ELSEVIER

Gene 241 (2000) 101–105

GENE
AN INTERNATIONAL JOURNAL ON GENES, GENOMES AND EVOLUTION

www.elsevier.com/locate/gene

Interspecific evolution in plant microsatellite structure

Marianne Barrier [a], Elizabeth Friar [b], Robert Robichaux [c], Michael Purugganan [a,*]

[a] *Department of Genetics, Box 7614, North Carolina State University, Raleigh, NC 27695, USA*
[b] *Rancho Santa Ana Botanic Garden and Claremont Graduate University, 1500 N. College Ave., Claremont, CA 91711, USA*
[c] *Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA*

Received 7 July 1999; received in revised form 6 September 1999; accepted 12 October 1999
Received by V. Sgaramella

Abstract

Several intragenically linked microsatellites have been identified in the floral regulatory genes *A. sandwicense* APETALA1 (*ASAP1*) and *A. sandwicense APETALA3/TM6* (*ASAP3/TM6*) in 17 species of the Hawaiian and North American Madiinae (Asteraceae). Thirty-nine microsatellite loci were observed in the introns of these two genes, suggesting that they are hotspots for microsatellite formation. The sequences of four of these microsatellites were mapped onto the phylogenies of these floral regulatory genes, and the structural evolution of these repeat loci was traced. Both nucleotide substitutions and insertion/deletion mutations may be responsible for the formation of perfect microsatellites from imperfect repeat regions (and vice versa). © 2000 Published by Elsevier Science B.V. All rights reserved.

Keywords: Adaptive radiation; Floral regulatory genes; Hawaiian silversword alliance; Imperfect microsatellites; Simple sequence repeats

1. Introduction

Microsatellites, or simple sequence repeats (SSRs), are short sequence elements composed of tandem repeat units one to seven base pairs (bp) in length (Tautz, 1989). Microsatellites are present in high numbers in mammals and are randomly distributed with a density of approximately one microsatellite present every 10–15 kb of sequence (Tautz, 1989). These simple sequence repeats are also present in plant genomes, but appear to be less abundant than in mammalian or insect systems (Van Treuren et al., 1997). Microsatellite sequences possess high mutation rates, estimated at 10^{-2}–10^{-3} per locus per gamete per generation (Tautz, 1989). These repeat sequences have been shown to be highly polymorphic within and between species, a property that has permitted their application as molecular markers in population genetics (Goldstein et al., 1999), systematics (Goldstein and Pollock, 1997), and genome mapping (Weissenbach et al., 1992). Microsatellite instability is also implicated in several human diseases (Kunkel, 1993) and may be associated with variation in gene regulation (Meloni et al., 1998).

Relatively little is known about the origins and interspecific evolution of microsatellite loci. Documented examples of microsatellite formation suggest that a variety of mutations, including repeat number amplification (Schlotterer and Tautz, 1992), simple nucleotide substitutions, and insertion/deletion events, can contribute to the formation of simple sequence repeats (Estoup et al., 1995; Messier et al., 1996). Studies on the origins and fates of sequence repeats may provide insights into evolutionary trends in microsatellite behavior and assist investigators in identifying polymorphic markers that can be utilized across different taxa. One approach to investigating patterns of microsatellite evolution focuses on the analysis of between-species diversification of sequence repeat structures among a large number of closely related taxa.

We have isolated the floral regulatory genes *A. sandwicense APETALA1* (*ASAP1*) and *A. sandwicense APETALA3/TM6* (*ASAP3/TM6*) from members of the Hawaiian silversword alliance and their nearest North American tarweed relatives (Asteraceae: Heliantheae-Madiinae). These floral regulatory genes contain a large number of sequence repeat regions, many of which can

Abbreviations: *ASAP1*, *A. sandwicense APETALA1*; *ASAP3/TM6*, *A. sandwicense APETALA3/TM6*; bp, base pairs; SSR, simple sequence repeats.

* Corresponding author. Tel.: +1-919-515-1761; Fax: +1-919-515-3355.
E-mail address: michaelp@unity.ncsu.edu (M. Purugganan)

PII: S0378-1119(99)00463-1

be considered microsatellite sequences. Phylogenetic analyses of the gene sequences and mapping of the structure of the associated microsatellites provide a framework to examine interspecific evolution in plant microsatellite structure.

2. Materials and methods

2.1. Gene sampling

The floral regulatory genes *ASAP3/TM6* and *ASAP1* were isolated from 10 species of the Hawaiian silversword alliance and seven species of North American tarweeds as previously described (Barrier et al., 1999). The Hawaiian species were selected to represent each of the four major lineages in the silversword alliance as previously identified from rDNA ITS trees (Baldwin and Robichaux, 1995; Baldwin, 1996). Four North American species (*Madia bolanderi*, *M. nutans*, *Raillardiopsis scabrida*, and *R. muirii*) were chosen to represent each of the four major lineages in the *Madia/Raillardiopsis* group as also identified from rDNA ITS trees (Baldwin, 1996). Three other North American tarweed species (*Adenothamnus validus*, *Raillardella pringlei*, and *Osmadenia tenella*) are known to fall outside the clade comprising *Madia/Raillardiopsis* and the silversword alliance (Baldwin, 1996) and were included to serve as the outgroups in the analyses.

The PCR primers ASAP3-2 and ASAP3-3R were designed to allow amplification of a region spanning exons 1–4 of the *ASAP3/TM6* gene; the two copies of this gene in the Hawaiian species were discriminated from one another by size (Barrier et al., 1999). For *ASAP1*, gene-specific primer pairs (ASAP1-F1A/ASAP1-RB and ASAP1-AF/ASAP1-R) were constructed to amplify sequences from exons 3 to 8 of different duplicate copies of the gene in the Hawaiian species. Only single copies of the gene were identified in the North American species. The primers were used in PCR amplifications using the error-correcting rTth polymerase formulation (Perkin-Elmer) in a standard buffer with 40 cycles of 95°C for 1 min, 55°C for 1 min, and 72°C for 4 min. The nucleotide error rate for this formulation is less than 1 bp in 7 kb of sequence (unpublished observations). PCR-amplified DNA was cloned using the TA cloning kit (InVitrogen) and sequenced using automated sequencers (Iowa State University Sequencing Facility, NCSU DNA Sequencing Facility). Sequencing was carried out using nested primers, with multiple sequencing reactions conducted for divergent sequences. All sequence changes were rechecked visually against sequencing chromatograms, and are deposited in GenBank (Accession Nos. AF147210 to AF147258).

2.2. Evolutionary analyses

Nucleotide sequences were aligned visually. Phylogenetic analyses were conducted using maximum parsimony techniques implemented in PAUP* 4.0d54 (Swofford 1998). Both substitution and insertion/deletion (indel) differences were used and weighted equally in the analyses, with the indels separately coded (as additional characters) to reflect non-independence of continuous gaps. Identified microsatellite sequences were excluded from the analyses. Parsimony analyses were conducted using the heuristic search procedure, with random taxon addition (10 replicates), tree–bisection–reconnection branch swapping, and MULPARS in effect. Clade support was estimated by parsimony analysis of 500 bootstrap replicates of the data set using the search procedures outlined above. Microsatellite sequences were mapped onto the phylogenies under maximum parsimony, using MacClade under ACCTRAN conditions (Maddison and Maddison, 1992).

3. Results and discussion

The floral regulatory genes *ASAP1* and *ASAP3/TM6* were isolated from 17 members of the Hawaiian and North American Madiinae [Asteraceae] (Barrier et al., 1999). There are two copies of *ASAP1* and *ASAP3/TM6* (*A* and *B* copies) in species of the Hawaiian silversword alliance, while only a single copy of these loci is found within members of the North American *Madia/Raillardiopsis* group. Phylogenetic analyses lead us to conclude that the duplicate copies of these two loci in the Hawaiian silversword alliance arose from an interspecific hybridization event between species within the North American *R. scabrida* and *R. muirii* lineages (Barrier et al. 1999). Likelihood estimates based on data from the rDNA internal transcribed spacer (ITS) locus suggest that the most recent common ancestor of the Hawaiian silversword alliance existed 5.2 ± 0.8 million years ago (Ma), contemporaneous with the origin of the Island of Kaua'i (Baldwin and Sanderson 1998). In contrast, the earliest date for the diversification of the North American Madiinae appears to be in the mid-Miocene 15 Ma (Baldwin and Sanderson 1998).

Introns of the *ASAP1* and *ASAP3/TM6* genes appear to harbor a number of sequence repeat motifs containing four or more mono-, di- or trinucleotide repeats. There does not appear to be a consensus regarding the minimum number of repeats that defines a microsatellite sequence; for our purposes, we recognize a sequence motif as a microsatellite locus if it contains at least six repeat units. In more than 3 kb of common aligned sequence for the *ASAP1* locus in both Hawaiian and North American species, at least 102 repeat regions have

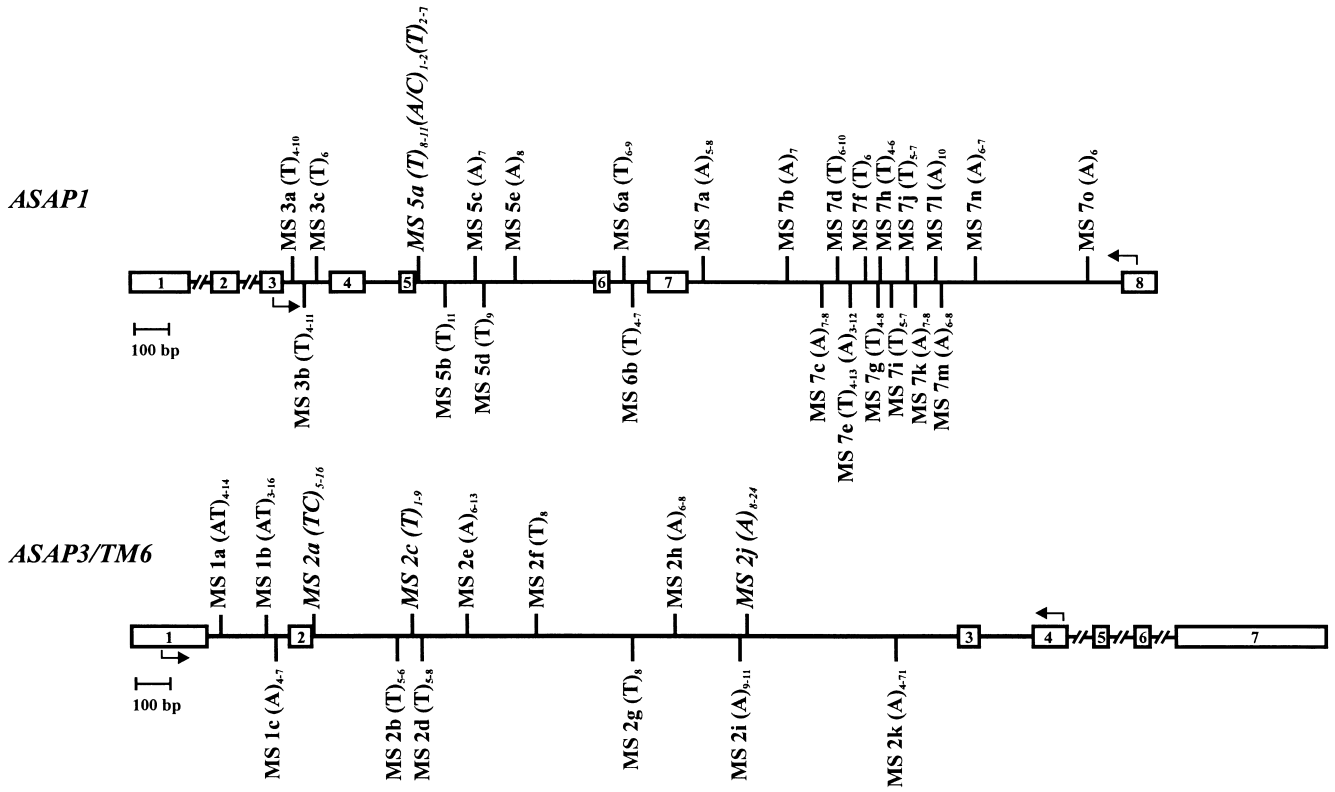

Fig. 1. Structures of the *ASAP1* and *ASAP3/TM6* genes. Exons are shown as numbered boxes. The relative positions of identified microsatellites are indicated. A microsatellite is identified if it contains at least six repeat units in one species. The microsatellites discussed in this paper are shown in italics. The positions of the PCR primers used in sequence amplification are marked by arrows (Barrier et al. 1999).

been identified. Twenty-five of these have at least six tandem repeats and are therefore defined as microsatellite loci (see Fig. 1). The 2.7 kb of common aligned sequence for *ASAP3/TM6* yields at least 46 repeat regions, 14 of which are microsatellites (see Fig. 1). Twenty-three of 25 microsatellites found in *ASAP1* and 11 of the 14 microsatellites found in *ASAP3/TM6* consist of $(A)_n$ or $(T)_n$ mononucleotide repeats. The other three microsatellites from *ASAP3/TM6* contain $(AT)_n$ or $(TC)_n$ dinucleotide repeats. The large number of repeat regions in the introns of these two loci suggests that these genes may be hotspots for microsatellite formation.

Microsatellite loci can be classed as perfect or imperfect (Estoup et al., 1995). Imperfect microsatellites contain nucleotide repeats that are interrupted by one or more non-repeat nucleotides. Structural interruptions decrease the number of tandem repeats and may stabilize microsatellite loci, rendering them less prone to slippage mutations (Van Treuren et al., 1997). We observe several cases of structural evolution of imperfect to perfect microsatellites (and vice versa). Mutational steps that govern the interspecific dynamics of microsatellite loci were inferred by reconstructing the phylogenies of the genes that contain microsatellite sequences (Barrier et al., 1999) and mapping the structures of the embedded simple sequence repeats and their associated flanking regions onto the phylogenies. For four simple sequence repeat regions (see Fig. 2), we can document insertions/deletions or substitutions of nucleotide interruptions in microsatellite loci.

Perfect microsatellite loci may arise from imperfect repeat sequences by the removal of nucleotide interruptions. One example is the *ASAP3/TM6* MS (microsatellite) 2a locus, which is a perfect dinucleotide $(TC)_n$ microsatellite in the *ASAP3/TM6-B* gene in the Hawaiian species (see Fig. 2B). This microsatellite displays a greater degree of between-species variability ($n=$ 6–16) than its imperfect microsatellite orthologue in the *ASAP3/TM6-A* gene in the Hawaiian species, all of which share the same $(TC)_4C(TC)_1$ structure (see Fig. 2B). In this case, an insertion of a T nucleotide upstream of the C interruption appears to have led to the formation of the perfect, uninterrupted microsatellite locus.

A similar pattern is observed in the *ASAP1* MS 5a locus, where removal of nucleotide bases that interrupt the contiguity of repeat regions results in the formation of longer repeat tracts (see Fig. 2A). MS 5a is a compound microsatellite and consists of a $(T)_n(C/A)_1(T)_m\ (C)_{1-2}(T)_p$ interrupted repeat region in the *ASAP1* gene of the North American species. The *ASAP1-B* gene of the Hawaiian species, however, con-

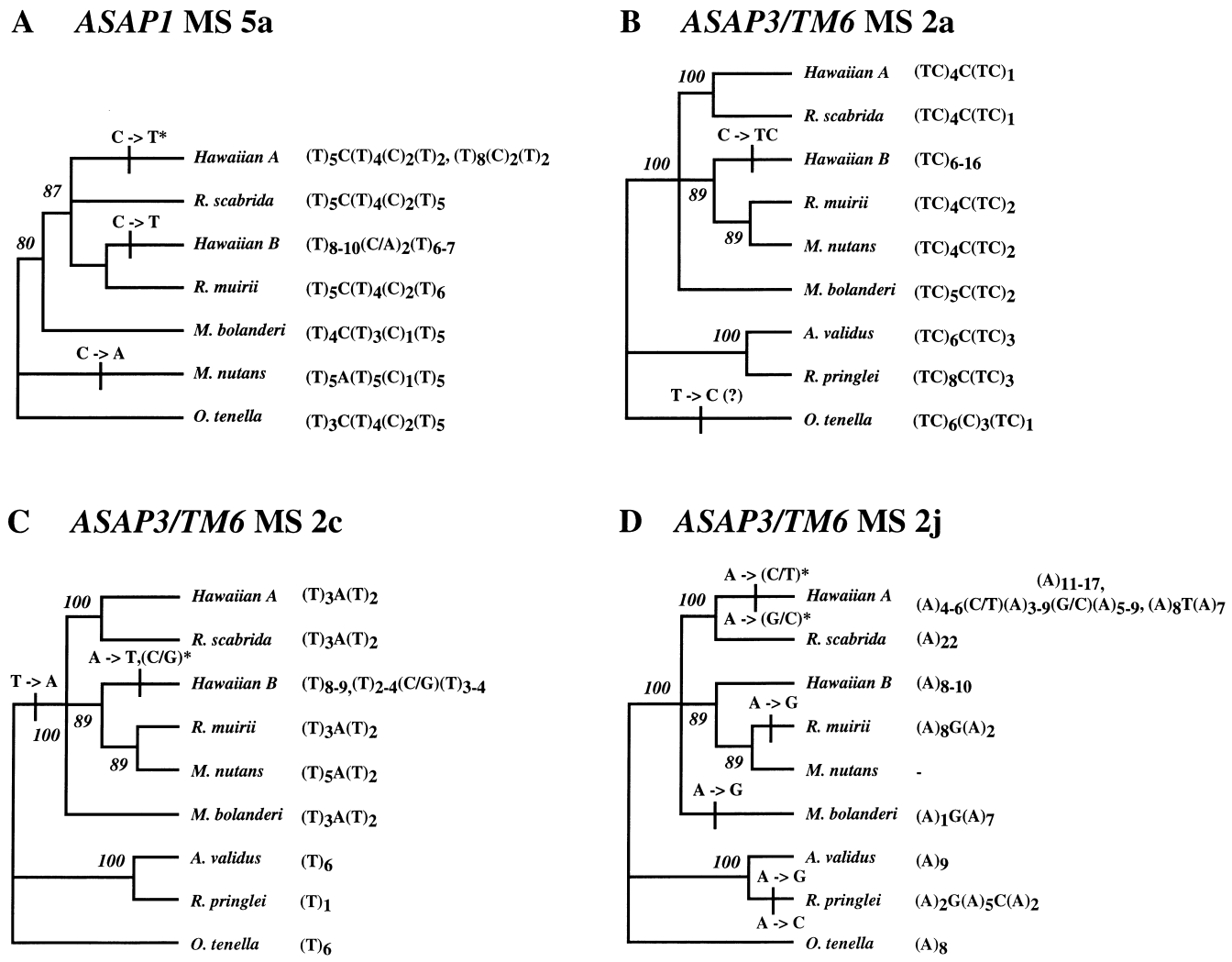

Fig. 2. Evolution of four *ASAP1* and *ASAP3/TM6* microsatellite loci. The structures of the microsatellite loci are mapped onto the respective *ASAP1* and *ASAP3/TM6* gene phylogenies (Barrier et al., 1999). The variable microsatellite sequences were excluded when the gene phylogenies were reconstructed. The Hawaiian genes are indicated as either *A* or *B* copy. The inferred mutations represent the most parsimonious changes for the structures of these loci. Mutations with an asterisk indicate that the changes are observed in only some members of the clade. A dash indicates that a deletion event has removed that microsatellite locus. In the *ASAP1* MS 5a locus, we assume that the first C interruption is the ancestral state. In the *ASAP3/TM6* MS 2j locus, we assume that the perfect $(A)_n$ microsatellite is the ancestral state. Bootstrap percentage values from 500 replicates of the data are given in italics next to the nodes.

tains a longer, variable $(T)_n$ tract at the 5′ end of this complex microsatellite. From the phylogeny, we infer that the imperfect $(T)_{3-5}C(T)_{3-5}(C)_{1-2}(T)_{5-6}$ was the founder sequence, and that loss of the first C interruption in the ancestral copy of the *ASAP1-B* locus resulted in a longer $(T)_n$ microsatellite tract in the Hawaiian species. This has occurred at least twice in the Hawaiian taxa; at least one Hawaiian species has an *ASAP1-A* gene that also exhibits a loss of the first C interruption.

Several clear instances of evolution of imperfect, interrupted microsatellites from perfect loci are also observed. In the *ASAP3/TM6* MS 2c locus, the members of the *Madia/Raillardiopsis* group and the Hawaiian silversword alliance contain an A interruption that is absent in the $(T)_n$ repeat sequences in the outgroup species (see Fig. 2C). Interestingly, there is a secondary loss of this interruption in the *ASAP3/TM6-B* gene in some of the Hawaiian species. The evolution of structural interruptions may occur independently several times in species groups. There is evidence, for example, of multiple gains (or losses) of interruptions in the *ASAP3/TM6* MS 2j microsatellite locus (see Fig. 2D).

It is unclear in most of these cases (as in most other cases examined) whether the gain or loss of the interrupted base(s) occurs via insertion/deletion events or simple point mutations. Previous studies on microsatellite loci in primates (Messier et al., 1996) and bees (Estoup et al., 1995) have highlighted the role that both simple nucleotide substitutions and insertion/deletion events play in the formation and evolution of perfect

microsatellite loci. Our results also suggest that the structural evolution of plant loci may proceed via multiple molecular mechanisms.

Acknowledgements

The authors thank Bruce Baldwin for providing us with some DNA samples for this study, and J. Mark Porter for helpful discussions. This work was supported by a grant from the National Science Foundation to M.D.P. and R.H.R. and from the Alfred P. Sloan Foundation to M.D.P.

References

Baldwin, B.G., Robichaux, R.H., 1995. Historical biogeography and ecology of the Hawaiian silversword alliance: New molecular phylogenetic perspectives. In: Wagner, W.L., Funk, V.A. (Eds.), Hawaiian Biogeography: Evolution on a Hot Spot Archipelago. Smithsonian Institution Press, Washington, DC, pp. 259–287.

Baldwin, B.G., 1996. Phylogenetics of the California tarweeds and the Hawaiian silversword alliance (Madiinae: Heliantheae sensu lato). In: Hind, D.J.N., Beentje, H.J. (Eds.), Compositae: Systematics. Royal Botanic Gardens, Kew, UK, pp. 377–391.

Baldwin, B.G., Sanderson, M.J., 1998. Age and rate of diversification of the Hawaiian silversword alliance (Compositae). Proc. Natl. Acad. Sci. USA 95, 9402–9406.

Barrier, M., Baldwin, B.G., Robichaux, R.H., Purugganan, M.D., 1999. Interspecific hybrid ancestry of a plant adaptive radiation: Allopolyploidy of the Hawaiian silversword alliance (Asteraceae) inferred from floral homeotic gene duplications. Mol. Biol. Evol. 16, 1105–1113.

Estoup, A., Tailliez, C., Cornuet, J.M., Solignac, M., 1995. Size homoplasy and mutational processes of interrupted microsatellites in two bee species, *Apis mellifera* and *Bombus terrestris* (Apidae). Mol. Biol. Evol. 12, 1074–1084.

Goldstein, D.B., Pollock, D.D., 1997. Launching microsatellites: A review of mutation processes and methods of phylogenetic inference. J. Hered. 88, 335–342.

Goldstein, D.B., Roemer, G.W., Smith, D.A., Reich, D.E., Wayne, R.K., 1999. The use of microsatellite variation to infer population structure and demographic history in a natural model system. Genetics 151, 797–801.

Kunkel, T.A., 1993. Slippery DNA and human diseases. Nature 365, 207–208.

Maddison, W., Maddison, D., 1992. MACCLADE 3.0: Analysis of phlogeny and character evolution. Sinauer Press, Sunderland, MA.

Meloni, R., Albanese, V., Ravassard, P., Treilhou, F., Mallet, J., 1998. A tetranucleotide polymorphic microsatellite, located in the first intron of the tyrosine hydroxylase gene, acts as a transcription regulatory element in vitro. Hum. Mol. Genet. 7, 423–428.

Messier, W., Li, S.H., Stewart, C.B., 1996. The birth of microsatellites. Nature 381, 483.

Schlotterer, C., Tautz, D., 1992. Slippage synthesis of simple sequence DNA. Nucleic Acids Res. 20, 211–215.

Swofford, D.L., 1998. PAUP* 4.0 Beta. Sinauer Associates, Sunderland, MA.

Tautz, D., 1989. Hypervariability of simple sequence repeats as a general source for polymorphic DNA markers. Nucleic Acids Res. 17, 6463–6471.

Van Treuren, R., Kuittinen, H., Karkkainen, K., Baena-Gonzalez, E., Savolainen, O., 1997. Evolution of microsatellites in *Arabis petraea* and *Arabis lyrata*, outcrossing relatives of *Arabidopsis thaliana*. Mol. Biol. Evol. 14, 220–229.

Weissenbach, X., Gyapay, G., Dib, C., Vignal, A., Morisset, J., Millasse, P., 1992. A second-generation linkage map of the human genome. Nature 359, 794–801.