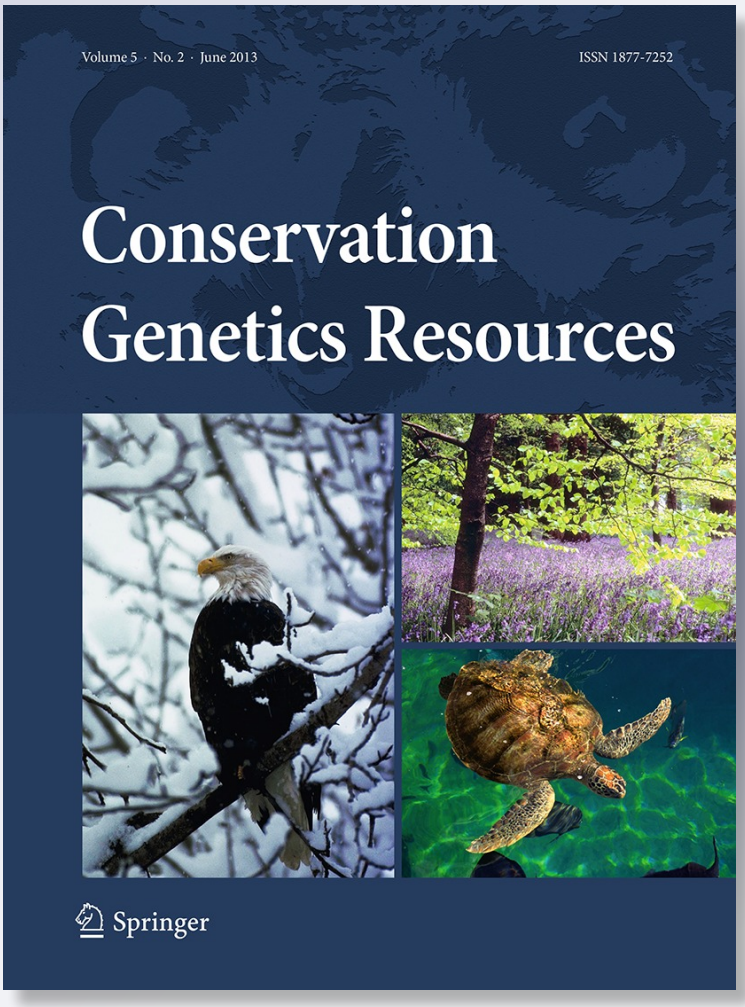
Large-scale isolation of Eastern spiny mouse Acomys dimidiatus *microsatellite loci through GS-FLX 454 titanium sequencing*

Sanad Alfadala, Deborah A. Dawson, Gavin J. Horsburgh, Jerzy M. Behnke, Anna Bajer, Eman M. E. Mohallal, Samy Zalat & Jon Slate

Conservation Genetics Resources

ISSN 1877-7252
Volume 5
Number 2

Conservation Genet Resour (2013)
5:519-524
DOI 10.1007/s12686-012-9842-z



Springer

Conservation Genet Resour (2013) 5:519–524
DOI 10.1007/s12686-012-9842-z

TECHNICAL NOTE

Large-scale isolation of Eastern spiny mouse *Acomys dimidiatus* microsatellite loci through GS-FLX 454 titanium sequencing

Sanad Alfadala · Deborah A. Dawson · Gavin J. Horsburgh · Jerzy M. Behnke · Anna Bajer · Eman M. E. Mohallal · Samy Zalat · Jon Slate

Received: 9 December 2012 / Accepted: 13 December 2012 / Published online: 18 December 2012

Abstract We isolated and characterized Eastern spiny mouse, *Acomys dimidiatus* microsatellite loci. A microsatellite-enriched library was created and *A. dimidiatus* fragments sequenced using 454 sequencing. In total, 1,221 primer-designable microsatellite sequences were identified. We designed primer sets for 20 loci. Loci were characterized in *A. dimidiatus* individuals from a semi-isolated desert wadi (valley) in St. Katherine Protectorate, Egypt. After initial trails, 18 microsatellite loci were genotyped in 67 mice. The number of alleles displayed in the 18 markers ranged from three to nine (mean = 6) with mean expected and observed heterozygosities of 0.63 and 0.65, respectively. All 18 selected loci were in Hardy–Weinberg equilibrium ($P > 0.01$). These markers will be used to investigate the fine-scale spatial patterns of genetic diversity and divergence of *A. dimidiatus* populations. The isolated loci are of potential utility in other murines, including 260 threatened species.

Keywords Eastern spiny mouse · Pyrosequencing · Microsatellite markers · Muridae · Rodent Simple sequence repeats (SSRs)

Electronic supplementary material The online version of this article (doi:10.1007/s12686-012-9842-z) contains supplementary material, which is available to authorized users.

S. Alfadala (✉) · D. A. Dawson · G. J. Horsburgh · J. Slate
Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK
e-mail: bop10sma@sheffield.ac.uk

J. M. Behnke
School of Biology, University of Nottingham, Nottingham NG7 2RD, UK

A. Bajer
Department of Parasitology, Institute of Zoology, University of Warsaw, ul. Miecznikowa 1, 02-096 Warsaw, Poland

E. M. E. Mohallal
Desert Research Centre, 1 Mataf El Matareya St, El Matareya, Cairo, Egypt

S. Zalat
Department of Zoology, Suez Canal University, I smailia, Egypt

S. Zalat
Faculty of Arts and Sciences, Taibah University El-Ula Branch, El-Ula, Saudi Arabia

The Eastern spiny mouse, *Acomys dimidiatus* (Rodentia: Muridae), inhabits adjacent desert wadis (valleys) in St. Katherine Protectorate, Sinai, Egypt but populations differ in parasite burdens (Behnke et al. 2004). The markers we report here will be used to study the fine-scale spatial genetic structure of Eastern spiny mouse populations and how heterozygosity influences host susceptibility to infection. The large number of *A. dimidiatus* microsatellite sequences isolated are likely to be of utility in other species of the same genus such as the endemic *Acomys minous* (Bates 1994). Additionally, by aligning the sequences with their homologues in the house mouse *Mus mus* genome (following the approach of Dawson et al. 2010), markers can be created with enhanced cross-species utility, so enabling studies of conservation genetics in Muridae more broadly, 36 % of which are threatened species (≈260 species; Myers et al. 2006).

A microsatellite-enriched library was constructed from a male spiny mouse (ID: 8183) from Wadi Tlah, St. Katherine Protectorate, Egypt. Genomic DNA was extracted from the mouse tail using an ammonium acetate precipitation method (Nicholls et al. 2000) and the library was made using the

Table 1 Summary of Eastern spiny mouse *Acomys dimidiatus* microsatellite marker isolation

	Singleton	Contig	Total
Number of 454 sequences obtained. (Six different enriched libraries (6 species) were tagged and sequenced together on a quarter of a plate)	–	–	14,585
Number of sequences remaining after the removal of the *Sau-LA/B* linker sequences (Royle et al. 1992) and poly-A tails. Seqman NGen 2.0.0. parameters: MerLength = 5; minMerMatch = 3; MinTrimLength = 15	–	–	14,206 (97.4 %)
Number of sequences assembled in contigs or remaining unassembled as singletons using SeqMan NGen v1.2 (DNAStar, Inc); assembly criteria included minimum overlap match of 81 bases and 90 % similarity for assignment into contigs (with a minimum of two sequences aligned per contig). The minAveLowQual parameter was set to 14	9,576 (67.4 %)	4,630 reads assembled into 1,676 contigs, (32.6 %)	14,206
Number of sequences remaining after Sputnik software was used to search for microsatellites). Sputnik parameter settings were sputnik –u 2 –v 5 –s 20 –p –L 20 –F infile.fas > outfile.fas Sequences with a repeat purity of at least 90 % were retained	4,408 (46.0 %)	1,676 (32.2 %)	6,084 (42.8 %)
Number of sequences remaining after removing sequences with a flanking region of 20 base-pairs or less on both sides flanking the repeat region, which we deemed not primer designable	933 (21.2 %)	288 (17.2 %)	1,221 (20.1 %)
Number of primer sets designed	10	10	20
Number of primer sets amplifying	9	10	19
Number of primer sets polymorphic	9	10	19
Number of primer sets deviating from Hardy–Weinberg equilibrium	1	0	1
Number of primer sets deemed suitable for use in population structure and parentage studies	8[a]	10	18

[a] After preliminary genotyping of 24 individuals, two loci, whose primer sets were designed from singleton sequences, *Adim019* and *Adim020* (EMBL accession numbers: HE994160 and HE994161) were excluded from further use. *Adim019* displayed the same two alleles in all individuals (hence deviating from Hardy–Weinberg equilibrium) and *Adim020* did not amplify a product

The primer set for *Adim019*: (F): [6-FAM]GCATATGGGCAGCATTAAGTAG & (R): TTTCACACGACGGTATTTCC The primer set for *Adim020*: (F): [6-FAM]AAGGCTTGGCCAGTATTAAGC & (R): TGGACAAGCTCCAATCAATG Details of the remaining 18 validated markers are provided in Table 2

enrichment approach of Armour et al. (1994). The library was enriched for the following di- and tetranucleotide microsatellite motifs and their complements separately: $(AC)_n$, $(AG)_n$, and $(GATA)_n$, $(TTTC)_n$, $(GTAA)_n$, and $(CTAA)_n$, which had been denatured and bound to magnetic beads following Glenn and Schable (2005). Following enrichment, the dinucelotide- and tetranucleotide-enriched fragments were PCR amplified separately, in three separate reactions for each, to obtain sufficient DNA (ca 5 µg) for sequencing. Each 25 µl PCR contained 2.0 µl dinucleotide- or tetranucleotide-enriched DNA, 1× reaction buffer (Bioline), 25 µg/ml BSA, 150 µM dNTPs, 0.5 µM *Sau-L-A* linker/primer (Royle et al. 1992), 2.0 mM $MgCl_2$ and 1 unit of DNA *Taq* polymerase (Bioline). The PCR program used was as in Glenn and Schable (2005). The three dinucleotide and three tetranucleotide PCRs were pooled together. The resultant mixed dinucleotide/tetranucleotide-enriched DNA was purified using a QIAquick PCR purification column (Qiagen) and eluted in 40 µl to create a concentration of ca 125 ng/µl. DNA concentration was measured on the Nanodrop 8000 (Thermo Scientific). The pooled PCR-amplified enriched fragments were sequenced without shearing by 454 GS FLX. Six different enriched libraries (6 species) were tagged and sequenced together on a quarter of a plate.

A total of 14,585 sequences was obtained from 454 sequencing. Data were cleaned and assembled using Seqman NGen 2.0.0 (DNASTAR, Inc); for parameter settings see Table 1. Microsatellite repeats were detected using a modified version of SPUTNIK (http://wheat.pw.usda.gov/ITMI/EST-SSR/LaRota (again, see Table 1 for parameter settings)A total of 933 singletons and 288 contigs remained that each contained a microsatellite of at least ten repeat units and had sufficient suitable flanking sequence to allow primers to be designed to amplify the repeat region. The 288 contigs were submitted to the EMBL database (EMBL accession numbers HE994150–HE994159 and HE994162–HE994439). We used Primer3 (Rozen and Skaletsky 2000) to design primer sets from 20 unique sequences (*Adim0001–Adim020;* Table 2). The primer design criteria

Table 2 Characterization of 18 autosomal Eastern spiny mouse (*Acomys dimidiatus*) microsatellite loci [a]

Locus [b]	EMBL accession number	Repeat motif	Primer sequence (5′–3′)	Tm (°C)	Expected allele size of cloned individual 8183 (bp)[c]	Observed genotype of mouse 8183 (bp)	Allele size range (bp)	N_A	H_O	H_E	P_{HWE}	Est. null allele freq
Adim001	HE994142	(TATC)7	F:[6-FAM]CCCTGCATCCTCTAGTCACC	59.7	247	247, 259	243–267	6	0.75	0.75	0.58	– 0.01
			R:AGCAGCCTGGGTTACATGAT	59.6								
Adim002	HE994143	(TATC)11.. (TATC)2.. (TATC)2	F:[HEX]GCAGCAGCTATGTGGACTTG	59.6	200	200, 200	176–208	8	0.71	0.69	0.69	– 0.02
			R:GATGTGTACTGACAGAAGCATGG	59.7								
Adim003	HE994144	(CA)6.. (CA)13	F:[6-FAM]TGTGTTCAGAAATACCACTTCTATTC	57.9 57.8	199	183, 198	180–203	9	0.93	0.83	0.15	– 0.06
			R:CTGGGTTTGGTACCTGGAC									
Adim004	HE994145	(GATA).. (GATA).. (GATA)7.. (GATA)2	F:[6-FAM]GAGTTTGTAGCTCTTAGTGCTGGAC	60.0 60.2	225	222, 222	218–226	3	0.30	0.27	0.61	– 0.04
			R:TGTTCATGGTGAGAACCTCAAG									
Adim005	HE994146	(GATA).. (TG).. (TG)5.. (GATA)11.. (TG).. (TG)3.. (TG)3.. (TG)8	F:[HEX]CGGTAAGACCATAGAAGGTTTG	57.8	245	243, 246	231–251 [f]	9	0.75	0.76	0.95	0.00
			R:AGTTGACGAGTTGACATCAGTG	57.9								
Adim006	HE994147	(CA)17.. (CA)3.. (CA)3(CG)4..(TG)2	F:[HEX]GAAATGTCTGTGTTTGGTCGTG	60.5	147[d]	131, 133	125–139	6	0.81	0.77	0.04	– 0.02
			R:ACCATAGGGATTTCCAACCAG	60.0								
Adim007	HE994148	(CA)12.. (CA)	F:[HEX]CGACTTCTGGCCTCTACATGG	62.1	175	171, 171	163–173	4	0.54	0.60	0.15	0.05
			R:CAGGGCACTGGAATGAAGC	61.8								
Adim008	HE994149	(GTCT)4.. (TATC)12	F:[HEX]GCAATACAGCCGTACCTTGC	60.7	161	142, 155	142–167	5	0.69	0.77	0.13	0.05
			R:TGTCAAATGCATTCATGACAAAC	60.8								
Adim009[e]	HE994150	(TATC)6.. (TATC).. (TATC).. (TATC).. (GTCT)4	F:[HEX]GATCGATCACAATAACTAGGAAAGT	47.8	231	235, 243[e]	235–258	6	0.51	0.57	0.02	0.04
			R:GCGGAAGCTTAGATTTAAAGG	(57.1)[e] 57.4	(235)[e]							
Adim010	HE994151	(TTTTCC)2.. (TTTC)15	F:[6-FAM]AGTGGTGGCCTGCCTTAG	58.8	236	239, 241	223–245	9	0.81	0.83	0.47	0.01
			R:ACAGTGACAGCCAACAGAGC	59.1								
Adim011	HE994152	(TATC)10	F:[6-FAM]TCACAGTACAGTGGGTAATGCAG	60.1	382	378, 382	378–393	5	0.78	0.74	0.71	– 0.03
			R:CAGGAGAATTAAGCATCCAAGG	60.1								
Adim012	HE994153	(GT)9.. (GT)8	F:[HEX]CCAAACTTAGATGTTGACAAGCTG	60.2	299	309, 311	292–311	6	0.78	0.68	0.48	– 0.08
			R:TCAAATCTTCCTATGTTCAATGGAG	60.9								
Adim013	HE994154	(TTTC)5.. (TTTC)3	F:[HEX]CAGTCGAAAGAAGACAGGACAG	59.1	241	235, 243	230–255 [f]	9	0.82	0.83	0.97	0.00
			R:AGCACTTTGCTAACATGCATAAG	58.7								

Table 2 continued

Locus [b]	EMBL accession number	Repeat motif	Primer sequence (5′–3′)	Tm (°C)	Expected allele size of cloned individual 8183 (bp)[c]	Observed genotype of mouse 8183 (bp)	Allele size range (bp)	N_A	H_O	H_E	P_{HWE}	Est. null allele freq
Adim014	HE994155	(AGG)3.. (AGG)2.. (AGG)4.. (AGG)4.. (AGG)2.. (AGG)4.. (AGG)2.. (AGG)3	F:[HEX]CCTAATTTGGTCAGGTGTTGTAGC	61.1	394[d]	296, 296	284–296	3	0.36	0.30	0.25	–
			R:ATGAGCATTGTTTGGCTTGC	61.2								0.10
Adim015	HE994156	(TATC)8.. (TATC)2	F:[6-FAM]CTCTGTCCTGATGGAGGAGTC	58.8	241	222, 222	222–242	6	0.75	0.71	0.02	–
			R:CCAGCCTAGCTAGATGTTTGC	59.2								0.04
Adim016	HE994157	(GT)10.. (GT)3.. (GT)3.. (GT)3	F:[6-FAM]GTAGCCTCGGCTGACCTG	59.5	175	177, 177	177–192	3	0.27	0.24	0.63	–
			R:TGTTGGTCAATACCTCCATATCC	60.0								0.07
Adim017	HE994158	(CA).. (CA)11.. (CA)3.. (CA)	F:[6-FAM]CTTGCAAGCTCATGGGAAG	59.5	131	127, 127	121–127	4	0.30	0.29	1.00	–
			R:GTTCACTCCCACCCTTCTTG	59.6								0.04
Adim018	HE994159	(TATC)11.. (TATC)3.. (TATC)2	F:[HEX]TTCACTCATATGTGGAATTTAGACC	58.5	192	189, 193	185–212	7	0.81	0.73	0.89	–
			R:TCTCTTTGGAACCCTTTAGTCC	58.8								0.03

T_m melting temperature of primer sequence (calculated using Primer3; Rozen and Skaletsky 2000), N_A number of alleles observed, H_O observed heterozygosity, H_E expected heterozygosity, P_{HWE} probability of deviation from Hardy–Weinberg equilibrium (HWE)

[a] Loci were genotyped in 67 *A. dimidiatus* individuals (37 males and 30 females) sampled at the Wadi Tlah, St. Katherine Protectorate, South Sinai, Egypt. Every locus was typed in every individual (following reruns). All loci were autosomal based on the amplification of a proportion of heterozygotes in males and females

[b] The first eight primer sets (*Adim001–Adim008*) were designed from singleton sequences and the remainder from contigs

[c] The expected allele size of cloned individual 8183 was based on the sequenced allele

[d] The smaller difference between the observed and expected allele size for *Adim006* may be due to (singleton) sequencing error; whereas f*or Adim014*, it could be due to or problems with contig assembly or perhaps due to a different locus being amplified to that cloned

[e] The forward primer of locus *Adim009* includes the 5′ pigtail "GATC" (shown underlined). The calculation of the forward primer's melting temperature and the expected (shown in brackets) and observed allele sizes include the 4 bp of the pigtail

[f] Some allele sizes at the indicated locus differed by 1 bp

used included a maximum 0.5 °C difference between the forward and reverse primers, possession of a G/C clamp and a maximum of three consecutive mononucleotide bases.

Loci were initially genotyped in 24 individuals also sampled from Wadi Tlah. For amplification, each reaction contained 1 µl air-dried DNA at ≈10 ng/µl, 1 µl of primer mix (fluorescently labelled forward and reverse) at 0.2 µM and 1 µl QIAGEN multiplex PCR mix (QIAGEN Inc). The PCR program used consisted of an initial denaturation at 95 °C for 15 min, followed by 35 cycles of 30 s at 94 °C, 90 s at 57 °C, and 60 s at 72 °C, followed by a final extension at 60 °C for 30 min. PCR products were diluted 1:358 with double-deionized H_2O before resolving them in an ABI 3730 48-well capillary DNA Analyser (Applied Biosystems, California, USA). Allele scoring was performed using GeneMapper v3.7 (Applied Biosystems, California, USA). Finally, all individuals possessing alleles differing by one base-pair at a locus were genotyped at least twice, along with additional control individuals to check for allele sizing errors. Observed and expected heterozygosities were calculated and null allele frequencies estimated using CERVUS v3.0.3 software (Kalinowski et al. 2007), and deviation from the Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium were assessed with GENEPOP v4.0.10 (Raymond and Rousset 1995; Rousset 2008). Markers amplifying a product of the expected size, which were polymorphic and adhered to HWE, were used to genotype additional mice from the same population, totalling 67 individuals.

After preliminary genotyping of 24 individuals, two loci that were both developed from singletons, *Adim019* and *Adim020* (Table 1), were excluded from further use, as the former displayed the same two alleles in all individuals and the latter did not amplify. The remaining 18 loci (*Adim0001–Adim018*) possessed three to nine alleles (mean = 6) in 67 individuals (Table 2). The allele sizes amplified for each locus matched those expected based on the individual that was cloned and sequenced (± 2 bp), except for *Adim006* and *Adim014* (14 and 101 bp smaller than expected respectively; Table 2). All loci had very low estimated null allele frequencies (below 10 %) and were in HWE ($P > 0.01$; Table 2). Observed heterozygosities ranged from 0.24 to 0.83 (Table 2). Deviation from linkage equilibrium was detected between two pairs of loci: *Adim007–Adim015* ($P < 0.001$), and *Adim005–Adim011* ($P = 0.006$). However, neither pair displayed linkage disequilibrium in a different population ($P > 0.05$; 57 individuals typed; data not shown).

A higher number of validated markers was obtained from contig sequences than singletons (100 vs. 80 %; Table 1). However, the difference in success was relatively low and only ten primer sets of each category were tested; therefore, singleton sequences should not be discarded or overlooked. Many (more) primer-designable microsatellite sequences were obtained from singleton sequences compared to contigs (933 vs. 288), and these may be especially useful for obtaining higher genome coverage or identifying markers in a region of specific interest. We therefore provide the uncharacterized singleton microsatellite sequences as supplementary information (Supplementary Table 1).

Acknowledgments We thank Mohammed Shaker of St. Katherine Protectorate for the permission to collect spiny mice and the staff of the Rangers office for providing vehicles and drivers that enabled access to remote locations. The fieldwork was supported by a grant from the Teaching Quality Enhancement Fund of the University of Nottingham, for which we are most grateful. Marker development and genotyping was supported by the UK Natural Environment Research Council (NERC) and performed at the NERC Biomolecular Analysis Facility at the Universities of Liverpool and Sheffield. We thank Christian Bourne, John Kenny and Xuan Li for 454 sequencing at the NERC Biomolecular Analysis Facility at the University of Liverpool. Andy Krupa provided genotyping advice and Francis Gilbert helped with logistics in Egypt. We thank Terry Burke for project support. SA is funded by the Government of Kuwait and thanks Prof. M. Bou-Resli (retired) and Drs A. Al-Lanqawi, F. Al-Balool, F. Alsidrawi (retired), M. Al-Salameen and S. Al-Mohanna for their endless moral support. Finally, the fieldwork could not have been undertaken without the enthusiastic participation of undergraduate students from the School of Biology at the University of Nottingham.

References

Armour JAL, Neumann R, Gobert S, Jeffreys AJ (1994) Isolation of human simple repeat loci by hybridization selection. Hum Mol Genet 3:599–605

Bates PJ (1994) The distribution of *Acomys* (Rodentia: Muridae) in Africa and Asia. Israel J Zool 40:199–214

Behnke JM, Harris PD, Bajer A et al (2004) Variation in the helminth community structure in spiny mice (*Acomys dimidiatus*) from four montane wadis in the St Katherine region of the Sinai Peninsula in Egypt. Parasitology 129:379–398

Dawson DA, Horsburgh GJ, Küpper C et al (2010) New methods to identify conserved microsatellite loci and develop primer sets of high utility—as demonstrated for birds. Mol Ecol Res 10: 475–494

Glenn TC, Schable NA (2005) Isolating microsatellite DNA loci. Methods Enzymol 395:202–222

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error and increases success in paternity assignment. Mol Ecol 16: 1099–1106

Myers P, Espinosa R, Parr CS, Jones T, Hammond GS, Dewey TA (2006) Family Muridae. The Animal Diversity Web. http://www.arlis.org/docs/vol1/52386062/muridae.html#b4b85fac760ab2d318cdd0cbe91a1662 Accessed 24 October 2012

Nicholls JA, Double MC, Rowell DM, Magrath RD (2000) The evolution of cooperative pair breeding in thornbills *Acanthiza* (Pardalotidae). J Avian Biol 31:165–176

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenism. J Hered 86: 248–249

Rousset F (2008) Genepop ′007: a complete re-implementation of the GENEPOP software for Windows and Linux. Mol Ecol Res 8:103–106

Royle NJ, Hill MC, Jeffreys AJ (1992) Isolation of telomere junction fragments by anchored polymerase chain reaction. Proc R Soc Lond B 247:57–61

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics methods and protocols: methods in molecular biology. Humana Press, Totowa, pp 365–386