and all reads containing primers indicated as 'potentially duplicated' were excluded from further analysis. Within each of the data sets, we then searched for duplicated reads as well as up- and downstream primer sequences occurring more than once and excluded all respective reads from further analyses. A schematic overview of the workflow is depicted in Fig. 1. In addition, an R script to facilitate this analysis is available in the supplementary online materials. We are aware that, with this conservative approach, we might have excluded reads containing potentially usable microsatellites. Nonetheless, we explicitly decided to use this approach to demonstrate the simplicity and efficiency in developing potentially high quality microsatellite markers. On the other hand, we emphasize that in case of too low



**Figure 1** Workflow used for primer development. Dark grey boxes with white writing represent files, light grey boxes indicate programs and functions while dark grey boxes with black writing indicate work steps. An R script to facilitate this analysis can be found in the supplementary online materials as well as on dryad.

numbers of suitable loci for the respective study, one may come back to this step at a later point and check reads in detail for additional suitable microsatellite sequences or apply less stringent primer quality standards.

Subsequently, we calculated and compared the ratios of microsatellites for which primer development was possible with the number of reads and performed a comparison between organismic groups and sequencing sample types. Further, we grouped the pure species data sets without a prior repeat enrichment step applied in the following three broad organismic groups: ‘fungi’ (combining true fungi and Oomycetes), ‘plants’ and ‘birds’. In addition, we recorded the number of repeats per study species and compared mean values of potential di-, tri- and tetranucleotide microsatellites between the three taxonomic groups for both  $\geq 5$  and  $\geq 8$  repeats. Most studies advise to use only microsatellites with more than eight repeats, even prefer microsatellites with 11–16 repeats (van Asch *et al.*, 2010). However, since one of our study groups is fungi, where long repeats are generally more difficult to detect, we show comparisons both for  $\geq 5$  repeats and  $\geq 8$  repeats. For the comparisons among species, we standardized the number of reads to 1/16 of a full run for all data sets.

For species with primers already fully tested, we additionally indicate the number of primer pairs independently tested in the respective labs, the number of primer pairs that successfully amplified, the number of polymorphic loci and the number of loci finally chosen for analysis (Table 2). Repeat motif lengths of markers in use were then compared. As testing approaches varied between labs, a respective comparison needs to be handled with care. With the exception of the *B. alpinus* data set, which was analysed by Ecogenics (Switzerland), and the *E. quercinus* data set, which was analysed by Geno-

screen (France), primers for all other species were developed in the respective labs. Primers for *A. alpina* (Buehler *et al.*, 2011), *F. sultana* (Hartikainen & Jokela, 2012), *L. pulmonaria* (Werth *et al.*, in press), *S. grossum* (Keller *et al.*, 2012) and *T. minima* (Csencsics *et al.*, 2010) are published. In the case of *F. sultana*, a two-step approach to develop microsatellites originating from the target organism was used, as DNA was not extracted from a single species but from several bryozoan colonies together with their gut content (Hartikainen & Jokela, 2012).

### Statistical analyses

To test for differences in numbers of detected microsatellites, two univariate analyses of variance (ANOVAS) were performed, each for potential microsatellites with  $\geq 5$  repeats and  $\geq 8$  repeats. In these analyses, the dependent variable was the number of potential microsatellites, standardized for 1/16 run and  $\ln(x + 2)$ -transformed. Independent variables were organismic group (fungi, plants and birds) and microsatellite motif (di-, tri-, or tetranucleotides). A Spearman’s rank correlation was used to test for the effect of genome size on the number of potential microsatellites for each microsatellite motif separately. We ran separate analyses for potential microsatellites with  $\geq 5$  repeats and  $\geq 8$  repeats. The two repeat enriched data sets and the *F. sultana* data set (mixture of bryozoan colonies and gut contents) were excluded from these analyses. Finally, Spearman’s rank correlations were applied to test for the effect of genome size on the proportion of polymorphic microsatellites, calculated as the percentage of initially amplified loci, and amplified microsatellites, calculated as the percentage of the loci initially tested. For these last two correlation analyses, we only

**Table 2** Overview of the change in number of microsatellite markers tested, amplified (as percentage of the number microsatellites initially tested), polymorphic (as percentage of the number amplified) for each species. The total number of reads is given according to the original proportion of run sequenced. Testing of the microsatellites for *Phytophthora gonapodyides* and *P. citricola* s.s. is not yet finished and therefore some numbers are not available (na).

| Broad grouping | Species                            | Total number of reads | Number of microsatellites tested | Number of microsatellites amplified (% of tested) | Number of microsatellites polymorphic (% of amplified) |
|----------------|------------------------------------|-----------------------|----------------------------------|---|--|
| Fungi          | <i>Lobaria pulmonaria</i>          | 181 541               | 16                               | 14 (87.5%)  | 14 (100%)  |
|                | <i>Phytophthora gonapodyides</i>   | 14 857                | 24                               | na  | na   |
|                | <i>Phytophthora multivora</i>      | 20 716                | 24                               | 24 (100%)   | 18 (75%)   |
|                | <i>Phytophthora plurivora</i>      | 16 162                | 36                               | 35 (97.2%)  | 13 (37.1%)   |
|                | <i>Phytophthora citricola</i> s.s. | 15 009                | na                               | na  | na   |
|                | <i>Phytophthora pini</i>           | 18 465                | 13                               | 12 (92.3%)  | 8 (66.7%)  |
| Plants         | <i>Arabis alpina</i>               | 40 367                | 34                               | 28 (82.4%)  | 25 (89.3%)   |
|                | <i>Typha minima</i>                | 76 692                | 30                               | 30 (100%)   | 26 (86.7%)   |
|                | <i>Pinus cembra</i>                | 37 496                | 37                               | 31 (83.8%)  | 11 (35.5%)   |
|                | <i>Taxus wallichiana</i>           | 119 508               | 81                               | 48 (59.3%)  | 12 (25%)   |
| Invertebrates  | <i>Baetis alpinus</i>              | 7075                  | 12                               | 10 (83.3%)  | 5 (50%)  |
|                | <i>Stethophyma grossum</i>         | 52 693                | 50                               | 45 (90%)  | 10 (22.2%)   |
|                | <i>Fredericella sultana</i>        | 196 025               | 30                               | 20 (66.7%)  | 16 (80%)   |
| Vertebrates    | <i>Lagopus muta</i>                | 33 362                | 22                               | 19 (86%)  | 9 (47%)  |
|                | <i>Pyrrhura orcesi</i>             | 80 093                | 16                               | 7 (44%)   | na   |
|                | <i>Sylvia atricapilla</i>          | 43 939                | 51                               | 40 (78%)  | 23 (58%)   |
|                | <i>Eliomys quercinus</i>           | 33 461                | 12                               | 11 (91%)  | 7 (87.5%)  |

used the 14 species where the microsatellite markers were actually tested in the lab. All analyses were carried out using IBM SPSS (Version 20.0. Armonk, NY: IBM Corp.).

## Results

### Overall 454 pyrosequencing results

As run sizes in this study ranged from 1/32 (*E. quercinus*) up to 1/4 (*L. pulmonaria*), including five pooled species in 2/16 of a full run (*Phytophthoras* Table 1), the number of reads per species varied considerably. Average read lengths ranged from 140 bp in *B. alpinus* to 388 bp in *F. sultana*, while the mean read length was 340 bp  $\pm$  41.2 SE. On average, there were 41 486  $\pm$  5110 SE reads per 1/16 run (40 925  $\pm$  12 015 SE including the two repeat enriched data sets).

### Microsatellite detection

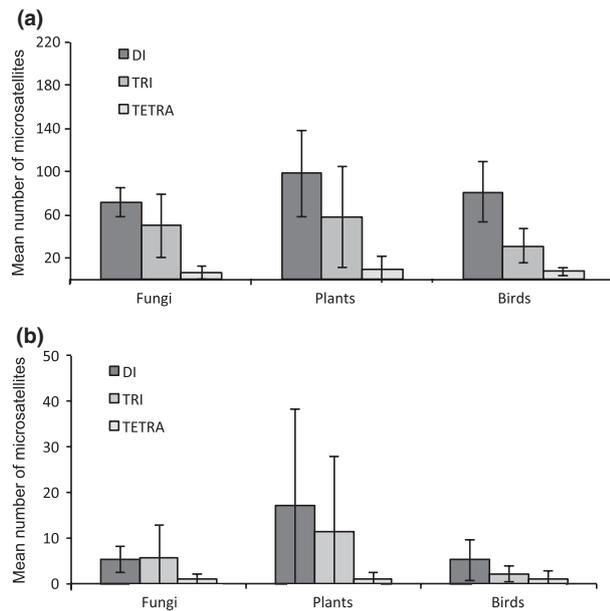
Detection of reads containing microsatellite loci including flanking regions suitable for primer development was feasible for all 17 analysed data sets. For the data sets standardized to 1/16 run, the numbers of potential dinucleotide microsatellite loci with  $\geq 5$  repeats ranged from 50 (*L. muta*) to 163 (*S. grossum*) and from 0 (*P. pini* and *L. muta*) to 48.5 (*T. minima*) for  $\geq 8$  repeats (Table S2A, Fig. 3A). For trinucleotide loci, numbers ranged from 17 (*L. muta*) to 163 (*S. grossum*) for  $\geq 5$  repeats

and from 0 (*P. gonapodyides*, *P. citricola* s.s. and *P. orcesi*) to 36 (*T. minima*) for  $\geq 8$  repeats (Table S2B, Fig. 3B). For tetranucleotide motifs, numbers ranged from 0 (*P. gonapodyides*, *P. citricola* s.s.) to 29 (*T. minima*) for  $\geq 5$  repeats and from 0 (*P. gonapodyides*, *P. citricola* s.s., *P. pini*, *A. alpina*, *L. muta*, *P. orcesi*, *T. wallichiana* and *P. cembra*) to 3.5 (*T. minima*) for  $\geq 8$  repeats (Table S2C, Fig. 3C). Sequencing data of *T. minima* and *S. grossum* yielded exceptionally high numbers of microsatellite motifs compared to the other species (Table S2).

### Differences in microsatellite repeat motif and length among organismic groups

From now on we will only refer to microsatellite loci for which primers could be developed according to the standard requirements of PRIMER3 as implemented in MSATCOMMANDER.

The two ANOVAS revealed a significant effect of microsatellite motif on the number of potential microsatellites with  $\geq 5$  repeats ( $F_{2,38} = 46.62$ ,  $P < 0.001$ ) as well as for microsatellites with  $\geq 8$  repeats ( $F_{2,38} = 6.17$ ,  $P = 0.006$ ). In potential microsatellites with  $\geq 5$  repeats, dinucleotide motifs were the most frequent, followed by tri- and tetranucleotides (Fig. 2). In potential microsatellites with  $\geq 8$  repeats, the pattern was less clear, but tetranucleotides were always the least frequent in all three organismic groups. Neither the organismic group ('fungi', 'plants' and 'birds') nor the interaction term (organismic group  $\times$  microsatellite



**Figure 2** Mean number of di-, tri- and tetranucleotide motifs with potential primers to be designed (according to MSATCOMMANDER default criteria) converted to 1/16 of a full 454 run, depicted for each taxonomic group: fungi (5 species), plants (4 species) and bird (3 species). (a) Including microsatellites with  $\geq 5$  repeats and (b) including microsatellites  $\geq 8$  repeats. Error bars =  $\pm 1$  SE.

motif) significantly influenced the number of potential microsatellites in both analyses.

### Microsatellite repeat length and relation to genome size

There was no significant effect of genome size on the number of potential microsatellites for each microsatellite motif and potential microsatellites with  $\geq 5$  repeats, and  $\geq 8$  repeats for all analysed species (Fig. 3). Moreover, the last two Spearman's rank correlation tests showed a significant effect of genome size on the final proportion of polymorphic microsatellites ( $n = 14$ ,  $\rho = -0.54$ ,  $P = 0.048$ ), but not on the proportion of amplified microsatellites ( $n = 14$ ,  $\rho = -0.46$ ,  $P = 0.098$ ). In other words, for species with smaller genomes, more polymorphic microsatellites could be detected. However, this result was mainly influenced by the low abundance of potential microsatellites in the three species with the by far largest genomes, *S. grossum*, *P. cembra* and *T. wallichiana* (Table 2). The number of microsatellites amplified as the proportion of all microsatellites initially tested ranged from 59.3% (*T. wallichiana*) to 100% (*P. multivora* and *T. minima*). The number of microsatellites showing polymorphic alleles, as a proportion of the number of microsatellites that had amplified, ranged from 22.2% (*S. grossum*) to 100% (*L. pulmonaria*; Table 2).

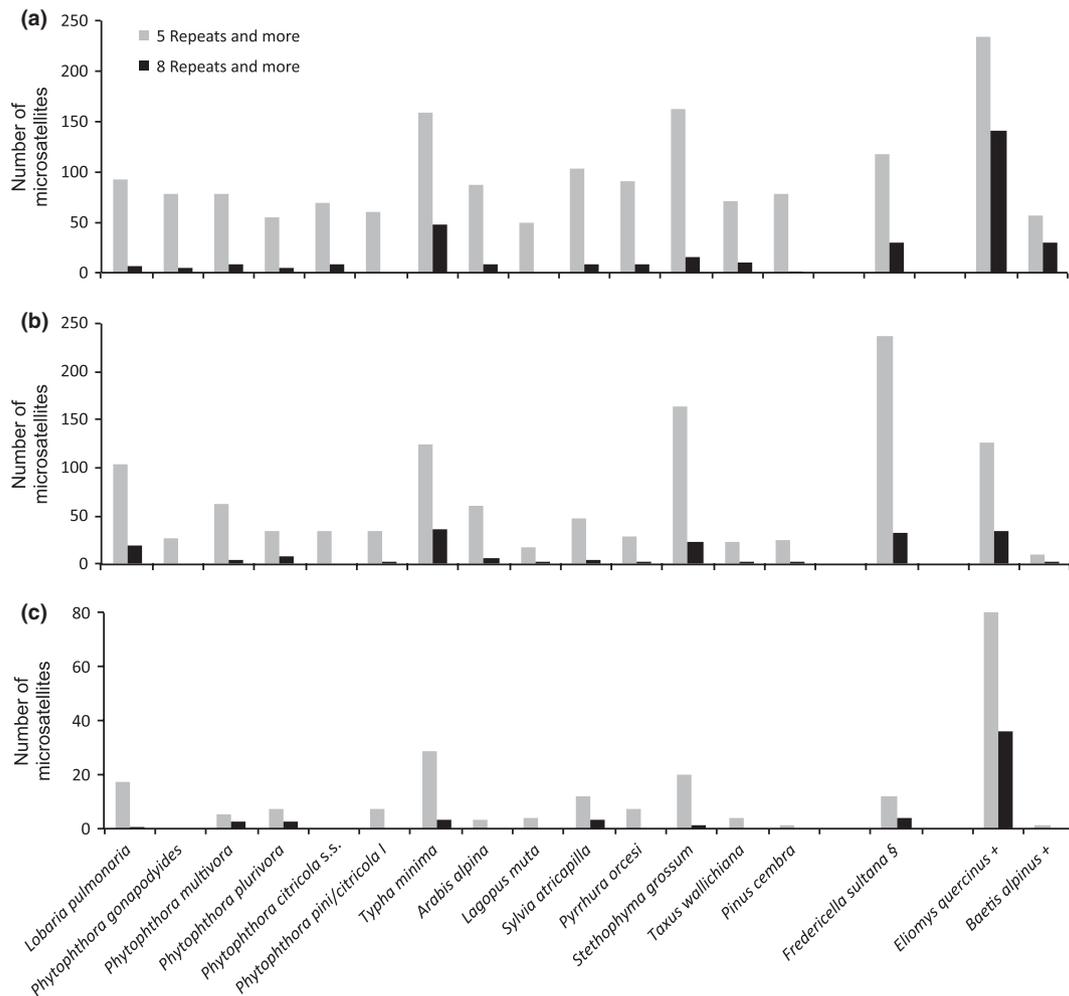
## Discussion

By employing a user friendly approach to analyse 454 pyrosequencing data sets with commonly used, free software for microsatellite marker identification, we successfully in silico developed sufficient numbers of microsatellite markers for population genetic studies in all 17 investigated species. In comparison with the numbers of lab-tested microsatellite markers, we showed that a high percentage of tested microsatellite markers were actually amplifiable and polymorphic. These results highlight that the development of microsatellites from 454 pyrosequencing data is generally straightforward and easy. However, there were diverse challenges encountered in particular species.

### Fungi – small genomes and low repeat numbers

Fungi are generally thought to be difficult in microsatellite marker development due to their small genomes (approximate c-values: 0.03 to 0.082 pg in the species studied here) and the rather short lengths of microsatellites detected with traditional marker development (Dutech *et al.*, 2007). With the advent of Next Generation Sequencing technology, the identification and characterization of microsatellites became more feasible also in fungi and Oomycetes, and microsatellite markers can be developed without a prior repeat enrichment step. Nonetheless, comparing fungal and nonfungal microsatellite markers developed using 454 pyrosequencing, our results are in accordance with prior evidence from traditional marker development showing fungal microsatellites tending to be rather short (mostly  $< 8$  repeats; Fig. 3; Dutech *et al.*, 2007). Nevertheless, for *L. pulmonaria*, we found 14 microsatellites with eleven or more repeats which showed substantial polymorphism. Hence, it might generally not be good advice to follow recommendations for microsatellite development using Next Generation Sequencing when working with fungi. Most previous studies advise to only use perfect microsatellites containing  $\geq 8$  repeats (Guichoux *et al.*, 2011). However, depending on the fungal species, only few reads containing microsatellites with  $\geq 8$  repeats may be available in the Next Generation Sequencing data set. For instance, for several *Phytophthora* species we found most microsatellites to be rather short (Table S2). Nonetheless, for *P. plurivora* and *P. multivora*, both microsatellites with  $\geq 8$  (up to 16 repeats) as well as  $< 8$  repeats were developed and found to be highly polymorphic.

Previously encountered problems concerning the small genome of fungi and the rather low polymorphism in developed microsatellite loci (Dutech *et al.*, 2007) do not seem to be a big problem in studies where microsatellites were developed using pyrosequencing. Also, the average number of potential microsatellite markers was not lower than in other organismic groups (Fig. 1). We



**Figure 3** Sums of (a) di-, (b) tri- and (c) tetranucleotide microsatellites with potential primers to be designed for each study species, converted to 1/16 of a full 454 run, depicted with increasing genome size except for repeat enriched and mixed samples (left to right, for genome sizes see Table 1). Grey bars show microsatellite motifs with  $\geq 5$  repeats, black bars with  $\geq 8$  repeats. +: repeat enriched samples; §: mixed sample.

agree with previous literature (Dutech *et al.*, 2007) that developing loci with more than eight repeats for fungi is a challenging task, even if Next Generation Sequencing technology is employed. However, also loci with fewer repeats showed the required level of polymorphism.

### Large genome size

In species with large genomes, the frequency and length of microsatellites is generally higher than in species with smaller genomes (Tóth *et al.*, 2000; Hancock, 2002). Although the proportion of the genome investigated when ordering 1/16 of a full 454 pyrosequencing run is much lower than in species with smaller genomes, we did not detect a difference in microsatellite length or frequency for species with larger genomes. Nevertheless, it has been suggested that amplification

success of microsatellite primers is negatively related to genome size (Garner, 2002). For instance in *S. grossum*, which has a comparatively large genome, a large amount of potential primers had to be discarded because they produced multiple PCR products. This has already been reported in previous studies using traditional microsatellite development (e.g. Van De Vliet *et al.*, 2009). The large amount of repetitive motifs throughout the genome, typically occurring in large genomes (Hancock, 2002) might cause nonspecific binding of primers. To identify such repetitive primer sequences, Primer3 (Rozen & Skaletsky, 2000) offers the option to screen primers for repetitive sequences by comparison with repetitive element banks. However, for most nonmodel organisms, there is no repetitive element bank of a closely related species available. Primer development for conifer species with large

genomes (i.e. *P. cembra*, *T. wallichiana*) has shown to be especially challenging (Parchman *et al.*, 2010). Even though we found plenty of microsatellite motifs within the genome, there were only few with flanking regions that were suitable for optimal primer development. The examples analysed here indicate that microsatellite development for species with large genomes (especially for conifers) is still more cost- and time-consuming than for other species, as more primers need to be tested to close with a reasonably high number of polymorphic and reliable markers. Accordingly, it might be necessary to order more than 1/16 plate of a full 454 pyrosequencing run to come by with enough microsatellite markers or to change strategy and focus on transcriptome sequencing instead, which sometimes is more appropriate than whole genome Next Generation Sequencing for species with large genomes (Parchman *et al.*, 2010; Guichoux *et al.*, 2011).

### Mixed culture

A particular challenge is the use of nonpure genomic DNA, for example, in species with intracellular parasites or when pure DNA is difficult to obtain in large enough quantities for 454 pyrosequencing. This possibly results in 454 pyrosequencing reads belonging to multiple species. Similarly, separating the gut contents from target organism tissue may be complicated in small invertebrates. Likewise, the presence of abundant epibionts or symbionts complicates microsatellite development. In the case of the bryozoan *F. sultana*, a two-step approach to identify microsatellites originating from the target organism was applied (Hartikainen & Jokela, 2012). Here, to achieve sufficient quantities of high quality genomic DNA for 454 pyrosequencing, several whole colonies were extracted, including guts with potentially contaminating microorganisms. First, loci that did not amplify consistently across small, single colony specimens collected from diverse sites were not selected for further testing. It is unlikely that the same epibionts and food items would have been present in all sites, allowing for coarse screening against contaminant loci. Second, the final set of loci was verified to be of target origin by amplifying from pure DNA, extracted from nonfeeding resting stages (statoblasts). Obtaining small quantities of pure DNA, sufficient for PCR amplification of microsatellite loci, may be possible by careful dissection or selection of nonfeeding larvae/eggs or other noncontaminated stages. Using such a two-step approach is largely enabled by the large number of potential loci identified by 454 pyrosequencing and would have been difficult to achieve using traditional microsatellite marker development.

In other cases, blast searches of obtained 454 pyrosequencing reads might be another option to assign reads to different species, although the presence of long repeat regions within the sequence reduces the number

of informative bases, and for many nonmodel organisms, sequence data availability in public databases is limited. To minimize problems from mixtures of target and nontarget organisms, careful selection of tissue to be sequenced, followed by verification steps such as described above, are recommended.

### Direct versus enriched 454 pyrosequencing

Lepais & Bacles (2011) compared direct and repeat enriched 454 pyrosequencing for *Acacia harpophylla* (approximate genome size of acacias: 1.20–2.13 pg/2C; Gallagher *et al.*, 2011). In the repeat enriched data set, they found 2.2% of the reads yielding microsatellite markers with developable primers, while in the data set obtained without prior repeat enrichment it was only 0.5% of the reads. Such a substantial difference between the two approaches could not be detected in this study. Comparing the results of 15 direct 454 pyrosequencing runs with two repeat enriched 454 pyrosequencing data sets, we found no obvious differences concerning the final number of polymorphic microsatellites developed (Fig. 3). However, this is most probably due to the fact that we compared different species and organismic groups while Lepais & Bacles (2011) compared the two methods within one species.

A drawback of repeat enrichment is the need for a priori choice of repeat motifs and thus the biased sample of the genome, which hinders the use of such data sets for purposes other than microsatellite identification. For future microsatellite development in species with large genomes knowledge of whether repeat enrichment optimizes microsatellite mining in these species is needed.

### Pooling different species in one run

We could not detect significant disadvantages of pooling several individually tagged species within the same run compared with single species runs (Figs 2 and 3). However, we only analysed one such data set and the overall read numbers were slightly lower than average when standardized to 1/16 of a full run, but this might be a characteristic of Oomycetes. Nonetheless, all developed microsatellites were highly polymorphic. Jennings *et al.* (2011) and Takayama *et al.* (2011) also successfully developed microsatellite markers for different species pooled in one run. Pooling samples can be an ideal option if the budget is limited and no further analysis other than microsatellite development shall be conducted. Especially for small labs, pooling species and sharing the cost for a 454 pyrosequencing run together with affiliated labs, followed by in house primer development could be a promising option to receive their own microsatellites tailored for their study species. Anyways, some commercial suppliers tend to barcode all libraries and load them on one plate without subdi-

vision, as for example, 1/16 of a plate is tedious to fill. Therefore, a careful evaluation of the specific methods used by different suppliers is recommended.

## Conclusions

The recent rise of Next Generation Sequencing and especially the 454 pyrosequencing technology has made it feasible to identify high quality microsatellite markers for any nonmodel species without any prior knowledge of the species' genome, including species of ecological, evolutionary or conservation concern.

On the basis of the results of this study, we recommend running 1/16 of a full plate per species. However, if the approximate genome size of the study species is known to be rather large, it is advisable to order a higher proportion of a full 454 pyrosequencing run (i.e. 2/16 or 1/4). A repeat enrichment step might be considered if the taxon under study is known to be problematic for microsatellite marker development. Furthermore, pooling species is a valid option, which can be used to reduce costs. In case of repeatedly analysing 454 pyrosequencing data in the same manner, users might also write a R, Perl or Python script in order to create a robust pipeline for data analysis instead of using MS Excel only. In summary, the development of species-specific, polymorphic microsatellite markers for nonmodel species is no longer a time and resource consuming bottleneck.

## Acknowledgments

We are thankful to Christian Rellstab for helpful discussions and input for data analysis as well as to Felix Gugerli, Rolf Holderegger, Simone Prospero and two anonymous referees for critical and helpful comments on the manuscript. CNS was supported by the European Cooperation in Science and Technology (COST Switzerland; COST Action FP0801). We also acknowledge financial support from the Swiss National Science Foundation (grant JRP IZ70Z0\_131338/1 to JPG, grant 3100AO-105830 to SW and grant 31003A\_1276346/1 to CC).

## References

- Abdelkrim, J., Robertson, B.C., Stanton, J.-A.L. & Gemmill, N.J. 2009. Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. *Biotechniques* **46**: 185–192.
- Ansell, S.W., Grundmann, M., Russell, S.J., Schneider, H. & Vogel, J.C. 2008. Genetic discontinuity, breeding-system change and population history of *Arabis alpina* in the Italian Peninsula and adjacent Alps. *Mol. Ecol.* **17**: 2245–2257.
- Armaleo, D. & May, S. 2009. Sizing the fungal and algal genomes of the lichen *Cladonia grayi* through quantitative PCR. *Symbiosis* **49**: 43–51.
- van Asch, B., Pinheiro, R., Pereira, R., Alves, C., Pereira, V., Pereira, F. *et al.* 2010. A framework for the development of STR genotyping in domestic animal species: characterization and population study of 12 canine X-chromosome loci. *Electrophoresis* **31**: 303–308.
- Barbara, T., Palma-Silva, C., Paggi, G.M., Bered, F., Fay, M.F. & Lexer, C. 2007. Cross-species transfer of nuclear microsatellite markers: potential and limitations. *Mol. Ecol.* **16**: 3759–3767.
- Bertolino, S., Amori, G., Henttonen, H., Zagorodnyuk, I., Zima, J., Juškatis, R. *et al.* 2008. 'Eliomys quercinus'. *IUCN Red list of Threatened Species*. Version 3.1. Available at: www.iucnredlist.org.
- Buehler, D., Graf, R., Holderegger, R. & Gugerli, F. 2011. Using the 454 pyrosequencing-based technique in the development of nuclear microsatellite loci in the alpine plant *Arabis alpina* (Brassicaceae). *Am. J. Bot.* **98**(5): e103–e105.
- Casalegno, S., Amatulli, G., Camia, A., Nelson, A. & Pekkarienen, A. 2010. Vulnerability of *Pinus cembra* L. in the Alps and the Carpathian mountains under present and future climates. *Forest Ecol. Manag.* **259**: 750–761.
- Castoe, T.A., Poole, A.W., Gu, W., de Koning, J., Daza, J.M., Smith, E.N. *et al.* 2010. Rapid identification of thousands of copperhead snake (*Agkistrodon contortrix*) microsatellite loci from modest amounts of 454 shotgun genome sequence. *Mol. Ecol. Resour.* **10**: 341–347.
- Csencsics, D., Brodbeck, S. & Holderegger, R. 2010. Cost-effective, species-specific microsatellite development for the endangered dwarf bulrush (*Typha minima*) using next-generation sequencing technology. *J. Hered.* **101**: 789–793.
- Doležel, J., Bartoš, J., Voglmayr, H. & Greilhuber, J. 2003. Nuclear DNA content and genome size of trout and human. *Cytometry* **51A**: 127–128.
- Dolgener, N., Schroeder, C., Schneeweiss, N. & Tiedemann, R. 2012. Genetic population structure of the Fire-bellied toad *Bombina orientalis* in an area of high population density: implications for conservation. *Hydrobiologia* **689**: 111–120.
- Dutech, C., Enjalbert, J., Fourneier, E., Delmotte, F., Barrès, B., Carlier, J. *et al.* 2007. Challenges of microsatellite isolation in fungi. *Fungal Genet. Biol.* **44**: 933–949.
- Erwin, D.C. & Ribeiro, O.K. 1996. *Phytophthora Diseases Worldwide*. APS Press, St. Paul, MN.
- Faircloth, B.C. 2008. MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Mol. Ecol. Resour.* **8**: 92–94.
- Farjon, A. 1998. *World Checklist and Bibliography of Conifers*. Royal Botanic Gardens, Kew.
- Gallagher, R.V., Leishman, M.R., Miller, J.T., Hui, C., Richardson, D.M., Suda, J. *et al.* 2011. Invasiveness in introduced Australian acacias: the role of species traits and genome size. *Divers. Distr.* **17**: 884–897.
- Gardner, M.G., Fitch, A.J., Bertozzi, T. & Lowe, A.J. 2011. Rise of the machines - recommendations for ecologists when using next generation sequencing for microsatellite development. *Mol. Ecol. Resour.* **11**: 1093–1101.
- Garner, W.J. 2002. Genome size and microsatellites: the effect of nuclear size on amplification potential. *Genome* **45**: 212–215.
- Goldstein, D.B. & Schlotterer, C. 1999. *Microsatellites: evolution and applications*. Oxford University Press, Oxford.
- Greilhuber, J. 1986. Severely distorted Feulgen-DNA amounts in *Pinus* (Coniferophytina) after nonadditive fixations as a

- result of meristematic self-tanning with vacuole contents. *Can. J. Genet. Cytol.* **28**: 409–415.
- Greilhuber, J., Borsch, T., Müller, K., Worberg, A., Porembski, S. & Barthlott, W. 2006. Smallest angiosperm genomes found in Lentibulariaceae, with chromosomes of bacterial size. *Plant Biol.* **8**: 770–777.
- Grover, A., Aishwarya, V. & Sharma, P.C. 2012. Searching microsatellites in DNA sequences: approaches used and tools developed. *Physiol. Mol. Biol. Plants* **18**: 11–19.
- Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Leger, P., Lepais, O. *et al.* 2011. Current trends in microsatellite genotyping. *Mol. Ecol. Resour.* **11**: 591–611.
- Hallingbäck, T. & Martinsson, P.O. 1987. The retreat of two lichens, *Lobaria pulmonaria* and *L. scrobiculata* in the district of Gästene (SW Sweden). *Windahlia* **17**: 27–32.
- Hancock, J.M. 2002. Genome size and the accumulation of simple sequence repeats: implications of new data from genome sequencing projects. *Genetica* **115**: 93–103.
- Hartikainen, H. & Jokela, J. 2012. Characterisation of polymorphic microsatellite loci in the freshwater bryozoan *Fredericella sultana*. *Conserv. Genet. Resour.* **4**: 475–477.
- Horáček, I. 1986. Fossil records and chorological status of dormice in Czechoslovakia, Part I: *Glis glis*, *Eliomys quercinus*. *Folia Musei Rerum Naturalium Bohemiae Occidentalis, Plzen, Zoologica* **24**: 49–59.
- Humpesch, U.H. 1979. Life-cycles and growth-rates of *Baetis* spp (Ephemeroptera, Baetidae) in the laboratory and in 2 stony streams in Austria. *Freshwater Biol.* **9**: 467–479.
- Hurtado, L.A., Santamaria, C.A. & Fitzgerald, L.A. 2012. Conservation genetics of the critically endangered Saint Croix ground lizard (*Ameiva polops* Cope 1863). *Conserv. Genet.* **13**: 665–679.
- Jennings, T.N., Knaus, B.J., Mullins, T.D., Haig, S.M. & Cronn, R.C. 2011. Multiplexed microsatellite recovery using massively parallel sequencing. *Mol. Ecol. Resour.* **11**: 1060–1067.
- Jung, T. & Burgess, T.I. 2009. Re-evaluation of *Phytophthora citricola* isolates from multiple woody hosts in Europe and North America reveals a new species, *Phytophthora plurivora* sp. nov. *Persoonia* **22**: 95–110.
- Keller, D., Jung, E. & Holderegger, R. 2012. Development of microsatellite markers for the wetland grasshopper *Stethophyma grossum*. *Conserv. Genet. Resour.* **4**: 507–509.
- Kopatz, A., Geir Eiken, H., Hagen, S.B., Ruokonen, M., Esparza-Salas, R., Schregel, J. *et al.* 2012. Connectivity and population subdivision at the fringe of a large brown bear (*Ursus arctos*) population in North Western Europe. *Conserv. Genet.* **13**: 681–692.
- Kroon, L.P.N.M., Brouwer, H., de Cock, A.W.A.M. & Govers, F. 2011. The Genus *Phytophthora* Anno 2012. *Phytopathology* **102**: 348–364.
- Leitch, I.J., Hanson, L., Winfield, M., Parker, J. & Bennett, M. D. 2001. Nuclear DNA C-values complete familial representation in gymnosperms. *Ann. Bot.* **88**: 843–849.
- Lepais, O. & Bacles, C.F.E. 2011. Comparison of random and microsatellite-enriched shotgun pyrosequencing for microsatellite discovery and single multiplex PCR optimization in *Acacia harpophylla* F. Muell. *Ex Benth. Mol. Ecol. Resour.* **11**: 711–724.
- Malausa, T., Gilles, A., Meglécz, E., Blanquart, H., Duthoy, S., Costedoat, C. *et al.* 2011. High-throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries. *Mol. Ecol. Resour.* **11**: 638–644.
- Manel, S., Poncet, B.N., Legendre, P., Gugerli, F. & Holderegger, R. 2010. Common factors drive adaptive genetic variation at different spatial scales in *Arabidopsis thaliana*. *Mol. Ecol.* **19**: 3824–3835.
- Manoel, R.O., Alves, P.F., Dourado, C.L., Gaino, A.P.S.C., Freitas, M.L.M., Moraes, M.L.T. *et al.* 2012. Contemporary pollen flow, mating patterns and effective population size inferred from paternity analysis in a small fragmented population of the Neotropical tree *Copaifera langsdorffii* Desf. (Leguminosae-Caesalpinioideae). *Conserv. Genet.* **13**: 613–623.
- Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J. S., Bemben, L.A. *et al.* 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**: 376–380.
- Mitchell-Jones, A.J., Amori, G., Bogdanowicz, W., Kryštufek, B., Reijnders, P.J.H., Spitzenberger, F. *et al.* 1999. *The atlas of European mammals*. Academic Press, London.
- Monnerat, C., Thorens, P., Walter, T. & Gonseth, Y. 2007. *Rote Liste der gefährdeten Arten der Schweiz: Heuschrecken*. BAFU, Bern, Switzerland.
- Müller, N. 1991. Verbreitung, Vergesellschaftung und Rückgang des Zwergrohrkolbens (*Typha minima* Hoppe). *Hoppea* **50**: 19.
- Okamura, B., Hartikainen, H., Schmidt-Posthaus, H. & Wahli, T. 2011. Life cycle complexity, environmental change and the emerging status of salmonid proliferative kidney disease. *Freshwater Biol.* **56**: 735–753.
- Oxford, E. 1996. The great famine – when a blight struck the Irish potato crop in the 1840s, it precipitated a famine that caused more than a million deaths and the exodus of that many more to America. *Am. Hist.* **31**: 52–61.
- Parchman, T., Geist, K., Grahnen, J., Benkman, C. & Buerkle, C.A. 2010. Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. *BMC Genomics* **11**: 180.
- Prunier, P., Garraud, L., Köhler, C., Lambelet-Haueter, C., Selvaggi, A. & Werner, P. 2010. Distribution et régression de la petite massette (*Typha minima*) dans les Alpes. *Bot. Helv.* **120**: 43–52.
- Reid, K., Hoareau, T. B. & Bloomer, P. 2012. High-throughput microsatellite marker development in two sparid species and verification of their transferability in the family Sparidae. *Mol. Ecol. Resour.* **12**: 740–752.
- Rolshausen, G., Segelbacher, G., Hobson, K.A. & Schaefer, H. M. 2009. Contemporary evolution of reproductive isolation and phenotypic divergence in sympatry along a migratory divide. *Curr. Biol.* **19**: 2097–101.
- Rozen, S. & Skaletsky, H.J. 2000. Primer3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (S. Kraetzet & S. Misener, eds), pp. 365–386. Humana Press, Totowa.
- Rushforth, K.D. 1987. *Conifers*. Helm, London.
- Santana, Q.C., Coetzee, M.P.A., Steenkamp, E.T., Mlonyeni, O.X., Hammond, G.N.A., Wingfield, M.J. *et al.* 2009. Microsatellite discovery by deep sequencing of enriched genomic libraries. *Biotechniques* **46**: 217–223.
- Sartori, M. & Landolt, P. 1999. *Atlas de distribution des éphémères de Suisse (Insecta, Ephemeroptera)*. Centre Suisse de Cartographie de la Faune, Neuchâtel, Switzerland.
- Schopf, T.J.M. 1983. A genomic library and the genome size for the cheilostome bryozoan *Watersipora*. In: *Bryozoa*:

- Ordovician to recent* (C. Nielsen & G.P. Larwood, eds), pp. 285–292. Olsen & Olsen, Fredensborg.
- Schultze-Motel, W. 1986. *Arabis alpina*. In: *Illustrierte Flora von Mitteleuropa, Band 4 Teil 1* (H.J. Conert, U. Hamann, W. Schultze-Motel & G. Wagenitz, eds), pp. 248–251. Parey, Berlin.
- Selkoe, K.A. & Toonen, R.J. 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecol. Lett.* **9**: 615–629.
- Takayama, K., Lopez, P., Konig, C., Kohl, G., Novak, J. & Stuessy, T.F. 2011. A simple and cost-effective approach for microsatellite isolation in non-model plant species using small-scale 454 pyrosequencing. *Taxon* **60**: 144: 2–1449.
- Tóth, G., Gaspari, Z. & Jurka, J. 2000. Microsatellites in different eukaryotic genomes: Survey and analysis. *Genome Res.* **10**: 967–981.
- Tyler, B., Tripathy, S., Grunwald, N., Lamour, K., Ivors, K., Garbelotto, M. *et al.* 2006. Genome sequence of *Phytophthora ramorum*: implications for management. In: *Proceedings of the sudden oak death second science symposium: the state of our knowledge* (S.J. Frankel, P.J. Shea & M.I. Haverty, tech. coords, eds), pp. 157–158. Gen. Tech. Rep. PSW-GTR-196. Pacific Southwest Research Station, Forest Service, U.S. Department of Agriculture, Albany, CA.
- Van De Vliet, M.S., Diekmann, O.E., Serrao, E.A. & Beja, P. 2009. Isolation of highly polymorphic microsatellite loci for a species with a large genome size: sharp-ribbed salamander (*Pleurodeles waltli*). *Mol. Ecol. Resour.* **9**: 425–428.
- Vance, S. 1996. Morphological and behavioural sex reversal in mermithid-infected mayflies. *Proc. R. Soc. B* **263**: 907–912.
- Wang, R., Farrona, S., Vincent, C., Joecker, A., Schoof, H., Turck, F. *et al.* 2009. PEP1 regulates perennial flowering in *Arabis alpina*. *Nature* **42**: 3–428.
- Werth, S., Cornejo, C. & Scheidegger, C. In press. Characterization of microsatellite loci in the lichen fungus *Lobaria pulmonaria* (Lobariaceae). *APPS*.
- Wood, T. S. & Backus, B. T. 1992. Differentiation of North American and European forms of *Fredericella sultana* (Blumenbach) (Ectoprocta: Phylactolaemata). *Hydrobiologia* **237**: 185–193.
- Zalapa, J.E., Cuevas, H., Zhu, H., Steffan, S., Senalik, D., Zeldin, E. *et al.* 2012. Using next-generation sequencing approaches to isolate simple sequence repeat (microsatellite) loci in the plant sciences. *Am. J. Bot.* **99**: 193–208.

## Supporting information

Additional Supporting Information may be found in the online version of this article:

**Table S1** Species-specific information on the number of microsatellites tested and the numbers of polymorphic microsatellites chosen to work with.

**Table S2** Number of 454 pyrosequencing reads containing dinucleotide, trinucleotide and tetranucleotide motifs sorted by approximate genome size.

Data deposited at Dryad: doi: 10.5061/dryad.22p7m

Received 12 October 2012; accepted 19 November 2012