Taking advantage of genomic data to develop reliable microsatellite loci in Trypanosoma brucei

Chaz Hyseni

Molecular Ecology Resources

1

Taking advantage of genomic data to develop reliable microsatellite loci in *Trypanosoma brucei*.

Mark Sistrom[1], Richard Echodu[2], Chaz Hyseni[1], John Enyaru[3,] Serap Aksoy[4], Adalgisa Caccone[1]

[1] Department of Ecology and Evolutionary Biology, Yale University, 21 Sachem Street, 06511, New Haven, CT, USA

[2] Faculty of Science, Gulu University, Gulu, Uganda

[3] School of Biological Sciences Makerere University, Uganda

[4] Department of Epidemiology of Microbial Diseases, Yale University School of Public Health, New Haven, CT, USA

Keywords: microsatellites, population genetics, pathogens, Africa

Abstract

Taking advantage of the annotated genome of *Trypanosoma brucei*, a unicellular eukaryotic parasite which causes sleeping sickness across sub-Saharan Africa, this study develops 11 variable microsatellite loci, which reliably amplify in poor and unknown quality DNA samples. We provide insights into effective screening methods to develop microsatellite loci from genomic resources and a set of markers to study regional population variation in this important parasite.

Introduction

The use of genomic data and resources to develop microsatellite markers for population-scale studies is rapidly becoming standard procedure, as it is increasingly cost- and time-effective, especially in taxa for which marker development has been traditionally challenging (Abdelkrim 2009; Meglécz *et al.* 2010; Santana 2010). Representing such a challenging organism is *Trypanosoma brucei*, a eukaryotic, single cell parasite, whose distribution across sub-Saharan Africa is restricted by that of its host, the tsetse fly (*Glossina sp*.) (Brun *et al.* 2010). *Trypanosoma brucei* is divided into three subspecies based on human infectivity and geographic origin, not necessarily reflecting genetic or evolutionary distinction. Relationships between and within these subspecies are incompletely understood (Balmer *et al.* 2011; Gibson 1986; Koffi *et al.* 2009). *Trypanosoma brucei brucei* (*Tbb*) causes disease in animals but is unable to infect humans; *T. b. rhodesiense* (*Tbr*) and T. *b. gambiense* (*Tbg*) causes human African trypanosomiasis (HAT in eastern and southern (*Tbr*) and western (*Tbg*) Africa, which causes significant deaths and risk for epidemics across 36 countries (Hotez & Kamath 2009; W.H.O. 2002). There are no vaccines for prevention of HAT and high toxicity is associated with the drugs used for treatment (Brun *et al.* 2010; Jannin& Cattand 2004). The diagnostic feature of *Tbr* is the presence of a serum resistance associated (SRA) gene (Gibson 2005), which has been shown to confer *Tbr* with its ability to infect humans (Van Xong *et al.* 1998). Aside from this crucial functional distinction, previous studies have shown a lack of differentiation between *Tbb* and *Tbr* at neutral loci (Balmer *et al.* 2011).

Population genetic studies have yielded insights into the evolutionary processes that shape *T. brucei* across its distribution and provided important context for HAT epidemiology (Balmer *et al.* 2011; Koffi *et al.* 2009; MacLeod 2001; Symula *et al.* 2012). However, the processes affecting *T. brucei* at regional and local scales, which are important in the context of regional outbreaks, are understudied. A challenge for studying regional and local scales in *T. brucei* is that field samples are difficult to acquire and obtained as small amounts of blood from infected human and livestock hosts. Because these samples have highly variable trypanosome concentration and are contaminated with host cells, DNA extractions are difficult to quantify (Ahmed *et al.* 2011). It is therefore crucial that primers and conditions for the amplification of genetic markers be reliable. While microsatellite loci have been developed for *T. brucei* in the past (Balmer *et al.* 2006; Biteau *et al.* 2000; Truc *et al.* 2002), many of these loci do not have perfect repeats and thus potentially do not evolve in a stepwise manner. Additionally, reliable amplification using template DNA of unknown quality has been unacceptably poor.

Here we use the *Tbb* genome TREU927/4 (Berriman *et al.* 2005) to identify microsatellite markers for population genetic investigations in *T. brucei* and test the reliability of amplification in these loci, along with their utility in two populations of *T. brucei* in Uganda.

Materials and Methods:

Initial screening of primers to evaluate amplification in *T. brucei* was conducted on five DNA samples from cryopreserved laboratory isolates (Appendix 1) extracted using standard phenol-chloroform protocols. These samples were quantified using a Qubit 2.0 flourometer (Life Technologies, USA) to verify template concentration and

confirm their utility as positive controls. We sampled 51 isolates from Kaberamaido (Human and bovine) and 26 isolates from Tororo (human, zebu, hippo, and bovine) (Appendix 2) collected as two drops of blood on FTA cards (Whatman, UK). Whole genomic DNA was extracted using a Qiagen QiaAMP micro DNA kit (Qiagen, Germany), according to the manufacturer's guidelines.

We used the TREU927/4 *Tbb* genome (Berriman *et al.* 2005) as reference data for the detection of microsatellite markers. The eleven megabase chromosomes of the *Tbb* genome were analyzed as a separate fasta files using MISA (http://pgrc.ipk-gatersleben.de/misa/) under default settings. We employed MISA to design primers using its Primer3 (Rozen 2000) module around suitable microsatellite loci. We selected a subset of 50 markers for further analysis based on repeat number, variation of expected product lengths to enable multiplexing, and a range of different motifs (Appendix 3).

Due to high attrition in the 50 markers initially tested, we used a secondary marker identification approach. First, we used SciRoKo v3.4 (Kofler et al. 2007) to find perfect di-, tri- and tetra-nucleotide repeats at least 40-45 bp long (≥20 di-, ≥15 tri- and ≥10 tetra-nucleotide repeats) with ≥200 bp of flanking sequence on either side. The resulting sequences (microsatellite plus flanks) were analyzed in MsatCommander v1.0.8 (Faircloth 2008) to identify the markers that allowed for primer design. Primer design was conducted using modified settings in the Primer3 module of Msatcommander. Modifications were: GC content range - 45-75% and optimal melting temperature = 62C (minimum 60C and maximum 65C). Primers were then aligned to the TREU 927/4 genome to eliminate markers that aligned to multiple sites, and to confirm expected genomic locations and product length, using the MUSCLE (Edgar 2004) plugin in

Geneious v5.6.4. (Biomatters, New Zealand). This filtering strategy led to the selection of a subset of 20 loci for further analysis based on repeat number, variation of expected product lengths to enable multiplexing, and inclusion of a range of different repeat motifs (Appendix 4).

Primers were initially amplified in the five isolate samples using a Qiagen type-it microsatellite PCR kit (Qiagen, Germany), using 4ul of PCR mix, 0.5ul of each primer and approximately 100ng of template gDNA. PCR reactions were conducted using an Eppendorf mastercycler pro thermocycler (Eppendorf, Germany) and a touch-down PCR protocol with an initialization step of 95C for 4 minutes, followed by ten touch down cycles of 95C for 30 seconds, 60-50C for 25 seconds and 72C for 30 seconds, an additional 30 cycles of 95C for 30 seconds, 50C for 25 seconds and 72C for 30 seconds, and a final extension step of 72C for 20 minutes. Detection of product in these five samples was confirmed using a 1.5% agarose electrophoresis gel.

Markers which reliably amplified in the laboratory isolates were screened on field samples to confirm amplification in lower quality DNA samples. The same PCR protocols were performed for field samples with the exception that approximaltey 500ng of gDNA template was used – as the vast majority of DNA in each sample was likely to be host DNA. Products were cleaned using an ExoSAP PCR cleanup kit (Affymetrix, USA) according to the manufacturer's instructions, and analyzed using an ABI 3730xl DNA analyzer (Life Technologies, USA).

Amplification of a scorable product was verified by eye for each locus and allele size calling was conducted using GeneMarker v2.2.0 (SoftGenetics, USA). Raw allele sizes were exported from GeneMarker and Tandem v1.0.8 (Matschiner 2009) was used to

create bin sets for the data and allocate individual raw allele scores into bins. Genepop v4.0.10 (Rousset & Rousset 2008) was used to calculate number of alleles (N_a), observed (H_o) and expected (H_e) heterozygosity and to conduct tests of linkage disequilibrium in the two test populations using 10,000 dememorizations and 1,000 batches.

Results and Discussion

Initial screening detected 258 microsatellite loci following filtering and primer design. Even with the ability to screen the entire *T. brucei* genome, this discovery rate of microsatellites is extremely low when compared with typical discovery rates in other organisms for which there is often significantly less genomic reference data (e.g. 454 shotgun sequences - (Lepais & Bacles 2011; Malausa 2011)). This may be explained by the small genome size of *T. brucei* (26 mb) and the low proportion of non-coding regions across its genome, meaning that fewer suitable microsatellite markers are present than in other organisms. Additionally, the repetitive nature of many regions within the *T. brucei* genome (Berriman *et al.* 2005) suggests that many of the microsatellites present are in regions where the flanking sequences are unsuitable for the design of primers.

Of the 50 loci selected under MISA's default settings, 22 showed reliable amplification in the 5 laboratory isolate samples and were therefore screened across the field samples. Unfortunately, none of these loci yielded reliable amplification in field isolates. The second round of 20 markers designed under more stringent parameters yielded a total of 14 loci which reliably amplified across the laboratory isolate samples. Eleven of these loci yielded a scorable PCR product and were tested in field samples. When tested across two populations, 9 showed an amplification rate of over 50% in the field samples. Details of these eleven loci are shown in Table 1. Table 2 provides diversity estimates for each locus.

The more intensive screening undertaken during the second round of marker selection excluded the majority of makers for which amplification was unreliable with our unknown and likely poor template quality and quantity. Many of the microsatellites detected in the second round of primer design were removed due to their location in repetitive variable surface glycoprotein pseudo-gene libraries (Pays *et al.* 2004). This suggests that the success of amplification followed by subsequent failure to yield a scorable product in many of the primer pairs designed in the first round of screening could be due to primers binding to multiple sites across these repetitive and highly variable regions of the *T. brucei* genome. This highlights the importance of using stringent settings for marker screening in challenging organisms.

A total of 82 alleles (1-9 alleles per locus per site) were detected across the 11 loci. The locus Tryp 66 was invariant in the Tororo sample set, corresponding with a low amplification rate, however this locus was variable in the Kaberamaido sample set, suggesting it may have utility in multi-population studies. Higher allelic diversity observed in the Tororo samples, despite the lower sample size at this site, may be driven by the higher diversity of hosts from which samples were obtained, a hypothesis requiring further testing with an expanded dataset.

Most loci deviated from HWE in both populations (Table 2), a not unexpected finding for this organism. This organism has a high prevalence of clonal reproduction (MacLeod 2001; MacLeod *et al.* 2000), which leads to HWE violations and a heterozygote deficiency. No pairwise comparisons of loci showed significant linkage disequilibrium.

Our study highlights how intensive screening of loci *in silico* can significantly improve the efficiency over the use of a standalone microsatellite discovery method. These loci will be useful for both the analysis of the fine scale patterns of genetic differentiation in *T. brucei* to gain insights into the currently unpredictable nature of regional outbreaks associated with *Tbr* in discrete foci in western Uganda (Berrang-Ford *et al.* 2006) and for broader population genetic studies of *T. brucei* in other regions (Beadell et al., 2010).

Acknowledgements

This work was funded by an NIH R21 grant AI094615-01 awarded to A.C. and S.A. We thank Oliver Balmer for advice on manuscript preparation. Ethical clearance for samples used was obtained from the Ministry of Health Uganda Government with permission from Uganda National Council of Science and Technology under Project-HS361.

References

Abdelkrim J (2009) Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. *BioTechniques* **46**, 185.

Ahmed H, MacLeod E, Hide G, Welburn S, Picozzi K (2011) The best practice for preparation of samples from FTA(R)cards for diagnosis of blood borne infections using African trypanosomes as a model system. *Parasites & Vectors* **4**, 68.

Balmer O, Beadell JS, Gibson W, Caccone A (2011) Phylogeography and Taxonomy of *Trypanosoma brucei*. *PLoS Negl Trop Dis* **5**, e961.

Balmer O, Palma C, Macleod A, Caccone A (2006) Characterization of di-, tri- and tetranucleotide microsatellite markers with perfect repeats for Trypanosoma brucei and related species. *Molecular Ecology Notes* **6**, 508-510.

Berrang-Ford L, Berke O, Abdelrahman L, Walter-Toews D & McDermott J (2006) Spatial Analysis of sleeping sickness, southeastern Uganda 1970-2003. *Emerg Infect Dis* 12(5): 813-820

Berriman M, Ghedin E, Hertz-Fowler C*, et al.* (2005) The Genome of the African Trypanosome Trypanosoma brucei. *Science* **309**, 416-422.

Biteau N, Bringaud F, Gibson W, Truc P, Baltz T (2000) Characterization of Trypanozoon isolates using a repeated coding sequence and microsatellite markers. *Molecular and Biochemical Parasitology* **105**, 187-202.

Brun R, Blum J, Chappuis F, Burri C (2010) Human African trypanosomiasis. *The Lancet* **375**, 148-159.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792-1797.

Faircloth BC (2008) msatcommander: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular Ecology Resources* **8**, 92-94.

Gibson WC (1986) Will the real Trypanosoma b. gambiense please stand up. *Parasitology Today* **2**, 255-257.

Gibson WC (2005) The SRA gene: the key to understanding the nature of Trypanosoma brucei rhodesiense. *Parasitology* **131**, 143.

Guo SW, Thompson EA (1992) Performing the Exact Test of Hardy-Weinberg Proportion for Multiple Alleles. *Biometrics* **48**, 361-372.

Hotez PJ, Kamath A (2009) Neglected Tropical Diseases in Sub-Saharan Africa: Review of Their Prevalence, Distribution, and Disease Burden. *PLoS Negl Trop Dis* **3**, e412.

Huelsenbeck JP, Andolfatto P (2007) Inference of Population Structure Under a Dirichlet Process Model. *Genetics* **175**, 1787-1802.

Jannin J, Cattand P (2004) Treatment and control of human African trypanosomiasis. *Current Opinion in Infectious Diseases* **17**, 565-571.

Koffi M, De Meeûs T, Bucheton B*, et al.* (2009) Population genetics of Trypanosoma brucei gambiense, the agent of sleeping sickness in Western Africa. *Proceedings of the National Academy of Sciences* **106**, 209-214.

Kofler R, Schlötterer C, Lelley T (2007) SciRoKo: A new tool for whole genome microsatellite search and investigation. *Bioinformatics* **23**, 1683-1685

Lepais O, Bacles CFE (2011) Comparison of random and SSR-enriched shotgun pyrosequencing for microsatellite discovery and single multiplex PCR optimization in Acacia harpophylla F. Muell. Ex Benth. *Molecular Ecology Resources* **11**, 711-724.

MacLeod A (2001) The detection of geographical substructuring of Trypanosoma brucei populations by the analysis of minisatellite polymorphisms. *Parasitology* **123**, 475.

MacLeod A, Tweedie A, Welburn SC*, et al.* (2000) Minisatellite marker analysis of Trypanosoma brucei: Reconciliation of clonal, panmictic, and epidemic population genetic structures. *Proceedings of the National Academy of Sciences* **97**, 13442-13447.

Malausa T (2011) High-throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries. *Molecular Ecology Resources* **11**, 638.

Matschiner M (2009) TANDEM: integrating automated allele binning into genetics and genomics workflows. *Bioinformatics* **25**, 1982.

Meglécz E, Costedoat C, Dubut V*, et al.* (2010) QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics* **26**, 403-404.

Pays E, Vanhamme L, Pérez-Morga D (2004) Antigenic variation in Trypanosoma brucei: facts, challenges and mysteries. *Current Opinion in Microbiology* **7**, 369-374.

Rousset F, Rousset (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources* **8**, 103.

Rozen S (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods in molecular biology* **132**, 365.

Santana QC (2010) Microsatellite discovery by deep sequencing of enriched genomic libraries. *BioTechniques* **46**, 217.

Truc P, Ravel S, Jamonneau V, N'Guessan P, Cuny G (2002) Genetic variability within Trypanosoma brucei gambiense: evidence for the circulation of different genotypes in human African trypanosomiasis patients in Côte d'Ivoire. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **96**, 52-55.

Van Xong H, Vanhamme L, Chamekh M*, et al.* (1998) A VSG Expression Site–Associated Gene Confers Resistance to Human Serum in Trypanosoma rhodesiense. *Cell* **95**, 839-846.

W.H.O. (2002) World Health Report 2002. World Health Organization, Geneva.

Locus	Forward Primer	Reverse Primer	Motif	Size range	Chromosome	Start	End
Tryp67	GTTGCTGAGGTGCAACTGG	GTCGTCAGGCACCAAAACG	(GTT)	151-178	7	555,385	555,536
Tryp65	GGAGGTAAACTTGATTCGGGTG	ACGACAACAGCGACAAAGC	(ATT)	207-234	9	1,136,322	1,136,513
Tryp54	AGTCGGCGTGATGGTACTC	TTCAGCCCACAAACAACCG	(AAAT)	144-176	10	3,222,378	3,222,497
Tryp59	GAGGCAATCGCAGTGTGTG	CGCACGTTTCACCATCCTC	(GT)	209-225	9	2,303,901	2,304,108
Tryp61	ACTCGCGACAGACCATGAG	ACAGGAGAGTGTTGTGAGTG	(ATT)	179-215	11	4,952,204	4,952,417
Tryp51	TGACCCGTGAGAAGTGAAC	GCGCATCTACAGGCATAGAC	(ATT)	187-238	9	2,210,345	2,210,491
Tryp53	GTACAGCCACGTGCAAACC	TGTACACAATCGGGTGGATG	(AC)	200-254	7	777,975	778,203
Tryp66	TCCTCGTACCTTTTCTCTCAC	ACGAAATTTAGGTGTGAAAGCTG	(ATT)	384-396	5	74,638	74,802
Tryp55	AATTCAACCCCAACAGCCC	CTCGTTCAATGACTTGCCCC	(GT)	208-246	5	978,318	978,479
Tryp52	GCATCATTGACGTCGACCC	TAACAACCACTGGGACCGC	(GT)	201-231	11	4,353,481	4,353,663
Tryp62	AAGGCGACCAACTTCAACC	GTTGTCATCGGCTTGCTCC	(AC)	153-177	11	2,146,659	2,146,793

Table 1: Summary of marker properties. Chromosome indicates the chromosome on which the marker is found and Start and End indicate its location on the TREU927 reference genome.

	Kaberamaido							Tororo						
Marker	*N*	*Ar*	N_a	H_e	H_o	F_{is}	*P*	*N*	*Ar*	N_a	H_e	H_o	F_{is}	*P*
Tryp51	29	56.9	7	0.28	0.16	0.43	>0.01	26	96.3	9	0.80	0.48	0.40	>0.01
Tryp52	33	64.7	9	0.45	0.47	-0.05	>0.01	26	96.3	8	0.75	0.74	0.01	>0.01
Tryp53	33	64.7	5	0.36	0.47	-0.33	0.07	24	88.9	9	0.63	0.67	-0.06	0.02
Tryp54	34	66.7	2	0.33	0.35	-0.06	1.00	26	96.3	5	0.52	0.33	0.36	>0.01
Tryp55	34	66.7	8	0.33	0.04	0.88	>0.01	27	100.0	7	0.65	0.59	0.09	0.01
Tryp59	28	54.9	4	0.31	0.45	-0.49	>0.01	24	88.9	3	0.29	0.19	0.37	0.05
Tryp61	21	41.2	4	0.24	0.12	0.51	>0.01	12	44.4	3	0.25	0.00	1.00	>0.01
Tryp62	27	52.9	3	0.08	0.00	1.00	>0.01	27	100.0	6	0.66	0.85	-0.29	>0.01
Tryp65	39	76.5	3	0.22	0.14	0.39	0.02	26	96.3	4	0.61	0.78	-0.28	>0.01
Tryp66	5	9.8	3	0.06	0.04	0.38	0.24	19	70.4	1	0.00	0.00	NA	NA
Tryp67	34	66.7	3	0.04	0.02	0.50	0.02	26	96.3	6	0.71	0.56	0.22	>0.01

Table 2: Summary of marker properties and variation for data separated into districts of collection. N is the number of individuals successfully amplified for each locus, Amplification rate is the proportion of successful amplifications. Ho is the observed level of heterozygosity, He is the expected level of heterozygosity under Hardy-Weinberg equilibrium conditions, Fis is the Fisher's exact test score for each locus and P is the associated P value for this score.

Isolate	Subspecies	Country	Host	Date collected
LVBG3N	Tbb	Kenya	Bovine	1980
STIB213	Tbb	Tanzania	Hyena	1971
1829 (Aljo)	Tbg1	D.R. Congo	Human	1970
TH126	Tbg2	Ivory Coast	Human	1978
LVH56	Tbr	Kenya	Human	1983
STIB809	Tbr	Ethiopia	Human	1967

Appendix 1: Details of laboratory isolate samples initially used to confirm marker amplification.

17

Sample	Host	Location
KaberamaidoH101,	Human	Kaberamaido
KaberamaidoH103,	Human	Kaberamaido
KaberamaidoH104,	Human	Kaberamaido
KaberamaidoH105,	Human	Kaberamaido
KaberamaidoH106,	Human	Kaberamaido
KaberamaidoH107,	Human	Kaberamaido
KaberamaidoH108,	Human	Kaberamaido
KaberamaidoH110,	Human	Kaberamaido
KaberamaidoH113,	Human	Kaberamaido
KaberamaidoH115,	Human	Kaberamaido
KaberamaidoH116,	Human	Kaberamaido
KaberamaidoH118,	Human	Kaberamaido
KaberamaidoH121,	Human	Kaberamaido
KaberamaidoH122,	Human	Kaberamaido
KaberamaidoH123,	Human	Kaberamaido
KaberamaidoH124,	Human	Kaberamaido
KaberamaidoH125,	Human	Kaberamaido
KaberamaidoH128,	Human	Kaberamaido
KaberamaidoH129,	Human	Kaberamaido
KaberamaidoH130,	Human	Kaberamaido
KaberamaidoH131,	Human	Kaberamaido
KaberamaidoH170,	Bovine	Kaberamaido
KaberamaidoH351,	Human	Kaberamaido
KaberamaidoH353,	Human	Kaberamaido
KaberamaidoH354,	Human	Kaberamaido
KaberamaidoH355,	Human	Kaberamaido
KaberamaidoH356,	Human	Kaberamaido
KaberamaidoH359,	Human	Kaberamaido
KaberamaidoH360,	Human	Kaberamaido
KaberamaidoH361,	Human	Kaberamaido
KaberamaidoH362,	Human	Kaberamaido
KaberamaidoH459,	Human	Kaberamaido
KaberamaidoH461,	Human	Kaberamaido
KaberamaidoH462,	Human	Kaberamaido
KaberamaidoH463,	Human	Kaberamaido
KaberamaidoH465,	Human	Kaberamaido
KaberamaidoH468,	Human	Kaberamaido
KaberamaidoH470,	Human	Kaberamaido
KaberamaidoH476,	Human	Kaberamaido
KaberamaidoH478,	Human	Kaberamaido
KaberamaidoH479,	Human	Kaberamaido
KaberamaidoH483,	Human	Kaberamaido

KaberamaidoH487,	Human	Kaberamaido
KaberamaidoH489,	Human	Kaberamaido
KaberamaidoH490,	Human	Kaberamaido
KaberamaidoH491,	Human	Kaberamaido
KaberamaidoH492,	Human	Kaberamaido
KaberamaidoH494,	Human	Kaberamaido
KaberamaidoH498,	Human	Kaberamaido
KaberamaidoH505,	Human	Kaberamaido
TororoH285, 1721	Hippo	Tororo
TororoH578, 1541	Bovine	Tororo
TororoH579, 0000	Human	Tororo
TororoH582, 1721	Bovine	Tororo
TororoH596, 1721	Human	Tororo
TororoH601, 1631	Human	Tororo
TororoH604, 1631	Human	Tororo
TororoH605, 1631	Human	Tororo
TororoH606, 1571	Human	Tororo
TororoH607, 1721	Human	Tororo
TororoH608, 1541	Human	Tororo
TororoH612, 1631	Human	Tororo
TororoH617, 1541	Human	Tororo
TororoH618, 1631	Human	Tororo
TororoH620, 1631	Human	Tororo
TororoH838, 1541	Human	Tororo
TororoH840, 1541	Human	Tororo
TororoH847, 1541	Human	Tororo
TororoH850, 1541	Human	Tororo
TororoH854, 1541	Zebu	Tororo
TororoH857, 1631	Zebu	Tororo
TororoH858, 1541	Zebu	Tororo
TororoH862, 1541	Zebu	Tororo
TororoH866, 1541	Human	Tororo
TororoH868, 1541	Human	Tororo
TororoH870, 1541	Human	Tororo

Appendix 2: Details of field samples in which markers were screened.

SSR	FORWARD PRIMER1 (5'-3')	Tm(°C)	size	REVERSE PRIMER1 (5'-3')	Tm(°C)	size	PRODUCT1 size (bp)
(AAC)16	ACAAAAGACGACAAAACGGG	60.008	20	TGAGTGTTTGCAGCTGTCCT	59.622	20	100
(AAT)15	GAAAGCCCAAGAgTAAGGGG	60.068	20	TTGCGCTGCTTTATGTTACG	60.038	20	107
(CCCTAA)5	CTAACCCTAACCCtAACCCTAATCC	60.627	25	TTATCGGGTTCAGGTGTTTC	57.517	20	100
(GAGT)7	GCGAAAACTAAATGGCCTGA	60.209	20	ACTCCTTCACTTTCCCCTGT	57.677	20	114
(AAAC)12	AATTTGGTGAAAAGCCAACG	59.975	20	TACTCCTCCACCGCAACTCT	59.867	20	135
(ATATAA)13	GTGCGGAAACCCTATTCAAa	59.938	20	CACGCTCTCTCACACACACA	59.624	20	155
(TATT)10	GAAAAaGAaGCAGCAGCACC	60.14	20	TCCTAATCCAGCTCCCACAC	60.073	20	159
(TTA)22	TTACTGCCCTCATTATAATATTCGT	57.463	25	GGGAGTGTTGTGAGTGTGtGa	59.613	21	163
(TAA)25	CAACAAGTGCAGTTTGAGTGTG	59.454	22	GGtCTTCACCACCACTCTTCTT	59.645	22	168
(TTAGGC)6	GGTAGGTAGGTTAGTTAGGGGG	57.237	22	cCgCTGCATTGCTAAGAAAA	61.381	20	171
(TAT)17	TGCACACTTAcAaCACTCTCCT	57.539	22	tTCGCCTATGGGTGAATAGTT	58.593	21	182
(TTA)24	CCCTCGGTTGAGTGTTGTTT	60.005	20	CAtTTCACCACCATCCAGTG	59.806	20	182
(CATTAT)10	CTCACGCACTCACAGCACTT	60.251	20	TGGTGAAATGGTATGAAAATTGa	59.221	23	184
(TTTA)8	CACGCGTTGTACTTTCCCTT	60.168	20	GCTGGAGGAGAGAAGGTGTG	59.986	20	188
(TCACCT)11	TCATGCTCGGTCTACCCTTC	60.218	20	TCGTCGTTATGTGGAATGGA	59.924	20	191
(TATTTA)6	ATAAAAATCAGCGGCAGCAT	59.71	20	TATTATCATCGCCGTGGACA	59.914	20	193
(TAT)21	CGCATATACACACACTCACAGC	59.298	22	AAAGTGGCCTACATGTGCAA	59.199	20	194
(GGAA)7	CAGATAGCGGAATGGGAAAG	59.662	20	TCCTCCTCCTCCTCCTCTTC	59.883	20	197
(TGTT)9	GCAaCAACGGGAACACTTTT	60.015	20	ACAGAAAGCGAGCGAGAAAG	59.898	20	199
(AC)22	ATGTGGTAAGGCGAAACGTC	60	20	ACCCAATCATTTTCCTTCCC	59.996	20	199
(GT)20	ACTTTTGTTAGCGAGCCGAA	60.018	20	AGACGAAGCCTCGGTATGTG	60.277	20	206
(GCGACT)6	AAGGTCTCTGCGGTTGTGAT	59.727	20	CTCCTCTCAACAAGTCCCCA	60.229	20	202
(TAAA)10	TCGTTGATTGGGACCCTAAG	59.926	20	ACCGGTTTCACCACTTTCAG	60.005	20	211
(TAA)26	GGACAAATCGCCATCAGATT	59.9	20	GaAAAACACGTACGGGAAGG	59.471	20	220
(GGGTTA)9	GTGTTTCGGgTTcAGGtGTT	59.867	20	ATAAGTGCCAGCCTTGTGCt	59.904	20	221
(CA)20	GCACGCATACACAGATGGAG	60.296	20	GGTGTGGGTTACGGTTTGTC	60.134	20	225
(CA)28	CACCTTCAACACATGCAACA	59.137	20	CAGCaACAgGcCACACTAAG	59.51	20	229

20

(AC)31	GCAAACCAGTCGGTTCAGTT	60.156	20	TTTGTACACAATCGGGTGGA	59.816	20	228
(GGAT)8	AAAaGGGAAATGCTGCTGAA	59.823	20	GGAGGAAACAAAGGGGAAAG	59.912	20	229
(AAAT)9	CCAGTTTCCAGCAGGGTAAA	60.103	20	TTGTTATTCGCTACCGGGAC	59.96	20	230
(AAAT)8	CGACGAAAGCGATCATGTAA	59.833	20	TGCACACTTcCCCCTTCTAT	59.55	20	232
(TG)22	GCTGCGACAGTTGTTGGTTA	59.911	20	TGCTCTGCTTGCTTGAAAGA	60.011	20	237
(TG)21	CGTACGGTTGCTAACCCAGT	60.052	20	ACACACACACAcACACACACAC	58.828	22	237
(GTTT)9	TGTGTTTGTGTGTCTGTGCG	60.412	20	GaAATGGTGGGACCAAAAGA	59.767	20	236
(TAT)29	TCAATTTAGTACTCTTCACCACCA	57.88	24	AAGGTTGATAGGGTACGGAAAG	58.585	22	241
(GTTAGG)11	GGAGTTTTtGTGGAaGTGGG	59.425	20	AAGtTGAAACAATCCGcgAC	60.118	20	244
(ACCCTA)15	CGAAACACCTAAGCCCGATA	60.089	20	TTGtTTAGCAGCTGCATTCG	60.155	20	247
(CA)21	GTGCCGGTAGTTGTGATTGA	59.572	20	CACACACTAACGCATACGCA	59.379	20	251
(GTTT)10	GAAGGCATAAGTCGGGTTGA	60.074	20	CGCGACCAACGATAGGTATT	59.982	20	249
(AAAC)10	TGGAAATGACATCAAGGAAACA	60.344	22	ATTTTGTTACACGGCGCCTA	60.508	20	257
(TATG)29	AAAACGATCAGAAATGGCGT	59.574	20	CCGAGAAACAAACACAAGCA	59.881	20	260
(TATT)10	TTTCCAAACACCCACAAACA	59.834	20	GATAgaAgGCGaAGcCaACA	60.352	20	260
(TTTA)15	AATGCAATGCACTGAAGGAA	59.276	20	TTTTGAACTTTGCGCATCTG	59.992	20	264
(CTAACC)16	CCAaCCcTAACCCTAACCcT	59.21	20	TCCCCGAAAAATTTGAACTG	59.91	20	268
(TTG)33	GGAGAAACGGCAACAATCAt	59.939	20	CACACATTTGAATAACGGCG	59.992	20	274
(GGTTAG)10	TTCCCcGAAAAaTTGAACTG	59.91	20	ATACGGAGGAGCtGCGTAAA	59.867	20	275
(ACACAA)19	AACTCCCTCGGGATGATTTT	59.766	20	ATTGAACACGTCGGAGAGGT	59.579	20	276
(AAACA)6	TTCAGCCCACCACAATGTAA	59.964	20	TGCGGATGTGCGTTAATAAG	59.725	20	279
(ACAA)7	ATAACAGGTGGGAaGGGgAA	60.549	20	TTGCTTCAGTCATTTCGCTC	59.152	20	280
(TTTG)7	GCATTTGCATCCATCAAAGTT	59.959	21	GCGGGGaAAACAATCAAATA	59.773	20	274

Appendix 3: Details of markers selected for the first round of development protocols.

21

SSR	FORWARD PRIMER1 (5'-3')	Tm(−C)	size	REVERSE PRIMER1 (5'-3')	Tm(−C)	size	PRODUCT 1 size (bp)
(TCCT)7	GCCTCATGTTAcTtCCCCC	59.351	19	TCAATAGaAACGGAGGGGAT	58.452	20	259
(TAAAT)7	ATCGATAATGCTGCCCAAAA	60.424	20	TCtCGTCCTCTCTTCACTCCA	60.12	21	211
(GTTG)10	CGCGATAAGAAAGGTTCGAG	59.975	20	CCCAGACACGTATGATGTGC	59.992	20	245
(AAAAG)5	GTAGCCATTCGTCCGTCATT	59.962	20	TTTCCATCGTCATGATTCCA	59.859	20	146
(TG)21	tGTTTtGAGGTTGTGGGTGA	59.976	20	TGACTTGCCCCTTTTCAATC	60.051	20	268
(AATG)9	ACAAACACGTGATGGACGAA	60.008	20	CCCCtTTTCAAACGAACAAA	59.947	20	274
(TTTA)7	AACGATTGCTTCCTGATTGC	60.221	20	CGTCCCACAGAGAAGAGAGG	59.978	20	193
(AAAAC)6	TGCCTGAGGAAGAAGCAAAT	59.955	20	TTCCTTTTGGTTTGGTTTGG	59.811	20	269
(AGAAA)6	CATGTTTTGTTGCTGCTGCT	60.058	20	TCAGGCACCACATGACAACT	60.162	20	266
(TTTA)8	GGAATGAGGGAAACAGCAAG	59.67	20	TGTCTCCATTCATTACGGCA	60.073	20	197
(ATT)19	AAATTTGTATTTTAGTTGCAAGCG	58.896	24	ACTTTTCGGGTGTGGCTATG	59.993	20	267
(ATAA)9	AGCGCATACAAAAAGGaTGG	60.096	20	tGTGTGCGTGAGAATTCGTT	60.31	20	143
(TTTG)13	ATCGTTTTCGTTTTtGtCGC	60.117	20	ATTTTTGTGTTCCCGCTCAC	59.978	20	215
(CACT)9	tCcTTTTGATTTCCCACTGC	60.051	20	TGATAGGAGAGTGTTGTGAGtGTG	59.325	24	150
(ACT)22(TAT)16	TTCATTtGTtATTCACCTCTACTCG	58.811	25	AGTGTTGTGAGTGTCCGCAT	59.183	20	236
(AGAA)8	GAAGAGGGACCAAATGTCCA	59.903	20	CGAGTACACACACGGAAACG	60.213	20	232
(ACT)22(TAT)16	TTCATTtGTtATTCACCTCTACTCG	58.811	25	AGTGTTGTGAGTGTCCGCAT	59.183	20	236
(GTTG)10	CGCGATAAGAAAGGTTCGAG	59.975	20	CCCAGACACGTATGATGTGC	59.992	20	245
(TG)29	TCCCCcTACTAGCGGTCTTT	60.089	20	CAACCACCAGCAGGGACTAC	60.567	20	248
(AAT)25	GAGCAGTAATAAtgaTAATGATGgTAA	57.101	27	CACACTCACAACACTCTCCTATCA	59.325	24	254
(ATT)19	CACACACTCTCAACACTGTCCT	57.836	22	TGTACAAGATTGtGGCaAATTGA	60.403	23	258
(TCCT)7	GCCTCATGTTAcTtCCCCC	59.351	19	TCAATAGaAACGGAGGGGAT	58.452	20	259
(TATT)7	tCCCTTtCATTACTTAAGGCAG	57.605	22	AGGCAGAGTTGGTGCTTCAT	59.874	20	259
(TTG)15	GCTCGAGTATGATCCGGTGT	60.104	20	CATAGAGTCGTCAGGCACCA	59.855	20	271
(AATG)9	ACAAACACGTGATGGACGAA	60.008	20	CCCCtTTTCAAACGAACAAA	59.947	20	274

Appendix 4: Details of markers selected for the second round of development protocols.