



A genome-wide investigation of microsatellite mismatches and the association with body mass among bird species

Haiying Fan and Weibin Guo

Department of Ecology, College of Life Sciences, Wuhan University, Wuhan, China

ABSTRACT

Mutation rate is usually found to covary with many life history traits of animals such as body mass, which has been readily explained by the higher number of mutation opportunities per unit time. Although the precise reason for the pattern is not yet clear, to determine the universality of this pattern, we tested whether life history traits impact another form of genetic mutation, the motif mismatches in microsatellites. Employing published genome sequences from 65 avian species, we explored the motif mismatches patterns of microsatellites in birds on a genomic level and assessed the relationship between motif mismatches and body mass in a phylogenetic context. We found that small-bodied species have a higher average mismatches and we suggested that higher heterozygosity in imperfect microsatellites lead to the increase of motif mismatches. Our results obtained from this study imply that a negative body mass trend in mutation rate may be a general pattern of avian molecular evolution.

Subjects Evolutionary Studies, Genetics, Genomics, Zoology

Keywords Microsatellites, Heterozygosity, Evolution, Mutation rate, Birds

INTRODUCTION

It has long been recognized that the molecular evolutionary rates always covary with many life history traits of animals. Numerous studies have documented a negative relationship between the rate of molecular evolution and body mass (*Nabholz, Glémin & Galtier, 2008; Bromham, 2011; Amos & Filipe, 2014*), where genes in small-bodied species are likely to evolve faster than those in large-bodied species. This has been readily explained by higher number of mutation opportunities per unit time (generation length hypothesis, *Li et al., 1996*) or higher mutation probability in a round of DNA replication due to higher metabolic rate (metabolic rate hypothesis, *Mindell et al., 1996*) in small-bodied species. Although the precise reason for the pattern is not clear at present, to determine the universality of this pattern, we need to study additional form of genetic mutation besides mitochondrial DNA or nuclear 'genes' which are most frequently used. The first to consider is the fastest evolving components of the genome such as microsatellites.

Microsatellites, also known as simple sequence repeats (SSRs), are tandem repeats of simple nucleotide motifs, which have wide coverage in eukaryotic and prokaryotic genomes (*Tóth, Gáspári & Jurka, 2000; Ellegren, 2004; Adams et al., 2016*). One feature of

Submitted 15 November 2017

Accepted 21 February 2018

Published 14 March 2018

Corresponding author

Haiying Fan,
fanhaiying1989@126.com

Academic editor

William Amos

Additional Information and
Declarations can be found on
page 9

DOI 10.7717/peerj.4495

© Copyright
2018 Fan and Guo

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

microsatellites is that they have a high mutation rate (10^{-7} to 10^{-3} mutations per locus per generation), leading to high heterozygosity and extensive length polymorphisms (*Kruglyak et al., 2000*). It has long been assumed that the major cause of variation of microsatellite repeats is replication slippage (*Kornberg et al., 1964; Bhargava & Fuentes, 2010*), which will increase or decrease repeat copy numbers in microsatellites. Specifically, when it creates a loop in one of strands, a slippage error occurs. If the loop is formed in the replicating strand, it will introduce an insertion. If the loop is in the template strand, a deletion will emerge. Several mathematical models of microsatellite evolution have been proposed to represent the mutation processes of microsatellites, such as stepwise mutation model (SMM) of *Ohta & Kimura (1973)*, which suggests that mutation in microsatellite loci occurs by one repeat unit at a time.

Many studies on microsatellites have explored the frequencies, abundance and polymorphism of microsatellites in the genomes (*Wang et al., 2014; Qi et al., 2015; Adams et al., 2016*). Few, if any, have correlated these microsatellite characters to the life history traits of a species. Specifically, microsatellites are hypothesized to experience a life cycle: start short (birth) and expand predictably due to mutation bias (expansion) until they become unstable and either collapse or degrade through internal point mutations (contraction and death) (*Chambers & MacAvoy, 2000; Buschiazzo & Gemmell, 2006*). Life history traits of species are expected to have an influence on the life cycle —‘birth’, expansion, and ‘death’—of microsatellites in the genome (*Amos & Filipe, 2014*). For example, in smaller species, higher mutation rate allows the ‘birth’ and expansion of microsatellites faster, due to higher mutation rate and slippage rate. Since the death rate is lower than the birth rate, microsatellites tend to accumulate in the genome (*Buschiazzo & Gemmell, 2006*). In that, the smaller species harbour a higher frequency of microsatellites across the genome, which has been proved in mammals (*Amos & Filipe, 2014*).

It is well known that except for repeat copy number variation, a microsatellite (e.g., ATATATATAT) also suffers from nucleotide substitutions and insertion/deletion mutations, hence becoming imperfect (e.g., ATATATCATAT: AT repeat with an insertion of C). Perfect and imperfect microsatellites are thus defined. It has been found that genomes possess a relatively small but significant number of imperfect microsatellites (*Brinkmann et al., 1998*). Mismatch variation of imperfect microsatellites is critical for their maintenance in the genome and imperfect microsatellites are more stable compared to perfect microsatellites since the former is less prone to slippage mutation (*Sturzeneker et al., 1998*). Several previous studies have already revealed the genome-wide motif imperfection pattern among species (e.g., *Behura & Severson, 2015*). Nevertheless, our understanding of motif mismatches in imperfect microsatellites is still very limited and their correlation with life history traits remains to be revealed and explained.

In this study, we used 65 avian genome sequences, employing SciRoKo (*Kofler, Schlötterer & Lelley, 2007*) to search SSRs in the whole genome. We chose avian genomes for this study because microsatellites have been widely used in population genetics of bird species, yet the pattern of microsatellites mismatches in birds is still not well understood, mostly owing to the lack of avian genomic information. With the advance of whole genome sequencing, evolution of microsatellites is attracting attention from researchers. With the

genome-wide microsatellites data in hand, we presented the first detailed comparative study of microsatellites, aiming to reveal the patterns of motif mismatches across different bird species and to help understand the relationship with life history traits.

MATERIALS & METHODS

Genome sequences and body mass

We downloaded FASTA files of the 65 avian genomes from NCBI and GigaDB (<http://dx.doi.org/10.5524/101000>). These avian species represent nearly all of the major clades of living birds. We compiled data from the original and secondary references and the world-wide web about the mean body mass of adult males and females (Table S1). If a mean value was not provided for a species, we took the median of the range. Where separate body masses were given for males and females, the average value of the masses was calculated.

Identification of microsatellites

We searched microsatellites in each genome sequences using SciRoKo 3.4, a simple sequence repeats (SSRs) identification program (Kofler, Schlötterer & Lelley, 2007), with the default parameters (minimum score = 15 and mismatch penalty = 5) in the mismatched modes. In addition, we used different parameters to search SSRs (minimum score = 15 and mismatch penalty = 3, minimum score = 10 and mismatch penalty = 5) considering changing in parameters would affect the results of this study. Specially, the motif mismatches refer to the number of base mismatches of an imperfect microsatellite compared with its idealized perfect counterpart. For example, the string TACTACTAGTACTAC, is a trinucleotide repeat with five repeats and, by comparison with its idealized perfect counterpart (consensus repeat), it has a mismatch of 1. The number of mismatches of each microsatellites as well as their length for each genome was used for different comparative analyses across the species.

Statistical analyses

In this study, we used phylogenetic generalized least-squares regression (PGLS) (Freckleton, Harvey & Pagel, 2002) implemented in the package 'ape' (Paradis, Claude & Strimmer, 2004) to control shared ancestry (for the script used, see Fig. S1). We used the evolutionary tree of the 48 bird species estimated by Jarvis *et al.* (2014) as a backbone topology, and used the phylogenetic information provided by Jetz *et al.* (2012) to add the remaining 17 species (for the resulting phylogeny, see Fig. S2). In order to achieve the statistical requirements for linearity and normality, adult average mass were log₁₀-transformed prior to analysis. Average mismatches was reciprocal transformed. GC content was arcsine square root transformed.

Firstly, we computed some basic statistics on characteristics of microsatellite loci in 65 bird genomes (Tables S2–S5). Secondly, to better understand the occurrence of motif mismatches in bird genomes, we determined the frequency of microsatellites of 20 bp that either lacks mismatches or harbours one mismatch for each species (Table S6). 20 bp was used because that the average length of perfect microsatellites of 65 birds is 20 bp. Then we computed the ratio of imperfect (mismatch = 1) repeats frequency to perfect (mismatch

= 0) ones. Then, we employed a PGLS, treating the ratio of imperfect (mismatch = 1) to perfect (mismatch = 0) repeats as a dependent variable and body mass as an explanatory variable. Thirdly, we explored whether or not the extent of motif mismatches is related to genomic abundance of imperfect microsatellites. We first calculated the probability of per-site mismatches (the total number of mismatches divided by total lengths of all loci) in each genomes. Then the expected number of mismatches was determined based on the length and compared with the observed number of mismatches in each imperfect microsatellite (Table S7). The first paired sample *t*-test was conducted between the numbers of microsatellites harbouring more mismatches than expected and that of carrying fewer mismatches than expected. The second paired sample *t*-test was performed between imperfect repeats that have a length of at least 30 bp and have either <3 or ≥3 mismatches (Table S8). Finally, to test whether differences of mismatches in imperfect microsatellites link with body mass, we fitted a PGLS analysis with average mismatches of imperfect SSRs as dependent variable and body mass as a predictor. The average mismatches of imperfect SSRs in individual genomes was estimated as the sum of mismatches divided by the number of imperfect microsatellites (Table S1). Average mismatches was used because it indicates the mismatches in an ‘average’ imperfect SSR. For controlling the probability that GC content will have a potential influence on microsatellite mismatches, we added it to the models as a predictor variable. Taking di-, tri-, tetra-, penta- and hexamers as the five classes of repeats, we repeated the PGLS analysis in each repeat type. Since the mutations in the mononucleotide repeats tend to cause the emergence of a new motif of other repeat type, we excluded it from our analysis. All statistical analyses were conducted with R 3.1.2 (R Core Team, 2014).

RESULTS

Characteristic of microsatellite loci in 65 avian species

In total, 11803896 SSR loci with a minimum length of 15 bp were identified from 65 avian genome assemblies, and were classified into mono-, di-, tri-, tetra-, penta- and hexanucleotide SSRs according to the motif length (Table S2). Among these, mononucleotide SSRs are the most abundant (42.3%) type, followed distantly by tetra- (18.8 %) and pentanucleotide SSRs (17.1%) (Table S2; Fig. S3). The SSR abundance composition and SSR density of the birds varies greatly among species, with the maximum value in *Anas platyrhynchos* (416,040 counts; 376.49 counts/Mb) and the minimum value in *Melopsittacus undulatus* (81,643 counts; 73.07 counts/Mb). Additionally, the SSR abundance composition are predicted by genome size ($\beta \pm SE = 1.56 \pm 0.64$, $t = 2.44$, $P = 0.017$, $R^2 = 0.09$).

Frequency of imperfect microsatellites in bird genomes

The number of imperfect microsatellites varies among the birds species and the imperfect repeats account for 15–27% of all microsatellites searched from the genome assemblies of the 65 bird species as shown in Table S3. The imperfect repeats represented less than 0.2% of the genome sequence in most of these birds except four species (*Anas platyrhynchos*, *Calypte anna*, *Columba livia*, *Picoides pubescens*) (Fig. S4). The data in Table S3 and Fig. S4

shows that the frequency of imperfect microsatellites in bird genomes appears substantial variation among these species. It was observed that *Anas platyrhynchos* has a higher percentage of imperfect microsatellites than other bird species. Moreover, the proportion of imperfect repeats varies differentially among species, to some extent, depending on the motif size of microsatellites. Specifically, the paired sample *t*-test results indicated the di-, tri- and hexanucleotide SSRs have an increased rate of motif mismatches compared with all other types of motif size (Table S4). Furthermore, it seems that this pattern is conserved among different avian species.

The occurrence of motif mismatches

Imperfect microsatellites are longer than perfect microsatellites in each species (37 vs 20 bp; $t = 33.334$, $df = 64$, $P < 0.001$; Table S5). The PGLS analysis revealed that the small-bodied species has a higher ratio of imperfect (mismatch = 1) to perfect (mismatch = 0) repeats of 20bp than large-bodied species ($\beta \pm SE = -0.006 \pm 0.002$, $t = 2.86$, $P = 0.006$, $R^2 = 0.12$).

The accumulation of mismatches in imperfect microsatellites and genomic abundance

The paired sample *t*-test revealed that the microsatellites harboring mismatches higher than expected has significantly lower abundance than that carrying mismatches lower than expected (13,381 versus 25,138 counts; $t = 22.651$, $df = 64$, $P < 0.001$; Table S7), implying that the imperfect microsatellites which containing more mismatches have lower abundance in the genome. We also found that loci with three or more number of mismatches are less common than that have less than three mismatches (7,364 vs 18,361 counts; $t = 11.316$, $df = 64$, $P < 0.001$; Table S8).

Correlation between body mass and average motif mismatches

We found that on a whole genome scale, the average body mass accounts for 28.2% of the variation in average mismatches of imperfect SSRs (Table 1, Fig. 1). Body mass also has a significantly negative correlation with microsatellites mismatches in five motif length classes (Table 1, Fig. 2). This negative correlation remains significant when adding GC content to the regression models. Inclusion of GC content only enhances the model's explanatory power slightly except in tetra- and pentanucleotide SSRs. When we used different parameters including minimum score 15 and mismatch penalty 3 and a minimum score of 10 and mismatch penalty 5 to search microsatellites in the genomes, the results of repeated analyses were highly consistent (Table S9). This confirmed that our observations were not influenced by the search parameters of microsatellites.

DISCUSSION

In the present study, we did a genome-wide search of microsatellites using SciRoKo with the same parameters to ensure that the program can search all possible microsatellites with the same probability for every genome. Microsatellites search results showed that the frequency of microsatellites varies extensively among species. We have also found a positive relationship between microsatellites abundance and genome size among 65 bird species, which is consistent with earlier studies (e.g., Hancock, 1996). After providing a

Table 1 Result for the relationship between average mismatches and body mass fitted in PGLS analyses.

Type	Model	R^2	Coefficients					
			Body mass			GC content		
			$\beta \pm SE$	t	P	$\beta \pm SE$	t	P
All	BM	0.282	0.015 ± 0.003	4.974	<0.001			
	BM + GC	0.340	0.015 ± 0.003	5.102	<0.001	-2.533 ± 1.090	2.324	0.023
Di	BM	0.279	0.022 ± 0.005	4.932	<0.001			
	BM + GC	0.332	0.022 ± 0.004	5.033	<0.001	-2.940 ± 1.255	2.342	0.022
Tri	BM	0.257	0.015 ± 0.003	4.664	<0.001			
	BM + GC	0.293	0.015 ± 0.003	4.711	<0.001	-2.044 ± 1.149	1.778	0.080
Tetra	BM	0.290	0.019 ± 0.004	5.072	<0.001			
	BM + GC	0.393	0.019 ± 0.003	5.382	<0.001	-4.074 ± 1.257	3.241	0.002
Penta	BM	0.123	0.011 ± 0.004	2.972	0.004			
	BM + GC	0.205	0.011 ± 0.004	3.050	0.003	-3.272 ± 1.296	2.525	0.014
Hexa	BM	0.254	0.013 ± 0.003	4.634	<0.001			
	BM + GC	0.268	0.013 ± 0.003	4.620	<0.001	-1.129 ± 1.049	1.076	0.286

Notes.

Key to symbols: All, all imperfect microsatellites; Di, Tri, Tetra, Penta, Hexa, means imperfect microsatellites with different repeat type; BM, Body mass; GC, GC content.

general description of the basic characteristics of microsatellites, we particularly focused on comparing the motif mismatches of imperfect microsatellites to body mass across bird species in a phylogenetic context.

We found a negative relationship between body mass and the ratio of frequency of imperfect repeats (mismatch = 1) to perfect (mismatch = 0) ones with the same length 20bp among the species. Moreover, it is known that mutations in microsatellites shorter than a critical length are generally gain or loss of single repeat units which cannot disturb the repeat tract (*Buschiazzo & Gemmell, 2006*). Whereas when it reached a critical length, mismatch was introduced, a perfect microsatellite became imperfect. Here, our result implied that the introduction of motif mismatches in imperfect microsatellites is significantly associated with the nature of point mutation in microsatellites. In small-bodied species, since more perfect microsatellites suffer from the introduction of mismatches due to the higher mutation rate, a larger number of imperfect microsatellites relative perfect ones can be observed.

We observed that the microsatellites harbouring mismatches higher than expected have lower abundance than that carrying mismatches lower than expected. Consistent with this result, we also found that the microsatellites ≥ 30 bp and < 3 mismatches have lower abundance than that ≥ 30 bp and > 3 mismatches, indicating that mismatches of motifs is a key determinant leading to a paucity of long imperfect microsatellites in the genome. That is to say, mismatches would stabilize the repeat array and impede the further expansion. When the extent of mismatches reached saturation point, the repetition pattern is interrupted, leading the microsatellites to degeneration and death. (*Taylor, Durkin & Breden, 1999; Harr & Schlotterer, 2000; Yamada et al., 2002; Vowles & Amos, 2006*). Although the exact

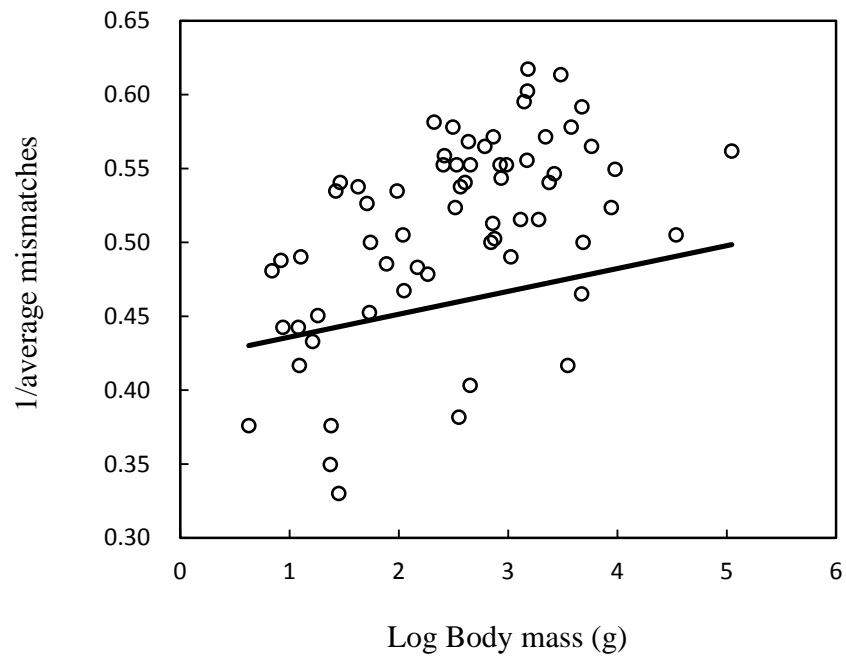


Figure 1 Regression scatterplot of the inverse of the average mismatches of imperfect SSRs on the log of body mass in whole genome scale.

Full-size DOI: 10.7717/peerj.4495/fig-1

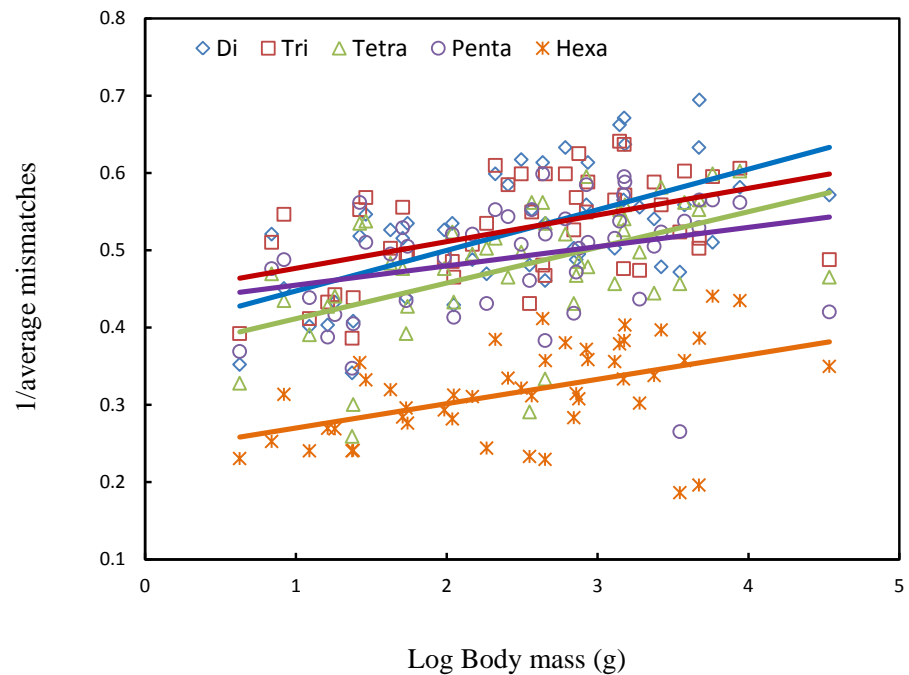


Figure 2 Regression scatterplot of the inverse of the average mismatches of imperfect SSRs in five classes of repeat type on the log of body mass.

Full-size DOI: 10.7717/peerj.4495/fig-2

details of death is still poorly understood, the relative number of older mismatches in an 'average' microsatellite is likely to reflect the mutability during its lifetime. It can be further confirmed by the finding that the average mismatches of imperfect SSRs decreases with increasing body mass.

Our results that higher average mismatches of imperfect SSRs in small-bodied species support a correlation between mutation rate and life history traits. The pattern is usually explained by a generation length model, where smaller species evolve faster due to higher number of mutation opportunities per unit time (*Li et al., 1996*). In addition, body mass might affect the mutation rate through a link with metabolic rate and/or body temperature, which can directly change the mutation probability in a round of DNA replication (*Mindell et al., 1996*). Apart from these two key hypotheses, a rising hypothesis which proposes mutation rates are influenced by heterozygosity (*Amos, 2010*) can better explain the intrinsic correlation of motif mismatches with body mass. Smaller species have larger number of imperfect microsatellites which has been demonstrated by our data ($\beta \pm SE = -0.008 \pm 0.002$, $t = 4.008$, $P < 0.001$; $R^2 = 0.203$). Meanwhile, more heterozygous sites at these imperfect microsatellites can be expected. Recognition and 'repair' of heterozygous sites during synapsis will cause additional rounds of DNA replication which in turn provide more opportunities for mutations (*Amos, 2011*) and introduce more motif mismatches at imperfect microsatellite sites. Therefore, a negative relationship between body mass and motif mismatches can be observed. We suggest that heterozygote instability hypothesis, which is supported by increasing evidence (*Drake, 2007; Masters et al., 2011; Amos, 2013; Amos, 2016*), could provide a potential link between body mass and motif mismatches. However, further studies are needed in order to examine carefully whether homologous imperfect microsatellites are generally more prone to introduce mismatches in smaller species with a detail comparison between sister species.

CONCLUSIONS

In conclusion, the present study is the first effort to explore the motif mismatch patterns of microsatellites in birds on a genomic level. The results we obtained from this study provide support for the long-standing correlation between mutation rate and life history traits and suggest that a negative body mass trend in mutation rate may be a general pattern of avian molecular evolution.

ACKNOWLEDGEMENTS

We thank Xin Lu, Hongtao Xiao, Guoyue Zhang, Changcao Wang, Qingchen Zhang and Juanjuan Rao for data collection, statistical advice and insightful discussions. We also thank William Amos and Andrew Clarke for helpful suggestions and two anonymous referees for comments on earlier versions of this manuscript.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The authors received no funding for this work.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Haiying Fan conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper.
- Weibin Guo performed the experiments, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper.

Data Availability

The following information was supplied regarding data availability:

The raw data is provided in the [Supplemental Files](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.4495#supplemental-information>.

REFERENCES

- Adams RH, Blackmon H, Reyes-Velasco J, Schield DR, Card DC, Andrew AL, Waynwood N, Castoe TA. 2016. Microsatellite landscape evolutionary dynamics across 450 million years of vertebrate genome evolution. *Genome* 59:295–310 DOI 10.1139/gen-2015-0124.
- Amos W. 2010. Heterozygosity and mutation rate: evidence for an interaction and its implications. *BioEssays* 32:82–90 DOI 10.1002/bies.200900108.
- Amos W. 2011. Population-specific links between heterozygosity and the rate of human microsatellite evolution. *Journal of Molecular Evolution* 72:215–221 DOI 10.1007/s00239-010-9423-2.
- Amos W. 2013. Variation in heterozygosity predicts variation in human substitution rates between populations, individuals and genomic regions. *PLOS ONE* 8:e63048 DOI 10.1371/journal.pone.0063048.
- Amos W. 2016. Heterozygosity increases microsatellite mutation rate. *Biology Letters* 12:20150929 DOI 10.1098/rsbl.2015.0929.
- Amos W, Filipe LNS. 2014. Microsatellite frequencies vary with body mass and body temperature in mammals, suggesting correlated variation in mutation rate. *PeerJ* 2:e663 DOI 10.7717/peerj.663.
- Behura SK, Severson DW. 2015. Motif mismatches in microsatellites: insights from genome-wide investigation among 20 insect species. *DNA Research* 22:29–38 DOI 10.1093/dnares/dsu036.

- Bhargava A, Fuentes FF. 2010.** Mutational dynamics of microsatellites. *Molecular Biotechnology* **44**:250–266 DOI [10.1007/s12033-009-9230-4](https://doi.org/10.1007/s12033-009-9230-4).
- Brinkmann B, Klintschar M, Neuhuber F, Huhne J, Rolf B. 1998.** Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *American Journal of Human Genetics* **62**:1408–1415 DOI [10.1086/301869](https://doi.org/10.1086/301869).
- Bromham L. 2011.** The genome as a life-history character: why rate of molecular evolution varies between mammal species. *Philosophical Transactions of the Royal Society B-Biological Sciences* **366**:2503–2513 DOI [10.1098/rstb.2011.0014](https://doi.org/10.1098/rstb.2011.0014).
- Buschiazzo E, Gemmell NJ. 2006.** The rise, fall and renaissance of microsatellites in eukaryotic genomes. *BioEssays* **28**:1040–1050 DOI [10.1002/bies.20470](https://doi.org/10.1002/bies.20470).
- Chambers GK, MacAvoy ES. 2000.** Microsatellites: consensus and controversy. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology* **126**:455–476 DOI [10.1016/S0305-0491\(00\)00233-9](https://doi.org/10.1016/S0305-0491(00)00233-9).
- Drake JW. 2007.** Too many mutants with multiple mutations. *Critical Reviews in Biochemistry and Molecular Biology* **42**:247–258 DOI [10.1080/10409230701495631](https://doi.org/10.1080/10409230701495631).
- Ellegren H. 2004.** Microsatellites: simple sequence with complex evolution. *Genetics* **5**:435–445 DOI [10.1038/nrg1348](https://doi.org/10.1038/nrg1348).
- Freckleton RP, Harvey PH, Pagel M. 2002.** Phylogenetic analysis and comparative data: a test and review of evidence. *American Naturalist* **160**:712–726 DOI [10.1086/343873](https://doi.org/10.1086/343873).
- Hancock JM. 1996.** Simple sequences and the expanding genome. *BioEssays* **18**:421–425 DOI [10.1002/bies.950180512](https://doi.org/10.1002/bies.950180512).
- Harr B, Schlotterer C. 2000.** Long microsatellite alleles in *Drosophila melanogaster* have a downward mutation bias and short persistence times, which cause their genome-wide underrepresentation. *Genetics* **155**:1213–1220.
- Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SYW, Faircloth BC, Nabholz B, Howard JT, Suh A, Weber CC, Da Fonseca RR, Li JW, Zhang F, Li H, Zhou L, Narula N, Liu L, Ganapathy G, Boussau B, Bayzid MS, Zavidovych V, Subramanian S, Gabaldon T, Capella-Gutierrez S, Huerta-Cepas J, Rekepalli B, Munch K, Schierup M, Lindow B, Warren WC, Ray D, Green RE, Bruford MW, Zhan XJ, Dixon A, Li SB, Li N, Huang YH, Derryberry EP, Bertelsen MF, Sheldon FH, Brumfield RT, Mello CV, Lovell PV, Wirthlin M, Schneider MPC, Prosdocimi F, Samaniego JA, Velazquez AMV, Alfaro-Nunez A, Campos PF, Petersen B, Sicheritz-Ponten T, Pas A, Bailey T, Scofield P, Bunce M, Lambert DM, Zhou Q, Perelman P, Driskell AC, Shapiro B, Xiong ZJ, Zeng YL, Liu SP, Li ZY, Liu BH, Wu K, Xiao J, Yinqi X, Zheng QM, Zhang Y, Yang HM, Wang J, Smeds L, Rheindt FE, Braun M, Fjeldsa J, Orlando L, Barker FK, Jonsson KA, Johnson W, Koepfli KP, O'Brien S, Haussler D, Ryder OA, Rahbek C, Willerslev E, Graves GR, Glenn TC, McCormack J, Burt D, Ellegren H, Alstrom P, Edwards SV, Stamatakis A, Mindell DP, Cracraft J, Braun EL, Warnow T, Jun W, Gilbert MTP, Zhang GJ. 2014.** Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**:1320–1331 DOI [10.1126/science.1253451](https://doi.org/10.1126/science.1253451).
- Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO. 2012.** The global diversity of birds in space and time. *Nature* **491**:444–448 DOI [10.1038/nature11631](https://doi.org/10.1038/nature11631).

- Kofler R, Schlötterer C, Lelley T. 2007.** SciRoKo: a new tool for whole genome microsatellite search and investigation. *Bioinformatics* **23**:1683–1685 DOI [10.1093/bioinformatics/btm157](https://doi.org/10.1093/bioinformatics/btm157).
- Kornberg A, Bertsch LL, Jackson JF, Khorana HG. 1964.** Enzymatic synthesis of deoxyribonucleic acid: XVI. Oligonucleotides as templates and the mechanisms of their replication. *Proceedings of the National Academy of Sciences of the United States of America* **51**:315–323 DOI [10.1073/pnas.51.2.315](https://doi.org/10.1073/pnas.51.2.315).
- Kruglyak S, Durrett R, Schug MD, Aquadro CF. 2000.** Distribution and abundance of microsatellites in the yeast genome can be explained by a balance between slippage events and point mutations. *Molecular Biology and Evolution* **17**:1210–1219 DOI [10.1093/oxfordjournals.molbev.a026404](https://doi.org/10.1093/oxfordjournals.molbev.a026404).
- Li WH, Ellsworth DL, Krushkal J, Chang BH, Hewett-Emmett D. 1996.** Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis. *Molecular Phylogenetics and Evolution* **5**:182–187 DOI [10.1006/mpev.1996.0012](https://doi.org/10.1006/mpev.1996.0012).
- Masters BS, Johnson LS, Johnson BGP, Brubaker JL, Sakaluk SK, Thompson CF. 2011.** Evidence for heterozygote instability in microsatellite loci in house wrens. *Biology Letters* **7**:127–130 DOI [10.1098/rsbl.2010.0643](https://doi.org/10.1098/rsbl.2010.0643).
- Mindell DP, Knight A, Baer C, Huddleston CJ. 1996.** Slow rates of molecular evolution in birds and the metabolic rate and body temperature hypotheses. *Molecular Biology and Evolution* **13**:422–426 DOI [10.1093/oxfordjournals.molbev.a025601](https://doi.org/10.1093/oxfordjournals.molbev.a025601).
- Nabholz B, Glémin S, Galtier N. 2008.** Strong variations of mitochondrial mutation rate across mammals—the longevity hypothesis. *Molecular Biology and Evolution* **25**:120–130 DOI [10.1093/molbev/msm248](https://doi.org/10.1093/molbev/msm248).
- Ohta T, Kimura M. 1973.** A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetics Research* **22**:201–204 DOI [10.1017/S0016672308009531](https://doi.org/10.1017/S0016672308009531).
- Paradis E, Claude J, Strimmer K. 2004.** APE: analysis of phylogenetics and evolution in R language. *Bioinformatics* **20**:289–290 DOI [10.1093/bioinformatics/btg412](https://doi.org/10.1093/bioinformatics/btg412).
- Qi WH, Jiang XM, Du LM, Xiao GS, Hu TZ, Yue BS, Quan QM. 2015.** Genome-wide survey and analysis of microsatellite sequences in bovid species. *PLOS ONE* **10**:e0133667 DOI [10.1371/journal.pone.0133667](https://doi.org/10.1371/journal.pone.0133667).
- R Core Team. 2014.** R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Available at <https://www.R-project.org/>.
- Sturzeneker R, Haddad LA, Bevilacqua RAU, Simpson AJG, Pena SDJ. 1998.** Polarity of mutation in tumor-associated microsatellite instability. *Human Genetics* **102**:231–235 DOI [10.1007/s004390050684](https://doi.org/10.1007/s004390050684).
- Taylor JS, Durkin JMH, Breden F. 1999.** The death of a microsatellite: a phylogenetic perspective on microsatellite interruptions. *Molecular Biology and Evolution* **16**:567–572 DOI [10.1093/oxfordjournals.molbev.a026138](https://doi.org/10.1093/oxfordjournals.molbev.a026138).
- Tóth G, Gáspári Z, Jurka J. 2000.** Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Research* **10**:967–981 DOI [10.1101/gr.10.7.967](https://doi.org/10.1101/gr.10.7.967).

- Vowles EJ, Amos W. 2006.** Quantifying ascertainment bias and species-specific length differences in human and chimpanzee microsatellites using genome sequences. *Molecular Biology and Evolution* **23**:598–607 DOI [10.1093/molbev/msj065](https://doi.org/10.1093/molbev/msj065).
- Wang JF, Qi HG, Li L, Zhang GF. 2014.** Genome-wide survey and analysis of microsatellites in the Pacific oyster genome: abundance, distribution, and potential for marker development. *Chinese Journal of Oceanology and Limnology* **32**:8–21 DOI [10.1007/s00343-014-3064-z](https://doi.org/10.1007/s00343-014-3064-z).
- Yamada NA, Smith GA, Castro A, Roques CN, Boyer JC, Farber RA. 2002.** Relative rates of insertion and deletion mutations in dinucleotide repeats of various lengths in mismatch repair proficient mouse and mismatch repair deficient human cells. *Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis* **499**:213–225 DOI [10.1016/S0027-5107\(01\)00282-2](https://doi.org/10.1016/S0027-5107(01)00282-2).