Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

ISSR and IRAP marker polymorphism association to its localization features in reference genomes of mammals and reptiles

Pavel Apashkin (pavel.ap.97@gmail.com)
Synergy University https://orcid.org/0000-0002-7351-8279
Ivan Blokhin
RGAU-MSHA im K A Timirâzeva: Rossijskij gosudarstvennyj agrarnyj universitet MSHA imeni K A Timirazeva
Valery Glazko
RGAU-MSHA im K A Timirâzeva: Rossijskij gosudarstvennyj agrarnyj universitet MSHA imeni K A Timirazeva
Tatyana Glazko
FGBNU Naucno-issledovatel'skij institut pusnogo zverovodstva i krolikovodstva imeni V A Afanas'eva

Research Article

Keywords: mobile genetic elements, microsatellites, G4 quadruplexes, microRNA, polymorphism

Posted Date: February 23rd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1322155/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Abstract

A well-known problem of population genetic studies is the selection of the most polymorphic genomic elements for genotyping to obtain the high resolution "portrait" of the population genetic structures. Such elements include microsatellite loci, regions of endogenous retroviruses in genomes of various taxa. The use these primers in polymerase chain reaction can allow the highly polymorphic polyloci spectra obtaining for detection of individual and population differences between animals. At the same time, the polymorphism of such spectra is dependent on the sequence chosen as a primer. A comparative analysis of the polymorphism of amplification product spectra in different taxa such as domestic dog *Canis familiaris* (Linnaeus, 1758) and sand lizard *Lacerta agilis* (Linnaeus, 1758) and primer co-localization near transposons, G4 quadruplexes, miRNA in reference genomes was carried out.

The study was carried out at the Federal State Budgetary Educational Institution of Higher Professional Education of the Russian Academy of Agricultural Sciences named after K.A. Timiryazev in 2020-2021. Microsatellites (ACC6T, THC6C, GAG6C) and endogenous retrovirus long terminal repeats (SIRE 1, Sabrina) were used as primers for genotyping. Further their co-localization in reference genomes with nucleotide sequences associated with increased variability (G4 quadruplexes, microRNA) was evaluated.

The studied species have a tendency to reduced polymorphism of genomic DNA fragments flanked by inverted microsatellite repeats and increased density of G4 quadruplexes localization in these fragments. This trend is realised in both species, despite the observed interspecific differences.

Introduction

Microsatellites and retrotransposons are the most polymorphic regions of the genome and convenient tools to describe the population-genetic structure of different species (Blokhin, Glazko, (2021); Kalendar, Glazko, (2002); Glazko et al., (2020)). Mobile genetic elements, their fragments, and recombination products between them occupy about half of the genomes in most species. In recent years, many works have presented examples of their involvement in horizontal transfers of genetic material even between animal and plant kingdoms (Ivancevic et al., (2018); Nishiyama, Ohshima, (2018); Zhang et al., (2014)), as well as a significant role in genomic reorganizations affecting various processes and structures of the organism (Ferrari et al., (2021); Xia et al., (2021)). The polymorphism of these genomic structures allows us to track changes in population and genetic structures across populations and species with high resolution (Alzohairy et al., (2014); Kalendar, (2021)). Homology sites to microsatellites, fragments of long terminal repeats of endogenous retroviruses are often used as primers in polymerase chain reaction for polylocus genotyping in order to compare genetic structures and control their dynamics and detect the influence of, in particular, environmental factors on them.

Such methods as amplification of genomic DNA fragments flanked by an inverted microsatellite repeat (Inter Simple Sequence Repeat - ISSR-PCR markers) or a homology site to the long terminal fragment of an endogenous retrovirus (Inter Retrotransposon Amplified Polymorphism - IRAP-PCR markers) have become quite widespread due to the speed of final results and ease of their interpretation (Kalendar et al., (2021)).

At the same time, one of the problems of using such markers in population genetic studies is the differences in the polymorphism spectra of the amplification products obtained in the polymerase chain reaction using different nucleotide sequences as primers. This leads to the need to select the most polymorphic systems each time a new group of organisms is studied. One approach to the selection of such markers can be expected to be the elucidation of co-localization in the genomes of homology regions to primers and nucleotide sequences contributing to the increased variability of such regions. The latter may include, in particular, transposons, as well as mutual positioning of guanines underlying the formation of DNA secondary structures such as G4 quadruplexes (Capra et al., (2010)). G4 quadruplexes are markers of local genomic instability and recombination susceptibility (Bohalova et al., (2021); Ruggiero, (2021)).

In view of the above, the present work attempts to compare the polymorphism of ISSR-PCR and IRAP-PCR markers detected in subjects such as domestic dog and jumping lizard in experimental studies, and co-localization in homology sites to the primers

used in the reference genomes of these species of transposons and G4 quadruplexes.

Material And Methods

The tails of the captured jumping lizard specimens were lifelong cropped, and their muscles served as a source of genomic DNA; buccal epithelium samples were taken from three dog breeds (English greyhound, Russian wolfhound, and Hortaya greyhound). The polymorphism of genomic DNA fragments flanked by inverted microsatellite repeats (Inter Simple Sequence Repeat - ISSR-PCR markers) and long terminal repeats (LTR) of endogenous retroviruses (InterRetrotransposon Amplified Polymorphism - IRAP-PCR markers) were estimated as markers.

Genomic DNA was isolated from biosamples using a standard DNA Extran II kit (Syntol). Polymerase chain reaction (PCR) was performed on amplifier Terzik with the following parameters: primary denaturation (t=94ºC, 2 min), denaturation (t=94ºC, 30 sec), annealing (t=58ºC, 30 sec), elongation (t=72ºC, 2 min) - 40 cycles, final elongation (t=72ºC, 10 min).

The trinucleotide microsatellites $(ACC)_6$ T, $(TGC)_6$C, $(GAG)_6$ C and long terminal repeats sites of endogenous retroviruses LTR-SIRE1 (5' - GCAGTTATGCAAGTGGGATCAGCA- 3') and Sabrina 111 (5' - AAACAAGAACTGACACTTGGCACT- 3') were used as primers. Elements of Sabrina endogenous retrovirus family were first described in barley (Shirasu et al., 2000), LTR-SIRE1 - in soybean (Glazko et al., (2015)).

The amplification products were separated in 1.5% agarose gel in TAE buffer. Visualization was performed using a UV transilluminator. The size of DNA fragments was determined using the M27 molecular weight marker 100 bp+1.5 Kb+3 Kb (12 fragments from 100 to 3000 bp) (SibEnzyme, Russia). For each amplification product spectrum obtained with the respective primer, a matrix was constructed representing the presence or absence of a particular amplification product (amplicon), each of which was treated as a locus. In the absence of a DNA fragment of the appropriate length, such a genotype was evaluated as homozygous for the recessive allele. The proportion of polymorphic loci (% of polymorphic DNA fragments relative to the total number of amplicons detected in the spectrum of each primer) and the polymorphic information content (PIC) of the spectrum were evaluated as population-genetic characteristics. The PIC was calculated using the formula for the diallelic loci for which $PIC = 2f(1-f)$, where f is the frequency of one of the two alleles. Since ISSR-PCR markers have a dominant manifestation in the presence of the amplification product in the spectrum, the expected heterozygosity was calculated based on the frequency of homozygotes for the conditionally recessive allele (absence of a DNA fragment of the corresponding length). Statistical processing of the obtained data was performed according to the standard methodology (Plokhinsky, (1961)). Genetic distances were calculated using the GenAlEX program (https://biology-assets.anu.edu.au/GenAlEx/Welcome.html)

Homology sites with at least 90% identity to the primers used in the reference genomes of domestic dog and jumping lizard were searched using NCBI BLAST+ Standalone algorithms (https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/), the Python programming language, and the Biopython library (https://biopython.org/). The search for inverted repeats within 100 to 2000 bp was performed for the detected sites. In the absence of inverted repeats, the search was performed using a 100% homology site flanked by 500 nucleotides (total length 1025 bp). From the sites found for retrotransposons and microsatellites with at least 90% identity, one site was randomly selected where mobile genetic elements, quadruplexes, and microRNAs were searched for. The presence and number of mobile genetic elements, quadruplexes, and microRNAs were assessed using CENSOR (https://www.girinst.org/censor/index.php), QGRS Search Options (https://bioinformatics.ramapo.edu/QGRS/analyze.php).

Results

The highest number of amplifiable fragments in the spectra of the jumping lizard and domestic dog and the highest polymorphism rates (on average per amplicon spectrum of one primer) were found for IRAP-PCR markers compared to ISSR-PCR markers (Table 1).

Table 1
Results of genome analysis of the sand lizard *Lacerta agilis* (Linnaeus, 1758)

Primer	Number of fragments to be amplified	DPL spectra of amplifiable fragments, %	Number of homology sites in the reference genome	Sequences between inverted repeats of length from 100 to 2000 n.	Number of mobile genetic elements	Number of quadruplexes (without overlaps)	Number of quadruplexes (with overlaps)	MicroRNA
					In regions of the reference genome about 1,000 nucleotides long			
SIRE 1	16	88	17	0	2	6	98	14
Sabrina	19	100	21	0	1	6	1	43
$(ACC)_6T$	16	75	3314	21	1	8	658	36
$(TGC)_6C$	8	87,5	12350	476	2	17	1015	13
$(GAG)_6C$	11	36,4	37088	3216	2	13	1756	26

Based on the data presented in Tables 1 and 2, the polymorphism of retrotransposons and microsatellites does not depend on the homology sites to mobile genetic elements identified nearby. In Table 1, in the reference genome of the jumping lizard, 2 of them were found at sites near the fragment SIRE 1, as well as at sites $(GAG)_6C$, although the polymorphism of the primer SIRE 1 spectra is 100%, and that of $(GAG)_6C$ spectra is 36.4% (Table 1). The same situation is observed when estimating the number of mobile genetic elements in the canine genome: 3 mobile genetic elements were detected in the microsatellite $(TGC)_6C$ sites and 2 in the Sabrina retrotransposon sites with the polymorphism of the spectra at 17% and 100%, respectively (Table 2).

Table 2
Results of genome analysis of the domestic dog *Canis familiaris* (Linnaeus, 1758)

Primer	Number of fragments to be amplified	DPL spectra of amplifiable fragments, %	Number of homology sites in the reference genome	Sequences between inverted repeats of length from 100 to 2000 п.ы.	Number of mobile genetic elements	Number of quadruplexes (without overlaps)	Number of quadruplexes (with overlaps)	MicroRNA
					In regions of the reference genome about 1,000 nucleotides long			
SIRE 1	21	100	6	0	1	2	3	21
Sabrina	18	100	5	0	2	2	9	21
$(ACC)_6T$	19	95,2	6422	0	1	5	70	131
$(TGC)_6C$	6	17	4055	15	3	5	12	395
$(GAG)_6C$	18	83	16204	468	2	13	3789	312

The sites of the lizard genome flanked by inverted repeats of the SIRE1 and Sabrina retrotransposons, despite their high polymorphism revealed in the experimental studies, show the lowest number of overlapped and non-overlapped quadruplexes (98 and 6 in the retrotransposon sites SIRE 1 and 43 and 1 in the Sabrina retrotransposon) is found in the lizard genome fragments flanked by inverted repeats of the SIRE 1 and Sabrina retrotransposons, despite their high polymorphism revealed in the experimental studies. The maximum number of G4 quadruplexes without overlaps (17) was found in the DNA site flanked by the inverted microsatellite repeat $(TGC)_6C$. The most G4 quadruplexes with overlaps were found in the DNA fragment flanked by $(GAG)_6C$ (1756), which is expected due to the presence of repetitive GG nucleotides in the primer itself. At the same time, the polymorphism of the amplification product spectra obtained in experimental studies using these microsatellites as primers was 87.5% and 36.4%, respectively (Table 1).

In the reference genome of the domestic dog, the highest number of quadruplexes with overlaps was found in the spectrum sites of microsatellite $(GAG)_6C$ (13 and 3789), with the polymorphism of its spectra in the experiment being 83%. In the DNA sites flanked by the other primers, the number of quadruplexes was significantly lower, although the polymorphism of their spectra in the experimental studies exceeded that of $(GAG)_6C$ (the exception was microsatellite $(TGC)_6C$ with a level of polymorphism of its spectra of 17%) (Table 2). Thus, one can note a certain tendency towards the association of a relatively higher number of quadruplexes with a low polymorphism of the primer spectrum.

Significant differences in the frequency of occurrence of homology sites to microRNA sequences in the nucleotide sequences of reference genomes under consideration are noteworthy: in the domestic dog such sites occur an order of magnitude more frequently than in the jumpy lizard, in genomic regions flanked by inverted microsatellite repeats (Tables 1, 2).

Discussion

In the reference genomes of different species, the frequency of microsatellite $(GAG)_6C$ is the highest among the studied sequences; the frequency of homology sites to long terminal repeats of two retroviruses, SIRE 1 and Sabrina, is the lowest, but it is not associated with the results of experimental studies of polymorphism spectra of their flanking genomic fragments (Tables 1, 2). Thus, there is no direct relationship between co-localization of homology sites to them in reference genomes and mobile genetic elements as well as their polymorphism in experimental studies (Tables 1, 2). At the same time, a certain negative association between the frequency of co-localization of G4 quadruplexes in the regions of reference genomes flanked by sequences used as primers in the experiment and the polymorphism of their spectra in polymerase chain reaction was found in both species (Tables 1, 2). The marked differences between the frequency of G4 quadruplexes in the considered regions of domestic dog and jumping lizard reference genomes are consistent with the data on the species specificity of their representation (Marsico et al., (2019)).

It is known that the source of microRNAs is mobile genetic elements and this is the essential pathway of their influence on the formation of regulatory networks that determine gene expression profiles and their dynamics in multicellular organisms (Ali et al., (2021)).

Interestingly, our studies revealed a certain negative association between the frequency of homology sites to microRNA and the density of G4 quadruplexes in the considered fragments of the reference genomes of domestic dog and jumping lizard (Tables 1, 2). Only the sites flanked by the inverted repeat of the microsatellite $(GAG)_6C$, the frequency of which is significantly higher than the other considered microsatellites in the reference genomes of both species, fall out of this association (Tables 1, 2). The observed interspecific differences in the frequency of homology sites to microRNAs and G4 quadruplexes in the studied regions of the reference genomes of domestic dog and jumping lizard are not consistent with the notion that these two genomic characteristics are generally closely related to each other (Chan et al., 2018).

The nucleotide sequence of microsatellite (GAG)n refers to DNA regions potentially predisposed to the formation of triplexes (purine-pyrimidine tracks) that play an essential role in the formation of secondary DNA structures, DNA:RNA binding, chromatin organization, and gene expression regulation (Postepska-Igielska et al., 2020). The diversity of functional involvement of such sequences in the structural and functional organization of the genome can apparently explain the differences in their distribution in the reference genomes of both species from other microsatellite sequences.

In general, comparative analysis of polymorphism of genomic DNA fragments of two species in experimental studies of population-genetic structures using ISSR-PCR and IRAP-PCR markers and distribution of homology sites to nucleotide sequences used as primers in PCR in reference genomes of domestic dog and jumping lizard suggests a certain association between G4 quadruplex densities and polymorphism of such markers.

G-quadruplexes are known to be an integral part of complex regulatory systems, closely related to the movement of retrotransposons in mammals (Sahakyan et al., (2017)). Evidence is accumulating that there is a delicate balance between genome instability caused by G4 quadruplexes and its repair processes stimulated by the same G4 quadruplexes. It is suggested that this is what ensures the stability of G4 structures in vitro and in vivo (Pavlova et al., (2021)), their certain conservatism

(Capra et al., (2010); Zybaĭlov et al., (2013)). High density of G4 quadruplexes in the "hot spots" of mononucleotide substitutions, which does not always coincide with the increased recombination frequency, indicates the possible influence of nucleotide context and various features of DNA secondary structure formation (Glazko et al., (2021)).

Our studies suggest that despite the polyfunctionality and complexity of the distribution of the considered nucleotide motifs, the preliminary analysis of reference genomes on the distribution of homology sites to potential primers and their associations with the frequency of G4 quadruplexes can improve the efficiency of ISSR-PCR and IRAP-PCR markers in population genetic studies.

Declarations

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Author Contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Ivan Blokhin, Valery Glazko, Tatyana Glazko and Pavel Apashkin. The first draft of the manuscript was written by Pavel Apashkin and Ivan Blokhin, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Ethics approval

All biomaterials used in the study were provided by the Russian State Agrarian University - Timiryazev Moscow Agricultural Academy.

References

1. AHMED M. ALZOHAIRY et al. (2014). Retrotransposon-based molecular markers for assessing genomic diversity. *Functional Plant Biology* 41: 781-789 https://doi.org/10.1071/FP13351
2. ALI A., HAN K. & LIANG P. (2021). Role of Transposable Elements in Gene Regulation in the Human Genome. *Life (Basel)* 11(2): 118. https://doi.org/10.3390/life11020118
3. ATMA M. IVANCEVIC et al. (2018). Horizontal transfer of BovB and L1 retrotransposons in eukaryotes. *Genome Biology*. B https://doi.org/10.1186/s13059-018-1456-7
4. BLOKHIN I.G. & GLAZKO V.I. (2021). Genetic and phenotypic characteristics in assessing population differentiation of the jumping lizard. *South of Russia: Ecology, Development*. 16(3): 47-58. https://doi.org/10.18470/1992-1098-2021-3-47-58
5. BO XIA et al. (2021). The genetic basis of tail-loss evolution in humans and apes. *bioRxiv.* https://doi.org/10.1101/2021.09.14.460388
6. BOHÁLOVÁ N. et al. (2021). Tracing dsDNA Virus-Host Coevolution through Correlation of Their G-Quadruplex-Forming Sequences. *International Journal of Molecular Sciences.* 2021, 22, 3433. https://doi.org/10.3390/ijms22073433
7. CAPRA J. A. et al. (2010). G-Quadruplex DNA Sequences Are Evolutionarily Conserved and Associated with Distinct Genomic Features in Saccharomyces cerevisiae. *PLoS Computational Biology*. No. 6(7). P. 1-13. e1000861 https://doi.org/10.1371/journal.pcbi.1000861
8. CHAN KL. et al. (2018). Structural analysis reveals the formation and role of RNA G-quadruplex structures in human mature microRNAs. *Chemical Communication*. 54(77): 10878-10881. https://doi.org/10.1039/c8cc04635b

9. ERI NISHIYAMA & KAZUHIKO OHSHIMA. (2018). Cross-Kingdom Commonality of a Novel Insertion Signature of RTE-Related Short Retroposons. *Genome Biology and Evolution*. 10(6): 1471B–1483. https://doi.org/10.1093/gbe/evy098
10. FERRARI, R. et al. (2021). Retrotransposons as Drivers of Mammalian Brain Evolution. *Life.* 2021, 11, 376. https://doi.org/10.3390/life11050376
11. GLAZKO T.T. et al. (2020). Genomic "portrait" of some canine species obtained using issr-pcr and irap-pcr markers. *Rabbit And Fur Farming*. No. 1, PP. 28-39.
12. GLAZKO, V.I. & KOSOVSKY, G.Y. (2013). Structure of genes coding the envelope proteins of the avian influenza a virus and bovine leucosis virus*. Russian Agricultural Sciences*. 39, 511-515. https://doi.org/10.3103/S1068367413060074
13. GUIBLET WM. et al. (2021). Selection and thermostability suggest G-quadruplexes are novel functional elements of the human genome. *Genome Research*. 31(7): 1136-49. https://doi.org/10.1101/gr.269589.120
14. HUA-HAO ZHANG et al. (2014). Recurrent Horizontal Transfers of Chapaev Transposons in Diverse Invertebrate and Vertebrate Animals. *Genome Biology and Evolution*. 6(6): 1375-1386. https://doi.org/10.1093/gbe/evu112
15. KALENDAR R.V. & GLAZKO V.I. (2002). Types of molecular genetic markers and their application. Physiology and biochemistry of cultivated plants. Г’. 34. N 4. Г‘. 279-295.
16. MARSICO G. et al. (2019). Whole genome experimental maps of DNA G-quadruplexes in multiple species. *Nucleic Acids Research*. 47(8): 3862-3874. https://doi.org/10.1093/nar/gkz179
17. PAVLOVA, A.V. et al. (2021). Impact of G-Quadruplexes on the Regulation of Genome Integrity, DNA Damage and Repair. *Biomolecules*. 2021, 11, 1284. https://doi.org/10.3390/biom11091284
18. POSTEPSKA-IGIELSKA A., BLANK-GIWOJNA A. & GRUMMT I. (2020). Analysis of RNA-DNA Triplex Structures In Vitro and In Vivo. *Methods in Molecular Biology*. 2161: 229-246. https://doi.org/10.1007/978-1-0716-0680-3_16
19. RUGGIERO, E. et al. (2021). G-Quadruplex Targeting in the Fight against Viruses: An Update. *International Journal of Molecular Sciences*. 2021, 22, 10984. https://doi.org/10.3390/ijms222010984
20. RUSLAN KALENDAR, ALEXANDER MUTERKO, & SVETLANA BORONNIKOVA. (2021). Retrotransposable Elements: DNA Fingerprinting and the Assessment of Genetic Diversity. *Molecular Plant Taxonomy: Methods and Protocols.* Methods in Molecular Biology, vol. 2222, https://doi.org/10.1007/978-1-0716-0997-2_15
21. SAHAKYAN AB. et al. (2017). G-quadruplex structures within the 3' UTR of LINE-1 elements stimulate retrotransposition. *Nature Structural & Molecular Biology*. 24(3): 243-247. https://doi.org/10.1038/nsmb.3367
22. ZYBAĬLOV B.L. et al. (2013). G4-quadruplexes and genome instability*. Mol Biol (Mosk).* 47(2):224-31. Russian. https://doi.org/10.7868/s0026898413020183