AGRICULTURAL AND FOOD CHEMISTRY

Statistical Correlations of Primer Thermodynamic Stability ΔG° for Enhanced Flax ISSR-PCR Cultivar Authentication

IVO WIESNER* AND DANA WIESNEROVÁ

The Institute of Plant Molecular Biology, Academy of Sciences of the Czech Republic, Branišovská 31, České Budějovice, CZ 370 05, Czech Republic

We revealed four statistically significant correlations related to inter-simple-sequence repeat (ISSR) patterns: (1) between thermodynamic free energy ΔG° of ISSR primer sequence and PCR reamplification intensity (dA_i), (2) between free energy ΔG° of ISSR primer sequence and PIC coefficient quantifying the polymorphism of ISSR patterns, (3) and (4) between free energy ΔG° of anchor sequence of primer and the number of total, and polymorphic bands in ISSR patterns, respectively. Methodological recommendations for effective ISSR primer design were inferred based on revealed correlations. In particular, free energy of ISSR primer sequence is recommended to be $\Delta G^{\circ} > 160 \text{ kJ/mol of interaction and free energy of flanking anchor sequence in primer to be around <math>\Delta G^{\circ} = 28 \text{ kJ/mol of interaction to produce ISSR patterns displaying maximum polymorphism of flax germplasm.}$

KEYWORDS: Cultivar authentication; flax germplasm; inter-microsatellite repeats; ISSR-PCR; thermodynamic free energy ΔG°

INTRODUCTION

Inter-simple-sequence-repeat (ISSR) PCR is a fast and cheap approach to identification of unknown, yet nonsequenced genome and is widely used for ecological, population, or phylogenetic studies (1-3), as well as in crop varietal authentication (4).

Oligonucleotide primers used in ISSR-PCR are usually designed by random playing with sequences of microsatellite core motifs. There is only one respected rule, which is the adjunction of flanking anchor to either the 5' or 3' end, which increases the marker informativeness and improves the quality of resulting ISSR patterns (5) required for effective computer-assisted interpretation.

Here we report about revealed statistically significant correlations between the thermodynamic parameter free energy ΔG° of ISSR primer sequence and parameters of gel ISSR patterns related to band polymorphism. These correlations enabled the formulation of new recommendations for more effective design of ISSR primers used in flax (*Linum usitatissimum* L.) genome identification and varietal authentication.

MATERIAL AND METHODS

Plant Material and DNA Samples. Four model flax cultivars (cv. Venika, cv. Wiko, cv. Biltstar, and cv. Ariane) were obtained from the collection of AGRITEC Research, Breeding, Services Ltd. Šumperk, Czech Republic, and from the collection of Sempra Prague, Breeding Station Slapy u Tábora, Czech Republic.

DNA samples were extracted from the roots of three-days old seedlings germinated on moisten filter paper in the dark box following the modified DNA extraction protocol (6). Two independent DNA samples were analyzed per each plant accession to identify and remove irreproducible bands before each ISSR pattern was transformed into standardized digital form (1/0 presence/absence classification) representative for each accession. For purposes of this analysis, reproducible band was defined as a band invariantly present or absent in both repetitions of gel pattern. Only reproducible bands were then taken into representative pattern of an accession. Any band present in only one of the two repetitions was considered as an instable product of random amplification. Even if a band was found instable only in one accession, it was sufficient reason for its elimination also from standardized digital patterns of all other accessions. Finally, each plant accession was represented by one standardized digital band pattern made up only from stable bands.

ISSR-PCR Analyses. We recently accommodated original protocol of ISSR–PCR (7–9) modified by insertion of reamplification step (rISSR-PCR). The second round PCR was performed using aliquots of the reaction product from the first round as a template. PCR reaction mixture (25 μ L volume) contained 25 μ L of 10× PCR buffer, 4 mM MgCl₂, 2.5 mM dNTP mix, 0.2 μ M primer, 20 ng of template DNA (or 1 μ L reaction mixture from the first round) and 1 U TaKaRa polymerase (TaKaRa Shuzo, Japan). PCR thermocycling was carried out on a Perkin-Elmer 480 thermal cycler using the following profile: 1 cycle of 94 °C for 3 min followed by 35 cycles of 94 °C for 60 s, annealing temperature in the first round and second reamplifying round of ISSR-PCR 55 °C, and 60 °C, respectively, finished by 1 cycle of 72 °C for 3 min. The set of 21 anchored primers were used in rISSR-PCR reactions (**Table 1**).

Amplification products were analyzed in 1.5% agarose ($1 \times TAE$ buffer) with ethidium bromide visualization (8, 9). DNA ladder 100 bp (New England BioLabs, Beverly, MA) was used as DNA size marker.

^{*} To whom correspondence should be addressed. Tel.: +420 38 777 5501. Fax: +420-3853 10 356. E-mail: nfix@jcu.cz.

Table 1. ISSR Primers Targeting Regions of Microsatellite Spacers in Flax^a

no.	name	sequence	$\min \Delta G^{\circ}$	$\operatorname{med}\Delta G^{\circ}$	$\max\Delta G^\circ$	Tm
1	3PCT4	5'-VRV [CT]6	70.29	79.08	93.72	36.3
2	3PCT5	5'-VRV [TG]6	71.55	81.17	88.70	36.3
3	3PCT6	5'-VSS [GATA] ₆	144.35	153.97	158.16	45.9
4	PCT6	5′-[GATA] ₆		112.97 ^b		49.9
5	3PCT1S	5′-YHY [GA]₀	69.87	75.31	93.72	35.0
6	3PCTanS	5'-SSS [GA]6	91.63	93.72	95.81	39.5
7	3PCTanW	5'-WWW [GA] ₆	66.94	71.96	76.57	31.3
8	PCT1an3W	5'-[GA] ₁₅ WWT	189.54	193.72	197.07	55.4
9	PCT1an3S	5'-[GA]15 SSG	204.6	207.94	210.04	59.1
10	PCT2an3W	5'-[CT]15 WWT	187.02	195.39	197.07	55.4
11	PCT2an3S	5'-[CT] ₁₅ SSG	205.85	209.20	211.29	59.1
12	PCT1noan	5'-[GA] ₁₅		194.14 ^b		55.4
13	PCT2noan	5'-[CT] ₁₅		194.14 ^b		55.4
14	PCTAT	5′–N7 [AT]10	102.09	122.59	174.05	38.0
15	PCTAA	5'-N7[A]20	156.48	176.98	228.45	38.0
16	TETRA1	5'-KKB NVS S [ATCT] ₆	159.41	179.07	210.04	49.9
17	TETRA2	5'-KKB NVS S [GGCT]6	264.43	284.93	315.89	65.8
18	TETRA3	5'-KKB NVS S [CTTT] ₆	197.90	216.31	247.27	49.9
19	TETRA4	5'-KKB NVS S [GACA] ₆	182.84	201.25	232.21	57.8
20	TETRA5	5'-KKB NVS S [CTAT] ₆	162.76	181.17	212.13	49.9
21	HEPTA1	5'-KKB NVS S [CCC T AAA] ₄	254.39	272.80	303.76	57.5

^{*a*} ISSR primer sequences are presented as sequence signatures, which specify sequence degeneration on a nucleotide position. Therefore each synthesized ISSR primer is actually a mixture of individual oligonucleotide sequences. Then, min ΔG° (kJ/mol of interaction) stands for the lowest value of the free energy ΔG° in the series of ΔG° s of individual oligonucleotide sequences defined by primer sequence signature with degeneration. med ΔG° (kJ/mol of interaction) stands for the statistical median value of the free energy ΔG° in the series of ΔG° s of oligonucleotide sequences signature with degeneration. med ΔG° (kJ/mol of interaction) stands for the statistical median value of the free energy ΔG° in the series of ΔG° s of oligonucleotide sequences defined by primer sequence signature with degeneration. max ΔG° (kJ/mol of interaction) stands for the highest value of the free energy ΔG° in the series of ΔG° s of oligonucleotide sequences defined by primer sequence signature with degeneration. Tm (°C) stands for the average melting temperature of the mixture of ISSR primer oligonucleotides defined by one sequence signature. ^{*b*} Indicates the presence of only single G° value of ISSR primer when there is no degeneration in sequence signature. There is only a single oligonucleotide synthesized as an ISSR primer instead of a mixture.

ISSR patterns on gels were electronically documented in TIFF format on UVP GDS 5000 system (UVP Ltd., Cambridge, England) using 312 nm UV transilluminator. Detection system was optimized to maximum sensitivity.

Statistical Parameters. The parameter "polymorphism information content" (PIC) was calculated according to Botstein et al. (10) as: PIC_j = $n(1 - \sum P^2_{ij})/(n-1)$ where *n* is the sample size, and P_{ij} the frequency of the *i*-th pattern revealed by the ISSR primer *j* summed across all patterns revealed by the primer *j*.

To quantify the degree of reamplification achieved by individual primers in reamplification—rISSR-PCR protocol, the semiquantitative parameters were further defined from electronic TIFF image of each ISSR pattern. First, the relative amplification intensity (A_i) was calculated as: $A_i = (1 - L_b/L_p)$. 100 [%], where L_b is the total lightness of pixels integrated per area of rectangle S_b (arbitrary units from the interval $\langle 0;255\rangle$); S_b is the rectangular area delineated in TIFF image so that it sampled the background lightness of JSSR pattern, this area did not cover any ISSR band; L_p is the total lightness of pixels integrated per area of rectangle S_p (arbitrary units from the interval $\langle 0;255\rangle$); and S_p is the minimized rectangular area just covering all bands of an ISSR pattern. Always $S_p = S_b$ for correct calculations of A_i .

ISSR-PCR reaction mixtures from both first and second PCR round were separated always on the same gel. Also computer measurement of lightness L_b , and L_p of corresponding ISSR pattern was performed always on one TIFF image. Lightness of pixels (L_b , L_p) integrated per rectangular area S_p , and S_b was measured using UVP Gel Documentation System software (UVP Life Sciences, Cambridge, UK).

The reamplification intensity on patterns of rISSR-PCR was defined as the difference dA_i of A_i taken from gel pattern of the second PCR round with annealing temperature 60 °C (A_{i60}) and A_i taken from gel pattern of the first PCR round with annealing temperature 55 °C (A_{i55}): $dA_i = A_{i60} - A_{i55}$. Statistical analysis was calculated using Statistica ver. 6.0 software package (Statsoft, Tulsa, OK).

RESULTS AND DISCUSSION

ISSR-PCR Analysis. Sequence composition of ISSR primers is known to be an essential factor that influences resulting quality of ISSR fingerprinting pattern (11), but no detailed rules were derived except the effective insertion of anchors on either the 5' or 3' end of a primer (5).

We wanted to reveal other possible rules, if they exist, which would reduce the arbitrariness of ISSR primer design. Therefore, we inspected various theoretically meaningful correlations, and especially those correlations between thermodynamic stability ΔG° of sequence of ISSR primer and parameters quantifying polymorphism of ISSR pattern produced by these primers. Below are discussed only those correlations, which were proven to be statistically significant (12).

We screened a set of 21 ISSR primers for their capacity to produce polymorphisms in flax ISSR patterns. Altogether 213 individual gel bands were generated, from which 43 bands (20.2%) were polymorphic on the genetic background of four flax accessions. These data represented 10.1 bands produced per primer with two polymorphic bands per primer in average. These data were further searched for statistically significant correlations.

Statistical Correlations. Study of thermodynamic nearestneighbor base interactions in DNA sequence revealed that the base sequence and not the base composition alone determines oligonucleotide stability. The thermodynamic library was published based on the nearest-neighbor base interactions (13). These data were proven to be valid for any DNA sequence and thus enabled predictive theoretical calculations of an oligomere stability and temperature-dependent behavior, including their behavior during PCR thermocycling. The calculations arranged to thermodynamic library were shown to be in excellent agreement with corresponding values determined experimentally (13). Therefore, we used this library also for calculations in this study.

Using the thermodynamic library, we calculated relative thermodynamic stability (free energy ΔG°) from the sequence of each of 21 ISSR primers used in this study and tested meaningful correlations (**Table 1**).

Beginning with the basic statistics, we calculated the mean free energy (ΔG°) of the set of all 21 ISSR primers averaged per a primer as $\Delta G^{\circ} = 167.19$ kJ/mol of interaction with minimum of $\Delta G^{\circ} = 73.06$ kJ/mol (primer 3PCTanW, **Table 1**), and maximum $\Delta G^{\circ} = 286.75$ kJ/mol of interaction (primer TETRA2, **Table 1**), with the coefficient of variation ($CV_{(\Delta G^{\circ})}$) = 36.78%.

Correlation Between Free Energy ΔG° of Primer and PCR A_i . We revealed prominent impact of the inclusion of PCR reamplification step on the final quality of ISSR-PCR patterns depending on a particular primer used for amplification (data not shown). Therefore, we expanded basic ISSR-PCR protocol for the second reamplification PCR round to fully exploit the marker potential of the ISSR-PCR method. By inspecting gel patterns from this reamplification protocol (rISSR-PCR, Figures 1 and 2), we revealed a statistically significant correlation between free energy of ISSR primer (ΔG°) and dA_i (Figure 3).

The described correlation shows that within the range of lower ΔG° values, the increase of free energy ΔG° caused by individual base changes in primer sequence brings about the desirable increase of PCR amplification intensity A_i . Higher A_i implicates higher detection sensitivity for displaying more

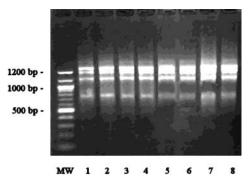


Figure 1. ISSR pattern (agarose gel with ethidium bromide staining) using the primer 3PCTanW. (5'-WWW [GA]₆) having the lowest value of free energy ΔG° among all primers tested. Only a minor impact of reamplification on the difference of PCR amplification intensity (*dA*_i) is demonstrated (1st run in lines 1–4 in comparison to 2nd reamplification run in lines 5–8). 1, Venika (1st round); 2, Wiko (1st round); 3, Biltstar (1st round); 4, Ariane (1st round); 5, Venika (2nd round); 6, Wiko (2nd round); 7, Biltstar (2nd round); 8, Ariane (2nd round).

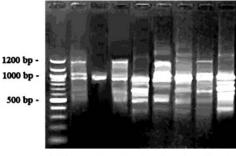




Figure 2. ISSR pattern (agarose gel with ethidium bromide staining) using the primer TETRA2. (5'-KKB NVS S [GGCT]₆) having the highest value of free energy ΔG° among all primers tested. Pronounced impact of reamplification on the difference of PCR amplification intensity (*dA*_i) as well as the appearance of new bands during the 2nd reamplification run are demonstrated (1st run in lines 1–4 in comparison to 2nd reamplification run in lines 5–8). 1, Venika (1st round); 2, Wiko (1st round); 3, Biltstar (1st round); 4, Ariane (1st round); 5, Venika (2nd round); 6, Wiko (2nd round); 7, Biltstar (2nd round); 8, Ariane (2nd round).

pattern polymorphisms. This correlation weakens at $\Delta G^{\circ} > 160$ kJ/mol, which indicates that ΔG° functions as a limiting factor of A_i at its low values during reamplification PCR protocol. On the basis of this correlation, we can formulate the following recommendation for ISSR primer design: free energy of ISSR primer should be $\Delta G^{\circ} > 160$ kJ/mol of interaction for efficient production of pattern polymorphism in flax by reamplification PCR protocol.

Correlation Between Free Energy ΔG° of Primer and PIC **Coefficient.** Further, we found correlation between free energy ΔG° of ISSR primer and PIC coefficient (**Figure 4**). ISSR primers producing only monomorphic ISSR patterns giving no marker value (PIC=0) were treated as statistical outliers in correlation analysis (*12*) and were removed from data set.

Polymorphic information content (PIC) is frequently used parameter of ISSR pattern, which effectively characterizes overall polymorphism (i.e., marker-information content of pattern generated by particular ISSR primer) (10). Revealed weaker correlation indicates the tendency of those primers of higher free energy ΔG° to generate ISSR patterns of higher polymorphism (i.e., patterns with higher PIC values). According

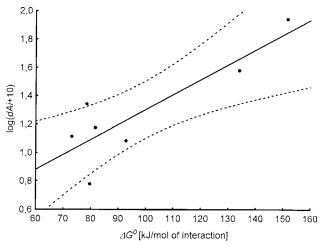


Figure 3. Positive correlation (Pearson r = 0.865, statistically significant at p = 0.019) between averaged free energy of ISSR primer (ΔG°) and increase of PCR amplification intensity after reamplification (dA_i) up to the threshold of $\Delta G^{\circ} = 160 \text{ kJ/mol}$ of interaction was revealed. This correlation may be approximated by equation $\log(dA_i + 10) = 0.011\Delta G^{\circ} + 0.246$. Confidence bands constructed at 95% confidential interval.

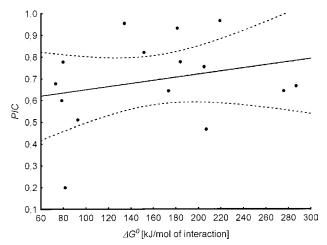


Figure 4. Weaker positive correlation (Pearson r = 0.253, statistically significant at p = 0.364) was revealed between averaged free energy (ΔG°) of ISSR primer and the PIC (polymorphic information content) parameter. This correlation may be approximated by the linear equation PIC = $7 \times 10^{-4} \Delta G^{\circ} + 0.576$. Confidence bands constructed at 95% confidential interval.

Table 2. Triple-Base Anchors of ISSR Primers^a

no.	sequence of anchor	$\operatorname{med}\Delta G^\circ$	avg ΔG°
1	5'-VRV	20.92	18.32
2	5′-VSS	28.03	24.52
3	5′-YHY	14.64	15.35
4	5′-SSS	28.03	26.99
5	5′-WWW	12.55	12.97
6	WWT-3′	14.22	13.60
7	SSG-3′	28.03	27.53

^a med ΔG° (kJ/mol of interaction) stands for the statistical median value of the free energy ΔG° in the series of ΔG° s of oligonucleotide sequences defined by anchor sequence signature with degeneration. avg ΔG° (kJ/mol of interaction) stands for the averaged value of the free energy ΔG° in the series of ΔG° s of oligonucleotide sequences defined by anchor sequence signature with degeneration. The codes for degenerated oligonucleotides are as follows: R = A,G; Y = C,T; S = G,C; W = A,T; V = G,A,C; H = A,T,C.

to this correlation, the recommendation to design ISSR primers of higher ΔG° up to $\Delta G^{\circ} = 300$ kJ/mol is supported if maximum polymorphism on ISSR pattern is targeted.

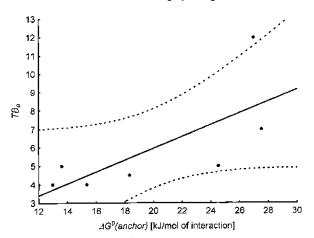


Figure 5. Positive correlation (Pearson r = 0.709, statistically significant at p = 0.075) between averaged free energy (ΔG°) of triple-base flanking anchor sequences of ISSR primers and total number of bands produced by all ISSR primers harboring the same anchor averaged per a primer (TB_a). The correlation may be approximated by the linear equation TB_a $= 0.321\Delta G^{\circ} - 0.450$. Confidence bands constructed at 95% confidential interval. Remaining two longer anchors (N₇ and KKB NVS S) were treated as outliers in this functional analysis.

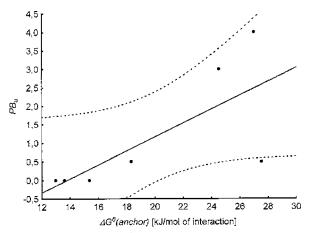


Figure 6. Positive correlation (Pearson r = 0.721, statistically significant at p = 0.067) between averaged free energy (ΔG°) of triple-base flanking anchors of ISSR primer sequences and the number of polymorphic bands produced by all ISSR primers harboring the same anchor averaged per a primer (PB_a). The correlation may be approximated by the linear equation PB_a = 0.188 ΔG° - 2.598. Confidence bands constructed at 95% confidential interval. The remaining two longer anchors (N₇ and KKB NVS S) were treated as outliers in this functional analysis.

Correlation Between Free Energy ΔG° of Anchor and the Number of Total and Polymorphic Bands in ISSR Patterns. Oligonucleotide anchor inserted on either the 5' or 3' end of ISSR primer strongly influences the quality of resulting ISSR pattern (5). Therefore, we calculated free energy ΔG° for each of seven triple-base anchor sequences inserted in flanking positions into ISSR primers (**Table 2**). In agreement with referred importance of anchors, we found two statistically significant correlations between free energy ΔG° of an anchor sequence and the number of total and polymorphic bands in ISSR patterns (**Figures 5** and **6**, respectively).

Both correlations demonstrated the tendency that the higher ΔG° of an anchor of ISSR primer the better is its "anchoring" ability during PCR cycling, which means more bands in

resulting pattern. Both these correlations bring about such recommendation for effective primer design, that the sequences of anchors flanking ISSR primer sequences should be designed with higher ΔG° values (around $\Delta G^{\circ} = 28$ kJ/mol) to get maximum polymorphism from flax ISSR pattern.

Conclusion. We can summarize that from the four described statistically significant correlations, the recommendation may be deduced that final free energy ΔG° of the sequence of designed ISSR primer should be kept at $\Delta G^{\circ} > 160$ kJ/mol of interaction for successful application of reamplification PCR protocol (rISSR-PCR), and the free energy of anchor sequence should near $\Delta G^{\circ} = 28$ kJ/mol to produce ISSR patterns with maximum polymorphism of flax germplasm. To comply with these criteria, the primer TETRA2 was best positioned among the set of 21 primers tested here on flax model germplasm.

LITERATURE CITED

- Barth, S.; Melchinger, A. E.; Lubberstedt, T. Genetic diversity in *Arabidopsis thaliana* L. Heynh. investigated by cleaved amplified polymorphic sequence (CAPS) and inter-simple sequence repeat (ISSR) markers. *Mol. Ecol.* 2002, *11*, 495–505.
- (2) Hodkinson, T. R.; Chase, M. W.; Renvoize, S. A. Characterization of a genetic resource collection for *Miscanthus (Saccharinae, Andropogoneae, Poaceae)* using AFLP and ISSR PCR. *Ann. Bot.* 2002, 89, 627–636.
- (3) Nagaraju, J.; Kathirvel, M.; Kumar, R. R.; Siddiq, E. A.; Hasnain, S. E. Genetic analysis of traditional and evolved Basmati and non-Basmati rice varieties by using fluorescence-based ISSR-PCR and SSR markers. *Proc. Natl. Acad. Sci. U.S.A.* 2002, 99, 5836–5841.
- (4) Reddy, M. P.; Sarla, N.; Siddiq, E. A. Inter-simple-sequence repeat (ISSR) polymorphism and its application in plant breeding. *Euphyt.* 2002, 128, 9–17.
- (5) Zietkiewicz, E.; Rafalski, A.; Labuda, D. Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics* **1994**, *20*, 176–183.
- (6) Cullings, K. W. Design and testing of plant-specific PCR primer for ecological and evolutionary studies. *Mol. Ecol.* **1992**, *1*, 233– 240.
- (7) Vogel, J. M.; Scolnik, P. A. In DNA markers: Protocols, Applications, and Overviews; Caetano-Annoles, G., Gresshoff, P. M., Eds.; VCH Publications: New York, 1998, pp 133–150.
- (8) Wiesner, I.; Wiesnerová, D.; Tejklová, E. Effect of anchor and core sequence in microsatellite primers on flax fingerprinting patterns. J. Agricult. Sci. 2001, 137, 37–44.
- (9) Wiesner, I.; Wiesnerová, D. Polyacrylamide gel polymerization with adjustable gelation rate. *BioTechniques* 2002, *32*, 740– 742.
- (10) Botstein, D.; White, R. L.; Scolnick, M.; Davis, R. W. Construction of a genetic linkage map in man using restriction fragment length polymorphism. *Am. J. Hum. Genet.* **1980**, *32*, 314–331.
- (11) Morgante, M.; Olivieri, A. M. PCR-amplified microsatellites as markers in plant genetics. *Plant J.* **1993**, *3*, 175–182.
- (12) Daniel, W. W. Biostatistics: A foundation for analysis in the health sciences; John Wiley and Sons Inc.: New York, 1999, p 159.
- (13) Breslauer, K. J.; Frank, R.; Blocker, H.; Marky, L. A. Predicting DNA duplex stability from the base sequence. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 3746–3750.

Received for review March 10, 2003. Revised manuscript received March 1, 2004. Accepted March 3, 2004. This research was supported by the Grant Agency of the Czech Republic, project no. 521/03/0019.

JF034231B