Lepidopteran microsatellite DNA: redundant but promising

De-Xing Zhang

State Key Laboratory of Integrated Management of Pest Insects & Rodents, Institute of Zoology, Chinese Academy of Sciences, 25 Beisihuan-Xi Road, Beijing 100080, PR China

The development of microsatellite DNA markers for use with butterflies and moths is extremely difficult for no apparent reason. New findings demonstrate that many lepidopteran microsatellite sequences exist in multiple copies in the genome, and have similar or almost identical flanking regions. These findings provide a compelling explanation for the low efficiency of microsatellite isolation in Lepidoptera, and might also shed light on the evolutionary dynamics of microsatellite sequences in these insects and other organisms.

Microsatellites are short tandem repeat DNA sequences with one to six nucleotides as the repeating unit. Owing to their high level of polymorphism, reliability and reproducibility in PCR and genotyping, these sequences are the most powerful nuclear DNA markers available in population genetic and evolutionary studies, and have been popular in diverse research fields for over a decade. However, the isolation of MICROSATELLITE LOCI (see Glossary) for population genetic study in lepidopteran insects is difficult and challenging [1]. Currently, there is no convincing explanation of the reason for why this might be so. New findings by Meglécz *et al.* [2] suggest an answer to this conundrum, and might also shed light on the evolutionary dynamics of microsatellite sequences.

Why are microsatellite loci difficult to isolate in lepidopteran insects?

The abundance of microsatellite sequences varies considerably from taxon to taxon, with some groups, such as birds, having much lower frequencies in their genomes compared with other organisms (e.g. microsatellite frequencies in birds are more than three times lower than in humans) [3]. In Lepidoptera, including the 20 or so reported microsatellite studies of which I am aware, the number of polymorphic microsatellite loci isolated is no more than five per genomic library in 80% of cases (reviewed in [4]), compared with the higher number of loci characterized from other insects. For example, whereas 75 loci have been obtained from the honeybee *Apis mellifera* [5], only three loci were found in the clouded Apollo butterfly *Parnassius mnemosyne* using an identical experimental procedure [1,6]. It is also remarkable that whereas a microsatellite CLONING EFFICIENCY of 14% was achieved in the orthopteran migratory locust *Locusta migratoria*, only a 2.5% cloning efficiency was possible in the lepidopteran cotton bollworm *Helicoverpa armigera*, although both insects were studied in parallel in the same laboratory using identical experimental methods [7,8]. Such differences might indicate a reduced frequency of microsatellites in the lepidopteran genome.

The isolation of polymorphic microsatellite markers in Lepidoptera is further complicated by the occurrence of MICROSATELLITE DNA FAMILIES. In a recent paper, Meglécz *et al.* [1] reported that microsatellite sequences with similar or almost identical FLANKING REGIONS (thus existing as DNA families) were observed in butterfly species, but not in an aphid species that was examined in parallel with the use of an identical protocol. This strengthens an earlier and independent observation that >70% of microsatellite loci cloned from the cotton bollworm have one or both flanking regions that are formed of REPETITIVE DNA [7]. Thus, the association of microsatellites with repetitive sequences in the flanking regions appears to be a characteristic of lepidopteran microsatellites [4]. In most current genetic studies, only SINGLE-COPY MICROSATELLITES are used as diagnostic markers for detecting polymorphisms. This requires the flanking regions of microsatellites to be unique sequences in the genome, thus enabling specific primers to be designed. In turn, this enables the isolation of a specific DNA segment by PCR. Microsatellite DNA families are not suitable for use as genetic markers because primers designed in the repetitive flanking regions amplify simultaneously from multiple loci with

Glossary

Cloning efficiency: the proportion of the recombinant clones that contain the target sequences (e.g. microsatellites) in a DNA library; that is, the percentage of true positive clones in a library.
Flanking region: DNA sequences adjacent on either side of the target sequence (e.g. a microsatellite).
Microsatellite DNA family: a group of microsatellite loci with similar or identical flanking regions.
Microsatellite locus: a specific genomic region consisting of microsatellite DNA and its flanking regions.
Mobile elements (formally known as transposable elements): sequences that can move to different loci in the genome, such as retroelements and retrosequences.
Repetitive DNA: nucleotide sequences, such as mobile elements, that occur several times in the genome. The copy number of such sequences can vary from a few to over a million. Individual copies do not have to be identical in sequence, but they share high similarity.
Retroelements: one of the several types of mobile element; transpose by retroposition with the help of the reverse transcriptase encoded in their sequences.
Retrosequences: genomic sequences that have been derived from the reverse transcription of RNA and then integrated into the genome; they do not encode for reverse transcriptase.
Single-copy microsatellite: microsatellite loci the DNA sequences of which are present only once (i.e. are unique) in the genome.

Corresponding author: De-Xing Zhang (dxzhang@ioz.ac.cn).
Available online 4 August 2004

similar sequences, thus confounding the data obtained. This makes those microsatellite sequences with similar flanking regions in lepidopteran genomes useless for genetic analysis. Therefore, compared with other organisms, such as the migratory locust, microsatellite DNA in lepidopterans has two remarkable characteristics: (i) the overall low frequency; and (ii) high sequence redundancy. These characteristics provide an attractive explanation for the unusually low efficiency of polymorphic microsatellite isolation in lepidopteran insects.

Redundancy and evolutionary dynamics of lepidopteran microsatellites

Meglécz *et al.*'s observations prompt us to ask why so many of the microsatellite sequences that occur in the genomes of Lepidoptera are redundant, even though their overall frequency is low. Two important points are worth considering. The first is that microsatellites from a variety of organisms [9–12] have recently been found to be associated with MOBILE ELEMENTS, many of which are RETROELEMENTS and RETROSEQUENCES. This results in microsatellites that have similar or nearly identical flanking regions (similar to the situation observed in Lepidoptera [2,7]) that share homology with the mobile elements. Several authors have demonstrated that mobile elements can be involved in the genesis and genomic spread of microsatellites in organisms as diverse as primates [10] (including humans [11]), fruit flies [12] and plants [9]. The findings from lepidopteran microsatellites imply that similar mechanisms might be functioning in these insects. Because of insufficient data about mobile elements and repetitive DNA in Lepidoptera, it is difficult to ascertain the nature of the association of the observed repetitive microsatellite flanking sequences with any mobile elements. Nevertheless, Meglécz *et al.* were able to find sequence similarities between some flanking regions and certain retroviruses (such as the *Drosophila* retrovirus *Gypsy*) [2].

The second point is that, although the proportion of single-copy microsatellite sequences is lower in the genomes of Lepidoptera than it is in the genomes of other taxa (e.g. the migratory locust) [4], the percentage of microsatellite loci with repetitive flanking regions is high (e.g. >70% in the cotton bollworm, and 30% of all DNA family loci) [2,4,7]. If mobile elements are involved in lepidopteran microsatellite evolution, as is suggested for other organisms [9–12], we can divide the evolutionary dynamics of microsatellites into three sequential stages (Figure 1): (i) genesis; (ii) propagation and dispersal; and (iii) diversification. Thus, after a microsatellite is created (e.g. it can be part of a mobile element; see [12]), it multiplies and disperses in the genome via the transposition of mobile elements. This should lead to the presence of many microsatellite DNA sections with identical flanking regions in the genome. Divergence should then occur among these sequences as mutations accumulate independently. Given enough divergence time, DNA sequences of the originally identical flanking regions will become dissimilar, developing into single-copy microsatellite sequences. Therefore, if high similarity of the flanking regions exists among a group of microsatellite loci, and if relatively few microsatellites exist as single-copy sequences in the genome, it would suggest a short divergence time, indicating that these microsatellite sequences are at the stage of propagation and dispersal. The characteristics of microsatellites in Lepidoptera (i.e. low frequency in the genome but high redundancy with relatively few existing as single-copy sequences) imply that lepidopteran microsatellites are in an early stage of evolution; that is, a large proportion of these sequences have experienced recent propagation or multiplication in the genome. Thus, the flanking regions of microsatellites have not accumulated many mutations and remain similar in sequence, resulting in a low proportion of unique or diverged, single-copy sequences.

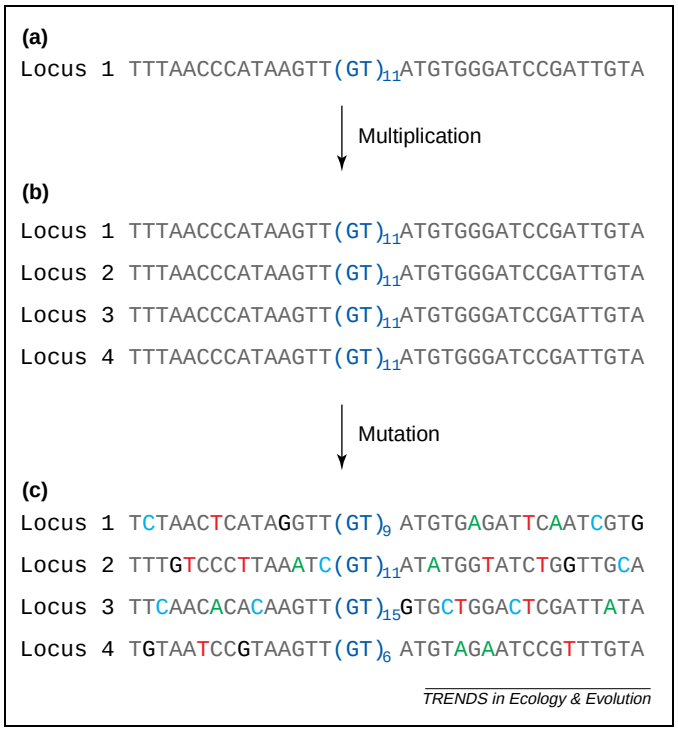


Figure 1. Hypothetical evolutionary dynamics of microsatellite DNA. Three sequential stages are proposed for the genomic evolution of microsatellite sequences, supposing that mobile elements are involved in their genomic multiplication and spreading: (i) genesis, (ii) propagation and dispersal, and (iii) diversification. After the formation of a microsatellite in a genomic region [**(a)**, shown as $(GT)_{11}$ in blue at locus 1], it multiplies and disperses in the genome via the transposition of mobile elements. This will lead to the presence of multiple copies of microsatellites with identical flanking regions [**(b)**, four copies shown] in the genome. As mutations [**(c)**, shown in green, blue, red and black] accumulate independently at different loci, DNA sequences of the originally identical flanking regions become dissimilar, eventually developing into unique, single-copy microsatellite sequences.

Future research

Given the unusual difficulty involved in their isolation, lepidopteran microsatellites have been little studied compared with, for example, those of hymenopterans [1], and there is little analysis of their molecular evolution. However, the striking characteristics of microsatellites in these organisms suggest that they are a system of choice for the in-depth study of the molecular evolution of microsatellite sequences. Their overall low frequency in the genome should simplify the analysis of their redundancy and genomic organization; the low proportion of single-copy microsatellite sequences provides an excellent opportunity to study the genesis and evolutionary

diversification of microsatellites. To do this, we first need to know more about these short sequence repeats by isolating more loci with enough genomic background information (i.e. enough sequence information about the flanking regions). We then need to characterize the repetitive flanking regions, for example, their genomic abundance, organization and any connection with mobile elements, because data from other organisms consistently suggest the coevolution of microsatellites with mobile elements. Phylogenetic-based analysis should then enable us to study the molecular evolution of microsatellite DNA families, testing the sequential hypothesis of evolutionary dynamics of microsatellites proposed above. Because the association of microsatellite with repetitive elements is not a sporadic phenomenon, and has been reported in a broad range of organisms (e.g. in plants [9,13], mammals [10,11] and fruit flies [12]), the outcome of such studies should be of general interest for understanding the molecular evolution of microsatellite sequences.

Another direction for research is to develop new methods for the efficient isolation of polymorphic microsatellite loci in these insects, taking into account the characteristics of these microsatellites reported so far. Such techniques will increase the efficiency in isolating microsatellite markers for population genetic and evolutionary studies, which has been the major limiting factor in the use of such markers in research.

Acknowledgements

This work was supported by an NSFC grant (30025008) and the CAS 'Bai Ren Ji Hua' Professorship. I thank Godfrey Hewitt for critical reading of the article, and Emese Meglécz and two anonymous referees for their valuable comments on an earlier version.

References

1 Nève, G. and Meglécz, E. (2000) Microsatellite frequencies in different taxa. *Trends Ecol. Evol.* 15, 376–377
2 Meglécz, E. *et al*. (2004) High similarity between flanking regions of different microsatellites detected within each of two species of Lepidoptera: *Parnassius apollo* and *Euphydryas aurinia*. *Mol. Ecol.* 13, 1693–1700
3 Primmer, C.R. *et al*. (1997) Low frequency of microsatellite in the avian genome. *Genome Res.* 7, 471–482
4 Ji, Y-J. and Zhang, D.X. (2004) Characteristics of microsatellite DNA in lepidopteran genomes and implications for their isolation. *Acta Zool. Sinica* 50, 608–614
5 Estoup, A. *et al*. (1993) Characterization of (GT)n and (CT)n microsatellites in two insect species: *Apis mellifera* and *Bombus terrestris*. *Nucleic Acids Res.* 21, 1427–1431
6 Meglécz, E. and Solignac, M. (1998) Microsatellite loci for *Parnassius mnemosyne* (Lepidoptera). *Hereditas* 128, 179–180
7 Ji, Y-J. *et al*. (2003) Polymorphic microsatellite loci for the cotton bollworm *Helicoverpa armigera* (Lepidoptera: Noctuidae) and some remarks on their isolation. *Mol. Ecol. Notes* 3, 102–104
8 Zhang, D.X. *et al*. (2003) Isolation, characterization and cross-species amplification of eight microsatellite DNA loci in the migratory locust (*Locusta migratoria*). *Mol. Ecol. Notes* 3, 483–486
9 Ramsay, L. *et al*. (1999) Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. *Plant J.* 17, 415–425
10 Arcot, S.S. *et al*. (1995) *Alu* repeats: a source for the genesis of primate microsatellites. *Genomics* 29, 136–144
11 Nadir, E. *et al*. (1996) Microsatellite spreading in the human genome: evolutionary mechanisms and structural implications. *Proc. Natl. Acad. Sci. U. S. A.* 93, 6470–6475
12 Wilder, J. and Hollocher, H. (2001) Mobile elements and the genesis of microsatellites in dipterans. *Mol. Biol. Evol.* 18, 384–392
13 Elsik, C.G. and Williams, C.G. (2001) Families of clustered microsatellites in a conifer genome. *Mol. Genet. Genom.* 265, 535–542

doi:10.1016/j.tree.2004.07.020

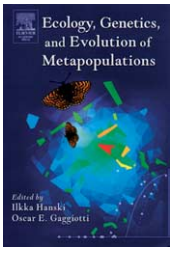
Book Reviews

Panmixia lost

Ecology, Genetics, and Evolution of Metapopulations edited by Ilkka Hanski and Oscar E. Gaggiotti, Elsevier Academic Press, 2004. US$54.95 / £36.99 (xix + 696 pages) ISBN 0123234484

David E. McCauley

Department of Biological Sciences, Vanderbilt University, Nashville, TN, 37235, USA



As reflected in the title of the recent edited volume by Hanski and Gaggiotti, the field of metapopulation biology has roots in three related disciplines: ecology, genetics and evolutionary biology. Studies of the ecology of metapopulations follow on from Levins' classic model [1] and were originally concerned with the conditions that enable species in a sub-divided habitat to persist in spite of frequent local extinctions. Population-genetic studies of metapopulations originated largely with Wright's discussion of how the local extinction and recolonization of populations might create a population structure favorable for 'shifting balance' evolution [2], and were later developed further with Slatkin's population-genetic models of subdivided populations that had been subject to local extinction and recolonization [3]. The study of adaptive evolution in metapopulations developed mainly from life-history theory, particularly from how dispersal strategies might evolve in unstable and patchy habitats [4]. These three approaches to the study of spatially subdivided and locally ephemeral populations combined to form the discipline of metapopulation biology during the late 1980s, as it became clear that many

Corresponding author: David E. McCauley (david.e.mccauley@vanderbilt.edu).
Available online 10 June 2004