

Discrepancy variation of dinucleotide microsatellite repeats in eukaryotic genomes

HUAN GAO^A, SHENGLI CAI^B, BINLUN YAN^A, BAIYAO CHEN^C and FEI YU^C

^a Jiangsu Key Laboratory of Marine Biotechnology, Huaihai Institute of Technology, Lianyungang 222005, China;

^b College of Fisheries and Life Science, Shanghai Ocean University, Shanghai, 201306, China;

^c Marine Fisheries Research Institute of Lianyungang, Lianyungang 222042, China

ABSTRACT

To address whether there are differences of variation among repeat motif types and among taxonomic groups, we present here an analysis of variation and correlation of dinucleotide microsatellite repeats in eukaryotic genomes. Ten taxonomic groups were compared, those being primates, mammalia (excluding primates and rodentia), rodentia, birds, fish, amphibians and reptiles, insects, molluscs, plants and fungi, respectively. The data used in the analysis is from the literature published in the Journal of Molecular Ecology Notes. Analysis of variation reveals that there are no significant differences between AC and AG repeat motif types.

Moreover, the number of alleles correlates positively with the copy number in both AG and AC repeats. Similar conclusions can be obtained from each taxonomic group. These results strongly suggest that the increase of SSR variation is almost linear with the increase of the copy number of each repeat motif. As well, the results suggest that the variability of SSR in the genomes of low-ranking species seem to be more than that of high-ranking species, excluding primates and fungi.

Key terms: Microsatellites, Eukaryotic genomes, Variation of SSR, Copy number, Meta-analysis.

INTRODUCTION

Microsatellites, also called simple sequence repeats (SSRs) or short tandem repeats (STRs), are a class of repetitive DNA sequences widespread in all of eukaryotic genomes (Tautz & Renz, 1984; Aishwarya, 2007) and most prokaryote genomes (Sreenu et al, 2003). The typical construction of SSRs consists of 1-6 short motifs repeating many times. It is widely accepted that short protomicrosatellites, which arise by chance in the genome, are expanded by replication slippage, a mutation process which is specific to microsatellite DNA (Schlötterer & Harr, 2000). Slippage mutations occur during DNA replication by displacement of the nascent strand, which subsequently realigns out of register. If DNA synthesis continues

on this misplaced DNA molecule, the repeat number of the microsatellite is altered (Tautz & Schlötterer, 1994). SSRs show a high degree of length polymorphism among individuals, and the variability degree in array length can be detected by the number of alleles.

If the above theory of slippage-mutation-recombination is right, we can assume that the abundance of some of the SSR repeat types in genomes should reflect the mutation ability of corresponding repeat types, i.e. the more abundant, the higher is the mutation rate. Additionally, we also conclude that over long macroevolutionary time periods, there must be significant differences in the aspect of the SSR mutation rate among species, particularly between superior creature and inferior creature. In fact, among major taxonomic

groups, microsatellites exhibit considerable variation in composition and allele length, but they also show considerable conservation within many major groups (Nikitina & Nazarenko, 2004; Grover et al, 2007; Gao & Kong, 2005; Edwards et al, 1998; Zhang et al, 2004). This variation may be explained by slow microsatellite evolution, so that all species within a group have similar patterns of variation, or by taxon-specific mutational or selective constraints (Ross et al, 2003). However, comparing and studying microsatellites across species may be problematic because of the biases that may exist among different isolation and analysis protocols and similar studies have so far been limited to some species belonging to the same genus (Ross et al, 2003) or individuals of the same species (Brandström & Ellegren, 2008). Nevertheless, through macro-data analysis and setting controls, the influence of ascertainment bias in SSR mutation comparison across taxonomic groups can be decreased to the maximum degree.

By analyzing the relationships between SSR motifs and their variation, we hope to address the questions of whether the variation of various microsatellite motifs is similar in genomes of different taxa and how the variation of SSR differs among repeat motif types. The results are expected to be useful for understanding the causes and consequences of genome evolution at microsatellite loci.

MATERIALS AND METHODS

Source of data

The microsatellite data (including repeat motifs and their corresponding number of alleles) were obtained from the papers published in the *Journal of Molecular Ecology Notes* (now renamed *Molecular Ecology Resources*) from 2001 to 2007. Mainly, there are two methods for obtaining SSR loci. One of the methods, mainly to screen loci of SSR, is to first establish the SSR-rich genome libraries, then to obtain the original sequences by sequencing the DNA clones, subsequently design the SSR

primer pairs and test the polymorphism of SSR loci using PCR technology among different individuals. The other method is to directly design the SSR primer pairs from known-sequences, such as EST sequences or some other known DNA sequences filled with SSR repeats. By the electrophoresis of PCR products in gels, the number of SSR alleles can be confirmed by calculating the number of amplified bands located at the expected size ranges of PCR amplification products. Hence, the resolving power of gel electrophoresis determines the number of alleles. In the published articles of *Molecular Ecology Notes*, there are three kinds of methods for detecting PCR products. The most frequently used method, and perhaps the most accurate, is that PCR amplification products are separated on an automated DNA sequencer, and alleles are sized with software, such as Gene Profiler (Scanalytics, Fairfax, VA, USA), Genescan software (ABI Prism) and Genotyper (ABI Prism) etc.

The second most frequently used method is that the PCR products are run on 6%-8% denaturing polyacrylamide gels stained with silver nitrate, and the length of the PCR products is estimated with 10-100 bp ladder markers. Additionally, PCR products may be separated by 2%-3% agarose gels, and visualized with UV light after ethidium bromide staining or by autoradiography. Considering the accuracy for the number of alleles, only data obtained from the first method is used to analyse the variation discrepancy of dinucleotide microsatellite repeats in this paper. Moreover, only counts of over 20 individuals were considered as valid data to be used in the statistic analysis.

The information of repeat motifs and their copy number, the number of alleles and their sequence sizes were inputs into the Excel sheets of Microsoft Office according to the classification of taxa.

Statistical criterion

Both perfect and imperfect repeats (only one motif, excluding the compound type) (Webber, 1990) of dinucleotide repeat types were considered to study the variation of

SSR. Considering base mutual partnership and the order difference of recording the first base of repeat units, similar repeat types could be sorted into a single type. For example, for dinucleotide repeat type, AT includes AT and TA; AC includes AC, CA, TG and GT; AG includes AG, GA, TC and CT; CG includes CG and GC respectively (Kong & Gao, 2005).

The taxonomic groups were as follows: primates, mammalia (excluding both primates and rodentia), rodentia, birds, fish, amphibians and reptiles, insects, molluscs, plants and fungi. In Whittaker's classification system (1969), Kingdom Fungi, Kingdom Plantae and Kingdom Animalia are paralleled. In Kingdom Animalia, the highest-ranking taxonomic group is Mammalia, including Primates, Artiodactyls and Rodentia et al, in which primates is the highest ranking evolutionary species. The others are respectively birds, amphibians and reptiles, fish, molluscs and insects in rank-decreasing order.

In order to decrease the error in which extremely low values of the number of alleles in some loci might affect the accuracy of the analysis, the detection of extreme values was carried out for each taxonomic group, using the following methods (Wang & Yang, 2005): after excluding the doubtful value (x_d) of the number of alleles, calculate the δ was calculated,

$$\delta = \frac{\sum_{i=1}^n |d_i|}{n} = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}. \text{ If } d = |x_d - \bar{x}| \geq 2.5\delta,$$

the doubtful value, x_d , would be discarded.

The comparative analysis of mutation slopes between perfect and imperfect SSRs was performed using covariance analysis. Analysis between different repeat motif types and copy numbers was done by correlation analysis. SSR analysis among taxonomic groups was completed by variance analysis. All of the above statistical analyses were performed by SPSS 14.0 software (SPSS Inc., Chicago, IL).

RESULTS

The categorization and component of repeats in total

In total, 4,960 dinucleotide microsatellite repeats were found in 674 organisms, in which the number of insects (144) is the largest, accounting for 21.36% of total organisms, and the others in decreasing order are plants, fish, birds, amphibians and reptiles, rodentia, mammalia, molluscs, fungi and primates (in table 1). The number of perfect and imperfect repeats was 4,212, representing 84.92% of the total of 4,960 dinucleotide repeats.

In dinucleotide repeats, the AC, 2833, accounting for 67.26% of total 4212 dinucleotide repeats, was the most abundant, and the second was AG, 1333, occupying 31.65% of total dinucleotide repeats. However, the number of AT and CG was rare, especially for CG, only 4 (in table 2).

TABLE 1

The distribution and component of dinucleotide repeats among taxonomic groups

Taxa	Total	Pr	Ma	R	B	F	AR	I	Mo	Pl	Fu
No. species	674	18	30	42	73	113	65	144	30	135	24
P(%)		2.67	4.45	6.23	10.83	16.77	9.64	21.36	4.45	20.03	3.56
No. di- in total	4960	224	258	370	408	1008	288	911	160	1206	127
P(%)		4.52	5.20	7.46	8.23	20.32	5.81	18.37	3.23	24.31	2.56
Di- with Single motif	4212	185	231	288	349	882	226	790	132	1014	115
P(%)		4.39	5.48	6.84	8.29	20.94	5.37	18.76	3.13	24.07	2.73

Notes: Pr, Ma, R, B, F, AR, I, Mo, Pl, and Fu is respectively referred to as Primates, Mammalia, Rodentia, Birds, Fishes, Amphibians and Reptiles, Insects, Molluscs, Plants and Fungis, and the same as below; di- is the dinucleotide repeats, and the same as below.

TABLE 2

The distribution characterization of dinucleotide repeat motif types

Taxa		Pr	Ma	R	B	F	AR	I	Mo	Pl	Fu	Total
No. repeats	AC	180	214	250	327	733	169	521	98	289	52	2833
	AG	5	13	38	22	140	56	265	34	709	51	1333
	AT		4			9	1	3		13	12	42
	CG							1		3		4
Total		185	231	288	349	882	226	790	132	1014	115	4212
Criterion of EV (\geq)		15	14	18	17	30	26	23	38	21	11	
No. repeats after deleted the EV		174	223	273	333	830	214	739	126	956	105	3973

Notes: EV is the abbreviation of Extreme Value, and is the unnormal number of alleles.

In each taxonomic group, the extreme value of number of alleles was tested and deleted from the corresponding data set. From table 2, we found the criterion of extreme value is quite different among different taxonomic groups, for example, it is ≥ 11 in the fungi, while it is ≥ 38 in molluscs. Thus, the number of repeats deleted was different in different taxonomic groups, which can be seen in table 2.

The comparison of perfect and imperfect SSR mutation slopes

According to Webber's criterion (Webber, 1990), SSR can be classified into three types: perfect, imperfect and compound. The origin mechanism of compound type SSR is more complicated than the perfect and imperfect types, which involve point mutation based on existing SSR repeats and subsequent slippage mutation (Kong & Gao, 2005). Thus, only perfect and imperfect repeats were considered to compare the variation of SSR across species. Nevertheless, there might be differences between perfect and imperfect repeat types in the aspect of mutation rates, such as AG being similar to AC in another species that may not be dinucleotide identical by descent. If the slopes of the regression equation, in which the independent variable and dependent variable are the copy number of repeat motifs and the number of alleles respectively, are identical for both perfect and imperfect repeat types in one given

taxonomic group, the mutation rates for both repeat types are expected to be identical. Otherwise, perfect and imperfect repeats must be considered to describe the variation character of SSR.

Using SPSS and regarding the copy number as an independent variable, covariance analysis showed that the slope difference between perfect and imperfect repeats were all of no significance ($P > 0.05$) (in table 3). Though no comparison was made in other species because data was insufficient (for example, less than 10 imperfect AC repeats in primates), the above results indicate that the mutation rate for perfect and imperfect repeats is identical. Thus, the data of perfect and imperfect repeats belonging to the same repeats were incorporated to analyze the variation character of dinucleotide repeats in the following text.

Relationships between number of alleles and repeat motif types

Here, only two kinds of repeats, AG and AC, were used to analyze the relationship between repeat motif types and their corresponding variation differences. The number of AT and CG repeats was too low to satisfy the requirement of data analysis. For example, only 10 AT repeats were found in fungi, and limited to only six kinds of copy number classes. The arithmetic mean of the number of alleles was regarded as the number of alleles corresponding to each

copy number (in Table 4). An independent-samples t-test was done to compare the means of number of alleles between AC and AG repeat types. Levene's Test for equality of variances showed that the variances of samples for AC and AG were equal ($F=1.988$, and $P=0.163>0.05$). By means of the two-tailed test, no significant differences were found between the two repeat types ($P=0.091>0.05$). Additionally, the variation comparisons between AC and AG repeats were also done in the following taxonomic groups respectively: fish, amphibian and reptiles, insects, molluscs, plants and fungi. The results showed that there were no significant differences between AC and AG repeats among above taxonomic groups.

Variation analysis among taxonomic groups

The copy number and the corresponding number of alleles in each taxonomic group for AG and AC repeats can be seen in table 5. Regarding taxonomic groups and the number of alleles as the fixed factor and dependent variable, respectively, a one-way ANOVA was completed to study the variation characteristics of AC and AG among taxa. The results show that there are significant differences among taxa for both of AC ($F=8.304$, $P=0.000$) and AG ($F=17.804$, $P=0.000$). In the case that the error of variance of the dependent variable was equal across groups by the Levene's test of equality of error variances, the method of Dunnett's T3 was used to complete the multiple comparisons among different taxonomic groups in both AC and AG repeats. For AC repeats, the comparison results that yielded significant differences are shown in table 6. In

decreasing order, the mean of number of alleles were molluscs, fish, insects, amphibians and reptiles, plants, rodentia, birds, primates, mammalia and fungi. From table 6, we can see that fungi was almost significantly different from the other taxonomic groups, excluding mammalia.

For AG repeats, the significant differences in the results of multiple comparisons are shown in table 7. Here, only six taxonomic groups are compared to each other, i.e. fish, amphibians and reptiles, insects, molluscs, plants and fungi. In decreasing order, the mean of number of alleles were molluscs, fish, plants, insects, amphibian and reptiles, and fungi. As with the AC repeats, fungi was also almost significantly different from the other taxonomic groups, excluding amphibians and reptiles. It was worth noting that the mean of number of alleles in molluscs had the highest value in both AC and AG repeats.

The correlation between the number of alleles and copy number in each taxonomic group

With the help of SPSS, the correlation analysis between the number of alleles and copy number (the original data is in table 5), was performed in order to explore whether there are differences among different organisms. The results are shown in table 8. From table 8 we can see that the number of alleles correlates to the copy number in almost all of taxonomic groups, excluding primates and fungi. However, a significant correlation relationship existed only in molluscs, plants and fungi for AG repeats.

DISCUSSION

The distribution of dinucleotide repeats

The abundance of each dinucleotide repeat motif type, such as AC, AG, AT and CG, is quite different in different eukaryotic genomes. For example, AC is usually the most abundant types in primates, mammalia, rodentia etc (Tóth et al, 2000). But, the number of AG repeats is more than that of AC repeats in plants, or almost equal

TABLE 3

The P value of slope sameness in three taxonomic groups for AC and AG repeats

Repeat motifs	Fishes	Insects	Plants
	P value	P value	P value
AC	0.075	0.085	0.082
AG	0.718	0.888	0.065

TABLE 4
The allele numbers in AC and AG repeats

copy number	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
No. alleles	AC	2.8	4.7	4.2	6.0	5.9	6.7	6.8	7.6	6.7	8.7	8.1	8.7	9.0	8.3	9.7	9.2	9.4	9.0	10.0	11.0	10.1	10.7	12.1	10.4	6.5	12.3	13.7	13.5	13.6	12.6	13.6	11.9	9.5	20.5
	AG	2.3	3.4	4.1	5.9	5.1	6.3	6.6	6.5	7.8	7.2	7.6	9.1	10.1	8.5	8.1	9.3	8.7	9.2	9.8	7.3	10.4	8.1	10.1	8.2	9.4	10.8	9.0	7.0	9.0	9.9	7.2	12.0	7.3	17.6

TABLE 5
The number of alleles corresponding to copy number in each taxonomic group

Copy number	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30		
P	AC				3.3	7.9	4.4	5.3	5.9	4.7	7.5	7.2	6.6	6.2	6.1	6.0	7.0	6.8	7.2	7.4	9.0	5.0	4.0	9.0			
Ma	AC	2.0	2.0	2.0	2.3	6.0	4.8	6.6	4.9	6.8	5.4	5.8	7.1	5.9	6.1	7.2	5.4	5.4	5.1	7.0	9.0	5.8	9.0				
R	AC	6.0	7.0	7.0	5.8	3.9	4.0	8.0	6.1	6.7	6.7	5.6	8.6	4.6	8.1	7.5	8.0	7.8	8.8	9.9	6.0	8.7	9.0	16.0			
B	AC	2.7	2.3	4.2	3.7	4.9	3.5	4.5	7.7	5.6	6.0	8.0	7.1	8.9	7.2	9.6	7.4	9.0	10.3	12.0	8.5	10.0	11.5	5.0	13.0	7.7	
F	AC	6.3	3.4	5.4	7.4	7.2	8.8	9.3	8.7	8.9	11.2	8.8	11.1	11.2	13.0	13.2	14.8	12.7	13.5	12.6	17.2	9.6	15.2	18.0	11.8	6.3	
	AG	6.0	2.0	3.9	5.3	9.2	7.9	9.9	6.8	6.9	6.7	6.0	12.0	13.3	14.7	9.3	8.6	8.0	11.4	12.5	8.3	4.0	6.3	5.6	5.0	23.0	
AR	AC	2.0	3.7	10.7	5.6	6.8	8.6	7.4	9.5	6.5	10.1	9.4	13.0	7.1	5.0	7.5	7.7	12.2	10.0	18.0	5.0	8.0	21.0	12.5	11.0	10.0	
	AG	2.3	3.6	5.0	3.5	6.3	4.3	6.0	11.0	5.5	4.0	5.5	10.8	16.0	15.5	4.0	10.0	10.0	8.0	10.0	6.0	14.8	1.0	7.0	4.0		
I	AC	8.2	5.9	6.9	6.6	5.2	7.7	6.7	8.9	6.9	8.9	9.8	11.7	11.5	7.9	12.7	9.1	12.3	9.0	13.0	10.3	14.8	11.0	13.2	8.5	6.0	
	AG	5.8	11.5	8.9	5.6	6.1	8.3	4.1	4.8	10.6	6.9	4.5	5.3	11.1	5.8	7.0	13.8	8.2	10.8	8.3	13.0	8.5	5.0	4.8	4.4	6.3	
Mo	AC	2.0	6.5	7.8	9.1	5.0	8.0	5.3	5.4	10.2	8.5	11.8	37.0	13.0	28.0	8.8	3.0	10.0	21.0	26.0	26.0	13.0	10.6	24.0			
	AG		9.7	9.3	11.0	14.3	17.7	11.0	11.0	36.0	17.0	24.0	33.0	24.5	16.0												
PI	AC	3.3	4.6	5.3	4.1	5.4	4.9	6.6	5.5	6.8	7.4	7.6	6.2	7.8	14.0	12.8	9.2	11.0	13.0	14.0	10.7	10.4	11.5	4.0	19.5		
	AG	3.1	3.8	5.0	4.2	5.8	6.0	5.7	6.8	7.1	7.3	8.3	9.1	8.0	8.3	7.3	8.5	8.9	8.7	9.4	7.1	10.8	10.8	12.3	13.0	9.8	
Fu	AC	2.8	3.7	3.7	2.8	3.0	5.5	5.7	2.9	2.7	4.0	9.0	3.0	3.0	3.0	5.0	7.0										
	AG		4.3	4.0	3.5	4.0	4.7	2.5	5.0	5.8	5.0	4.5	4.5	4.0	3.5	4.3	4.6	4.8	5.0	4.0	10.0	5.0	6.0	10.0	6.5		

Note: The number in the lines is the number of alleles.

TABLE 6
The significant difference among taxa for AC repeats

a	b	Mean Difference (a-b)	Std Error	Sig.
Primates	Fishes	-4.2990(*)	.81067	.000
	Insects	-2.9830(*)	.63011	.001
	Fungi	2.1500(*)	.57176	.031
Mammalia	Fishes	-5.0967(*)	.84790	.000
	Amphibians and Riptles	-3.6047(*)	.93608	.020
	Insects	-3.7807(*)	.67734	.000
Rodentia	Mollusc	-7.5162(*)	1.98288	.035
	Fishes	-3.2414(*)	.89498	.033
Birds	Fungi	3.2076(*)	.68609	.002
	Fishes	-3.4120(*)	.94003	.030
Fishes	Fungi	3.0370(*)	.74390	.009
	Primates	4.2990(*)	.81067	.000
Amphibian and Reptiles	Mammalia	5.0967(*)	.84790	.000
	Rodentia	3.2414(*)	.89498	.033
	Birds	3.4120(*)	.94003	.030
	Fungis	6.4490(*)	.86237	.000
	Mammalia	3.6047(*)	.93608	.020
Insects	Fungis	4.9570(*)	.94921	.000
	Primates	2.9830(*)	.63011	.001
Molluscs	Mammalia	3.7807(*)	.67734	.000
	Fungi	5.1330(*)	.69537	.000
	Mammalia	7.5162(*)	1.98288	.035
Plants	Fungis	8.8685(*)	1.98912	.007
	Fungi	4.3530(*)	.91943	.001

Note: * The mean difference is significant at the .05 level.

TABLE 7
The significant differences among taxa for AG repeats

a	b	Mean Difference (a-b)	Std Error	Sig.
Fishes	Molluscs	-9.5345(*)	2.60091	.030
	Fungi	3.4823(*)	.94535	.012
Amphibian and Reptiles	Molluscs	-10.8385(*)	2.59957	.011
Insects	Molluscs	-10.4625(*)	2.51639	.014
	Fungi	2.5543(*)	.67910	.008
Molluscs	Fishes	9.5345(*)	2.60091	.030
	Amphibian and Reptiles	10.8385(*)	2.59957	.011
	Insects	10.4625(*)	2.51639	.014
	Plants	10.2345(*)	2.50191	.017
	Fungi	13.0167(*)	2.47978	.002
Plants	Molluscs	-10.2345(*)	2.50191	.017
	Fungi	2.7823(*)	.62328	.001
Fungis	Fishes	-3.4823(*)	.94535	.012
	Insects	-2.5543(*)	.67910	.008
	Molluscs	-13.0167(*)	2.47978	.002
	Plants	-2.7823(*)	.62328	.001

Note: * The mean difference is significant at the .05 level.

TABLE 8

The results of correlation analysis between the number of alleles and copy number

	Taxa	P	Ma	Ro	B	Fi	AR	I	Mo	Pl	Fu
AC	Pearson correlation	0.374	0.748**	0.621**	0.790**	0.691**	0.526**	0.564**	0.542**	0.733**	0.332
	Sig.	0.104	0.000	0.002	0.000	0.000	0.007	0.003	0.008	0.000	0.209
AG	Pearson correlation					0.368	0.284	0.011	0.616*	0.910**	0.578**
	Sig.					0.070	0.179	0.959	0.025	0.000	0.004

Note: ** Correlation is significant at the 0.01 level (2-tailed).

to as the AC is in fungi (Tóth et al, 2000). Our statistical data (in table 2) is similar to what the above literature reported. Thus, the data used in our analysis is reliable, because it agrees with the natural distribution characteristics of dinucleotide repeats in eukaryotic genomes.

It is worth noting that AT repeats are usually very abundant in many taxonomic groups, such as human, plants, fungi (Tóth et al, 2000), shrimps (Kong & Gao, 2005), insects (Subirana & Messeguer, 2008; Prasad et al, 2005) etc. However, the number of AT repeats is scarce in the investigated taxonomic groups. The paucity of AT repeats might result from the isolated-SSR methods in which the (AC)_n and (AG)_n probes are used largely to screen the DNA fragment libraries (An et al, 2006; Funk et al, 2006). Nevertheless, the paucity of CG repeats resulted not only from the isolated-SSR methods in which the (CG)_n probes were scarcely used, but also their low contents in most of organism genomes, such as shrimps (Kong & Gao, 2005), fish (Edwards et al, 1998) et al.

The longer the repeat sequences, the higher the mutation rates

Microsatellite mutation rates range from 10⁻⁶ to 10⁻² per generation, and are thus significantly higher than base substitution rates (Schlötterer, 2000). Several potential factors that may contribute to microsatellite mutation rates in the evolutionary dynamics of microsatellites have been suggested: sequence of the repeat motif, repeat number, length of the repeat unit, flanking sequence, interruptions in the

microsatellite, recombination rate, transcription rate, etc (Schlötterer, 2000). As yet, some comparisons among different repeat types about mutation rates have been done, for example, the mutation rate of dinucleotide repeats was higher than that of tetranucleotide repeats (Chakraborty et al. 1997). On the contrary, a recent study showed the mean population mutation rates of microsatellites do not significantly differ in motifs of di-, tri-, and tetra-nucleotide repeats for rice (Gao & Xu, 2008). However, the variability of SSR among repeat motifs, such as between AT and AG repeats, are not yet fully understood among taxonomic groups, although Bachtrog et al's studies (2000) have shown that GT/CA microsatellites have a mutation rate 1.4 times as high as that of TC/AG and 3.0 times as high as that of AT/TA microsatellites from *Drosophila melanogaster*. Our results revealed there were no significant differences between AC and AG repeats in total. Moreover, the differences between AC and AG repeats were not significant in the following investigated taxonomic groups: fish, amphibians and reptiles, insects, molluscs, plants and fungi.

Although no significant differences for the variability of SSR between AG and AC repeat motif types were found, there were prominent positive correlations between copy number and number of alleles for AC (Pearson correlation coefficient is 0.858, and P=0.00) and AG (Pearson correlation coefficient is 0.704 and P=0.00) motifs. Many previous studies, including direct observations from pedigrees (Schlötterer et al, 1998), population surveys (Goldstein &

Clark 1995; Gao & Xu, 2008) and analysis of cloned microsatellites (Wierdl et al. 1997), have suggested that the mutation rate of microsatellite increases with the increase of repeat number. Here, we directly used number of alleles as the scale appraising the SSR variation, and arrived to the same conclusion. The results provided stronger support that the mutation rate of SSR increases with an increase in the copy number of repeat motifs.

Furthermore, the increase in the number of alleles is linear with the copy number of repeats (at least in the range of 5-39 copy numbers), which can be seen in figure 1. The results support the strand-slippage theories that are used to elucidate the genesis and development of SSR (Levinson & Gutman, 1987), since longer repeat sequences would result in more mutations. However, the growth of a microsatellite sequence was not infinite, and the evidence indicates that it would be prevented by the accumulation of base substitutions and point mutations (Kruglyak et al, 1998; Kruglyak et al, 2000). This viewpoint may explain why the correlation is not significant for AC repeats between number of alleles and the copy number in fungi and primates, where fungi belong to an inferior creature and primate belongs to the highest creature in the taxonomic category. For primates, the mutation rates of SSRs have held at a relatively steady level over long evolutionary time periods, however, the shorter repeats in fungi limit rapid mutation of SSR (Schlötterer et al, 2000).

Relationships between variation of SSR and evolution of genomes

Many studies (Grover et al, 2007; Tóth et al, 2000; Mrázek et al, 2007; Katti et al, 2001) have elucidated that the component and distribution of microsatellite repeats is varied in different genomes of organisms. These results reveal that SSR evolution in genomes might be closely related to the evolutionary rank of species. Schlötterer (2000) thought that species with short microsatellites, such as *D. melanogaster*, should have a lower microsatellite mutation rate than species with longer microsatellites. The hypothesis can be supported by the data of fungi in our studies, in which the microsatellite sequences are shorter than in other taxonomic groups. From table 6 and table 7, we also find that fungi is almost significantly different from the other taxonomic groups, excluding the mammalia, for AC repeats, and amphibians and reptiles for AG repeats.

Excluding fungi, it seems that the variability of SSR in the genomes of low-ranking species is more than that of high-ranking species. For example, the mean of the number of alleles for AC repeats in decreasing order were molluscs, fish, insects, amphibians and reptiles, plants, rodentia, birds, primates and mammalia; for AG repeats, they are molluscs, fish, plants, insects and amphibian and reptiles, in turn. An alternative explanation is that the genome sizes of low-ranking species is

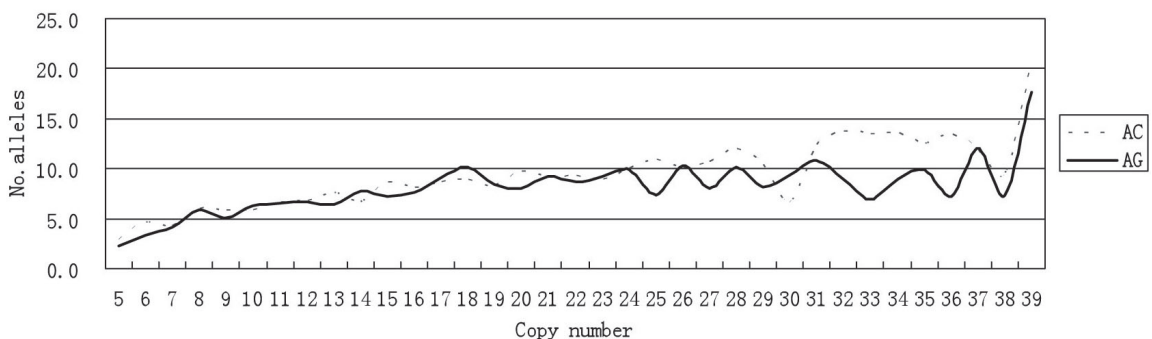


Figure 1: The increase in allele numbers following the increase of copy numbers

usually less than that of high-ranking species, and the length of microsatellites in low-ranking species is thus correspondingly less than that of high-ranking species. According to Lai and Sun's research(2003), when slippage mutations happen, the expansion of short microsatellite repeats occurs more frequently than long repeats, and the contraction of long microsatellite repeats contrarily occurs more frequently than short repeats. Thus, the variability of shorter SSRs in genomes would be more than that of longer SSRs.

Of course, the differences of SSR variation observed among taxonomic groups can result from a number of factors. For example, each taxonomic group includes many species, and there are also differences among related species (Schlötterer & Harr, 2000). Additionally, there are also differences in the aspects of age of the phylum, old or recent diversification, evolution of DNA repair systems, mating systems and so on, among species. A further and detailed study is apparently needed to validate the hypothesis, i.e. excluding fungi, it seems that the variability of SSR in the genomes of low-ranking species is more than that of high-ranking species.

ACKNOWLEDGMENTS

This work was supported by Leading Academic Discipline Project of Shanghai Municipal Education Commission, Project Number J50701 (Marine Biology), Open-end Funds of Jiangsu Key Laboratory of Marine Biotechnology, Huaihai Institute of Technology (2008HS003), and the Aquiculture Project of the Jiangsu Ocean and Fishery Bureau (k2008-6). We thank Dr. Jinling Cai and Jianguo Gao for helping to collect literature.

REFERENCES

AISHWARYA V, GROVER A, SHARMA PC (2007) EuMicroSatdb: a database for microsatellites in the sequenced genomes of eukaryotes. *BMC Genomics* 8: 225
 AN HS, KIM JW, PARK JY (2006) Microsatellite DNA

loci in the rock bream *Oplegnathus fasciatus*. *Molecular ecology notes* 6(1): 44-46
 BACHTROG D, AGIS M, IMHOF M, SCHLÖTTERER C (2000) Microsatellite variability differs between dinucleotide repeat motifs-evidence from *Drosophila melanogaster*. *Mol Biol Evol* 17(9): 1277-1285
 BRANDSTRÖM B, ELLEGREN H (2008) Genome-wide analysis of microsatellite polymorphism in chicken circumventing the ascertainment bias. *Genome research* 18(6): 881-887
 CHAKRABORTY R, KIMMEL M, STIVERS DN, DAVISON LJ, DEKA R (1997) Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. *Proc Natl Acad Sci USA* 94(3): 1041-1046
 EDWARDS YJ, ELGAR G, CLARK MS, BISHOP MJ (1998) The identification and characterization of microsatellites in the compact genome of the Japanese pufferfish, *Fugu rubripes*: perspectives in functional and comparative genomic analyses. *J Mol Biol* 278(4): 843-854
 FUNK CR, SCHMID-HEMPEL, R, SCHMID-HEMPEL P (2006) Microsatellite loci for *Bombus* spp. *Molecular ecology notes* 6(1): 83-86
 GAO H, KONG J (2005) The microsatellite and minisatellites in the genome of *Fenneropenaeus chinensis*. *DNA Sequence* 16(6): 426-436
 GAO LZ, XU HY (2008) Comparisons of mutation rate variation at genome-wide microsatellites: evolutionary insights from two cultivated rice and their wild relatives. *BMC Evol Biol* 8: 11
 GOLDSTEIN DB, CLARK AG (1995) Microsatellite variation in North American populations of *Drosophila melanogaster*. *Nucleic Acids Res* 23(19): 3882-3886
 GROVER A, AISHWARYA V, SHARMA PC (2007) Biased distribution of microsatellite motifs in the rice genome. *Mol Genet Genomics* 277(5): 469-480
 KATTI MV, RANJEKAR PK, GUPTA VS (2001) Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Mol Biol Evol* 18(7): 1161-1167
 KONG J, GAO H (2005) Analysis of tandem repeats in the genome of Chinese shrimp *Fenneropenaeus chinensis*. *Chinese Science Bulletin* 50(14): 1462-1469
 KRUGLYAK S, DURRET RT, SCHUG M, AQUADRO CF (1998) Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proc Natl Acad Sci USA* 95(18): 10774-10778
 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. *Mol Bio and Evol* 17(8): 1210-1219
 LAI Y, SUN F (2003) The Relationship Between Microsatellite Slippage Mutation Rate and the Number of Repeat Units. *Molecular Biology Evolution* 20(12): 2123-2131
 LEVINSON G, GUTMAN GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol* 4(3): 203-221
 MRÁZEK J, GUO X, SHAH A (2007) Simple sequence repeats in prokaryotic genomes. *Proc Natl Acad Sci USA*, 104(20): 8472-8477
 NIKITINA TV, NAZARENKO SA (2004) Human microsatellites: mutation and evolution. *Genetika* 40(10): 1301-1318
 PRASAD MD, MUTHULAKSHMI M, MADHU M, ARCHAK S, MITA K, NAGARAJU J (2005) Survey and analysis of microsatellites in the silkworm, *Bombyx mori*: frequency, distribution, mutations,

- marker potential and their conservation in heterologous species. *Genetics* 169(1): 197-214
- ROSS CL, DYER KA, EREZ T, MILLER SJ, JAENIKE J, MARKOW TA (2003) Rapid Divergence of Microsatellite Abundance Among Species of *Drosophila*. *Mol Biol Evol* 20(7): 1143-1157
- SCHLÖTTERER C (2000) Evolutionary dynamics of microsatellite DNA. *Chromosoma* 109(6): 365-371
- SCHLÖTTERER C, HARR B (2000) *Drosophila virilis* has long and highly polymorphic microsatellites. *Mol Biol Evol* 17(11): 1641-1646
- SCHLÖTTERER C, RITTER R, HARR B, BREM G (1998) High mutation rates of a long microsatellite allele in *Drosophila melanogaster* provides evidence for allele specific mutation rates. *Mol Biol Evol* 15(10): 1269-1274
- SREENU VB, ALEVOOR V, NAGARAJU J, NAGARAJARAM HA (2003) MICdb: database of prokaryotic microsatellites. *Nucleic Acids Res* 31(1): 106-108
- SUBIRANA JA, MESSEGUER X (2008) Structural families of genomic microsatellites. *Gene* 408(1-2): 124-132
- TAUTZ D, RENZ M (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Res* 12(10): 4127-4138
- TAUTZ D, SCHLÖTTERER (1994) Simple sequences. *Curr Opin Genet Dev* 4(6): 832-837
- TÓTH G, GÁSPÁRI Z, JURKA J (2000) Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res* 10(7): 967-981
- WANG DJ, YANG J (2005) *Design of Experiment and Statistics*. Beijing: China Agricultural University Press.
- WEBER JL (1990) Informativeness of Human(dC-dA)_n-(dG-dT)_n Polymorphisms. *Genomics* 7(4): 524-530
- WHITTAKER RH (1969) New concepts of kingdoms or organisms. Evolutionary relations are better represented by new classifications than by the traditional two kingdoms. *Science* 163(863): 150-160
- WIERDL M, DOMINSKA M, PETES TD (1997) Microsatellite instability in yeast: dependence on the length of the microsatellite. *Genetics* 146(3): 769-779
- ZHANG L, YUAN D, YU S, LI Z, CAO Y, MIAO Z, QIAN H, TANG K (2004) Preference of simple sequence repeats in coding and non-coding regions of *Arabidopsis thaliana*. *Bioinformatics* 20(7): 1081-1086

