Theor Appl Genet (2004) 108:550–557
DOI 10.1007/s00122-003-1458-y

ORIGINAL PAPER

S. M. Udupa · R. S. Malhotra · M. Baum

Tightly linked di- and tri-nucleotide microsatellites do not evolve in complete independence: evidence from linked $(TA)_n$ and $(TAA)_n$ microsatellites of chickpea (*Cicer arietinum* L.)

Received: 24 June 2003 / Accepted: 13 August 2003 / Published online: 16 October 2003

Abstract In order to understand the dynamics of microsatellite evolution, we have studied allelic variation at a closely linked $(TA)_n$ and $(TAA)_n$ microsatellite loci in 114 land races of chickpea (*Cicer arietinum* L.), sampled worldwide. These two loci are separated by 27 bp. The two loci showed a very high degree of polymorphism and hence the combined length with the genetic diversity of 0.93, 0.90 and 0.98 for $(TAA)_n$, $(TA)_n$ and the combined length, respectively. Using the variation data at the linked loci, a standardized index of linkage disequilibrium was also computed (I^S_A=0.092), which tests the null hypothesis of no linkage and was significant, indicating the presence of linkage disequilibrium. Furthermore, the dynamics of allelic variation showed that there is a threshold combined length, below which both $(TAA)_n$ and $(TA)_n$ loci evolve independently, and above which, if one locus increase in size, the other closely linked locus has a tendency to decrease its size and vice versa, without change in the overall ratio of $(TAA)_n$ and $(TA)_n$ allele sizes at the region. This result indicates that there are processes in the cell, which 'read' the combined size of the two loci both for proportion and length and determine the direction of tightly linked di- and tri-nucleotide repeat evolution.

Introduction

Microsatellite loci are short tandemly repeated motifs of 1–5 bases, and form a component of higher organism nuclear DNA and are hypervariable in length (Goodfellow 1992). Thus, microsatellites have become extremely popular molecular markers for their application in phylogenetic analysis (Struss and Plieske 1998; Alvarez et al. 2001; Matsuoka et al. 2002; Russell et al. 2003) and molecular mapping (Gupta et al. 1999; Baum et al. 2000; Harker et al. 2001; Udupa and Baum 2003) in various crop plants. The main mutational mechanism leading to variations in microsatellite length is replication slippage (Levinson and Gutman 1987; Tautz and Schlötterer 1994; Udupa et al. 1999; Udupa and Baum 2001). During replication of a microsatellite region, DNA strands may dissociate and then re-associate incorrectly. Renewed replication in this misaligned state leads to insertion or deletion of repeat units, thus altering allele length. Most of the observed variations in microsatellite length are by the ±1 repeat unit (Kruglyak et al. 1998; Udupa and Baum 2001).

For this reason, the stepwise mutation model (SMM) has often been used to model microsatellite evolution (Valdes et al. 1993; Zhivotovsky and Feldman 1995; Kimmel and Chakraborty 1996). In this model, the length of a microsatellite varies at a fixed rate independent of length and according to a symmetric random walk on the positive integers (Kruglyak et al. 1998). The problem with this model for the study of microsatellite length evolution is that a symmetric random walk does not converge to a stationary distribution, and the microsatellite locus is expected to attain arbitrarily high length (Kruglyak et al. 1998). However, Moran (1975) observed that the SMM would not predict a stationary distribution of lengths, but that the variance in allele length within a population of fixed size would stabilize. Some variants of the SMM address this problem by considering the constraint on microsatellite allele length that restricts microsatellite variation to an upper-length boundary (Bowcock et al. 1994; Feldman et al. 1997).

Although the upper ceiling of microsatellite length simplified the computations, which led to a stationary distribution of microsatellite lengths, it is not well understood how and why an upper length ceiling is seen at microsatellites. Dermitzakis et al. (1998) studied the pattern of variation at closely linked microsatellites $(GT)_n$ that are separated by 75 bp of the unique sequence in the fish, *Sparus aurata*, and provided some useful insight into

Communicated by R. Hagemann

S. M. Udupa · R. S. Malhotra · M. Baum (✉)
International Center for Agricultural Research
in the Dry Areas (ICARDA),
P.O. Box 5466, Aleppo, Syria
e-mail: M.Baum@cgiar.org
Fax: +963-21-2213490

the processes that may constrain microsatellite length variation. However, there is no information available yet regarding the pattern of variation at closely linked di- and tri-nucleotide microsatellites in both plant and animal genomes.

Chickpea (*Cicer arietinum* L.), an important world food legume crop, is self-pollinating (Malhotra and Singh 1986) and has 2n=16 chromosomes (Singh 1987). While generating sequence-tagged microsatellite site markers for this crop (Hüttel et al. 1999; Winter et al. 1999), we isolated a genomic clone with $(TAA)_n$ and $(TA)_n$ microsatellite arrays that are separated by only 27 bp of the unique sequence. Here, we report the allelic variation at those loci in a sample of 114 chickpea landraces originating from different parts of the world, and show that tightly linked tri- and di-nucleotide repeats (microsatellites) do not evolve in complete independence. There is a threshold size for the combined loci of the linked tri- and di- nucleotide repeat loci below which they evolve independently, and above which an increase in length of one locus had a negative effect on the growth of the other and vice versa, without any significant change in their length ratios.

Materials and methods

Plant materials

A set of 114 landraces (traditional cultivars not subjected to scientific selection) of cultivated chickpea including a cultivar ILC 3279, from which the microsatellites were isolated, were randomly selected for this study (Table 1). The seed materials were provided by the Genetic Resources Unit, ICARDA, Aleppo, Syria. These accessions were originating from 26 countries and their country-wise representation in this sample represent more or less their country representation in the Germplasm Bank. The accession numbers of the cultivars were designated by ICARDA. Details regarding collection sites and agronomic traits of the accessions are presented in Singh et al. (1991).

Molecular methods

The locus Ta42 is located on linkage group-5 of the chickpea genome (Udupa and Baum 2003). It has $(TAA)_n$ and $(TA)_n$ microsatellite arrays separated by 27 bp that are designated Ta42(TAA) and Ta42(TA), respectively. The combined region is called Ta42(TAA+TA). Primers were designed to amplify each locus separately and both together (Fig. 1). For the amplification of the Ta42(TAA) locus, primers OligoTa42(TAA)-F and OligoTa42(TAA)-R were used; for the Ta42(TA) locus, primers OligoTa42(TA)-F and OligoTa42(TA)-R; and for both loci, primers OligoTa42(TAA)-F and OligoTa42(TA)-R.

Total DNA was extracted from lyophilized leaf tissue of a single plant of each accession by a slightly modified CTAB protocol (Udupa et al. 1998, 1999) and used as a template for PCR amplification. The PCR amplification was performed in a Perkin-Elmer 9600 thermocycler in a 10-µl-volume reaction mix with 100 µM of each deoxynucleotide, 1 × PCR buffer, 15 pmol each of the primers, and 0.2 U of *Taq* DNA polymerase (Roche, Germany) and 50 ng of template DNA. The PCR cycles were programmed for 2 min initial denaturation at 94°C, followed by 35 cycles of each 20 s at 94°C, 50 s at 55°C and at 50 s for 60°C; and followed with a final extension of 5 min at 60°C. The PCR-amplified fragments were detected on an automated laser fluorescence sequencer (ALFexpress; Pharmacia Biotech). To allow this, one primer of each pair was labeled at the 5′ end with fluorescein (cy5). The gel electrophoresis and allele size determination were done as described by Udupa et al. (1999). The number of $(TAA)_n$ or $(TA)_n$ repeats within the microsatellite of all the accessions at the Ta42 locus were calculated based on the sequencing data of the 'ILC 3279' allele (Fig. 1), assuming that size differences of the

Table 1 The chickpea accessions

No.	Accession #	Origin	No.	Accession #	Origin
1	ILC 1317	Afghanistan	58	ILC 21	Jordan
2	ILC 1465	Afghanistan	59	ILC 24	Jordan
3	ILC 1523	Afghanistan	60	ILC 4	Jordan
4	ILC 1575	Afghanistan	61	ILC 4346	Jordan
5	ILC 1596	Afghanistan	62	ILC 7	Jordan
6	ILC 1635	Afghanistan	63	ILC 1782	Lebanon
7	ILC 1669	Afghanistan	64	ILC 1783	Lebanon
8	ILC 221	Afghanistan	65	ILC 1930	Lebanon
9	ILC 240	Afghanistan	66	ILC 563	Lebanon
10	ILC 604	Algeria	67	ILC 570	Lebanon
11	ILC 2402	Bulgaria	68	ILC 573	Lebanon
12	ILC 2657	Bulgaria	69	ILC 578	Lebanon
13	ILC 3869	Bulgaria	70	ILC 583	Lebanon
14	ILC 3709	Chile	71	ILC 3715	Mexico
15	ILC 4022	Chile	72	ILC 3743	Mexico
16	ILC 2364	Cyprus	73	ILC 6022	Mexico
17	ILC 5569	Cyprus	74	ILC 3823	Morocco
18	ILC 5580	Cyprus	75	ILC 3835	Morocco
19	ILC 546	Egypt	76	ILC 3836	Morocco
20	ILC 560	Egypt	77	ILC 219	Pakistan
21	ILC 2556	Ethiopia	78	ILC 220	Pakistan
22	ILC 3245	Ethiopia	79	ILC 5637	Pakistan
23	ILC 3767	Ethiopia	80	ILC 5649	Pakistan
24	ILC 195	Former USSR	81	ILC 5652	Pakistan
25	ILC 197	Former USSR	82	ILC 2299	Palestine
26	ILC 200	Former USSR	83	ILC 2300	Palestine
27	ILC 2506	Former USSR	84	ILC 241	Palestine
28	ILC 2665	Former USSR	85	ILC 2551	Portugal
29	ILC 2954	Former USSR	86	ILC 5351	Portugal
30	ILC 3279	Former USSR	87	ILC 127	Spain
31	ILC 5587	France	88	ILC 3423	Spain
32	ILC 5588	France	89	ILC 3517	Spain
33	ILC 5589	France	90	ILC 3519	Spain
34	ILC 152	Greece	91	ILC 33	Syria
35	ILC 2608	Greece	92	ILC 3312	Syria
36	ILC 1963	India	93	ILC 41	Syria
37	ILC 211	India	94	ILC 43	Syria
38	ILC 212	India	95	ILC 158	Tunisia
39	ILC 2413	India	96	ILC 161	Tunisia
40	ILC 584	India	97	ILC 1798	Tunisia
41	ILC 2350	Iran	98	ILC 1272	Turkey
42	ILC 286	Iran	99	ILC 1931	Turkey
43	ILC 322	Iran	100	ILC 3560	Turkey
44	ILC 336	Iran	101	ILC 449	Turkey
45	ILC 50	Iraq	102	ILC 460	Turkey
46	ILC 54	Iraq	103	ILC 470	Turkey
47	ILC 59	Iraq	104	ILC 483	Turkey
48	ILC 67	Iraq	105	ILC 491	Turkey
49	ILC 3401	Italy	106	ILC 500	Turkey
50	ILC 3406	Italy	107	ILC 502	Turkey
51	ILC 3555	Italy	108	ILC 503	Turkey
52	ILC 1	Jordan	109	ILC 512	Turkey
53	ILC 10	Jordan	110	ILC 513	Turkey
54	ILC 13	Jordan	111	ILC 518	Turkey
55	ILC 1789	Jordan	112	ILC 2514	USA
56	ILC 1932	Jordan	113	ILC 2517	USA
57	ILC 1793	Jordan	114	ILC 2518	USA

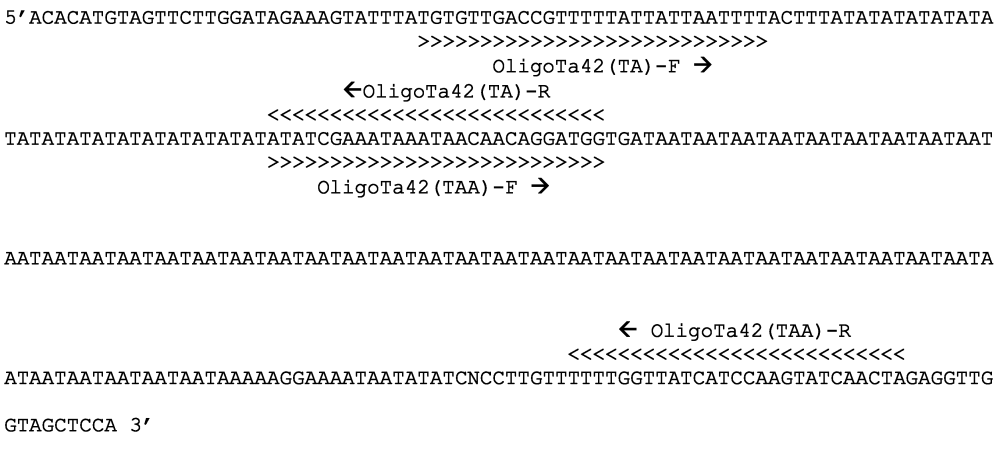

Fig. 1 Sequence of the Ta42 locus. The oligonucleotide primers used are indicated for loci Ta42(TAA) and Ta42(TA)

fragments are caused exclusively by alterations of the repeat number of the appropriate microsatellite sequence.

Data analysis

The estimate of genetic diversity (expected heterozygosity, *H*) was calculated according to Nei (1987): $n(1-\Sigma p^2)/(n-1)$, where *n* is the number of samples and *p* is the frequency of an allele. I^S_A is the "standardized index of association" which assesses the strength of association between loci, which is zero under linkage equilibrium, and was calculated according to Haubold et al. (1998, 2002) using the computer program LIAN 3.0 (Haubold and Hudson 2000). Analysis of variance, correlation and regression analysis were performed using the computer program SYSTAT ver. 7.0 (SPSS Inc.).

To estimate the threshold combined length, the combined allele sizes [Ta42(TAA+TA)] were arranged in increasing order, and a mean of the next ten consecutive allele sizes (sliding window) was worked out for each of the alleles, starting from the lowest allele size (overlapping classes). These mean values of the overlapping classes were plotted against respective correlation coefficients of $(TAA)_n$ vs $(TA)_n$ repeat numbers of the ten combined alleles. The threshold combined length was the point at which correlation coefficient crosses the *P*=0.01 significance level.

Results

Allelic variation at tightly linked Ta42(TA) and Ta42(TAA) microsatellite loci

The selected closely linked microsatellite loci Ta42(TA), with $(TA)_n$ repeats and Ta42(TAA) with $(TAA)_n$ repeats, occurred only once in the genome as indicated by amplification of a single polymorphic fragment in 114 landraces of chickpea (populations) with the tagged flanking primers. Both the microsatellite loci were highly polymorphic. The number of $(TAA)_n$ repeats ranged from 22 to 48, with mean of 35.62, and the number of $(TA)_n$ repeats ranged from 12 to 33, with a mean of 19.13. The average size of the locus Ta42(TAA) was significantly higher than Ta42(TA) (ANOVA, F=2,331.8; df=1, 226; P<0.0001). Other measures of central tendency, namely median and mode, also differed among the linked loci (Table 2).

Table 2 Allelic variation at closely linked $(TAA)_n$ and $(TA)_n$ microsatellite loci of chickpea[a]

Item	Ta42(TAA) locus	Ta42(TA) locus	Ta42(TAA+TA) (Haplotype)
Range			
In bp	72–144 bp	24–66 bp	98–177 bp
In repeat units	22–48	12–33	
Mean			
In bp	106.87	38.26	145.13
In repeat units	35.62	19.13	–
Median			
In bp	108.0	38.0	148.0
In repeat units	36	19	–
Mode			
In bp	197 and 203	92	250
In repeat units	38 and 40	18	
Standard deviation (bp)	13.305	7.284	15.728
Coefficient of variation (%)	12.49	19.03	10.83
Skewness	–0.420	0.701	–0.701
Kurtosis	0.030	1.977	0.823
No. of alleles	21	17	50[b] (71[c])
Genetic diversity	0.93	0.90	0.98[c]

[a] Original data is available upon request
[b] Based on fragment length in bp
[c] Based on the number of repeats

The combined size of the two loci [Ta42(TA+TAA)] ranged from 212 bp to 289 bp with three measures of central tendency, namely, mode (most frequent allele), mean and median, of 250 bp, 257 bp and 260 bp, respectively. Ta42(TA+TAA) is more variable than the individual linked loci as indicated both by the higher standard deviation and the coefficient of variation. The shape of the distribution (skewness and kurtosis) of the Ta42(TAA), Ta42(TA) and Ta42(TA+TAA) loci varied (Table 2; Fig. 2), and none of the distributions can be considered as normal (skewness = 0; kurtosis = 3).

The level of variation at microsatellite loci can be expressed in terms of the number of alleles and the genetic diversity at each locus. The number of alleles observed at each locus varied considerably, 21 in the case

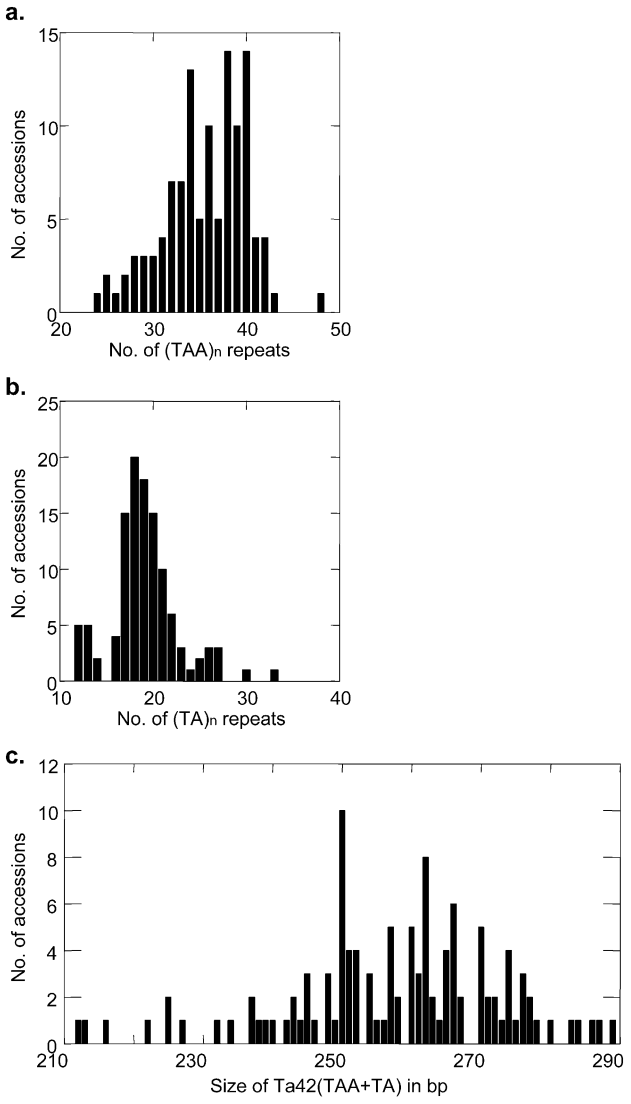

Fig. 2a–c Distributions of allele sizes for loci Ta42(TAA) (**a**), Ta42(TA) (**b**), and for the combined locus Ta42(TAA+TA) (**c**)

of the Ta42(TAA) locus, 17 in the case of the Ta42(TA) locus and 71 in the case of Ta42(TA+TAA). Consequently, the genetic diversity index was also higher for Ta42(TA+TAA) (0.98) compared to its constituent loci [0.93 for Ta42(TAA) and 0.90 for Ta42(TA)].

Linkage disequilibrium

Since the linked loci investigated in our study were separated just 27 bp apart (Fig. 1), we assessed the strength of association between these loci by calculating the standardized index of association, I^S_A, which is zero under linkage equilibrium. For our sample I^S_A=0.092, a value significantly higher than zero ($P<10^{-4}$), which strongly indicates the presence of linkage disequilibrium between the linked loci.

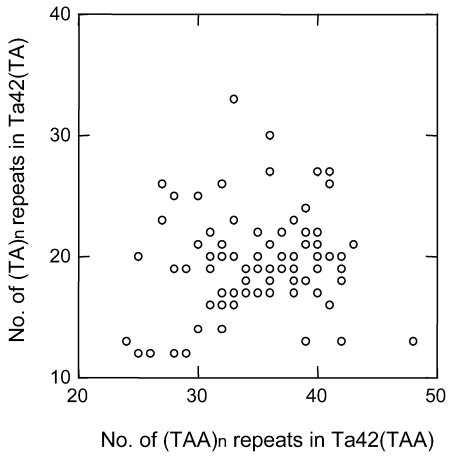

Fig. 3 Scatter plot of allele sizes (number of repeats) of Ta42(TAA) against Ta42(TA)

Relationship between the Ta42(TA) and Ta42(TAA) loci

To study the relationship between the two tightly linked loci, allele sizes of Ta42(TAA) were plotted against the respective allele sizes of Ta42(TA) (Fig. 3) and the correlation coefficient was estimated, which was positive but non-significant (r=0.08; P>0.05), suggesting that variation at one locus does not affect the other and vice versa. However, when the relationship between the individual allele size [(Ta42(TA) or Ta42(TAA)] and the allele size of Ta42(TA+TAA) was tested, the correlation coefficients were positive and highly significant, indicating that each locus contribute significantly for variation at the combined length. The correlation coefficient was higher between the $(TAA)_n$ locus and the combined sizes (r=0.894; P<0.0001) than between the $(TA)_n$ locus and the combined size (r=0.511; P<0.0001), indicating a higher contribution of the $(TAA)_n$ locus for variability of the combined size than the $(TA)_n$ locus.

To further study the dynamics of allelic variation at the linked loci, the data on the combined sizes of the microsatellite loci were arranged in increasing order, and then the data sets were split into two parts at the each point of the mode, mean and median. Altogether, there were six sub-sets of data. In all three cases, the first part contained allele sizes less than the point of interest (i.e. less than mode, mean or median), the second part included the allele size of the point of interest and above. Then, relationships among the two component allele sizes, namely, allele sizes of the $(TAA)_n$ locus and the $(TA)_n$ locus of the combined size of the microsatellites, were studied by estimating correlation coefficients within each of the data sets.

In the first part, in all three cases (i.e. allele sizes of the combined length less than each of the measures of central tendency), the correlation coefficient was very small and non-significant, revealing that the allele sizes of the $(TAA)_n$ locus and the $(TA)_n$ locus vary independently (randomly) at this combined locus range. On the other

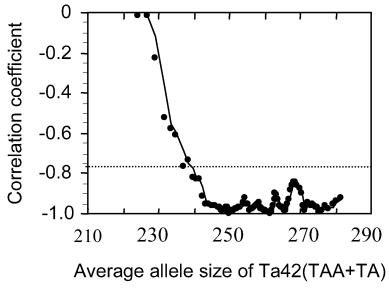

Fig. 4 Plot of the average allele size of Ta42(TAA+TA) vs correlation coefficient. See Materials and methods for the details

part, in all the remaining three cases, where the combined locus size was more than or equal to one of the measures of central tendency, showed strong significant negative correlation coefficients, indicating the allelic relationships at the corresponding $(TAA)_n$ and $(TA)_n$ locus. These results suggests that, greater than or equal to any one of the measures of central tendency for the combined length, the closely linked microsatellite arrays do not evolve independently from each other, and there seems to be a 'threshold' combined length. Above this threshold length, if one locus increases in size, the other closely linked locus has a tendency to decrease its size, and vice versa, to maintain the threshold level.

To estimate the 'threshold' combined length, the combined allele sizes were arranged in increasing order, and a mean of the next ten consecutive allele sizes (the sliding window) was worked out for each of the allele starting from the lowest allele size (overlapping classes). These mean values of the overlapping classes were plotted against respective correlation coefficients of the $(TAA)_n$ vs $(TA)_n$ repeat numbers of the ten combined alleles (Fig. 4). The correlation coefficient of the $(TAA)_n$ vs the $(TA)_n$ repeat numbers for the ten consecutive combined alleles at each of the average allele size was negative but not significant initially, and was decreasing with an increase in the average allele size, and crossed a highly significant level ($P<0.01$) at the average allele size of 239 bp. The point where the allele size crosses a highly significant level ($P=0.01$) of the correlation coefficient of $(TAA)_n$ vs $(TA)_n$ indicates the threshold size of the combined allele. At this point of the threshold average-allele size, the size of the combined allele ranged from 234 bp to 244 bp. This threshold length is lower than the mean, median and mode of the combined allele sizes. Above this threshold size of the combined allele, the correlations coefficients remained highly significant and were very close to −1 as the average allele size crossed 245 bp (Fig. 4).

The slope of the regression line (for the combined length in bp vs the number of repeats) of the $(TAA)_n$ repeats, for below the threshold size of the combined allele sizes, was significantly different (*t*-test value = 2.257; $P=0.026$) from the slope of the regression line of $(TAA)_n$ repeats for above the threshold size of the combined alleles, where as in case of $(TA)_n$ repeats it was non-significant (*t*-test value = 1.654; $P=0.101$) (Fig. 5). The slope of $(TAA)_n$ and $(TA)_n$ repeats was significantly different from each other (*t*-test value = 11.58, $P<0.001$) for the combined alleles above the threshold size; and the difference was non-significant for the combined alleles below the threshold sizes (*t*-test value = 0.489; $P=0.629$).

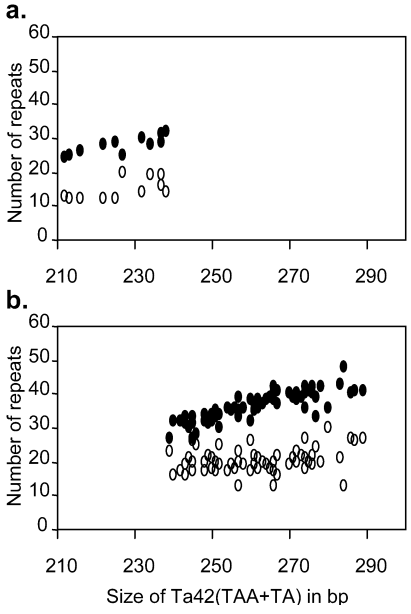

Fig. 5a,b Plot of allele sizes of Ta42(TAA+TA) vs the number of $(TAA)_n$ repeats (*filled circles*) or number of $(TA)_n$ repeats (*open circles*): **a** below the threshold length and **b** above the threshold length of Ta42(TAA+TA) allele-size ranges

Discussion

We have undertaken a systematic analysis of variation at linked $(TAA)_n$ and $(TA)_n$ microsatellites that are separated by 27 bp in 114 accessions of chickpea originating from 26 countries. The average size of the repeat array for $(TAA)_n$ and $(TA)_n$ was 35.62 and 19.13 repeats, respectively. This average size of repeat arrays is much higher than the threshold level of 11–12 repeats observed by Weber (1990), above which variability of the microsatellites increases dramatically.

Our analysis showed that linked $(TAA)_n$ and $(TA)_n$ microsatellites vary greatly in allele size. In general, the level of polymorphism of the linked $(TAA)_n$ and $(TA)_n$ was high, as indicated by a high number of alleles and high genetic diversity. The genetic diversity index and the number of alleles was highest for the combined size [Ta42(TAA+TA)], followed by the Ta42(TAA) and TA42(TA) loci. The high variation observed at the linked $(TAA)_n$ and $(TA)_n$ loci in this study is comparable to the results obtained previously in chickpeas for $(TAA)_n$ microsatellites (Udupa et al. 1999) and for $(TA)_n$ microsatellites (Hüttel et al. 1999).

The standard deviation of allele sizes and the shape of the distribution (skewness and kurtosis) of alleles of the

linked $(TAA)_n$, $(TA)_n$ and the combined length of (TAA+TA) loci also varied. None of the three allelic distributions followed a normal distribution. We categorized the $(TAA)_n$ and $(TA)_n$ alleles in two: less than the mode (most frequent allele size) and more than the mode. In the case of $(TAA)_n$, the distribution of allele sizes were biased towards smaller sizes than the mode, where as in the case of $(TA)_n$, distribution was biased towards more than the most frequent allele size. Population genetics theory predicts that the more frequent allele is usually the ancestral allele (Watterson and Guess 1977); hence, the bias observed, suggest that the predominant direction of mutation for $(TAA)_n$ was a change towards smaller sizes than the mode, and for $(TA)_n$ was a change towards larger sizes than the mode. Further, high variability and non-random distribution of allele sizes at the linked loci indicates that the linked $(TAA)_n$ and $(TA)_n$ microsatellite loci vary by chance, and any difference in allelic variation is due to the loci themselves.

We identified 21 and 17 alleles respectively for the linked $(TAA)_n$ and $(TA)_n$ loci with an average of 19 alleles per locus. If alleles were randomly associated with each other, there would be about 21×17 (=357) possible haplotypes. According to Clark et al. (1998), without recombination or recurrent mutation, the number of haplotypes should be less than the number of alleles. We observed 71 haplotypes (alleles) for the combined size of the loci, which was greater than the average number of alleles at the individual locus; which is an indication that some level of recombination and recurrent mutation occurred within these loci, if we follow Clark et al. (1998).

We examined the variability data for interlocus linkage disequilibrium, which may also enable us to learn more about the biology of recombination and mutations at linked microsatellite loci. Chickpea is 100% self-fertilizing (Malhotra and Singh 1986) and, as a consequence, effective recombination is restricted by the extreme reduction in heterozygosity. Large reductions in effective recombination in a species, in turn, widen the conditions for linkage to influence genetic diversity through the action of selective sweeps or background selection (Lin et al. 2002). These considerations would appear to predict: (1) homogeneity in patterns of genetic diversity, especially between closely linked loci, and (2) extensive linkage disequilibrium relative to cross-pollinated species (Lin et al. 2002).

In contrast to these expectations, the linked $(TAA)_n$ and $(TA)_n$ microsatellite variability data in chickpea reveal a high degree of genetic diversity and significant levels of linkage disequilibrium (I^S_A=0.092) that is nearly two-times higher than from those observed at *adh* loci in the wild barley (I^S_A=0.045, which is indistinguishable from linkage disequilibrium at maize *adh1*, an outbreeding grass species; Lin et al. 2002) and closely linked $(GT)_n$ repeats in the fish *S. aurata* (I^S_A=0.054, an outbreeding species; Dermitzakis et al. 1998); and a much lower (nearly half) linkage disequilibrium compared to the *Arabidopsis thaliana* 170-kb genomic region (I^S_A= 0.179), a selfing species (Haubold et al. 2002), indicating that recombination has played very little role in eroding linkage disequilibrium and generating variability at the linked $(TAA)_n$ and $(TA)_n$ loci.

High mutation rates could be another possible explanation for the observed high diversity at the closely linked microsatellite arrays. The genetic diversity of a population in mutation-drift equilibrium is a function of the product of the effective population size and the mutation rate (Schug et al. 1997). Since chickpea is a self-pollinating crop, its effective population size is expected to be much smaller than that of out-crossing species. Therefore, the high genetic diversity estimated at the $(TAA)_n$ and $(TA)_n$ loci indicates high mutation rates at these loci.

The difference in genetic diversity indicates different mutation rates between the two loci. The difference in mutation rates between the $(TAA)_n$ and $(TA)_n$ loci could be due to difference in the size of the locus (35.62 vs 19.13 repeats), and also due to the different type of the repeat motif. Although direct estimates of mutation rates are available in chickpea only for $(TAA)_n$ microsatellites (Udupa and Baum 2001), it is reasonable to conclude that the main cause for this high degree of variability at the linked $(TAA)_n$ and $(TA)_n$ microsatellites is due to the high rate of mutation at these loci. The high mutation rate has further contributed to lower linkage disequilibrium at the linked loci.

Our results showed that the two linked $(TAA)_n$ and $(TA)_n$ loci evolve independently from each other until they reach a combined threshold length; and above the threshold length, an increase in length of one microsatellite decreases the length of the other, which contradicts the expectation that two linked microsatellite loci evolve independently. The ratio of the number of $(TAA)_n$ repeats to the number of $(TA)_n$ repeats did not differ significantly below [ratio $(TAA)_n/(TA)_n$=1.92] and above [ratio $(TAA)_n/(TA)_n$=1.85] the threshold combined lengths (χ^2=0.0071, *df*=1; *P*=0.93). This kind of observed pattern of variation at the closely linked microsatellite loci could be generated by a number of mechanisms like the genetic bottleneck, natural selection and mutation bias.

The genetic bottleneck may result if a small number of haplotypes survive, whose allele sizes at the two loci show the observed pattern of variation, and if the population analyzed here are derived from those small populations. We ruled out this mechanism for three reasons: first, the samples of the present study are drawn randomly from wide geographical areas; second, the sample populations show wide variation for several phenotypic traits (Singh et al. 1991); and third, our previous work on allelic variation at the $(TAA)_n$ microsatellite loci in a worldwide sample of chickpea revealed a high degree of polymorphism (Udupa et al. 1999).

From our data on the relationships between allele size at the linked microsatellite loci, one can assume that natural selection is acting against the larger allele size of the microsatellite; as was also shown in the case of the

large tandemly repeated trinucleotide sequence in the gene *IT15* related to Huntington's chorea disease (Zhivotovsky et al. 1997) or in the case of the Fragile-X syndrome (Eichler et al. 1994) of humans. Although natural selection cannot be ruled out as a possible explanation, selection acting on the repeat arrays is less likely in our case; as the microsatellite loci used have originated from random genomic libraries and the flanking sequences did not show any significant homology to the known gene sequences in the existing data bases. Another possibility could be that selection at the linked loci could be much weaker than could be detected by the Hardy-Weinberg test, as observed by Dermitzakis et al. (1998) in the SA41 (GT+GT) locus in the fish, *S. aurata*.

Based on the observation that very long alleles (>50 repeat units) are rare in humans (Xu et al. 2000), both mutation rate and mutation biases are the constraints that restrict microsatellite variation to bound intervals (Bowcock et al. 1994; Feldman et al. 1997). This kind of restricted variation at microsatellites might be due to the equilibrium distributions of microsatellite repeat length, which resulted from a balance between point mutations and slippage events (insertion/deletion rates of repeats) at microsatellites (Kruglyak et al. 1998, 2000). A point mutation within a repeat unit interrupts the microsatellite repeat region, creating two shorter repeat regions, which are less mutable than the ancestral microsatellite. The evolutionary coupling of these point mutations (base substitutions) and the repeat expansion in the ancestral perfect microsatellite locus of Ta42 might have resulted in the 'birth' of another microsatellite repeat as observed in humans (Messier et al. 1996). However, with the available data, it is difficult to say which one of the two is an ancestral microsatellite of the Ta42 locus. As further point mutations accumulate in the repeat region, the repeat might have turned into unique, un-repetitive DNA, separating the $(TAA)_n$ and $(TA)_n$ stretches.

Several mutation biases were reported for microsatellites of chickpea (Udupa and Baum 2001), such as an excess of gains over losses of repeats for the short allele on mutation and an excess of loss over gains of repeats for long alleles to contract on mutation; and an increase in mutation rate with allele size, varietal and locus-specific mutation bias for mutation rate and direction might also be applicable to our results for closely linked di- and trinucleotide microsatellite loci.

In this study, we observed that below the threshold combined length [Ta42(TAA+TA)], the rate of change of $(TA)_n$ (slope = 0.22) and $(TAA)_n$ (slope = 0.20) repeats per base of the combined length, was positive and non-significant (*t*-test value = 0.489; P=0.629) and were evolving independently. The average length of $(TAA)_n$ and $(TA)_n$ below the threshold size was 28 and 14.6 repeats respectively. Here, it seems that cellular processes, such as replication slippage favouring independent growth of $(TAA)_n$ and $(TA)_n$, without changing significantly their relative size ratio, might have dominated. On the other hand, above the threshold combined length, the rate of change of $(TAA)_n$ (slope = 0.269) was 2.7-times significantly (*t*-test value = 11.585; P<0.001) higher than $(TA)_n$ (slope =0.095) and related negatively to each other (r=–0.25), with the ratio of the average size of $(TAA)_n$ and $(TA)_n$ remaining at the same level as that below the threshold combined length. At this range, the average size of $(TAA)_n$ and $(TA)_n$ were 36.5 and 19.7 repeats respectively, and are higher than average sizes at the linked loci below the threshold combined-size level. Therefore, mutation rates at both the linked loci are also higher at this level than at the linked loci below the threshold combined length. Furthermore, based on the lower allelic size of the $(TA)_n$ locus, one can predict that lower mutation rates compared to $(TAA)_n$ led to a lower incremental raise of $(TA)_n$ repeats for the unit rise in the combined length. These results indicate that $(TAA)_n$ evolved faster than $(TA)_n$ above the threshold combined length, which is in agreement with our previous observation that high mutation rates and mutation bias was towards the gain of repeats at the Ta42(TAA) loci (Udupa and Baum 2001). As a result, the $(TAA)_n$ array contributes more towards a change in the combined length than $(TA)_n$ and, in general, an increase in the length of the $(TAA)_n$ repeat had decreased $(TA)_n$ repeats, and vice versa, in order to stabilized the Ta42 repeat region. Here, cellular processes that could read the combined size, and the $(TAA)_n$ and $(TA)_n$ repeat ratio, accordingly determine the direction of evolution of linked di- and tri-microsatellites that might have dominated.

Our results on variability, at tightly linked di- and trinucleotide microsatellites, differed from that of two tightly linked di-nucleotide microsatellites in the fish, *S. aurata* (Dermitzakis et al. 1998). We observed the independent evolution of tightly linked $(TAA)_n$ and $(TA)_n$ microsatellites below the threshold sizes of the combined length in chickpea and negative correlation between the allele sizes of linked loci at above the threshold size of combined lengths; whereas in the fish, the linked di-nucleotide repeats showed negative correlation at all the levels. Secondly, the ratio between the sizes of $(TAA)_n$ and $(TA)_n$ repeats were maintained almost the same at all levels of variability, which was not observed at the linked di-nucleotide repeat variation in the fish. These differences clearly show that there is a different mechanism operating at the tightly linked $(TAA)_n$ and $(TA)_n$ repeats in chickpea than that operating at linked di-nucleotide repeats of the fish.

Although we did not rule out the possibility of natural section for the observed pattern of variation in chickpea, circumstantial evidence suggests that mutation bias played a greater role. Further studies towards the isolation of more tightly linked different types of microsatellite loci and their characterization, are required to obtain more information on the pattern of variation at linked microsatellites which enable us to understand the dynamics of microsatellites in plants.

Acknowledgments The authors' research was supported by grants to ICARDA from the German Federal Ministry of Economic Cooperation and Development (BMZ, Bonn, Germany) and the

Arab Fund for Economic and Social Development (AFESD, Kuwait).

References

Alvarez AE, van de Wiel CCM, Smulders MJM, Vosman B (2001) Use of microsatellites to evaluate genetic diversity and species relationships in the genus *Lycopersicon*. Theor Appl Genet 103:1283-1292

Baum M, Weeden NF, Muehlbauer F, Kahl G, Udupa SM, Eujayl I, Weigand F, Harrabi, M, Bouznad Z (2000) Marker technology for plant breeding. In: Knight R (ed) Linking research and marketing opportunities for pulses in the 21st century. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp 421–427

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368:455–457

Clark AG, Weiss KM, Nickerson DA, Taylor SL, Buchanan A, Stengård J, Salomaa V, Vartiainen E, Perola M, Boerwinkle E, Sing CF (1998) Haplotype structure and population genetic inferences from nucleotide-sequence variation in human lipoprotein lipase. Am J Hum Genet 63:595–612

Dermitzakis ET, Clark AG, Batargias C, Magoulas A, Zouros E (1998) Negative covariance suggests mutation bias in a two-locus microsatellite system in the fish *Sparus aurata*. Genetics 150:1567–1575

Eichler EE, Holden JJ, Popovich BW, Reiss AL, Snow K, Thibodeau SN, Richards CS, Ward PA, Nelson DL (1994) Length of uninterrupted CGG repeats determines instability in the *FMR1* gene. Nature Genet 8:88–94

Feldman MW, Bergman A, Pollock DD, Goldstein DB (1997) Microsatellite genetic distances with range constraints: analytic description and problems of estimation. Genetics 145:207–216

Goodfellow PN (1992) Variation is now the theme. Nature 359:777–778

Gupta PK, Varshney RK, Sharma PC, Ramesh B (1999) Molecular markers and their applications in wheat breeding. Plant Breed 118:369–390

Harker N, Rampling LR, Shariflou MR, Hayden MJ, Holton TA, Morell MK, Sharp PJ, Henry RJ, Edwards KJ (2001) Microsatellites as markers for Australian wheat improvement. Aust J Agric Res 52:1121–1130

Haubold B, Hudson RR (2000) LIAN 3.0: detecting disequilibrium in multilocus data. Bioinformatics 16:847–848

Haubold B, Travisano M, Rainey PB, Hudson RR (1998) Detecting linkage disequilibrium in bacterial populations. Genetics 150:1341–1348

Haubold B, Kroymann J, Ratzka A, Mitchell-Olds T, Wiehe T (2002) Recombination and gene conversion in a 170-kb genomic region of *Arabidopsis thaliana*. Genetics 161:1269–1278

Hüttel B, Winter P, Weising K, Choumane W, Weigand F, Kahl G (1999) Sequence-tagged microsatellite-site markers for chickpea (*Cicer arietinum* L.). Genome 42:210–217

Kimmel M, Chakraborty R (1996) Measures of variation at DNA repeat loci under a general stepwise mutation model. Theor Pop Biol 50:345–67

Kruglyak S, Durrett RT, Schug MD, Aquadro CF (1998) Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. Proc Natl Acad Sci USA 95:10774–10778

Kruglyak S, Durrett RT, 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 Biol Evol 17:1210–1219

Levinson G, Gutman GA (1987) Slipped-strand misrepairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4:203–221

Lin J-Z, Morrell PL, Clegg MT (2002) The influence of linkage and inbreeding on patterns of nucleotide sequence diversity at duplicate alcohol dehydrogenase loci in wild Barley (*Hordeum vulgare* ssp. *spontaneum*). Genetics 162:2007–2015

Malhotra RS, Singh KB (1986) Natural cross pollination in chickpea. Int Chickpea Newslett 14:4–5

Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez G J, Buckler E, Doebley J (2002) A single domestication for maize shown by multilocus microsatellite genotyping. Proc Natl Acad Sci USA 99:6080–6084

Messier W, Li S-H, Stewart C-B (1996) The birth of microsatellite. Nature 381:483

Moran PA (1975) Wandering distributions and the electrophoretic profile. Theor Pop Biol 8:318–30

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Russell JR, Booth A, Fuller JD, Baum M, Ceccarelli S, Grando S, Powell W (2003) Patterns of polymorphism detected in the chloroplast and nuclear genomes of barley landraces sampled from Syria and Jordan. Theor Appl Genet DOI 10.1007/s00122-003-1261-9

Schug MD, Mackay TFC, Aquadro CF (1997) Low mutation rates of microsatellites in *Drosophila melanogaster*. Nature Genet 15:99–102

Singh KB (1987) Chickpea breeding, In: Saxena MC, Singh KB (eds) The chickpea. CAB International, Willingford, Oxford, UK, pp 127–162

Singh KB, Holly L, Bejiga G (1991) A catalog of kabuli chickpea germplasm. ICARDA, Aleppo, Syria

Struss D, Plieske J (1998) The use of microsatellite markers for detection of genetic diversity in barley populations. Theor Appl Genet 97:308–315

Tautz D, Schlötterer C (1994) Simple sequences. Curr Opin Genet Dev 4:832–837

Udupa SM, Baum M (2001) High mutation rate and mutational bias at $(TAA)_n$ microsatellite loci of chickpea (*Cicer arietinum* L.). Mol Genet Genomics 265:1097–1103

Udupa SM, Baum M (2003) Genetic dissection of pathotype-specific resistance to ascochyta blight disease in chickpea (*Cicer arietinum* L.) using microsatellite markers. Theor Appl Genet 106:1196–1202

Udupa SM, Weigand F, Saxena MC, Kahl G (1998) Genotyping with RAPD and microsatellite markers resolves pathotype diversity in the ascochyta blight pathogen of chickpea. Theor Appl Genet 97:299–307

Udupa SM, Robertson LD, Weigand F, Baum M, Kahl G (1999) Allelic variation at $(TAA)_n$ microsatellite loci in a world collection of chickpea (*Cicer arietinum* L.) germplasm. Mol Gen Genet 261:354–363

Valdes AM, Slatkin M, Freimer NB (1993) Allelic frequencies at microsatellite loci: the stepwise mutation model revisited. Genetics 133:737–749

Watterson GA, Guess HA (1977) Is the most frequent allele the oldest? Theor Pop Biol 11:141–60

Weber JL (1990) Informativeness of human $(dC\text{-}dA)_n\cdot(dG\text{-}dT)_n$ polymorphisms. Genomics 7:524–530

Winter P, Pfaff T, Udupa SM, Hüttel B, Sharma PC, Sahi S, Arreguin-Espinoza R, Weigand F, Muehlbauer FJ, Kahl G (1999) Characterization and mapping of sequence-tagged microsatellite sites in the chickpea (*Cicer arietinum* L.) genome. Mol Gen Genet 262:90–101

Xu X, Peng M, Fang Z (2000) The direction of microsatellite mutations is dependent upon allele length. Nature Genet 24:396–399

Zhivotovsky LA, Feldman MW (1995) Microsatellite variability and genetic distances. Proc Natl Acad Sci USA 92:11549–11552

Zhivotovsky LA, Feldman MW, Grishechkin SA (1997) Biased mutations and microsatellite variation. Mol Biol Evol 14:926–933