RESEARCH ARTICLE Indian Journal of Animal Research, Volume 58 Issue 9: 1474-1479 (September 2024)

Complete Chromosome Wise Identification of SSRs in the Two Published Chicken Genome Assemblies

Jayakumar Sivalingam[1], R.P. Athe[1], T.K. Bhattacharya[1], R.N. Chatterjee[1], U. Raj Kumar[1], Satya Pal Yadav[1], K.S. Raja Ravindra[1], M. Balakrishnan[2], M.V. Chaudhari[3]

10.18805/IJAR.B-5254

ABSTRACT

Background: Microsatellites are short repeat motifs consisting 1-6 base pair (bp) and have a high degree of length polymorphism and are useful to study the genetic diversity. In Chicken, no micro satellite markers were demarcated to different chromosomes, *viz*., 29, 34, 35, 36, 37, 38 and 39 in the genome. In the present study, we carried out chromosome-wise identification of SSRs that will be of great use in diversity and gene mapping studies in chicken.

Methods: The perl script of MISA tool was used to screen the polymorphic microsatellites with different thresholds level.

Result: We identified a total of 397877 SSRs from the latest chicken genome assembly (GRCg7b), of which mononucleotides, dinucleotides, trinucleotides, tetranucleotides, pentanucleotides and hexanucleotides composed of 307453 (77.27%), 40991 (10.30%), 23059 (5.79%), 16723 (4.20%), 8127 (2.04%) and 1524 (0.38%), respectively.

Key words: Chicken, Diversity, Genome assembly, Microsatellite markers.

INTRODUCTION

Microsatellites are short repeat motifs consisting 1-6 base pair (bp) and can be present both in coding and non-coding regions in prokaryotic and eukaryotic genomes (Gupta *et al*., 1996; Tóth *et al*., 2000; Katti *et al*., 2001). They have a high degree of length polymorphism (Sawaya*et al*., 2013) and are regarded as an important tool for studying genetic diversity. In chicken, recently a genome assembly of cross of broiler hen and white leghorn cock (GRCg7b) was released in January 2021 of size 1053.33 Mb.The present genome assembly has mapped the reads to all the chicken chromosomes when compared to the previous assembly of Red Jungle Fowl (GRCg6a) of size 1065.37 Mb released in March 2018.So far in Chicken, no microsatellite marker was demarcated to different chromosomes, *viz*., 29, 34, 35, 36, 37, 38 and 39 in the genome. But, this deficiency can be sorted out using the current genome assembly (GRCg7b) which has all the chromosomes clearly demarcated. So, we carried out chromosome-wise identification of SSRs that will be of great use in diversity and gene mapping studies in chicken.

MATERIALS AND METHODS

Chicken genome assemblies; GRCg7b (GCA_016699485.1) and GRCg6a (GCA_000002315.3) were downloaded from the NCBI and were used for identification of SSRs by MISA tool (Beier*et al*., 2017). The perl script of MISA tool was used to screen the polymorphic microsatellites with different thresholds level (1-10 2-6 3-5 4-5 5-5 6-5 *i.e* Mononucleotide repeats with minimum 10 repeats, dinucleotide repeats with minimum 6 repeats, Trinucleotide repeats with minimum 5 repeats, Tetra nucleotide repeats with minimum 5 repeats, Pentanucleotide repeats with minimum 5 repeats, Hexa nucleotide repeats with minimum 5 repeats). The minimum distance between 2 compound SSRs was set at 100 bp in length. In the present genome assembly (GRCg7b), SSRs were also identified in the chromosome 39 as per the classification of the recent genome assembly.

[1]ICAR-Directorate of Poultry Research, Hyderabad-500 030, Telangana, India.
[2]ICAR-National Academy of Agricultural Research Management, Hyderabad-500 030, Telangana, India.
[3]School of Agricultural Sciences and Technology, SVKM's Narsee Monjee Institute of Management Studies (Deemed-to-University), Shirpur-425 405, Maharashtra, India.

Corresponding Author: Jayakumar Sivalingam, ICAR-Directorate of Poultry Research, Hyderabad-500 030, Telangana, India. Email: jeyvet@gmail.com

How to cite this article: Sivalingam, J., Athe, R.P., Bhattacharya, T.K., Chatterjee, R.N., Kumar U.R.,Yadav, S.P., Ravindra K.S.R, Balakrishnan, M. and Chaudhari, M.V. (2024). Complete Chromosome Wise Identification of SSRs in the Two Published Chicken Genome Assemblies. Indian Journal of Animal Research. 58(9): 1474-1479. doi: 10.18805/IJAR.B-5254.

Submitted: 12-10-2023 **Accepted:** 10-6-2024 **Online:** 22-07-2024

RESULTS AND DISCUSSION

We identified a total of 397877 SSRs from the latest chicken genome assembly (GRCg7b), of which mononucleotides, dinucleotides, trinucleotides, tetranucleotides, pentanucleotides and hexanucleotides composed of 307453 (77.27%), 40991 (10.30%), 23059 (5.79%), 16723 (4.20%), 8127 (2.04%) and 1524 (0.38%), respectively (Table 1). Similarly, in GRCg6a, a total of 384069 SSRs were identified, out of which

Table 1: Number and types of SSRs identified in GRCg7b genome assembly.

	GRCg7b											
Chromosome No.	Mono-nucleotide repeats		Di-nucleotide repeats		Tri-nucleotide repeats		Tetra-nucleotide repeats		Penta-nucleotide repeats		Hexa-nucleotide repeats	
	Perfect	Compound	Perfect	Compound	Perfect	Compound	Perfect	Compound	Perfect	Compound	Perfect	Compound
1	56343	5166	8092	756	4627	168	4239	388	1456	587	311	152
2	43691	3818	5810	519	3466	151	2755	249	1004	341	179	115
3	32483	2683	4034	402	2156	68	1626	117	531	164	80	40
4	26969	2134	3311	309	2021	146	1165	60	374	85	52	16
5	16483	1298	1998	201	1028	25	563	30	165	42	22	10
6	9363	636	1093	125	615	14	292	12	64	14	9	3
7	9649	771	1209	105	599	17	293	17	75	15	9	5
8	7937	640	974	83	535	16	291	11	78	14	10	2
9	5752	375	704	51	365	12	157	7	55	8	6	12
10	5045	361	641	63	309	9	115	3	64	5	5	4
11	5141	415	636	53	325	15	140	8	65	9	8	2
12	4612	332	583	57	312	16	127	17	66	6	4	1
13	3918	264	521	37	324	11	131	9	79	13	4	2
14	3342	255	431	50	286	4	78	3	66	8	8	5
15	2727	215	427	32	219	8	74	9	37	7	12	0
16	426	144	103	3	162	23	56	2	95	11	36	5
17	2192	199	315	24	192	8	69	3	50	2	6	2
18	2094	169	293	17	262	7	83	8	70	15	4	4
19	2090	160	286	21	187	9	48	3	90	7	6	1
20	3254	239	383	30	257	7	228	4	130	21	5	3
21	1413	146	198	8	156	6	41	4	37	4	5	0
22	1117	166	135	11	118	3	28	3	32	9	15	2
23	1277	109	170	16	157	7	61	8	38	16	4	1
24	1154	106	183	16	138	5	32	0	77	12	5	3
25	1193	272	89	9	164	13	24	5	74	33	4	1
26	1138	124	170	13	167	6	41	3	38	8	5	2
27	1434	256	217	13	188	11	44	10	47	19	9	4
28	1515	235	174	15	164	12	59	9	76	17	19	4
29	28	2	2	0	1	0	20	0	77	20	1	0
30	392	155	18	4	29	16	7	2	9	10	4	0
31	202	25	121	4	45	1	67	11	37	34	10	10
32	14	0	NF	NF	NF	NF	NF	NF	3	0	NF	NF
33	214	26	58	1	20	0	56	1	11	2	8	7

Table 1: Continue...

Table 1: Continue...

34	1554	491	131	28	117	7	26	5	38	12	10	4
35	134	22	10	1	15	0	2	0	3	0	2	1
36	100	37	2	0	17	0	2	0	0	1	NF	NF
37	13	0	6	0	8	0	NF	NF	4	0	NF	NF
38	490	194	14	1	44	3	7	2	16	0	2	0
39	151	65	5	1	6	1	3	2	1	1	2	0
W	1219	101	473	58	115	1	10	1	83	166	9	7
Z	24133	2250	3464	370	2246	71	2330	307	754	320	141	63
MT	1	0	0	0	0	0	0	0	0	0	0	0
Total	282397	25056	37484	3507	22162	897	15390	1333	6069	2058	1031	493

NF – Not found.

Table 2: Number and types of SSRs identified in GRCg6a genome assembly.

	GRCg7b											
Chromosome No.	Mono-nucleotide repeats		Di-nucleotide repeats		Tri-nucleotide repeats		Tetra-nucleotide repeats		Penta-nucleotide repeats		Hexa-nucleotide repeats	
	Perfect	Compound	Perfect	Compound	Perfect	Compound	Perfect	Compound	Perfect	Compound	Perfect	Compound
1	56176	4851	8148	740	4598	179	4219	401	1327	696	271	133
2	42989	3679	5860	479	3399	105	2768	258	874	405	148	78
3	32210	2573	4067	405	2153	71	1624	123	499	208	75	38
4	26748	2075	3314	306	1808	53	1196	64	303	104	49	16
5	16260	1246	2006	198	1022	32	573	32	155	57	24	7
6	9030	604	1092	118	568	16	262	5	56	11	8	7
7	9587	729	1221	109	603	23	289	15	71	22	10	2
8	7921	597	1004	85	533	16	295	11	62	24	7	2
9	5665	353	682	58	383	14	156	9	49	11	7	8
10	4852	332	629	59	317	8	109	1	49	10	8	3
11	5004	374	636	55	323	12	140	5	61	18	14	9

Table 2: Continue...

Table 2: Continue...

12	4587	307	575	68	305	15	110	10	56	8	6	0
13	3856	258	499	41	310	10	121	7	60	20	4	1
14	3226	233	426	42	280	6	74	5	53	8	9	4
15	2689	201	419	35	229	9	64	10	31	12	12	1
16	115	12	20	0	25	3	7	1	8	2	1	0
17	2123	171	305	25	193	8	74	4	71	8	8	8
18	2057	171	300	21	224	10	78	5	62	20	3	5
19	1962	161	289	20	182	8	44	4	26	5	6	0
20	2896	217	389	33	256	6	91	5	83	24	10	0
21	1379	146	192	10	155	6	35	4	37	4	5	1
22	936	104	130	9	96	3	19	2	30	6	9	0
23	1169	105	177	15	162	8	41	6	33	18	5	2
24	1131	101	179	16	136	5	26	0	18	7	4	1
25	777	152	75	8	107	11	18	3	36	13	2	0
26	1123	107	177	11	164	8	39	2	59	10	7	0
27	1334	163	210	17	194	8	54	8	54	19	6	1
28	1392	191	157	15	140	15	35	8	53	8	14	5
29	NF	NF	NF	NF	NF	NF	NF	NF	NF	NF	NF	NF
30	34	10	2	0	NF	NF	NF	NF	NF	NF	NF	NF
31	1	0	2	2	NF	NF	1	0	0	1	NF	NF
32	66	11	1	0	5	1	2	0	2	0	NF	NF
33	940	247	81	12	86	7	14	4	15	15	5	1
34	NF	NF	NF	NF	NF	NF	NF	NF	NF	NF	NF	NF
35	NF	NF	NF	NF	NF	NF	NF	NF	NF	NF	NF	NF
36	NF	NF	NF	NF	NF	NF	NF	NF	NF	NF	NF	NF
37	NF	NF	NF	NF	NF	NF	NF	NF	NF	NF	NF	NF
38	NF	NF	NF	NF	NF	NF	NF	NF	NF	NF	NF	NF
39	NF	NF	NF	NF	NF	NF	NF	NF	NF	NF	NF	NF
W	768	71	290	38	67	0	8	0	5	1	2	0
Z	23533	2134	3409	363	2201	85	2260	288	742	269	127	56
MT	1	0	0	0	0	0	0	0	0	0	0	0
Total	274537	22686	36963	3413	21224	761	14846	1300	5040	2044	866	389

NF-Not found.

mononucleotides, dinucleotides, trinucleotides, tetranucleotides, pentanucleotides and hexanucleotides composed of 297223 (77.38%), 40376 (10.51%), 21985 (5.72%), 16146 (4.20%), 7084 (1.84%), 1255 (0.32%) respectively (Table 2). The SSRs identified in this study can be used precisely for diversity analysis in chicken. In GRCg7b, the most frequent motifs in mono, di, tri, tetra, penta and hexa nucleotide repeats were T (129407 times; 45.85%), AT (9466 times; 25.25%), TTG (3112 times; 14.04%), AAAC (2523 times; 16.39%), AAAAC (321 times; 5.29%) and AAAAGA (42 times; 4.07%) respectively.

The occurrences of longer repeats are less common as in other species (Temnykh *et al*., 2001; Grover *et al*., 2007) while smaller repeats are predominant and this is because the mutation rates are higher in case of longer repeats and therefore less stable (Wierdl *et al*., 1997). The point mutation rates in different species determine the length distribution of microsatellites (Bell and Jurka, 1997; Kruglyak *et al*., 1998). The microsatellite arises due to the replication mistakes, whereas the point mutations whose rate varies within the genomes (Webster *et al*., 2006) breaks the repeats into two or more shorter ones and act in the opposite direction. In the perfect SSRs that contains the tandem repeats of the motifs increase in size due to replication events which then forms the compound SSRs due to mutation of the nucleotides resulting in variable length of the microsatellites over a long period of time (Messier *et al*., 1996; Primmer and Ellegren, 1998). The mono nucleotide repeats are more common than di, tri, penta, tetra and hexa nucleotide repeats across the genome (Table 1) which is in accordance with the previous report (Arora *et al*., 2013). In general, dinucleotide repeats are more common in the eukaryotic genome (Kariin and Burge, 1995; Shioiri and Takahata, 2001) and have been used in many livestock species for assessing the variability (Chatterjee *et al*., 2010a), genetic relatedness (Chatterjee *et al*., 2010b), diversity analysis and QTL mapping (Rajkumar *et al*., 2007; Atzmon *et al*., 2008). It is concluded that precise SSR microsatellites are spread over all the chromosomes based on the analysis of recently published chicken genome assembly (GRCg7b), which may be used for selection of population specific SSRs for genetic diversity and mapping studies in chicken.

CONCLUSION

From our studies, it is concluded that precise SSR microsatellites spread over all the chromosomes in the recently published chicken genome assembly (GRCg7b), which may be used for selection of population specific SSRs for genetic diversity and mapping studies in chicken.

Conflict of interest

All authors declare that they no conflict of interest.

REFERENCES

Arora, V., Iquebal, M.A., Rai, A., Kumar, D. (2013). In silico mining of putative microsatellite markers from whole genome sequence of water buffalo (*Bubalus bubalis*) and development of first BuffSatDB. BMC Genomics. 14: 1-8.

Atzmon, G., Blum, S., Feldman, M., Cahaner, A., Lavi, U., Hillel, J. (2008). QTLs detected in a multigenerational resource chicken population. Journal of Heredity. 99: 528-538.

Beier, S., Thiel, T., Münch, T., Scholz, U., Mascher, M. (2017). MISA-web: A web server for microsatellite prediction. Bioinformatics. 33: 2583-2585.

Bell, G.I. and Jurka, J. (1997). The length distribution of perfect dimer repetitive DNA is consistent with its evolution by an unbiased single-step mutation process. Journal of Molecular Evolution. 44: 414-421.

Chatterjee, R., Sharma, R.P., Bhattacharya, T.K., Niranjan, M., Reddy, B.L. (2010a). Microsatellite variability and its relationship with growth, egg production and immunocompetence traits in chickens. Biochemical Genetics. 48: 71-82.

Chatterjee, R.N., Bhattacharya, T.K., Dange, M., Rajkumar, U. (2010b). Assessment of genetic relatedness of crossbred chicken populations using microsatellite markers. Biochemical Genetics. 48: 727-736.

Grover, A., Aishwarya, V., Sharma, P.C. (2007). Biased distribution of microsatellite motifs in the rice genome. Molecular Genetics and Genomics. 277: 469-480.

Gupta, P.K., Balyan, H.S., Sharma, P.C., Ramesh, B. (1996). Microsatellites in plants: A new class of molecular markers. Current Science. 45-54.

Kariin, S. and Burge, C. (1995). Dinucleotide relative abundance extremes: A genomic signature. Trends in Genetics. 11: 283-290.

Katti, M.V., Ranjekar, P.K., Gupta, V.S. (2001). Differential distribution of simple sequence repeats in eukaryotic genome sequences. Molecular Biology and Evolution. 18: 1161-1167.

Kruglyak, S., Durrett, R.T., Schug, M.D., Aquadro, C.F. (1998). Equilibrium Distributions of Microsatellite Repeat Length Resulting from a Balance Between Slippage Events and Point Mutations. Proceedings of the National Academy of Sciences. 95: 10774-10778.

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

Primmer, C.R. and Ellegren, H. (1998). Patterns of molecular evolution in avian microsatellites. Molecular Biology and Evolution. 15: 997-1008.

Rajkumar, U., Gupta, B.R., Ahmed, N., Venktaramaiah, A., Reddy, A.R. (2007). Genetic variation and genetic diversity in chicken populations using microsatellite assay. Indian Journal of Animal Sciences. 77: 1194-1198.

Sawaya, S., Bagshaw, A., Buschiazzo, E., Kumar, P., Chowdhury, S., Black, M.A., Gemmell, N. (2013). Microsatellite tandem repeats are abundant in human promoters and are associated with regulatory elements. PloS One. 8: e54710.

Shioiri, C. and Takahata, N. (2001). Skew of mononucleotide frequencies, relative abundance of dinucleotides and DNA strand asymmetry. Journal of Molecular Evolution. 53: 364-376.

Temnykh, S., DeClerck, G., Lukashova, A., Lipovich, L., Cartinhour, S., McCouch, S. (2001). Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations and genetic marker potential. Genome Research. 11: 1441-1452.

Tóth, G., Gáspári, Z. And Jurka, J. (2000). Microsatellites in different eukaryotic genomes: survey and analysis. Genome Research. 10: 967-981.

Webster, M.T., Axelsson, E., Ellegren, H. (2006). Strong regional biases in nucleotide substitution in the chicken genome. Molecular Biology and Evolution. 23: 1203-1216.

Wierdl, M., Dominska, M., Petes, T.D. (1997). Microsatellite instability in yeast: Dependence on the length of the microsatellite. Genetics. 146: 769-779.