

1     **Deceptive combined effects of short allele dominance and stuttering: an example**  
2             **with *Ixodes scapularis*, the main vector of Lyme disease in the U.S.A.**

3  
4     Thierry De Meeûs<sup>1,\*</sup>, Cynthia T. Chan<sup>2,3</sup>, John M. Ludwig<sup>2,4</sup>, Jean I. Tsao<sup>5</sup>, Jaymin Patel<sup>2,6</sup>,  
5                     Jigar Bhagatwala<sup>2,7</sup>, and Lorenza Beati<sup>2</sup>.

6  
7     1. Intertryp, IRD, Cirad, Univ Montpellier, Montpellier, France.

8     2. The U.S. National Tick Collection, Institute for Coastal Plain Science, Georgia Southern  
9             University, Statesboro, GA, USA.

10    3. College of Agricultural & Environmental Sciences, University of Georgia, Athens, GA,  
11             USA.

12    4. Department of Microbiology, University of Georgia, Athens, GA, USA

13    5. Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI, USA

14    6. Division of Hospital Medicine, College of Medicine, University of Florida, Gainesville, FL,  
15             USA.

16    7. Medical College of Georgia, Augusta University, Augusta, GA, USA.

17  
18    \* Corresponding author: [thierry.demeeus@ird.fr](mailto:thierry.demeeus@ird.fr)

19  
20    Keywords: Microsatellite loci, short allele dominance, stuttering, heterozygote deficit,  
21             linkage disequilibrium, curing microsatellite data.

22  
23    Running title: Short allele dominance and stuttering  
24  
25  
26

27 **Abstract**

28 Null alleles, short allele dominance (SAD), and stuttering increase the perceived  
29 relative inbreeding of individuals and subpopulations as measured by Wright's  $F_{IS}$  and  $F_{ST}$ .  
30 Ascertainment bias, due to such amplifying problems are usually caused by inaccurate  
31 primer design (if developed from a different species or a distant population), poor DNA  
32 quality, low DNA concentration, or a combination of some or all these sources of  
33 inaccuracy. When combined, these issues can increase the correlation between  
34 polymorphism at concerned loci and, consequently, of linkage disequilibrium (LD) between  
35 those. In this note, we studied an original microsatellite data set generated by analyzing  
36 nine loci in *Ixodes scapularis* ticks from the eastern U.S.A. To detect null alleles and SAD  
37 we used correlation methods and variation measures. To detect stuttering, we evaluated  
38 heterozygote deficit between alleles displaying a single repeat difference. We  
39 demonstrated that an important proportion of loci affected by amplification problems (one  
40 with null alleles, two with SAD and three with stuttering) lead to highly significant  
41 heterozygote deficits ( $F_{IS}=0.1$ ,  $p$ -value<0.0001). This occurred together with an important  
42 proportion (22%) of pairs of loci in significant LD, two of which were still significant after a  
43 false discovery rate (FDR) correction, and some variation in the measurement of  
44 population subdivision across loci (Wright's  $F_{ST}$ ). This suggested a strong Wahlund effect  
45 and/or selection at several loci. By finding small peaks corresponding to previously  
46 disregarded larger alleles in some homozygous profiles for loci with SAD and by pooling  
47 alleles close in size for loci with stuttering, we generated an amended dataset. Except for  
48 one locus with null alleles and another still displaying a modest SAD, the analyses of the  
49 corrected dataset revealed a significant excess of heterozygotes ( $F_{IS}=-0.07$  as expected in  
50 dioecious and strongly subdivided populations, with a more reasonable proportion (19%)  
51 of pairs of loci characterized by significant LD, none of which stayed significant after the  
52 FDR procedure. Strong subdivision was also confirmed by the standardized  $F_{ST}'$  corrected  
53 for null alleles ( $F_{ST}'=0.19$ ) and small effective subpopulation sizes ( $N_e=7$ ).

54

55

## 56 **Introduction**

57 Null alleles, short allele dominance (SAD) and stuttering are frequent consequences  
58 of poor PCR amplifications, in particular for microsatellite markers. Amplification problems  
59 usually arise when primers are designed by using DNA of a different species or a distant  
60 population, when DNA is degraded, at too low of a concentration (Chapuis and Estoup,  
61 2007), or any combination of these listed causes.

62 Null alleles occur when a mutation on the flanking sequence of the targeted locus  
63 affects the hybridization of the corresponding primer, resulting in amplification failure.  
64 Heterozygous individuals with one null allele falsely appear to be homozygous, while  
65 homozygous individuals for null alleles are considered to be missing data.

66 SAD, also called large allele dropout (Van Oosterhout et al., 2004), known from  
67 minisatellite markers, was also discovered to occur in microsatellite loci (Wattier et al.,  
68 1998). In heterozygous samples, longer alleles are less successfully amplified than shorter  
69 alleles through competition for Taq polymerase. This can lead to misinterpreting  
70 heterozygous individuals as homozygous for the shortest allele (De Meeûs et al., 2007).

71 Stuttering is the result of inaccurate PCR amplification through Taq slippage of a  
72 specific DNA strand. This generates several PCR products that differ from each other by  
73 one repeat and can cause difficulties when discriminating between fake and true  
74 homozygotes, such as heterozygous individuals for dinucleotide microsatellite allele  
75 sequences with a single repeat difference.

76 Allelic dropout is akin to SAD, but occurs randomly to any allele irrespective of its  
77 size, and can affect both alleles of heterozygous individuals.

78 The consequence of these issues is a homozygous excess when compared to the  
79 expected Castle-Weinberg proportions (Castle, 1903; Weinberg, 1908) (classically known  
80 as Hardy-Weinberg, however, more accurately depicted as Castle-Weinberg; because the  
81 former was the first to derive it for two equipotent alleles in 1903 and the latter  
82 generalized the concept in January 1908, prior to Hardy in April 1908 (Hardy, 1908)), as  
83 measured by Wright's  $F_{IS}$  (Wright, 1965). These problems, like all others associated with  
84 amplification, are locus specific (Van Oosterhout et al., 2004; De Meeûs et al., 2007; De  
85 Meeûs, 2018a) and thus lead to locus specific variation (namely, an increase) of  $F_{IS}$ . A less  
86 well known, though well documented (Chapuis and Estoup, 2007; Séré et al., 2017b;  
87 Manangwa et al., 2019) effect of such amplification problems consists of an increase of  
88 Wright's  $F_{ST}$  (Wright, 1965) that is commonly used to measure the degree of genetic  
89 differentiation between subpopulations.

90 While an analytical cure exists for null alleles (Chapuis and Estoup, 2007; Séré et  
91 al., 2017a), such remediation is unavailable for SAD and stuttering. To the best of our  
92 knowledge, the impact of amplification problems on linkage disequilibrium (LD) between  
93 locus pairs has yet to be investigated. Problems with amplification can be expected to  
94 more commonly occur in individuals that display some kind of deviating DNA in terms of  
95 quantity or quality: flanking sequences that have accumulated mutations, samples  
96 containing weak DNA concentration, badly preserved DNA extracts, or a combination of  
97 these different problems. When combined, the effect of the occurrence of null alleles, SAD,  
98 and stuttering may artificially generate a positive correlation between allele occurrences at  
99 affected loci and then increase the perceived LD between them.

100 In this note, we utilize an original data set generated through the analysis of nine  
101 microsatellite loci in *Ixodes scapularis*, sampled across the eastern U.S. to show that the  
102 combined effect of SAD, stuttering, and null alleles can induce an increase in the number  
103 of locus pairs in significant LD. We then propose and test an efficient way to amend such  
104 data.

105

## 106 **Material and Methods**

### 107 *Sampling and DNA extraction*

108 Larvae, nymphs, and adults of *I. scapularis* were sampled indiscriminately from  
109 different sites across the eastern U.S. on different occasions, extending from November  
110 2001 to May 2014, by means of dragging and flagging the vegetation (Figure 1 and Table  
111 1) (Rulison et al., 2013).

112 Gravid females fall on the ground where they lay thousands of eggs at the same  
113 place that hatch as weakly mobile larvae. Larval collections can thus be composed of  
114 clusters of thousands of sisters and brothers within the same subsample (Kempf et al.,  
115 2011). Consequently, to avoid possible Wahlund effects, where a heterozygote deficit  
116 results from the admixture of individuals from genetically distant subpopulations (e.g. see  
117 (De Meeûs, 2018a)) (here families), we removed immature stages from the present study.  
118 The remaining 387 adult ticks were subdivided into cohorts, with each cohort comprised of  
119 samples collected across two consecutive years in the fall, the following winter and spring  
120 across the tick distribution range. This subdivision was based on observations showing  
121 that northeastern adults active in Fall can undergo winter quiescence and resume activity  
122 in spring (Yuval and Spielman, 1990).

123 Many publications have emphasized the importance of mitochondrial clades in  
124 different populations of *I. scapularis* across the U.S. (Norris et al., 1996; Qiu et al., 2002;

125 Sakamoto et al., 2014). Thus, to account for the mitochondrial clade representation and to  
126 (again) avoid possible Wahlund effects, all ticks were assigned clades by phylogenetic  
127 analysis of their 12S rDNA gene sequences. We identified 6 main clades in our dataset,  
128 the previously identified American clade was subdivided in two lineages (AMI and AMII),  
129 and the so-called southern clade was subdivided in 4 lineages (SOI, SOII, SOIII and SOIV)  
130 (Table 1).

131 In conclusion, the combination of Site-Clade-Cohort data defined 61 subsamples  
132 within the 387 individual adult ticks. Overall, 35 subsamples included a small number of  
133 individuals (1-4), 12 subsamples contained at least 10 individuals and 5 subsamples  
134 contained at least 20 individuals (Table 1). Because the smallest subsamples were  
135 expected to exert a negligible weight on our analyses, they were not eliminated.

136 Procedures for all DNA extractions followed modified published protocols (Beati and  
137 Keirans, 2001) with a DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA).

138

### 139 *Selection and characterization of microsatellite markers*

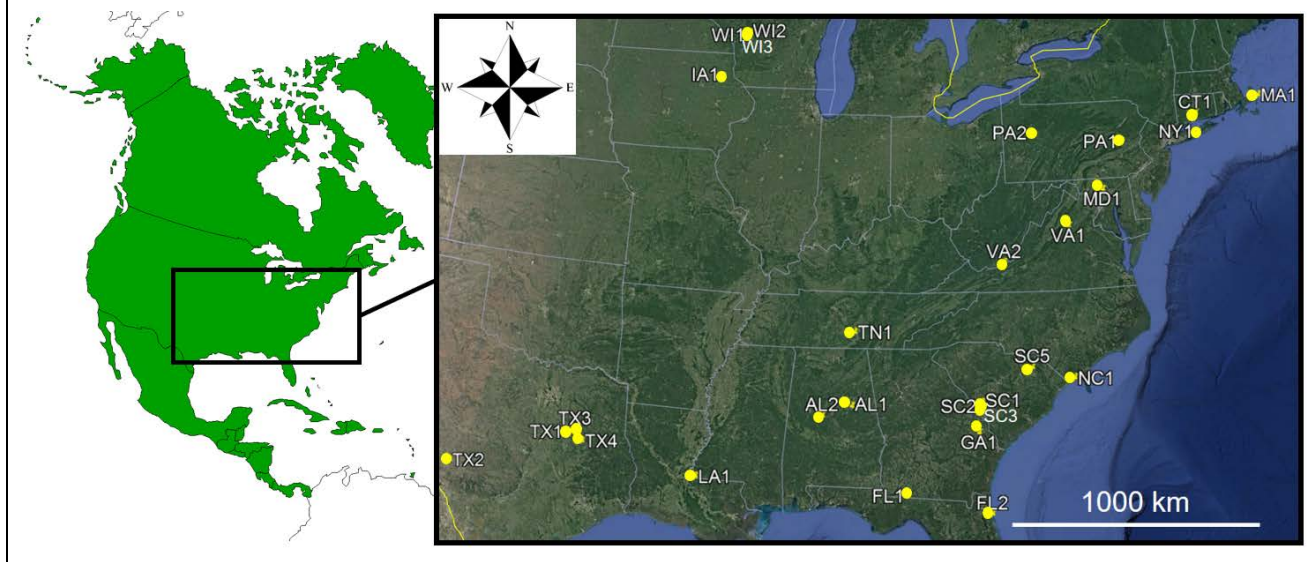
140 Thirteen of the first batches of genome sequences of *I. scapularis* that were  
141 accessioned by VectorBase ([www.vectorbase.org](http://www.vectorbase.org); Giraldo-Calderón *et al.* 2015) in  
142 GenBank (AC205653.1, AC205652.1, AC205650.1, AC205647.1, AC205646.1,  
143 AC205643.1, AC205642.1, AC205641.1, AC205638.1, AC205635.1, AC205634.1,  
144 AC205632.1, AC205630.1) were used to manually detect motifs with at least 6 repeats of  
145 AG, AT, CA, TA, TG, CT, GC, ACG, GTT, TTA, CAC, GAT, and AAAC. Primer pairs were  
146 selected in the flanking regions by using Oligo v.5 (Molecular Biology Insights, Colorado  
147 Springs, CO). DNA, extracted and pooled from six ticks from Connecticut, was used to test  
148 whether the selected primer sets successfully amplified fragments of the expected size.  
149 PCRs were performed using the 5-Prime Master PCR kit (5-Prime, Gaithersburg, MD) and  
150 a single touch-down amplification protocol consisting of 5 min. denaturation at 93°C; 5  
151 cycles: 20 sec. denaturation at 93°C, 20 sec. annealing at 55°C-1.5°C/cycle, 30 sec.  
152 elongation at 72°C; 30 cycles: 20 sec. denaturation at 93°C, 25 sec. annealing at 47°C, 30  
153 sec. elongation at 70°C; final extension at 70°C for 5 min. Amplicons were run on 4% E-  
154 gels (Life Technologies Co., Carlsbad, CA). The risk of having selected primers within  
155 repeated portions of the genome had to be considered due to the fact that large repeated  
156 genomic fragments are known to occur abundantly in *I. scapularis* (Gulia-Nuss et al.,  
157 2016). In order to confirm that the primers were amplifying the targeted loci, the amplicon  
158 of one randomly chosen tick was cloned with a TOPO-TA PCR cloning kit (Life  
159 Technologies Co, Carlsbad, CA) for each locus. Five cloned colonies were picked

160 randomly for each tick and the insert amplified and sequenced (DNA Analysis Facility on  
161 Science Hill, Yale University). Finally, as microsatellite primers are known to amplify  
162 sometimes more than one closely related species, the same set of primers was tested on  
163 DNA samples of *Ixodes ricinus*, *Ixodes pacificus*, and *Ixodes persulcatus* (LB, personal  
164 collection), all taxa belonging to the *I. ricinus* complex of ticks (Keirans et al., 1999).

165 The primer pairs that yielded amplicons of the expected size were then used to  
166 individually amplify a subset of 67 DNA samples from ticks, representative of the  
167 distribution area of *I. scapularis* in USA, and collected by flagging or dragging in Alabama  
168 (10 ticks), Georgia (15), Connecticut (16), Massachusetts (14), New York (2),  
169 Pennsylvania (2), and South Carolina (8). For these amplifications, forward primers were  
170 labeled with fluorescent dyes (Applied Biosystems, Thermo Fisher Scientific, CA) as listed  
171 in Table 2. The amplicons were sent to the DNA Analysis Facility on Science Hill (Yale  
172 University, New Haven, CT) for genotyping. The allele peaks were scored using  
173 GeneMarker (SoftGenetics, State College, PA). All data were recorded in an Excel  
174 spreadsheet for further ease of conversion.

175

176 Figure 1: Sampling sites for *Ixodes scapularis* from the eastern U.S.A. (State codes as in  
177 Table 1).



178

179

180

181  
182  
183

Table 1: State, site, GPS coordinates (decimal degrees), 12S clade membership, date of sampling, corresponding cohort membership and size of *Ixodes scapularis* adult subsamples (*N*) from the eastern U.S.A.

State	Site	Latitude	Longitude	Clade	Date	Cohort	<i>N</i>
Alabama	AL1	33.24	-86.13	AMI	2011 Jan	C7	21
	AL1	33.24	-86.13	SOI	2011 Jan	C7	6
	AL1	33.24	-86.13	AMII	2011 Jan	C7	18
	AL1	33.24	-86.13	SOII	2011 Jan	C7	6
	AL2	32.95	-87.14	AMI	2012 Dec	C9	3
	AL2	32.95	-87.14	AMII	2012 Dec	C9	2
Connecticut	CT1	41.35	-72.76	AMI	2001 Nov	C1	3
	CT1	41.35	-72.76	AMI	2003 Jun	C2	21
Florida	FL1	30.65	-84.21	SOII	2012 Dec	C9	4
	FL2	30.06	-81.37	AMI	2011 Jan	C7	2
	FL2	30.06	-81.37	AMII	2011 Jan	C7	3
	FL2	30.06	-81.37	SOI	2011 Jan	C7	1
	FL2	30.06	-81.37	SOII	2011 Jan	C7	4
Georgia	GA1	32.45	-81.78	AMI	2009 Dec	C5	11
	GA1	32.45	-81.78	SOI	2009 Dec	C5	5
	GA1	32.45	-81.78	SOII	2009 Dec	C5	17
	GA1	32.45	-81.78	SOIII	2009 Dec	C5	1
Iowa	IA1	42.67	-91.59	AMI	2007 May	C3	3
Louisiana	LA1	30.94	-91.46	AMI	2012 Dec	C9	3
	LA1	30.94	-91.46	AMII	2012 Dec	C9	2
	LA1	30.94	-91.46	SOIII	2012 Dec	C9	1
	LA1	30.94	-91.46	SOIV	2012 Dec	C9	2
Massachusetts	MA1	41.71	-69.92	AMI	2010 May	C5	6
Maryland	MD1	39.29	-76.88	AMI	2010 Oct	C7	11
North-Carolina	NC1	33.91	-78.39	AMI	2009 Jan	C4	22
	NC1	33.91	-78.39	SOI	2009 Jan	C4	4
	NC1	33.91	-78.39	SOIII	2009 Jan	C4	2
New-York	NY1	40.76	-72.83	AMI	2009 Oct	C5	10
Pennsylvania	PA1	40.67	-75.96	AMI	2010 Oct	C7	5
	PA2	41.06	-79.48	AMI	2014 May	C10	33
South-Carolina	SC1	33.33	-81.66	AMI	2011 Apr	C7	8
	SC1	33.33	-81.66	AMII	2011 Apr	C7	1
	SC1	33.33	-81.66	SOI	2011 Apr	C7	3
	SC1	33.33	-81.66	SOII	2011 Apr	C7	1
	SC1	33.33	-81.66	SOIII	2011 Apr	C7	1
	SC2	33.23	-81.73	AMI	2010 Dec	C7	4
	SC2	33.23	-81.73	AMII	2010 Dec	C7	8
	SC2	33.23	-81.73	SOI	2010 Dec	C7	4
	SC2	33.23	-81.73	SOII	2010 Dec	C7	10
	SC3	33.15	-81.61	SOII	2010 Dec	C7	2
	SC5	34.29	-79.87	AMI	2012 Dec	C9	3
	SC5	34.29	-79.87	SOII	2012 Dec	C9	2
Tennessee	TN1	35.37	-86.07	AMI	2010 Dec	C7	23
Texas	TX1	31.80	-96.23	AMI	2012 Dec	C9	5
	TX1	31.80	-96.23	SOI	2012 Dec	C9	2
	TX1	31.80	-96.23	SOIV	2012 Dec	C9	3
	TX2	30.24	-100.72	AMI	2012 Dec	C9	1
	TX2	30.24	-100.72	SOIV	2012 Dec	C9	5
	TX3	31.91	-95.9	AMI	2012 Dec	C9	3
	TX3	31.91	-95.9	SOI	2012 Dec	C9	2
	TX3	31.91	-95.9	SOIV	2012 Dec	C9	6
	TX4	31.59	-95.61	AMI	2012 Dec	C9	2
TX4	31.59	-95.61	SOI	2012 Dec	C9	2	
Virginia	VA1	38.29	-78.29	AMI	2010 Oct	C7	7
	VA2	37.32	-80.73	AMI	2012 Dec	C9	4
Wisconsin	WI1	43.95	-90.70	AMI	2011 Oct	C8	8
	WI1	43.95	-90.70	AMI	2012 Oct	C9	4
	WI2	44.04	-90.65	AMI	2011 Oct	C8	8
	WI2	44.04	-90.65	AMI	2012 Oct	C9	12
	WI3	44.02	-90.64	AMI	2011 Oct	C8	8
WI3	44.02	-90.64	AMI	2012 Oct	C9	3	

184

Table 2: List of primer sets used or developed for this study. PAC = *Ixodes pacificus*, PER = *I. persulcatus*, RIC = *I. ricinus*.

Locus	Repeat	Primer name	Primer sequences (5'→3')	Ref	Dye	Size)	NAl	Sp
IS1	(AG) <sub>10</sub>	Amy1-IsAG25a Amy2-IsAG25b	AAATGTCCGAACAGCCTTAT GCCCTTGAGTCTACCCACTA	Fagerberg <i>et al</i> (2001)	6 FAM	93-193	17	PAC/PER/RIC
IS3	(GTT) <sub>5</sub>	bac1d_a bac1d_b	GCAGATCTCTTTGGGCTAG AAGCTAAGGCGTTCGTTG	AC205653	VIC	76-100	7	none
IS4	(AT) <sub>21</sub>	bac1m_a bac1m_b	TGTCGGTTTTGATGCCAA GGCTCCATTCACCAGTC	AC205653	VIC	88-126	17	PAC/PER/RIC
IS5	(CA) <sub>9</sub>	bac3dh_a bac3dh_b	TGCCTGTGACGAAACCA TCTCCAAGAGATCTAGGTA	AC205650	NED	62-140	17	none
IS6	(TA) <sub>10</sub>	bac1j_a bac1j_b	TCTCCAAGAGATCTAGGTA ATCTGTTCAGTGGGCACA	AC205653	VIC	100-186	13	PER
IS7	(TA) <sub>11</sub>	bac1k_a bac1k_b	GGGACTGGACACACGA CTAGGTGGCGCAAGTC	AC205653	VIC	48-170	26	none
IS8	(CA) <sub>14</sub>	bac3s_a bac3s_b	CGTTTCAAAGTCGGAGA GATGTGAGGGCGTGGT	AC205650	PET	96-194	11	PER
IS9	(AAAC) <sub>5</sub>	bac4cef_a bac4cef_b	CGCCTTTTGTCCCAACC GACTAACAGCATTGGAGCA	AC205647	6 FAM	85-125	12	PER
IS10	(TTA) <sub>9</sub>	bac5cf_a bac5cf_b	TCCCCAACAAGATTGATG GAGACGACGTAGATTCTTG	AC205646	6 FAM	77-137	15	none
IS11	(TTA) <sub>6</sub>	bac5g_a bac5g_b	GCTTTAGCGGGCTGGT TACGTGAATACGTCCTTGG	AC205646	PET	81-165	12	PER
IS12	(TA) <sub>43</sub>	bac6a_a bac6a_b	GCAAGCTTCGCTATTCTC CAGTAATTTCCGATCGGTT	AC205643	6 FAM	111-229	26	none
IS13	(TA) <sub>22</sub>	bac6c_a bac6c_b	TAGGTACAAGAAAACGTGCT CAAGGTAATTGTTCTCGTCA	AC205643	NED	37-91	17	none
IS14	(TA) <sub>5</sub>	bac6d_a bac6d_b	CCTTGCCTTACATGGTT CGTACCAAACCAAAGCAAG	AC205643	HEX	57-105	13	PAC/PER/RIC
IS15	(AT) <sub>8</sub>	bac6e_a bac6e_b	TATTGTAACCGACGCTAGG GACAATCTCTACGCAAATCC	AC205643	NED	79-125	18	none
IS16	(CA) <sub>8</sub>	bac6f_a bac6f_b	CCCCCAAACACGCACA TTGCTTCATGCAGGGAAC	AC205643	VIC	80-106	12	RIC
IS17	(CA) <sub>6</sub>	bac7e_a bac7e_b	CCAGCATTTAACCCTCAAG TAGTGGGGTATGGCACTG	AC205642	HEX	139-197	12	PER/RIC
IS18	(TG) <sub>6</sub>	bac8a_a bac8a_b	GTAGGTACCCTAAGAAGGAT TTGAGGAAGCAGAATGTAGG	AC205641	6 FAM	75-195	16	PER/RIC
IS19	(CT) <sub>7</sub>	bac9a_a bac9a_b	AGAACCAGTTCAGCATTCC GAACATTTTCACGTGTTGC	AC205638	PET	94-166	6	PER
IS20	(GC) <sub>9</sub>	bac11a_a bac11a_b	CGCTCCCTTCGAAGTTC GAGAAGACAGTTTCCATCG	This study	HEX	76-106	13	PAC/PER/RIC
IS21	(ACG) <sub>6</sub>	bac11c_a bac11c_b	CGAATCGCGCACACTAG GCTGTGTTGCTGGTCAC	This study	NED	109-251	14	PAC/PER
IR27	(AC) <sub>9</sub>		ATACCCGTAGAACGAGAG GTTTTTCAAGATTTCCGCC	Delaye <i>et al</i> (2008)	6 FAM			RIC

Ref: Reference or GenBank accession number; Size: Approximate size range (bp, 67 individuals); Nal: Number of alleles (67 individuals); Sp: Cross-species amplification



188 *Genotyping*

189         Based on their degree of polymorphism, nine microsatellite loci (IR27, IS1, IS3,  
190 IS11, IS15, IS16, IS17, IS18, and IS19) were used for genotyping at the continental scale  
191 (Table 2). Of these, IR27 (Delaye et al., 1998) and IS1 (Fagerberg et al., 2001), were  
192 drawn from previously published studies. The loci were amplified and genotyped using the  
193 procedures described above, although PCR conditions had to be slightly optimized for  
194 markers IS11 and IS15 (touchdown annealing temperature decreased from 58°C to 50°C)  
195 and IR27 (touchdown annealing temperature decreased from 56°C to 53°C) (Table 2).

196

197 *Population genetics analyses*

198         The raw data set was coded and converted into all required formats using Create  
199 (Coombs et al., 2008).

200         To test for LD, we used the *G*-based test first described by Goudet et al. (Goudet et  
201 al., 1996) and adapted for contingency tables of locus pairs, with 15000 reshuffling of  
202 genotypes to get maximum precision and minimize possible *p*-values before false  
203 discovery rate corrections (see below). The *G* statistics obtained for each subsample were  
204 then summed over all subsamples to get a single statistic and hence, a single test across  
205 subsamples. This procedure was shown to be the most powerful (De Meeûs et al., 2009)  
206 and was implemented within Fstat 2.9.4 (Goudet, 2003) an updated version of the original  
207 1995 Fstat software (Goudet, 1995). There are as many tests as locus pairs and these  
208 tests are correlated (one locus is used as many times as there is any other locus). To take  
209 into account this repetition of correlated tests, we used Benjamini and Yekutieli (BY) false  
210 discovery rate procedure (Benjamini and Yekutieli, 2001) with R version 3.5.1 (R-Core-  
211 Team, 2018). To check if some loci were involved in a significant LD pair more often than  
212 by chance, as compared to the other loci, we also undertook a Fisher exact test with  
213 Rcmdr version 2.3-1 (Fox, 2005; Fox, 2007).

214         For a hierarchy with three levels (individuals, subsamples, and total sample), three  
215 *F*-statistics can be defined (Wright, 1965). *F<sub>IS</sub>* measures inbreeding of individuals relative  
216 to inbreeding of subsamples or relative deviation of observed genotypic proportions from  
217 local random mating proportions. *F<sub>ST</sub>* measures inbreeding of subsamples relative to total  
218 inbreeding or relative inbreeding due to the subdivision of the total population into several  
219 isolated subpopulations. *F<sub>IT</sub>* measures inbreeding of individuals relative to total inbreeding.  
220 Under the null hypothesis (panmixia and no subdivision), all these statistics are expected  
221 to be null. Otherwise, *F<sub>IS</sub>* and *F<sub>IT</sub>* can vary from -1 (one heterozygote class) to +1 (all  
222 individuals homozygous) and *F<sub>ST</sub>* from 0 (all subsamples share similar allele frequencies)

223 to +1 (all subsamples fixed for one or the other allele). These statistics were estimated with  
224 Weir and Cockerham's unbiased estimators (Weir and Cockerham, 1984) with Fstat.

225 In dioecious species (like ticks), heterozygote excess occurs over all loci (e.g. (De  
226 Meeûs et al. 2007)) and is proportional to subpopulation size ( $N_e = -1/(2 \times F_{IS}) - F_{IS}/(1 + F_{IS})$ )  
227 (Balloux, 2004). Therefore, the finding of homozygous excesses really represents a strong  
228 deviation from random mating expectations. Technical problems, like null alleles,  
229 stuttering, SAD or allele dropouts unevenly affects some loci, producing a positive  $F_{IS}$  with  
230 an important variation across loci and significant outliers (De Meeûs 2018). Significant  
231 departure from 0 of these  $F$ -statistics was tested with 10000 randomizations of alleles  
232 between individuals within subsample (deviation from local random mating test) or of  
233 individuals between subsamples within the total sample (population subdivision test). For  
234  $F_{IS}$ , the statistic used was  $f$  (Weir and Cockerham's  $F_{IS}$  estimator). To test for subdivision,  
235 we used the  $G$ -based test (Goudet et al. 1996) over all loci, which is the most powerful  
236 procedure when combining tests across loci (De Meeûs et al. 2009).

237 To compute 95% confidence intervals (95%CI) of  $F$ -statistics, we used the standard  
238 error of  $F_{IS}$  (StrdErrFIS) and  $F_{ST}$  (StrdErrFST) computed by jackknifing over populations, or  
239 5000 bootstraps over loci as described elsewhere (De Meeûs et al. 2007). For jackknives,  
240 the number of usable subsamples was restricted to subsamples with at least 5 ticks (23  
241 subsamples) (e.g. (De Meeûs, 2012) p 73).

242 In case of significant homozygote excess and LD we tried to discriminate  
243 demographic from technical causes with the determination key proposed by De Meeûs (De  
244 Meeûs 2018). Null alleles better explain the data if the StrdErrFIS becomes at least twice  
245 as high as StrdErrFST;  $F_{IS}$  and  $F_{ST}$  are positively correlated; and a positive correlation  
246 links  $F_{IS}$  and the number of missing data (putative null homozygotes). The significance of  
247 correlations was tested with a unilateral ( $\rho > 0$ ) Spearman's rank correlation test with R. The  
248 presence of null alleles was also verified with MicroChecker v 2.2.3 (Van Oosterhout et al.  
249 2004) and null allele frequencies estimated with Bookfield's second method (Brookfield  
250 1996). The adjustment between observed and expected numbers of missing data was  
251 tested with a unilateral exact binomial test in R with the alternative hypothesis being "there  
252 is not enough missing data as expected if heterozygote deficits were entirely explained by  
253 null alleles under panmixia". MicroChecker also seeks stuttering and SAD. Stuttering is  
254 detected when MicroChecker reveals an observed proportion of heterozygous individuals  
255 for alleles with one repeat difference significantly smaller than the expected value. The  
256 presence of stuttering was detected with the graphic output of MicroChecker (we ignored  
257 the comments panel that happened to contradict the graphic in some instances). We

258 considered that the observed deficit of heterozygous individuals for one repeat difference  
259 was a likely consequence of stuttering. Due to the small population sample sizes and  
260 bootstrapping procedure in MicroChecker, the statistical support ( $p$ -value) of this result  
261 was not always significant for all runs. Hence, we set the randomization at the maximum  
262 value (10000) and repeated the analysis three times to check for consistency. Stuttering  
263 was admitted when two out of three tests supported it. The occurrence of SAD was also  
264 checked with an unilateral ( $\rho < 0$ ) Spearman's rank correlation test between allele size and  
265  $F_{IT}$  as proposed by (Manangwa et al., 2019). This test is more powerful than with  $F_{IS}$  as  
266 was proposed earlier (Wattier et al., 1998; De Meeûs et al., 2004). If previous tests are not  
267 significant and if  $StrdErrFIS > StrdErrFST$ , then a Wahlund effect better explains the data  
268 (De Meeûs, 2018a), this is especially true in instances of significant LD (Manangwa et al.,  
269 2019). In these cases, a positive correlation between the number of times a locus is found  
270 in significant LD (NLD) and its total genetic diversity as measured by Nei's  $H_T$  (Nei and  
271 Chesser, 1983) (Spearman's correlation above 0.1) suggests an admixture of individuals  
272 from several subpopulations of the same species but with an important degree of genetic  
273 differentiation between admixed subpopulations (i.e. number of immigrants  $N_e m = 2$ ). If the  
274 correlation is negative and the proportion of significant LDs is above 40%, an admixture of  
275 different strongly divergent entities (e.g. species) better explains the data (Manangwa et  
276 al., 2019). We thus undertook a bilateral Spearman's test.

277 In some instances, the same null hypothesis was repeatedly tested (i.e. SAD was  
278 tested for each locus one by one). Repetition of independent tests were submitted to  
279 Benjamini and Hochberg (BH) correction (Benjamini and Hochberg, 2000) (computed with  
280 R), to check for robustness of significant  $p$ -values.

281 To our knowledge, when diagnosed, there is no analytical remedy for SAD or  
282 stuttering as there is for null alleles (Chapuis and Estoup, 2007; Séré et al., 2017a).  
283 Nevertheless, SAD can be cured by going back to the chromatograms of homozygous  
284 individuals and trying to find a micro-peak (see the Results and discussion section), with a  
285 larger size, ignored in the first reading. If enough profiles can be corrected this might  
286 salvage the incriminated locus.

287 Stuttering can be addressed by pooling alleles close in size. However, this  
288 procedure requires that none of the pooled allele groups is constituted of rare alleles only.  
289 Indeed, pooling rare alleles, usually found in heterozygosity with a more frequent allele,  
290 will tend to artificially generate heterozygous excesses. In order to avoid this  
291 consequence, each pooled group should contain at least one frequent allele (e.g. with  
292  $p \geq 0.05$ ).

293 In dioecious small populations, a heterozygote excess is expected. However, loci  
294 with null alleles may display heterozygote deficits. In such situations a bilateral test ( $F_{IS}$  is  
295 not different from 0) is needed and obtained as  $p_{bilateral}=p_{mini}+1-p_{maxi}$ , where  $p_{mini}$  is the  
296 minimum unilateral  $p$ -value (for heterozygote deficit or excess) and  $p_{max}$  is the maximum  
297 one.

298 Finally, a more accurate estimate of  $F_{ST}$  can be made for datasets with null alleles  
299 after recoding missing genotypes as homozygous for allele 999 (null allele) with the ENA  
300 method as implemented in FreeNA (Chapuis and Estoup, 2007). This estimate can then  
301 be corrected for polymorphism with the formula  $F_{ST}'=F_{ST}/(1-H_s)$  (Hedrick, 1999). This  
302 helped compute the number of immigrants as  $Nm=(1-F_{ST}')/(4F_{ST})$  (assuming an Island  
303 model of migration) (e.g. (De Meeûs et al., 2007)).

304

## 305 **Results and discussion**

### 306 *Primer selection and characterization of loci*

307 The inspection of the GenBank genomic sequences revealed the presence of 67  
308 short tandem repeated motifs. The program Oligo v.5 did not find suitable primers for 17 of  
309 them. Of the remaining 50 primer pairs, 22 amplified the pooled DNA sample and the sizes  
310 of the amplicons were approximately as expected. The cloned amplified inserts all  
311 contained the expected microsatellite repeats and flanking regions. The 22 primer sets  
312 consistently amplified DNA from the 67 *I. scapularis* ticks and some of them also amplified  
313 DNA of the other *Ixodes* species (Table 2). Finally, based on their polymorphism and ease  
314 of interpretation, nine loci were retained for the population genetics analyses: IR27 (from  
315 (Delaye et al., 1998)), IS1, IS3, IS11, IS15, IS16, IS17, IS18, and IS19 (from the present  
316 study), with 9.8, 3.4, 1.3, 8, 8.3, 4.9, 2.8, 5.7 and 4.1 % missing genotypes (blanks)  
317 respectively.

318

### 319 *Raw data analyses*

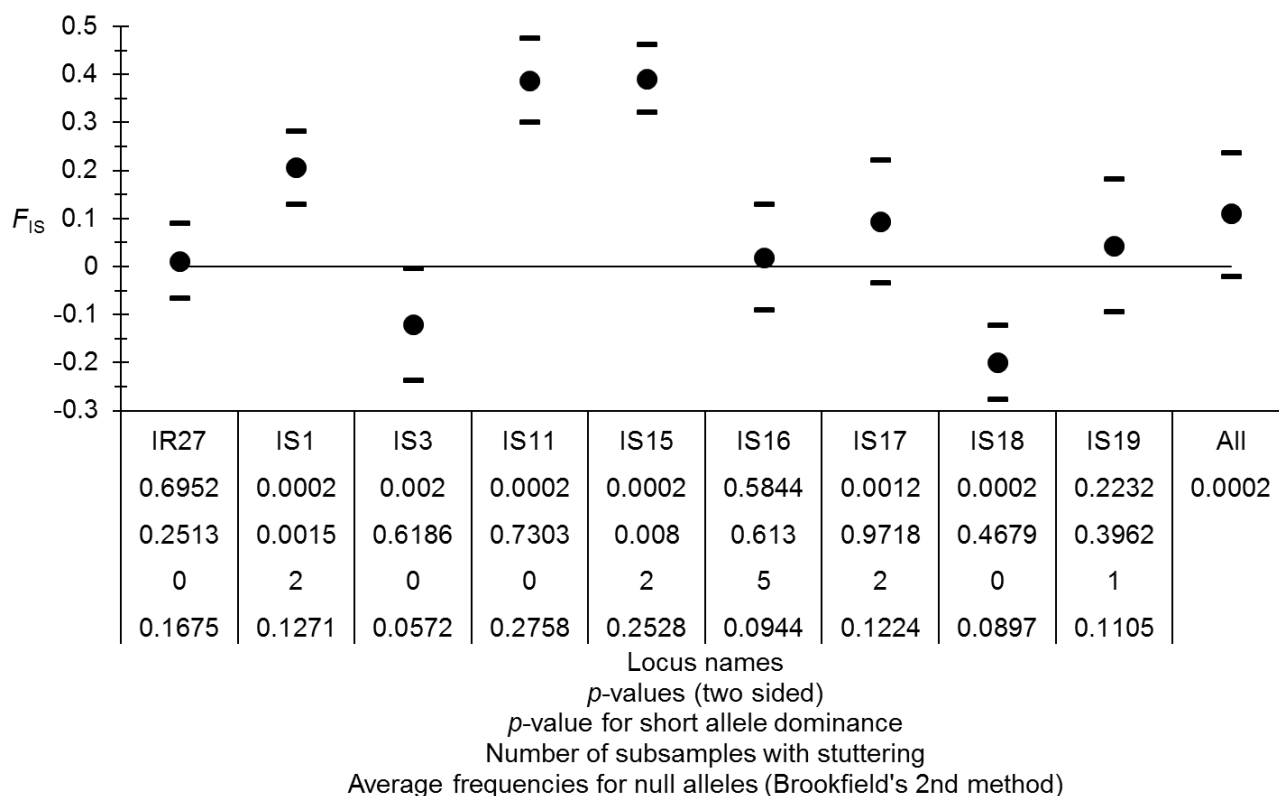
320 All data are available in the supplementary File S1.

321 There was a very important proportion of locus pairs in significant LD (~22%), with a  
322 negative correlation between NLD and  $H_T$  ( $\rho=-0.04$ ,  $p$ -value=0.9242). Two locus pairs  
323 remained in significant LD after BY correction: IR27-IS3 ( $p$ -value=0.0301) and IS11-IS16  
324 ( $p$ -value=0.0451). No single locus was found more often in significant LD than the others  
325 ( $p$ -value=0.09).

326 There was a highly significant heterozygote deficit ( $F_{IS}=0.111$ , in 95%CI=[-0.02,  
327 0.236,  $p$ -value<0.0002), with a considerable variation across loci (Figure 2).

328

329 Figure 2:  $F_{IS}$  values for each locus and averaged across those (All) of *Ixodes scapularis*  
 330 from the eastern U.S.A. with 95% jackknife confidence intervals over subsamples  
 331 (for each locus) and bootstraps over loci (All). Results of tests for panmixia, short  
 332 allele dominance, number of subsamples with stuttering for each locus, and null  
 333 allele frequencies are also indicated.



334

335

336  $StdErrFIS$  (0.07) was almost four times higher than  $StdErrFST$  (0.019); the  
 337 correlation between  $F_{IS}$  and  $F_{ST}$  was negative ( $\rho=-0.6166$ ,  $p$ -value=0.962) and the  
 338 correlation between  $F_{IS}$  and the number of blanks (missing genotypes) was positive but not  
 339 significant ( $\rho=0.1833$ ,  $p$ -value=0.3218). These results suggested that locus-specific effects  
 340 were involved. Nevertheless, null alleles only partly explained the observed pattern at best.  
 341 The substantial proportion of significant LDs suggested the existence of a possible strong  
 342 Wahlund effect, though a negative correlation between NLD and  $H_T$  with less than 40%  
 343 significant tests observed here would refute this interpretation (Manangwa et al., 2019).

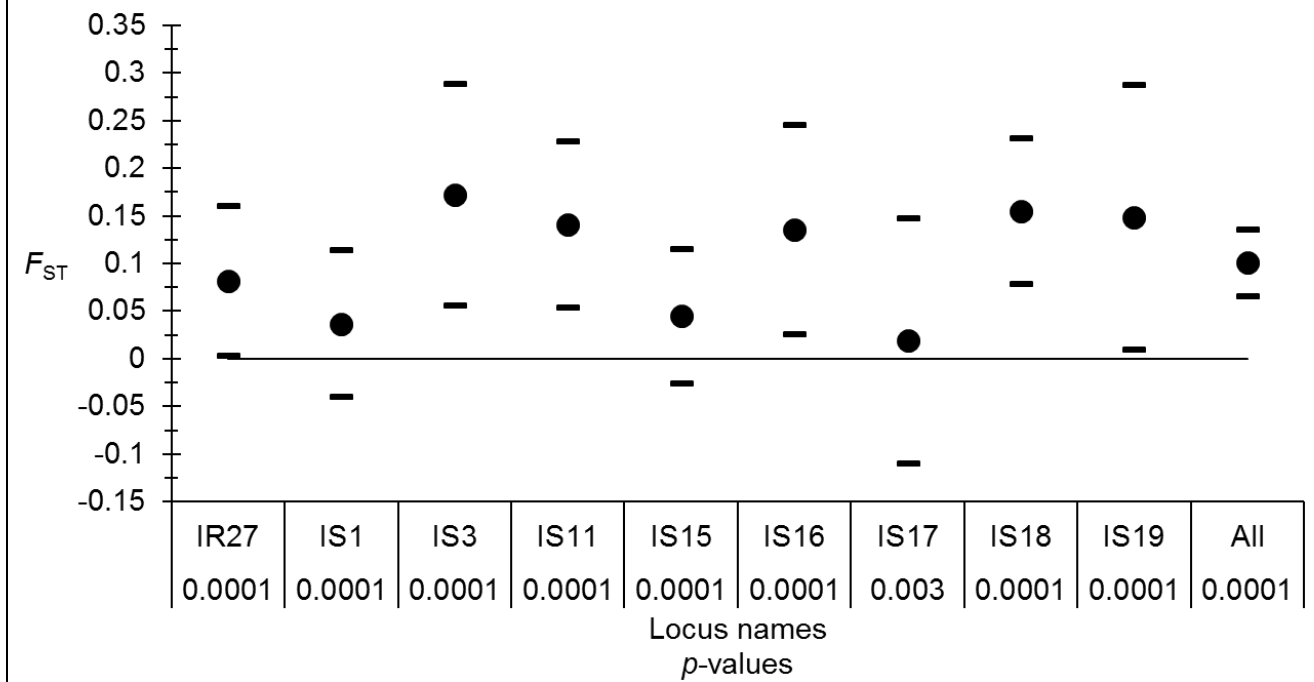
344 Small subsampling due to the partitioning of the data into cohorts and 12S clades  
 345 was expected to considerably lower the power of these tests, especially so after the rather  
 346 stringent BY procedure for the LD test series.

347 Subdivision was substantial ( $F_{ST}=0.101$ , 95%CI=[0.066, 0.136],  $p$ -value<0.0001),  
 348 and variable across loci and subsamples (Figure 3), but not meaningfully more variable

349 than expected (see Figure 6 in (De Meeûs, 2018b)). Some loci (IS1, IS15 and IS17)  
 350 displayed very low values (Figure 3). Selection (e.g. balanced selection) might have  
 351 produced the pattern observed, though evidence for this is weak (large confidence  
 352 intervals).

353

354 Figure 3:  $F_{ST}$  values for each locus and averaged across those (All) of *Ixodes scapularis*  
 355 from the eastern U.S.A. with 95% jackknife confidence intervals over subsamples  
 356 (for each locus) and bootstrap over loci (All). Results of tests for significant  
 357 subdivision ( $p$ -value) are also indicated.



358

359

360 Two loci displayed highly significant SAD (Figure 2): loci IS1 ( $\rho=-0.57$ ,  $p$ -  
 361 value=0.0015) and IS15 ( $\rho=-0.46$ ,  $p$ -value=0.008), which stayed significant after BH  
 362 correction (0.0135 and 0.036 respectively). Stuttering was diagnosed for four loci (IS1,  
 363 IS15, IS16, IS17 and IS19). According to Brookfield's second method, null allele  
 364 frequencies could range between 0.06 and 0.28 on average (Figure 2), but these  
 365 estimates do not correct for other errors and, as discussed above, null alleles only partly  
 366 explain the observed  $F_S$  and its variation across loci.

367 These heterozygote deficits and high proportion of significant LD may come from  
 368 amplification problems detected as SAD, stuttering and null alleles. Amplification problems  
 369 are expected to occur in individuals presenting an apostate DNA. This can happen if these  
 370 individuals belong to lineages that are significantly divergent from the majority; when their  
 371 DNA is partly degraded, at low concentration; or a combination of these different causes.

372 This may amplify a preexisting correlation between allele occurrences at different loci. One  
373 way to test for this is to study the correlation between the number of missing genotypes  
374 (blanks) of an individual and its number of heterozygous sites. This correlation is expected  
375 to be negative if amplification problems occur more often in some individuals than in others  
376 (Kaboré et al., 2011). A Spearman's rank correlation test with Rcmdr outputted  $\rho=-0.26$  ( $p$ -  
377 value $<0.0001$ ).

378

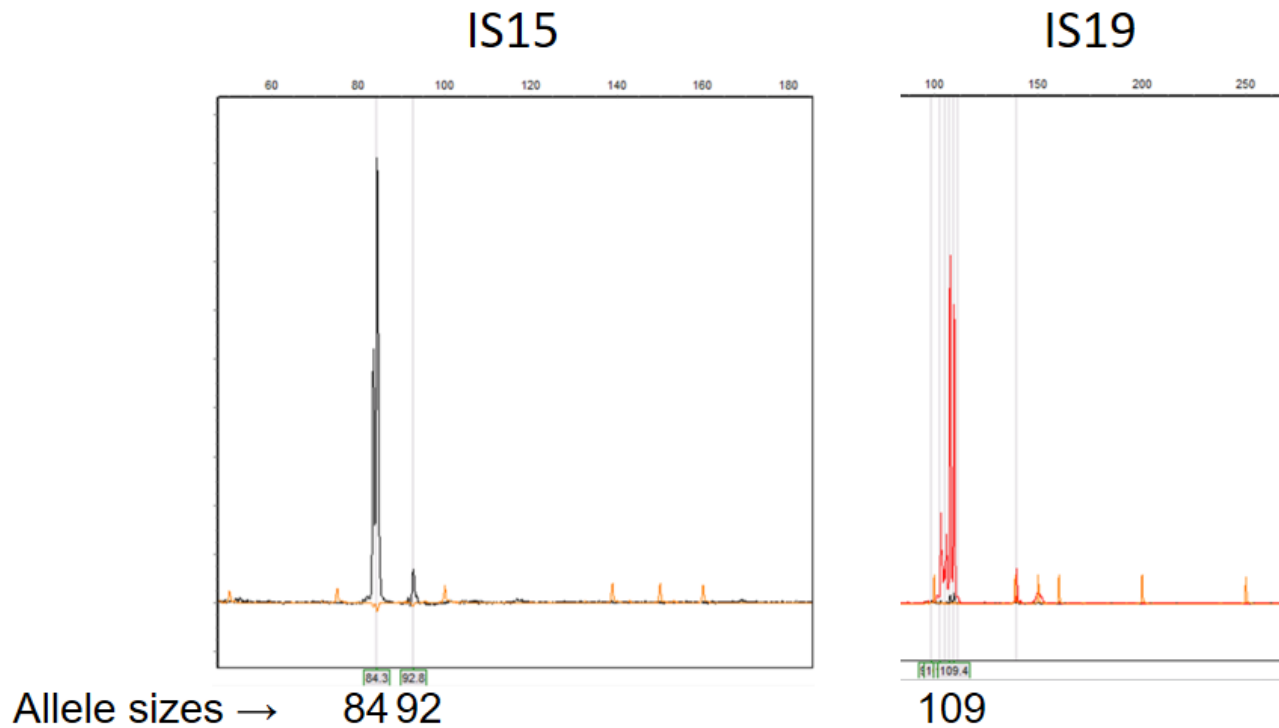
#### 379 *Cured data set*

380 To prevent the omission of five (out of nine) loci, due to stuttering, SAD and/or  
381 possible selection, we went back to the data. We first scanned the chromatograms for  
382 previously ignored micro-peaks in homozygous individuals at loci displaying SAD (IS1 and  
383 IS15) (Figure 4). For IS1 and IS15, SAD might well have explained why we also found  
384 stuttering at these locus (see below). We then tried to pool alleles close in size as  
385 described above for loci IS16, IS17 and IS19. For IS16, alleles 88, 90 and 92 were  
386 recoded as 94; and allele 96 as 98. For locus IS17, alleles 170 and 172 were recoded as  
387 174; alleles 178, 180, 182 and 184 were recoded as 186; alleles 188, 190 and 192 were  
388 recoded as 194; and alleles 196 and 198 as 200. Finally, for locus IS19, allele 91 was  
389 recoded as 93; alleles 97, 99, 101 and 103 as 105; and alleles 107 and 109 as 111. The  
390 obtained amended dataset was called "Cured dataset" (Supplementary file S1). Pooling  
391 alleles only increases homoplasy. Everything being equal, the effect of homoplasy on  $F_{IS}$   
392 or  $F_{ST}$  is equivalent to a mutation rate increase by a factor  $K/(K-1)$  (Rousset, 1996), where  
393  $K$  is the number of possible alleles. Thus, for microsatellite loci with many possible alleles,  
394 the effect is deemed negligible, especially for  $F_{IS}$ . Here, the resulting number of alleles per  
395 locus was 17 on average, with 5 and 27 as the outermost-lying values. The resulting  
396 homoplasy effect on  $F$ -statistics can thus be safely ignored.

397

398  
399  
400

Figure 4: Examples of an initially dismissed micro-peak that produced SAD at locus IS15 (the peak for allele 92 appears much smaller than for allele 84) and of stuttering at locus IS19 (stutters around allele 109 hides the possible presence of allele 111).



401  
402

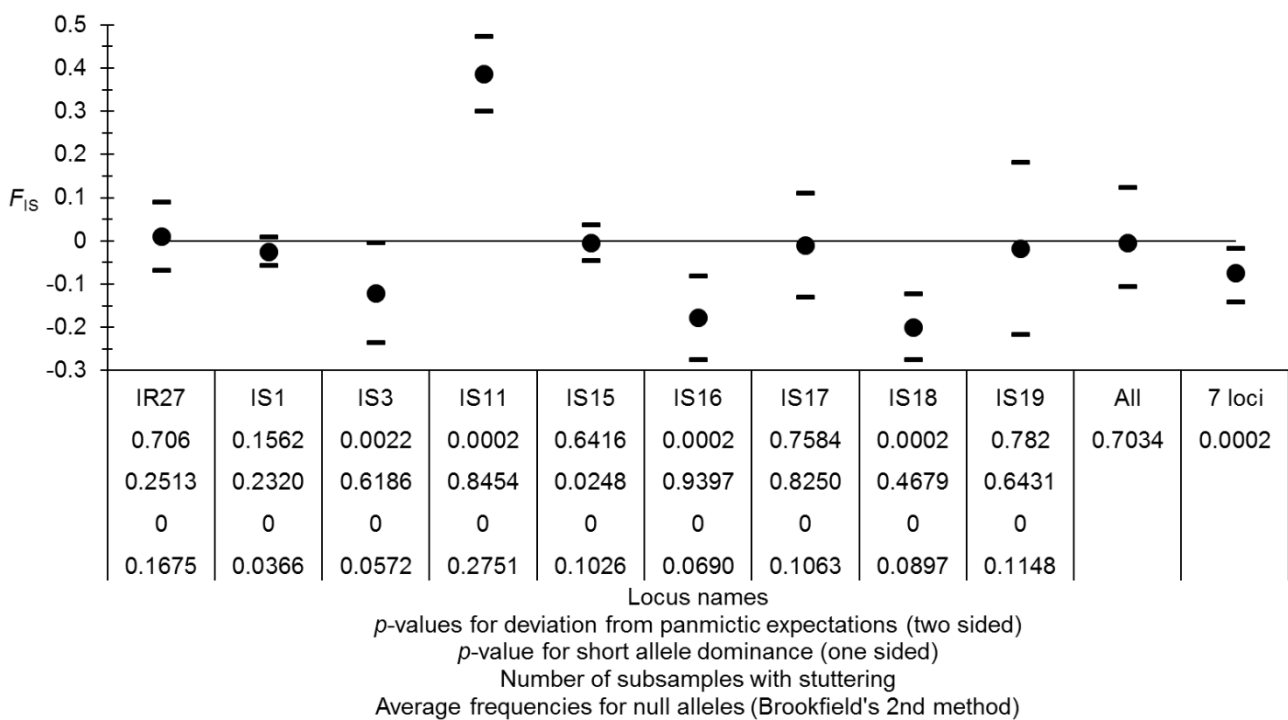
403 With the cured dataset, the proportion of locus pairs in significant LD dropped to  
404 19%, but more importantly, the smallest  $p$ -value after BY correction was 0.0797. The  
405 correlation between NLD and  $H_r$  became positive ( $\rho=0.57$ ,  $p$ -value=0.1056). This change  
406 may point towards a Wahlund effect (Manangwa et al., 2019). Nevertheless, this would be  
407 incompatible with the  $F_{IS}$  observed (see below) and with the fact that none of these tests  
408 remained significant after BY adjustment. A heterozygote deficit was no longer observed  
409 ( $F_{IS}=-0.004$  in 95%CI=[-0.105, 0.124], bilateral  $p$ -value=0.7034) (Figure 5).  
410  $StrdErrFIS=0.064$  was three times higher than  $StrdErrFST=0.022$ . There was no  
411 correlation between  $F_{IS}$  and  $F_{ST}$  ( $\rho=-0.27$ ,  $p$ -value=0.77). The correlation between number  
412 of missing genotypes and  $F_{IS}$  was positive, though marginally not significant ( $\rho=0.52$ ,  $p$ -  
413 value=0.0809). MicroChecker diagnosed null alleles in 11 subsamples for locus IS11 and  
414 in one subsample for loci IR39 and IS19, which correspondingly displayed relatively high  
415 and variable  $F_{IS}$  and substantial proportions of missing genotypes (see above and Figure  
416 5). IS15 still displayed SAD ( $\rho=-0.3814$ ,  $p$ -value=0.0248), though with much less intensity.  
417 In addition, after BH correction, the test was not significant any longer ( $p$ -value=0.2232).  
418 When IS11 and IS15 were removed from the dataset, global  $F_{IS}$  became significantly  
419 negative ( $F_{IS}=-0.074$  in 95%CI=[-0.142, -0.017],  $p$ -value=0.0002) as expected for small



420 subpopulations in dioecious species (Balloux, 2004). As for subdivision,  $F_{ST}$  remained  
 421 almost unaffected, even with the ENA method ( $F_{ST}=0.103$  in 95%CI=[0.067..0.14],  $p$ -  
 422 value<0.0001).

423

424 Figure 5:  $F_{IS}$  values for each locus, averaged across those (All), or over 7 loci without  
 425 amplification problems (IS11 and IS15) for *Ixodes scapularis* cured data from the  
 426 eastern U.S.A. with 95% jackknife confidence intervals over subsamples (for each  
 427 locus) and bootstraps over loci (All). Results of tests for panmixia, short allele  
 428 dominance, number of subsamples with stuttering for each locus, and null allele  
 429 frequencies are also indicated.



430

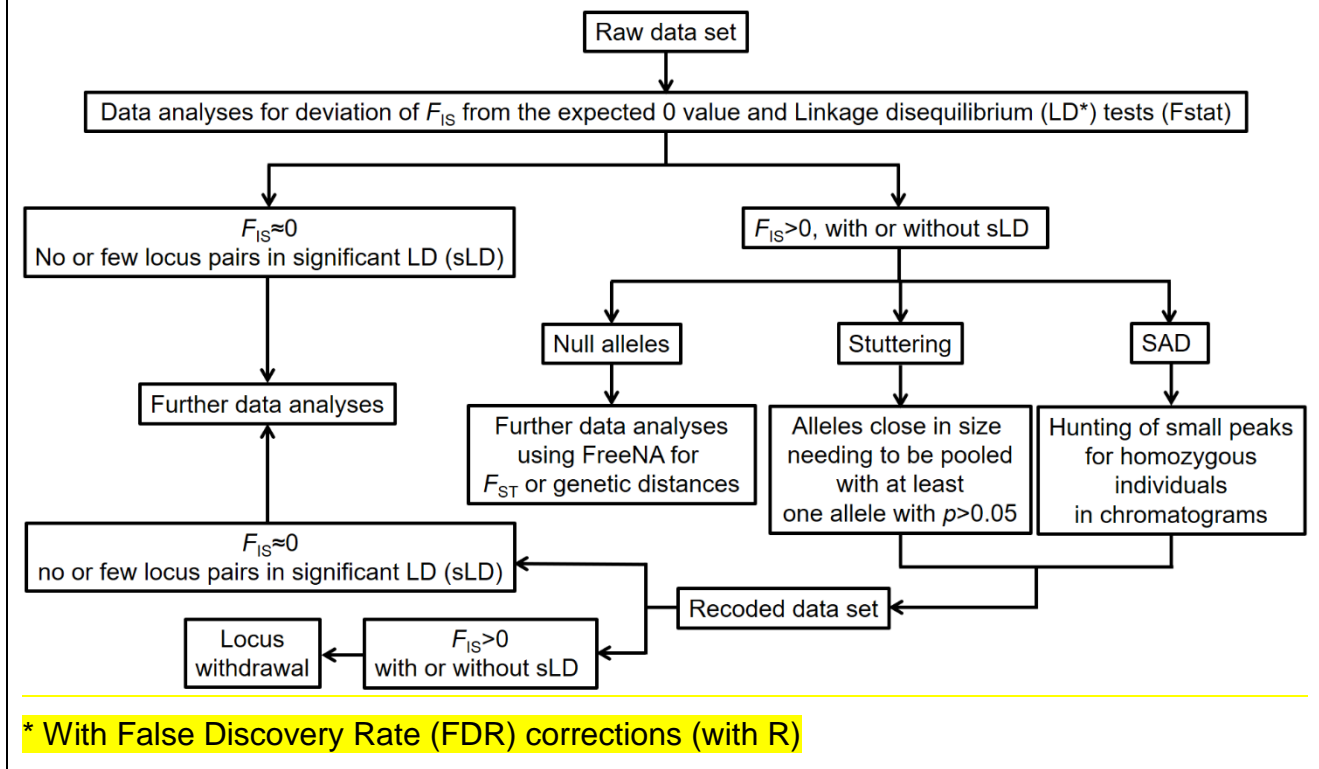
431

### 432 Conclusion

433 Combinations of amplification errors manifesting as null alleles, SAD or stuttering  
 434 lead not only to heterozygote deficits, but an overall increase in LD. In order to correct for  
 435 SAD, it may be useful to hunt for small and hard-to-detect peaks in chromatograms for  
 436 homozygous individuals and pool alleles close in size to correct for stuttering. Uncorrected,  
 437 these problems have the potential to lead to the unnecessary withdrawal of the “flawed”  
 438 loci. These proposed amendments, together with null allele management with the ENA  
 439 algorithm (Chapuis and Estoup, 2007), resulted in analyses revealing heterozygote  
 440 excesses, as expected in small dioecious subpopulations. These changes also reduced  
 441 the proportion of significant LD tests which, more importantly, were no longer significant

442 after BY correction. Such corrections can then lead to more accurate estimates of the  
 443 degree of population subdivision, as shown elsewhere for null alleles (Chapuis and  
 444 Estoup, 2007; Séré et al., 2017b). The different steps to be followed during population  
 445 genetics data analyses of such datasets are summarized as a flowchart in Figure 6.  
 446

447 Figure 6: Flowchart of the recommended steps in population genetics data analyses.



451 In our case, the relatively important global LD across loci is probably due to small  
 452 effective sizes of the *I. scapularis* subpopulations (Waples, 2006; Waples and Do, 2010).  
 453 Additionally, null alleles are still influencing the cured data, and predominantly affect  
 454 individuals that display some kind of deviating DNA (as explained in the introduction) and  
 455 may also contribute to inflate the perceived LD. In fact, the correlation between the number  
 456 of missing genotypes and the number of heterozygous sites was still significantly negative  
 457 in the cured data set ( $\rho = -0.32$ ,  $p\text{-value} < 0.0001$ ).

458 It is worthy to note that a reanalysis of the raw data set using MicroDrop 1.01 (Wang  
 459 et al., 2012) resulted in smaller  $F_{IS}$  and  $F_{ST}$  values and only three significant LD tests, none  
 460 of which stayed significant after BY correction (though the smaller  $p\text{-value} = 0.0556$  was  
 461 marginal). However, locus IS15 still displayed a significant SAD, ( $p\text{-value} = 0.0085$ , and  
 462 corrected  $p\text{-value} = 0.0765$ ) and stuttering still was detected once, twice and three times for  
 463 loci IS1, IS16 and IS17, respectively, and appeared once for locus IR27. MicroDrop takes  
 464 into consideration that heterozygote deficits and missing data are entirely due to allelic

465 dropouts. It is difficult to thoroughly understand these results since here we have  
466 convincing evidence that the amplification problems were due to SAD, stuttering and null  
467 alleles. This issue would require a full simulation study with different scenarios and a  
468 comparison of methods. That MicroDrop may efficiently cure not only allelic dropouts but  
469 any kind of amplification problem without any bias offers the potential for an extremely  
470 valuable tool for the scientific community, however, this remains to be ascertained.  
471 Nevertheless, MicroDrop did not entirely cure the data from SAD and stuttering, which  
472 does not advocate for the efficacy of the algorithm used in this software. As far as our *I.*  
473 *scapularis* data set is concerned, the cures proposed provided satisfactory results and  
474 additional useful tools to those already proposed in recently published papers on the  
475 detection and identification of causes of heterozygote deficits (Waples, 2015; De Meeûs,  
476 2018b; Waples, 2018; Manangwa et al., 2019).

477 Subdivision into small and isolated subpopulations was confirmed by the relatively  
478 small effective population size estimated from  $F_{IS}$  (without IS11 and IS15) with Balloux's  
479 method (Balloux, 2004) ( $N_e \approx 7$  in 95%CI=[4, 29] individuals), and the relatively important  
480 subdivision measurements between contemporaneous subsamples (to avoid temporal  
481 effects) in cohorts where these were possible (i.e. cohorts 5, 7, 8 and 9) (average  
482  $F_{ST} \approx 0.19$  in 95%CI=[0.1, 0.29]). This provided an estimate for an immigration rate of  
483  $Nm \approx 1$  individual per generation. These results contrast with those obtained from the  
484 closely related European *I. ricinus*, with no or weak LD in adults, displaying Wahlund  
485 effects at small scales, and with much weaker subdivisions (Delaye et al., 1997; De Meeûs  
486 et al., 2002; Kempf et al., 2010).

487 Finally, we still do not know if mitochondrial clades have any real biological meaning  
488 for *I. scapularis* population structure or systematics. This will be treated in detail in a  
489 further study that will also include immature stages.

490

#### 491 *Data availability*

492 The raw and cured datasets are available as “supplementary file S1” at:  
493 [http://www.t-de-meeus.fr/Data/DeMeeus-et-al-SAD&StutteringI-scapularisUSA-PCI-](http://www.t-de-meeus.fr/Data/DeMeeus-et-al-SAD&StutteringI-scapularisUSA-PCI-EvolBiol-TableS1.xlsx)  
494 [EvolBiol-TableS1.xlsx](http://www.t-de-meeus.fr/Data/DeMeeus-et-al-SAD&StutteringI-scapularisUSA-PCI-EvolBiol-TableS1.xlsx)

495

#### 496 **Acknowledgements**

497 This work was funded by NSF Grant # EF0914390 to L.B. and EEID EF-0914476 to  
498 J.T. We thank the members of the Lyme Gradient Consortium and many individuals who  
499 provided ticks. We are also grateful to Heather Walker, Jenny Dickson, Keely Duff, Laquita

500 Burton, Alysha Benn and Nina Griffin, and the several other undergraduate students who  
501 provided field and laboratory assistance. We would like to thank one anonymous referee  
502 and Eric Petit for their comments and suggestions that helped with improving the present  
503 manuscript.

504

505

## 506 **References**

507 Balloux, F., 2004. Heterozygote excess in small populations and the heterozygote-excess  
508 effective population size. *Evolution* 58, 1891-1900.

509 Beati, L., Keirans, J.E., 2001. Analysis of the systematic relationships among ticks of the  
510 genera *Rhipicephalus* and *Boophilus* (Acari : Ixodidae) based on mitochondrial 12S  
511 ribosomal DNA gene sequences and morphological characters. *J. Parasitol.* 87, 32-48.

512 Benjamini, Y., Hochberg, Y., 2000. On the adaptive control of the false discovery rate in  
513 multiple testing with independent statistics. *Journal of Educational and Behavioral*  
514 *Statistics* 25, 60-83.

515 Benjamini, Y., Yekutieli, D., 2001. The control of the false discovery rate in multiple testing  
516 under dependency. *Ann. Stat.* 29, 1165–1188.

517 Castle, W.E., 1903. The laws of heredity of Galton and Mendel, and some laws governing  
518 race improvement by selection. *Proc Am Acad Arts Sci* 39, 223-242.

519 Chapuis, M.P., Estoup, A., 2007. Microsatellite null alleles and estimation of population  
520 differentiation. *Mol. Biol. Evol.* 24, 621-631.

521 Coombs, J.A., Letcher, B.H., Nislow, K.H., 2008. CREATE: a software to create input files  
522 from diploid genotypic data for 52 genetic software programs. *Mol. Ecol. Res.* 8, 578–580.

523 De Meeûs, T., 2012. *Initiation à la génétique des populations naturelles: Applications aux*  
524 *parasites et à leurs vecteurs*. IRD Editions, Marseille.

525 De Meeûs, T., 2018a. Revisiting  $F_{IS}$ ,  $F_{ST}$ , Wahlund effects and null alleles. *J. Hered.* 109,  
526 446-456.

527 De Meeûs, T., 2018b. Revisiting  $F_{IS}$ ,  $F_{ST}$ , Wahlund effects, and Null alleles. *J. Hered.* 109,  
528 446-456.

529 De Meeûs, T., Beati, L., Delaye, C., Aeschlimann, A., Renaud, F., 2002. Sex-biased  
530 genetic structure in the vector of Lyme disease, *Ixodes ricinus*. *Evolution* 56, 1802-1807.

531 De Meeûs, T., Guégan, J.F., Teriokhin, A.T., 2009. MultiTest V.1.2, a program to  
532 binomially combine independent tests and performance comparison with other related  
533 methods on proportional data. *BMC Bioinformatics* 10, 443.

534 De Meeûs, T., Humair, P.F., Grunau, C., Delaye, C., Renaud, F., 2004. Non-Mendelian  
535 transmission of alleles at microsatellite loci: an example in *Ixodes ricinus*, the vector of  
536 Lyme disease. *Int. J. Parasitol.* 34, 943-950.

537 De Meeûs, T., McCoy, K.D., Prugnolle, F., Chevillon, C., Durand, P., Hurtrez-Boussès, S.,  
538 Renaud, F., 2007. Population genetics and molecular epidemiology or how to "débusquer  
539 la bête". *Infect. Genet. Evol.* 7, 308-332.

540 Delaye, C., Aeschlimann, A., Renaud, F., Rosenthal, B., De Meeûs, T., 1998. Isolation and  
541 characterization of microsatellite markers in the *Ixodes ricinus* complex (Acari : Ixodidae).  
542 *Mol. Ecol.* 7, 360-361.

543 Delaye, C., Béati, L., Aeschlimann, A., Renaud, F., De Meeûs, T., 1997. Population  
544 genetic structure of *Ixodes ricinus* in Switzerland from allozymic data: no evidence of  
545 divergence between nearby sites. *Int. J. Parasitol.* 27, 769-773.

546 Fagerberg, A.J., Fulton, R.E., Black, W.C., 2001. Microsatellite loci are not abundant in all  
547 arthropod genomes: analyses in the hard tick, *Ixodes scapularis* and the yellow fever  
548 mosquito, *Aedes aegypti*. *Insect Mol. Biol.* 10, 225-236.

549 Fox, J., 2005. The R commander: a basic statistics graphical user interface to R. *J. Stat.*  
550 *Software* 14, 1–42.

551 Fox, J., 2007. Extending the R commander by "plug in" packages. *R News* 7, 46–52.

552 Goudet, J., 1995. FSTAT (Version 1.2): A computer program to calculate F-statistics. *J.*  
553 *Hered.* 86, 485-486.

554 Goudet, J., 2003. Fstat (ver. 2.9.4), a program to estimate and test population genetics  
555 parameters. Available at <http://www.t-de-meeus.fr/Programs/Fstat294.zip>, Updated from  
556 Goudet (1995).

557 Goudet, J., Raymond, M., De Meeûs, T., Rousset, F., 1996. Testing differentiation in  
558 diploid populations. *Genetics* 144, 1933-1940.

559 Gulia-Nuss, M., Nuss, A.B., Meyer, J.M., Sonenshine, D.E., Roe, R.M., Waterhouse, R.M.,  
560 Sattelle, D.B., de la Fuente, J., Ribeiro, J.M., Megy, K., Thimmapuram, J., Miller, J.R.,  
561 Walenz, B.P., Koren, S., Hostetler, J.B., Thiagarajan, M., Joardar, V.S., Hannick, L.I.,  
562 Bidwell, S., Hammond, M.P., Young, S., Zeng, Q.D., Abrudan, J.L., Almeida, F.C., Ayllon,  
563 N., Bhide, K., Bissinger, B.W., Bonzon-Kulichenko, E., Buckingham, S.D., Caffrey, D.R.,  
564 Caimano, M.J., Croset, V., Driscoll, T., Gilbert, D., Gillespie, J.J., Giraldo-Calderon, G.I.,  
565 Grabowski, J.M., Jiang, D., Khalil, S.M.S., Kim, D., Kocan, K.M., Koci, J., Kuhn, R.J.,  
566 Kurtti, T.J., Lees, K., Lang, E.G., Kennedy, R.C., Kwon, H., Perera, R., Qi, Y.M., Radolf,  
567 J.D., Sakamoto, J.M., Sanchez-Gracia, A., Severo, M.S., Silverman, N., Simo, L., Tojo, M.,  
568 Tornador, C., Van Zee, J.P., Vazquez, J., Vieira, F.G., Villar, M., Wespiser, A.R., Yang,

569 Y.L., Zhu, J.W., Arensburger, P., Pietrantonio, P.V., Barker, S.C., Shao, R.F., Zdobnov,  
570 E.M., Hauser, F., Grimmelikhuijzen, C.J.P., Park, Y., Rozas, J., Benton, R., Pedra, J.H.F.,  
571 Nelson, D.R., Unger, M.F., Tubio, J.M.C., Tu, Z.J., Robertson, H.M., Shumway, M., Sutton,  
572 G., Wortman, J.R., Lawson, D., Wikel, S.K., Nene, V.M., Fraser, C.M., Collins, F.H.,  
573 Birren, B., Nelson, K.E., Caler, E., Hill, C.A., 2016. Genomic insights into the *Ixodes*  
574 *scapularis* tick vector of Lyme disease. *Nat Commun* 7.

575 Hardy, G.H., 1908. Mendelian proportions in a mixed population. *Science* 28, 49-50.

576 Hedrick, P.W., 1999. Perspective: Highly variable loci and their interpretation in evolution  
577 and conservation. *Evolution* 53, 313-318.

578 Kaboré, J., MacLeod, A., Jamonneau, V., Ilboudo, H., Duffy, C., Camara, M., Camara, O.,  
579 Belem, A.M.G., Bucheton, B., De Meeûs, T., 2011. Population genetic structure of Guinea  
580 *Trypanosoma brucei gambiense* isolates according to host factors. *Infect. Genet. Evol.* 11,  
581 1129-1135.

582 Keirans, J.E., Needham, G.R., Oliver, J.H., 1999. The *Ixodes (Ixodes) ricinus* complex  
583 worldwide. Diagnosis of the species in the complex, hosts and distribution, in: Needham,  
584 G.R., R., M., Horn, D.J., Welbourn, W.C. (Eds.), *Acarology IX: Symposia*. Ohio Biological  
585 Survey, Columbus, Ohio, pp. 341-347.

586 Kempf, F., De Meeûs, T., Vaumourin, E., Noel, V., Taragel'ová, V., Plantard, O., Heylen,  
587 D.J.A., Eraud, C., Chevillon, C., McCoy, K.D., 2011. Host races in *Ixodes ricinus*, the  
588 European vector of Lyme borreliosis. *Infect. Genet. Evol.* 11, 2043-2048.

589 Kempf, F., McCoy, K.D., De Meeûs, T., 2010. Wahlund effects and sex-biased dispersal in  
590 *Ixodes ricinus*, the European vector of Lyme borreliosis: new tools for old data. *Infect.*  
591 *Genet. Evol.* 10, 989-997.

592 Manangwa, O., De Meeûs, T., Grébaud, P., Segard, A., Byamungu, M., Ravel, S., 2019.  
593 Detecting Wahlund effects together with amplification problems : cryptic species, null  
594 alleles and short allele dominance in *Glossina pallidipes* populations from Tanzania. *Mol.*  
595 *Ecol. Res.* 19, 757-772.

596 Nei, M., Chesser, R.K., 1983. Estimation of fixation indices and gene diversities. *Ann.*  
597 *Hum. Genet.* 47, 253-259.

598 Norris, D.E., Klompen, J.S.H., Keiransand, J.E., Black, I.W.C., 1996. Population genetics  
599 of *Ixodes scapularis* (Acari: Ixodidae) based on mitochondrial 16S and 12S genes. *J.*  
600 *Med. Entomol.* 33, 78-89.

601 Qiu, W.G., Dykhuizen, D.E., Acosta, M.S., Luft, B.J., 2002. Geographic uniformity of the  
602 Lyme disease spirochete (*Borrelia burgdorferi*) and its shared history with tick vector  
603 (*Ixodes scapularis*) in the Northeastern United States. *Genetics* 160, 833-849.

604 R-Core-Team, 2018. R: A Language and Environment for Statistical Computing, Version  
605 3.5.0 (2018-04-23) Ed. R Foundation for Statistical Computing, Vienna, Austria,  
606 <http://www.R-project.org>.

607 Rousset, F., 1996. Equilibrium values of measures of population subdivision for stepwise  
608 mutation processes. *Genetics* 142, 1357-1362.

609 Rulison, E.L., Kuczaj, I., Pang, G., Hickling, G.J., Tsao, J.I., Ginsberg, H.S., 2013.  
610 Flagging versus dragging as sampling methods for nymphal *Ixodes scapularis* (Acari:  
611 Ixodidae). *J. Vector Ecol.* 38, 163-167.

612 Sakamoto, J.M., Goddard, J., Rasgon, J.L., 2014. Population and demographic structure  
613 of *Ixodes scapularis* Say in the eastern United States. *PLoS One* 9, e101389.

614 Séré, M., Thévenon, S., Belem, A.M.G., De Meeûs, T., 2017a. Comparison of different  
615 genetic distances to test isolation by distance between populations. *Heredity* 119, 55-63.

616 Séré, M., Thévenon, S., Belem, A.M.G., De Meeûs, T., 2017b. Comparison of different  
617 genetic distances to test isolation by distance between populations. *Heredity* 119, 55-63.

618 Van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M., Shipley, P., 2004. MICRO-  
619 CHECKER: software for identifying and correcting genotyping errors in microsatellite data.  
620 *Mol. Ecol. Notes* 4, 535-538.

621 Wang, C., Schroeder, K.B., Rosenberg, N.A., 2012. A maximum-likelihood method to  
622 correct for allelic dropout in microsatellite data with no replicate genotypes. *Genetics* 192,  
623 651-669.

624 Waples, R.S., 2006. A bias correction for estimates of effective population size based on  
625 linkage disequilibrium at unlinked gene loci. *Conserv Genet* 7, 167-184.

626 Waples, R.S., 2015. Testing for Hardy-Weinberg proportions: have we lost the plot? *J.*  
627 *Hered.* 106, 1-19.

628 Waples, R.S., 2018. Null alleles and  $F_{IS} \times F_{ST}$  correlations. *J. Hered.* In press.

629 Waples, R.S., Do, C., 2010. Linkage disequilibrium estimates of contemporary  $N_e$  using  
630 highly variable genetic markers: a largely untapped resource for applied conservation and  
631 evolution. *Evol. Appl.* 3, 244-262.

632 Wattier, R., Engel, C.R., Saumitou-Laprade, P., Valero, M., 1998. Short allele dominance  
633 as a source of heterozygote deficiency at microsatellite loci: experimental evidence at the  
634 dinucleotide locus Gv1CT in *Gracilaria gracilis* (Rhodophyta). *Mol. Ecol.* 7, 1569-1573.

635 Weinberg, W., 1908. Über den Nachweis der Verebung beim Menschen. *Jahresheft des*  
636 *Vereins für Vaterländische Naturkunde in Württemberg* 64, 368-382.

637 Weir, B.S., Cockerham, C.C., 1984. Estimating F-statistics for the analysis of population  
638 structure. *Evolution* 38, 1358-1370.

- 639 Wright, S., 1965. The interpretation of population structure by F-statistics with special  
640 regard to system of mating. *Evolution* 19, 395-420.
- 641 Yuval, B., Spielman, A., 1990. Duration and Regulation of the Developmental Cycle of  
642 *Ixodes dammini* (Acari: Ixodidae). *J. Med. Entomol.* 27, 196-201.
- 643
- 644