

Genome size and microsatellites: the effect of nuclear size on amplification potential

Trenton W.J. Garner

Abstract: Although the frequency of microsatellite DNA regions generally increases with increasing genome size, genome size has a negative effect on polymerase chain reaction (PCR) amplification. Thus, researchers developing sets of PCR primers, as is commonly done for microsatellite DNA regions, may encounter greater difficulty when working with species that have larger genomes. I investigated the effect of genome size on overall amplification success using data from nine different metazoan taxa. The proportion of primer sets that did not amplify PCR products was strongly and positively correlated with the haploid *C* value of the target species. Increasing genome size may affect amplification success negatively because of a decrease in target:nontarget DNA or by dilution of the available primer pool by nonspecific binding.

Key words: microsatellites, genome size, amplification success.

Résumé : Bien que la fréquence des microsatellites augmente généralement avec la taille du génome, la taille du génome a un effet négatif sur la réaction d'amplification à la polymérase (PCR). Ainsi, des chercheurs qui développent des amorces PCR, tel que cela se fait communément pour les microsatellites, peuvent rencontrer de plus grandes difficultés lorsqu'ils travaillent sur des espèces à grand génome. L'auteur a examiné l'effet de la taille du génome sur le taux de succès des amplifications en employant les données provenant de neuf taxons de métazoaires. La proportion des jeux d'amorces qui n'ont pas produit des amplicons était fortement et positivement corrélée avec le contenu en ADN de l'espèce à l'étude. L'accroissement de la taille du génome pourrait influencer négativement l'amplification en raison d'un accroissement de l'ADN non-cible par rapport à l'ADN cible ou encore par une diminution de la concentration d'amorces disponibles en raison de l'appariement non-spécifique.

Mots clés : microsatellites, taille du génome, succès de l'amplification.

[Traduit par la Rédaction]

Introduction

The ubiquity of microsatellites in all eukaryotic taxa, along with their highly mutable nature, has made them favored markers for population genetics studies and investigations of social systems. Several studies have examined the relationship between the frequency of microsatellite regions and total genome size (e.g., Kubis et al. 1998; Field and Wills 1998; Primmer et al. 1997). Microsatellite frequency generally correlates positively with genome size (Primmer et al. 1997; Hancock 1999), but the frequencies of different types of microsatellites vary widely between taxa and chromosomal types (Hancock 1999; Field and Wills 1998). Overall, the frequency of microsatellites detected in

eukaryotic taxa is greater than that predicted on the basis of nucleotide composition (Behe 1995).

Polymerase chain reaction (PCR) primers that amplify microsatellites are usually developed by screening sub-genomic libraries for repetitive regions. Primers specific for flanking regions are designed from sequenced positive clones and tested for amplification and polymorphism (e.g., Gautschi et al. 2000; Garner et al. 2000a, 2000b, 2001). Because genome size is positively correlated with the occurrence of microsatellite regions, the size of a genome should not affect the likelihood of finding suitable candidate regions for primer design. However, DNA concentration and content have an effect on the efficiency of single locus PCR amplification. Sufficient amounts of target region must be present and, conversely, sufficient quantities of primer must also be present for effective PCR amplification to occur (Hoelzel and Green 1992). Typically, template concentration is considered in terms of total nuclear DNA (Hoelzel and Green 1992), but if genome sizes are large, the relative proportion of target to nontarget DNA is reduced. Farrelly et al. (1995) investigated the potential effects that genome size could have on amplification success of partial 16S rDNA fragments. Their experimental manipulation of the relative

Received June 6, 2001. Accepted September 5, 2001.

Published on the NRC Research Press Web site at <http://genome.nrc.ca> on January 14, 2002.

Corresponding Editor: P.B. Moens.

T.W.J. Garner. Zoologisches Institut, Universität Zürich-Irchel, Winterthurerstraße 190, CH-8057 Zürich, Switzerland (e-mail: twjg@zool.unizh.ch).

Table 1. Species, genome sizes, and proportion of nonamplifying microsatellite primer sets.

Species	C value	Proportion
<i>Sepsis cynipsea</i>	0.791	0/11
<i>Scathophaga stercoraria</i>	0.965	2/17
<i>Rana temporaria</i>	4.31	14/39
<i>Rana lessonae</i>	6.37	9/36
<i>Rana latastei</i>	6.80	8/34
<i>Thamnophis sirtalis</i>	2.46	1/6
<i>Thamnophis elegans</i>	2.73	2/9
<i>Triturus alpestris</i>	25.59	18/37
<i>Triturus vulgaris</i>	25.62	15/34

Note: See text for description of how genome sizes were estimated.

proportions of target to nontarget DNA showed that a decrease in available rDNA template without a concurrent decrease in total template DNA corresponded with decreased amplification success (Farrelly et al. 1995).

If genome size has a negative effect on amplification success at a single locus then, by extension, the success rate of primer sets tested on organisms with large genomes should be lower than the success rate of primer sets tested on organisms with small genomes. If such a relationship exists, this could have serious consequences in terms of time and funding for the development of marker systems for organisms with large nuclear DNA complements. The practical implications prompted me to investigate the effect of genome size on nonamplification rates using information generated while developing microsatellite markers for nine eukaryotic organisms. These nine species span a broad spectrum of eukaryotic taxa: two dipterans (*Scathophaga stercoraria* and *Sepsis cynipsea*), three ranid frogs (*Rana lessonae*, *Rana latastei*, and *Rana temporaria*), two tritrid newts (*Triturus alpestris* and *Triturus vulgaris*), and two colubrid snakes (*Thamnophis sirtalis* and *Thamnophis elegans*).

Materials and methods

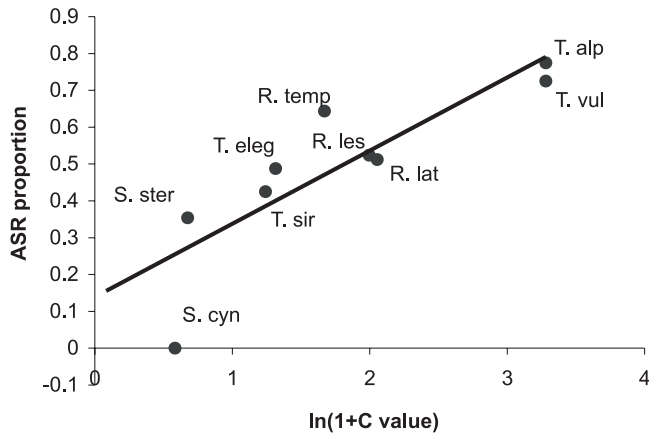
Enriched libraries specific for CA-repeat motifs were developed for each species and screened following the protocol of Gaustchi et al. (2000), with the exception of *T. alpestris*. For this species, two separate libraries were constructed, one specific for CA repeats and a second specific for CAGA repeats. Positive clones were sequenced in both directions using M13 forward and reverse primers, and the ABI Prism[®] BigDye[™] Terminator Cycle Sequencing Ready Reaction Kit, version 2.0 (Perkin-Elmer Applied Biosystems, Foster City, Calif.). Cleaned and precipitated sequencing reactions were analyzed using the ABI 377 automated sequencing system (Perkin Elmer Applied Biosystems) and assembled using Sequencher[™] 3.0 software (Gene Codes Corporation, Ann Arbor, Mich.). Nonredundant clones with sufficient flanking regions on either side of observed repetitive regions, both simple and complex, were analyzed using Primer3 software (Rozen and Skaletsky 1998), using the following alterations to the default primer design settings: maximum primer size was set to 27 nucleotides, maximum melting temperature (T_m) difference was set to 2.0, primer GC content was set at a minimum of 30%, and a GC clamp was required. In some cases, conditions were relaxed, but

those primer sets designed with relaxed conditions were excluded from the following analysis with the exception of those designed without a GC clamp. All candidate primers designed using the above criteria were assigned a suggested annealing temperature of no less than 57°C by the Primer3 program.

All candidate primer sets were tested for amplification using an initial denaturation step of 3 min at 94°C, followed by a 30 cycle thermotreatment. Each cycle consisted of 30 s, 94°C (denaturation); 1 min, 55°C (annealing); and 1 min, 72°C (extension). This was then followed by a final extension step of 2 min at 72°C. All candidate primer sets were tested using at least five (no more than eight) different test templates from individuals of the species from which the primers were developed. Individual PCRs were 10- μ L total volume and contained approximately 100 ng template DNA, 0.5 U *Taq* polymerase (Quantum-Appligene, Heidelberg, Germany), 10 mM Tris-HCl (pH 9.0), 50 mM KCl, 1.5 mM MgCl₂, 0.1% Triton X-100, 0.2 mg bovine serum albumin (BSA) (PCR Buffer, Quantum-Appligene), 100 μ M of each dNTP (Promega, Madison, Wis.), and 0.5 μ M of both forward and reverse primers. All completed PCRs were electrophoresed using one of the two following electrophoresis systems: (i) most PCRs were electrophoresed on 8%, nondenaturing, 14.5 \times 17 cm acrylamide gels overnight at 80 V, and the gels were then stained for 30 min using an ethidium bromide bath and visually scored for amplification products using a standard transilluminator system; (ii) the remaining PCRs were electrophoresed using the SEA 2000[™] advanced submerged gel electrophoresis apparatus and Spreadex[™] EL-300, EL-600 or EL-800 gels (Elchrom Scientific AG, Switzerland). Electrophoresis times ranged from 75–140 min and gels were subsequently stained with SYBR[®] Gold nucleic acid stain (Molecular Probes Europe BV, Leiden, The Netherlands). All primer sets were tested at least twice for potential amplification and polymorphism.

I classified primer sets into two separate categories based on amplification products. The first category included all primer sets that did not amplify discernable PCR products within 100 bp of the expected target region. The second included any detectable amplification patterns, including monomorphic bands in the expected size range, polymorphic bands within the expected size range, and multiband (more than two, i.e., nonspecific) banding patterns that included band(s) within the expected size range. I searched the Animal Genome Size (AGS) Database (Gregory 2001) for the haploid C values (pg) for each of the nine focal species. I could find published C values for all the vertebrate species included in this study and, in the cases of *T. sirtalis*, *T. alpestris*, *T. vulgaris*, *R. lessonae*, and *R. temporaria*, multiple, different, published values were present in the database. For these species, I used the average of all published values for analysis. Because I could find no value for the two dipteran species, I substituted the value of the nearest phylogenetic relative found in the AGS Database, as determined by browsing the Tree of Life. For *Scathophaga stercoraria*, the two possibilities were *Delia antiqua* (Anthomyiidae) or *Musca domestica* (Muscidae). I chose *Musca*, because its genome size was greater than *Delia* and therefore would result in a more conservative comparison. Again, I averaged the multiple entries for this species. For

Fig. 1. Regression of proportion of nonamplifying primer sets on genome size. Proportion is arcsine square-root (ASR) transformed and genome size is the $\ln(1 + C$ value (pg)



Sepsis cynipsea, no clear choice was present in the AGS Database. I selected *Drosophila nasutooides* (Drosophilidae) as the nearest representative with the largest genome size. Although my two dipteran C values are estimates, most dipteran species, with the exception of some of the Culcidae, exhibit haploid C values at least one order of magnitude less than that of any other taxon used in this study (e.g., Squamata). Therefore, variation between dipteran species is unlikely to substantially influence conclusions. I compared the proportion of category 1 (nonamplifying) primer sets for each species to the C value using simple linear regression. Proportions were arcsine square-root (AGR) transformed before regression analysis. Genome sizes were bimodally distributed at the extreme ends of the total distribution, so these values were \ln transformed. Two of the transformed C values proved to be negative, so the transformation was repeated using $1 + C$ value.

Results and discussion

C values used in this study ranged from 0.791 in *Drosophila nasutooides* (substituted for *Sepsis cynipsea*) to 25.62 for *Triturus vulgaris* (Table 1). Results of the regression clearly show a significant positive relationship between genome size and the proportion of primer sets not amplifying ($F = 18.534$, $P = 0.004$; Fig. 1). As I substituted C values of related species for the two dipterans, I removed these two species from the analysis and recalculated the regression for the remaining seven species. Again, the relationship between genome size and amplification success was positive and highly significant ($F = 16.895$, $P = 0.009$).

Although nonamplification of microsatellite loci often is because of a lack of conservation of priming sites (null alleles), this is more a factor when amplifying across species (Balloux et al. 1998) and is unlikely to affect all individuals within a species equally, as is the case here. Instead, I propose two mechanistic reasons that together could account for my result. First, as described by Farrelly et al. (1995), a reduction in target DNA relative to the total content of template DNA in a single PCR serves to reduce amplification efficiency. As genome sizes increase, this serves to dilute the proportion of available target DNA in a given volume of

template DNA. Second, as genome size increases, the number of regions complementary to any 20-bp region (the average size of primers generated for this study) also increases. A larger genome would then increase the amount of nonspecific binding of primers, a concurrent diminishing of the primer pool available for PCR, and aggravation of the dilution effect. Alteration of PCR conditions can have some effect on amplification success: increasing primer concentration above the standard $0.5 \mu\text{M}$ /primer does enable amplification in some cases (personal observation). However, this approach is by no means a panacea and has proven to rarely compensate for the dilution effect of large genomes (personal observation).

Orthopterans are regularly used in studies of sperm competition (e.g., Simmons and Achmann 2000; Tregenza and Wedell 1998), and at least one family within this taxon (Acrididae) exhibits elevated genome sizes (Gregory 2001). Caudate amphibians are proving to be model organisms for studying metapopulation dynamics (Jehle et al. 2001) and have long been focal organisms for studies of sexual selection and mating dynamics (Halliday 1998). This entire vertebrate order is characterized by enormous genomes (Sessions and Larson 1987). Microsatellites are indispensable tools for determining patterns of sperm use and paternity and have proven to be extremely valuable for evaluating gene flow and patterns of interpopulation structure. The results of this study signify that investigators who would pursue studies of caudates and other organisms with large genomes should be prepared to allocate greater effort to developing microsatellite marker systems than those investigators who instead work with species with relatively small genomes.

Acknowledgements

I thank D. Hosken, C. Primmer, and H.-U. Reyer for valuable comments. This project was funded by grants from the Swiss National Foundation (SNF 31-56902.99 to D.J. Hosken, SNF 31-55426.98 to J. Van Buskirk, and SNF 31-40688.94 to H.-U. Reyer).

References

- Balloux, F., Ecoffey, E., Fumagalli, L., Goudet, J., Wyttenbach, A., and Hausser, F. 1998. Microsatellite conservation, polymorphism, and GC content in shrews of the genus *Sorex* (Insectivora, Mammalia). *Mol. Biol. Evol.* **15**: 473–475.
- Behe, M.J. 1995. An overabundance of long oligopurine tracts occurs in the genome of simple and complex eukaryotes. *Nucleic Acids Res.* **23**: 689–695.
- Farrelly, V., Rainey, F.A., and Stackebrandt, E. 1995. Effect of genome size and *rrn* gene copy number on PCR amplification of 16S rRNA genes from a mixture of bacterial species. *Appl. Environ. Microbiol.* **61**: 2798–2801.
- Field, D., and Wills, C. 1998. Abundant microsatellite polymorphism in *Saccharomyces cerevisiae*, and the different distributions of microsatellites in eight prokaryotes and *S. cerevisiae*, result from strong mutation pressures and a variety of selective forces. *Proc. Natl. Acad. Sci. U.S.A.* **95**: 1647–1652.
- Garner, T.W.J., and Tomio, G. 2001. Microsatellites for use in studies of the Italian Agile frog, *Rana latastei* (Boulenger). *Conserv. Genet.* **2**: 77–80.

- Garner, T.W.J., Gautschi, B., Röthlisberger, S., and Reyer, H.-U. 2000a. A set of CA repeat microsatellite markers derived from the pool frog, *Rana lessonae*. *Mol. Ecol.* **9**: 2173–2174.
- Garner, T.W.J., Brinkmann, H., Gerlach, G., Meyer, A., Ward, P.I., Spörri, M., and Hosken, D. 2000b. Polymorphic DNA microsatellites identified in the yellow dung fly (*Scathophaga stercoraria*). *Mol. Ecol.* **9**: 2207–2208.
- Gautschi, B., Tenzer, I., Müller, J.P., and Schmid, B. 2000. Isolation and characterization of microsatellite loci in the bearded vulture (*Gypaetus barbatus*) and cross-amplification in three Old World vulture species. *Mol. Ecol.* **9**: 2193–2194.
- Gregory, T.R. 2001. The animal genome size database. (Available from <http://www.genomesize.com>).
- Halliday, T.R. 1998. Sperm competition in amphibians. *In* Sperm competition and sexual selection. *Edited by* T.R. Birkhead and A.P. Møller. Academic Press, London. pp. 465–502.
- Hancock, J.M. 1999. Microsatellites and other simple sequences: genomic context and mutational mechanisms. *In* Microsatellites: evolution and applications. *Edited by* D.B. Goldstein and C. Schlötterer. Oxford University Press, Oxford. pp. 1–9.
- Hoelzel, A.R., and Green, A. 1992. Analysis of population-level variation by sequencing PCR-amplified DNA. *In* Molecular genetic analysis of populations. *Edited by* A.R. Hoelzel. Oxford University Press, Oxford. pp. 159–187.
- Jehle, R., Arntzen, J.W., Burke, T., Krupa, A.P., and Hödl, W. 2001. The annual number of breeding adults and the effective population size of syntopic newts (*Triturus cristatus*, *T. marmoratus*). *Mol. Ecol.* **10**: 839–850.
- Kubis, S., Schmidt, T., and Heslop-Harrison, J.S. 1998. Repetitive DNA elements as a major component of plant genomes. *Ann. Bot.* **82**(Suppl. A): 45–55.
- Primmer, C.R., Raudsepp, T., Chowdhary, B.P., Møller, A.P., and Ellegren, H. 1997. Low frequency of microsatellites in the avian genome. *Genome Res.* **7**: 471–482.
- Rozen, S., and Skaletsky, H.J. 1998. Primer3. (Available from http://www-genome.wi.mit.edu/genome_software/other/primer3.html).
- Sessions, S.K., and Larson, A. 1987. Developmental correlates of genome size in plethodontid salamanders and their implications for genome evolution. *Evolution*, **41**: 1239–1251.
- Simmons, L.W., and Achmann, R. 2000. Microsatellite analysis of sperm-use patterns in the bushcricket *Requena verticalis*. *Evolution*, **54**: 942–952.
- Tregenza, T., and Wedell, N. 1998. Benefits of multiple mates in the cricket *Gryllus bimaculatus*. *Evolution*, **52**: 1726–1730.