4. RESULTS

4.1 Sampling sites

As described in the Materials and Methods, 30 sampling sites were identified in PAT (Fig. 4.1 and Table 4.1). In order to explore possible epidemiological implications, such as differential host use, specifically for the bloodmeal analysis (see 4.2.3), we chose sites in extensive forests (N=18; EXTF) and forest patches surrounding human settlements (N=12; PATF) (Table 4.1 and Fig 4.1). Instead, in order to interpret population genetics patterns, it was also important to distribute site locations according to relevant geographical barriers (i.e. mountain chains; rivers) that may influence gene flow between tick populations. In PAT, previous genetic studies on large mammalslike roe deer and red deer have shown that the Adige River Valley, that cuts north-south through the Trento-Alto Adige Region, is a major barrier for gene flow between populations on the West and East side of the river. Since these same species may have an important role in tick dispersal, sampling sites have also been grouped in East (13) and West (17) Province of Trento in some analyses (Table 4.1 and Fig. 4.1).

Hence, the sites were chosen from a range of altitudes, from the lowest, Ala, at 253 m a.s.l, to the highest, Lundo (San Giovanni), at 1264 m a.s.l., evenly distributed throughout PAT, with at least one site in each major valley in typical tick habitat. Ticks were mainly sampled in high stand mixed (coniferous and broad leaved forest) or broad leaved forests, highly favorable tick habitat (Rizzoli et al. 2009; Gray 1998); in two cases, Val Genova and Revò, conifers were predominant, but ground cover, maintaining a sufficient level of humidity for ticks survival, was guaranteed by a rich understory.

Table 4.1 Sampling site parameters. Area (EAST or WEST) is in reference to the Adige River (see Fig. 4.1). EXP: exposure; EXTF: Extensive forst; PATF: Forest patch near urban area. Tick sampling was carried out at each site for questing ticks (Q) or both questing and feeding ticks (Q & F).

Area	Site code	Site (locality)	Coordinates DMS		m a.s.l.	EXP	Brief site description for questing tick sampling	Forest type	Habitat type	Sampling type
EAST	TRA	Transacqua	46° 9'52.09"N	11°49'57.10"E	850	NW	mountain roadside in forest	high stand mixed forest	EXTF	Q & F
	CAO	Caoria	46°11'27.15"N	11°40'59.16"E	940	flat	rural area	meadows and high stand mixed forest	EXTF	Q
	ALA	Ala (Sdruzzinà)	45°44'2.85"N	10°58'29.26"E	253	N	walking path in forest	high stand broad-leaved	EXTF	Q
	VOL	Volano	45°54'55.71"N	11° 5'9.81"E	255	N	walking path in forest	coppices	EXTF	Q
	TRE	Trento (Parco Casteller)	46° 2'7.20"N	11° 8'21.10"E	400	W	urban park	coppices	PATF	Q
	GRI	Grigno	46° 0'42.37"N	11°37'23.61"E	254	flat	SIC area	coppices	EXTF	Q & F
	LEV	Levico	46° 1'10.77"N	11°17'55.83"E	533	S	walking path near the village	coppices	PATF	Q
	PER	Pergine	46° 3'28.68"N	11°12'38.41"E	570	S-E	walking path in forest nearby meadows	high stand broad-leaved forest	PATF	Q
	TEL	Telve	46° 4'12.39"N	11°27'49.05"E	831	SW	walking path in forest	high stand mixed forest	EXTF	Q
	TES	Tesino	46° 3'12.67"N	11°37'54.98"E	749	SW	walking path in rural area	meadows and high stand mixed forest	PATF	Q
	SEG	Segonzano	46°11'6.24"N	11°15'56.30"E	809	SW	walking path in forest	high stand broad-leaved forest	PATF	Q
	GIO	Giovo	46° 9'25.23"N	11° 9'31.92"E	556	-	walking path in forest (houses nearby)	high stand mixed forest	PATF	Q & F
	CVS	Cavalese	46°16'53.97"N	11°27'23.10"E	870	SW	walking path in forest nearby meadows	high stand mixed forest	PATF	Q

Table 4.1 Con'd

Area	Site code	Site (locality)	Coordinates DMS		m a.s.l.	EXP	Brief site description	forest type	habitat	sampling type
WEST	PIN	Pinzolo (Giustino)	46° 9'13.37"N	10°46'22.15"E	895	SW	walking path in forest (houses nearby)	high stand mixed forest	PATF	Q & F
	LUN	Lundo - San Giovanni	45°58'58.80"N	10°53'21.63"E	1264	SE	random sampling in forest	high stand mixed forest	EXTF	Q & F
	VGE	Val Genova	46°10'0.39"N	10°39'40.61"E	1130	S	walking path in forest	high stand mixed forest (mainly coniferous)	EXTF	Q
	TIO	Tione di Trento	46° 2'33.59"N	10°43'59.72"E	566	SW	walking path in rural area	high stand mixed forest	EXTF	Q
	CON	Condino (Calamara)	45°53'5.71"N	10°36'9.01"E	454	W	walking path in rural area	meadows and coppices	PATF	Q & F
	PDU	Passo Durone	46° 1'48.93"N	10°48'17.57"E	956	SE	walking path in forest nearby meadows	high stand mixed forest	EXTF	Q
	LED	Ledro	45°54'15.35"N	10°43'56.86"E	822	W	walking path from the village to the forest	meadows and high stand mixed forest	PATF	Q
	PIE	Pietramurata	46° 0'56.23"N	10°55'27.34"E	677	E	random sampling in forest	high stand broad-leaved	EXTF	Q
	LAM	Laghi di Lamar	46° 7'55.31"N	11° 3'50.47"E	730	SW	walking path along the lake	high stand broad-leaved	EXTF	Q
	CAD	Cadine	46° 5'51.77"N	11° 4'27.45"E	550	W	random sampling in forest	coppices	EXTF	Q & F
	CAV	Cavedine (Dos gaggio)	45°59'7.36"N	10°57'48.26"E	776	flat	walking path in forest	meadows and high stand broad-leaved forest	EXTF	Q & F
	BRE	Brentonico (Castione)	45°50'6.51"N	10°57'45.66"E	615	NW	roadside in rural area	coppices	EXTF	Q
	MOL	Molveno	46° 7'21.99"N	10°57'58.08"E	981	W	walking path along the lake	high stand beech forest	EXTF	Q
	DIM	Dimaro	46°19'6.88"N	10°52'0.03"E	950	N	walking path in forest (mountain houses nearby)	high stand mixed forest	PATF	Q
	MEZ	Mezzocorona	46°14'25.57"N	11° 3'33.16"E	270	flat	SIC area	coppices	EXTF	Q & F
	REV	Revò	46°25'5.90"N	11° 4'44.28"E	827	SE	walking path in forest (apple orchards nearby)	high stand mixed forest	PATF	Q & F
	VER	Vervò	46°18'32.33"N	11° 7'30.48"E	965	SE	walking path in forest	high stand mixed forest (mainly *Pinus* spp.)	EXTF	Q

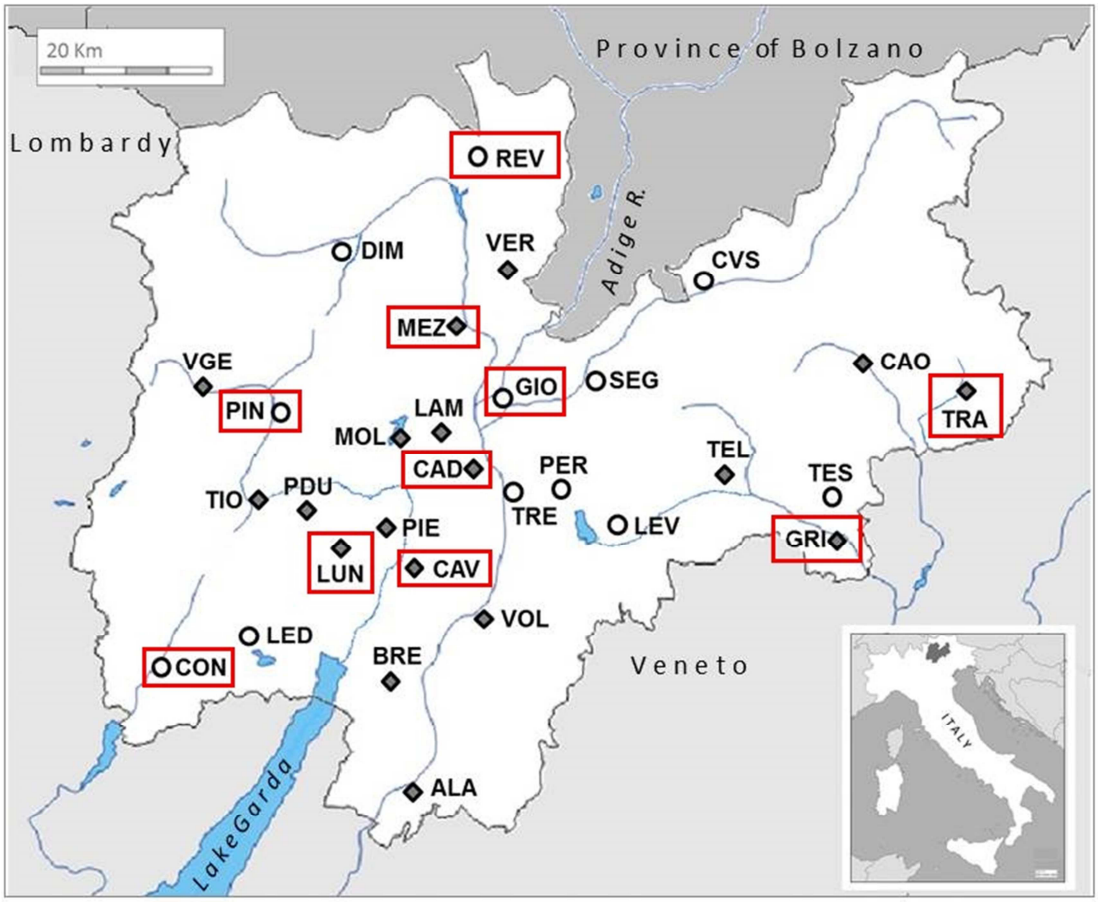


Fig. 4.1 Location of sampling sites in PAT (northeastern Italy: see inset). Note that the Adige River cuts longitudinally across the Province. Sites in which both questing and feeding tick were collected are marked by a red box. Open circles represent sites in forest patches (PATF), while solid diamonds correspond to sites in or at the edge of extensive forests (EXTF).

4.1.1 Questing tick sampling

Generally, the sampling target for each site was reached (see section 3.2.2), except for the sites VGE, DIM, LED, LUN and ALA, where density of adult ticks was particularly low (i.e. even after repeated sampling it was not possible to reach the goal of 12 adults individuals; Table 4.2). In total, 1887 ticks were sampled (314 females, 269 males and 1304 nymphs). Ticks density varied from 1/100 m^2 for TES to 63/100 m^2 for ALA.

Table 4.2 Questing tick sampling results (F=adult females; M=adult males and N=nymphs). For site codes, see Table 4.1.

Site code	Collection date	Tick density (ticks/100 m^2)	F	M	N	Total
ALA	10 May 2012	63/100	5	4	55	64
BRE	10 May 2012	17/100	15	11	54	80
CAD	07 May 2012	21/100	7	16	56	79
CAO	24 April 2013	18/100	10	12	54	76
CVS	23 May 2012	6/100	11	3	23	37
CAV	17 April 2012; 16 April 2013	-	24	14	102	140
CON	27 April 2012	55/100	18	3	48	69
DIM	02 June 2013	8/100	4	9	40	53
GIO	23 April 2013	17/100	11	10	56	77
GRI	24 April 2013	13/100	8	10	50	68
LAM	22 May 2012	46/100	11	10	41	62
LED	25 May and 22 June 2012	6/100	5	4	44	53
LEV	26 April 2012	19/100	15	12	37	64
LUN	14 May 2012	4/100	3	5	35	43
MEZ	23 April and 03 June 2013	-	11	10	37	58
MOL	24 May 2012	2/100	14	6	36	56
PDU	14 May 2012	-	10	7	27	44
PER	26 April 2012	9/100	4	8	38	50
PIE	07 May 2012	6/100	9	10	59	78
PIN	11 May 2012	4/100	8	5	32	45
REV	29 May 2012	20/100	12	12	30	54
SEG	23 May 2012	53/100	8	14	41	63
TEL	21 June 2012	3/100	7	6	21	34
TES	25 June 2012	1/100	10	10	37	57
TIO	11 May 2012	12/100	19	11	37	67
TRA	05 June 2012	19/100	8	9	43	60
TRE	16 April and 02 May 2013	16/100	24	15	53	92
VGE	26 May 2013	12/100	5	4	36	45
VER	30 May 2012	5/100	8	7	37	52
VOL	22 May 2012	20/100	10	12	45	67
		TOTALS	**314**	**269**	**1304**	**1887**

4.1.2 Feeding tick sampling

4.1.2.1 Trapping of small mammals

As expected, because of its ubiquity in all forest types, *Apodemus* spp. were captured in all sites, whereas *M. glareolus* was captured only at LUN, GRI and PRI sites (Table 4.3). A total of 865 ticks from *Apodemus* spp. were collected and the minimum target of 30 samples for each site was reached. For *Apodemus* spp., the highest tick load (number of ticks/parasitized animal) was found at CON, with a mean of 17 ticks/individual, mainly larvae, whereas at other sites load ranged between 2 (LUN) and 7 (MEZ). This result could related to the specific date of sampling (mid August) when the larval stage normally has a density peak (Tagliapietra et al., 2011). 95% (825/865) of the sample were larvae, while nymphs and adults represented only a small fraction of the ticks parasitizing small mammals. Other small mammals species were also captured; specifically in LUN, *Muscardinus avellanarius* was trapped carrying 2 larvae, as well as 9 *Sorex* individuals, carrying a total of 44 larvae.

Table 4.3 Feeding ticks collected from small mammals (l=larvae; n=nymphs and a=adults).

	Species									
	Apodemus spp.					*Myodes glareolus*				
Site	Ind*	l	n	a	total	Ind*	l	n	a	total
CAD	24	104	8	0	112	n.r.[#]				
CAV	33	172	12	0	184	n.r.				
LUN	17	29	2	2	33	14	21	2	0	23
GRI	78	-	-	-	> 200	4	-	-	-	>10
CON	8	134	3	0	137	n.r.				
PIN	39	115	1	0	116	n.r.				
REV	18	75	1	0	76	n.r.				
PRI	20	66	0	0	66	9	29	7	0	36
GIO	13	63	5	0	68	n.r.				
MEZ	11	67	6	0	73	n.r.				
Totals	261	825	38	2	865	27	50	9	0	59

*number of trapped animals parasitized by ticks; -: number of tick/stage not available; [#] n.r.: species not retrieved in this site.

4.1.2.2 Bird netting

In all sites other than MEZ, the three focal host species were present. *T. merula* and *T. philomelos* presented the highest tick load (number of ticks/parasitized animals) with 5 ticks/individual; *T. merula* carried twice as many nymphs as larvae, while *T. philomelos* was infested almost equally by larvae and nymphs (Table 4.4). On the other hand, *E. rubecula* carried a mean of fewer ticks (2 ticks/individual) and mainly larvae; consequently, for this species, I was unable to reach the target number of feeding ticks except where a high number of *E. rubecula* were trapped (CAV and GIO), and REV (a single individual carried 14 larvae). No adult ticks were found feeding on birds.

Table 4.4 Feeding ticks collected from birds for the most frequently netted species (tick l=larvae; n=nymphs and a=adults) at the 10 selected sampling sites in PAT.

Species	***Turdus merula***					***Turdus philomelos***					***Erithacus rubecula***				
Site	Ind*	l	n	a	total	Ind*	l	n	a	total	Ind*	l	n	a	total
CAD	14	23	61	0	84	13	33	17	0	50	8	15	6	0	0
CAV	97	·	-	-	428	7	4	19	0	23	27	26	15	0	41
LUN	7	6	16	0	22	4	-	-	-	37	2	1	1	0	2
GRI	8	3	29	0	32	7	13	36	0	49	3	4	2	0	6
CON	4	11	25	0	36	9	37	32	0	69	4	6	3	0	9
PIN	10	62	9	0	71	4	23	23	0	46	4	3	5	0	8
REV	15	43	21	0	64	6	25	4	0	29	8	34	1	0	35
PRI	9	8	29	0	37	7	15	2	0	17	9	19	3	0	22
GIO	8	25	37	0	62	10	17	33		50	20	63	7	0	70
MEZ	26	20	48	0	68	5	0	10	0	10	n.r.[#]				0
Totals	198	244	560	0	904	72	167	176	0	343	85	171	43	0	171

*number of trapped animals parasitized by ticks; -: number of tick/stage not available; [#] not retrieved.

In addition to these three species, a large variety of other birds were netted and found parasitized by ticks (Table 4.5), even though some of them are not ground feeding species. It is interesting to note that *Sylvia atricapilla* and *Parus major*, netted at a lower frequency in all the sites, were also frequently found infested by ticks, although with a low number (mean load 1 or 2 ticks/infested individual), while three *Garrulus glandarius* carried in average 6 ticks each, an higher load possibly related to the greater body size of this species (Marsot et al., 2012).

Table 4.5 Feeding ticks on bird species that were less frequently trapped and/or are not normally considered important tick hosts (l=larvae and n=nymphs).

Species	Ind*	l	n	Total
Sylvia atricapilla	16	16	6	22
Parus major	11	14	8	22
Garrulus glandarius	3	5	12	17
Luscinia megarhynchos	7	5	10	15
Poecile palustris	2	1	8	9
Troglodytes troglodytes	3	7	1	8
Parus scopaiola	5	1	7	8
Pyrrhula pyrrhula	2	0	5	5
Fringilla coelebs	4	2	2	4
Aegithalos caudatus	1	0	3	3
Turdus viscivorus	1	1	1	2
Phylloscopus collybita	1	1	1	2
Phoenicurus phoenicurus	2	1	1	2
Motacilla cinerea	1	0	2	2
Acrocephalus palustris	1	0	2	2
Sylvia borin	1	0	1	1
Sitta europaea	1	0	0	1
Phoenicurus ochruros	1	0	1	1
Muscicapa striata	1	1	0	1
Lanius collurio	1	1	0	1
Hippolais icterina	1	0	1	1
Coccothraustes coccothraustes	1	1	0	1

*number of trapped animals parasitized by ticks.

4.1.2.3 Collection of ticks feeding on large mammals

C. capreolus and *C. elaphus* provided most of the ticks from wild large mammals, with 1204 and 242 samples collected, respectively (Table 4.6). *C. capreolus* provided the highest number of ticks since population sizes permit many more hunting licenses for this species than for *C. elaphus* in PAT; in additional, there were high numbers of tick larvae on the collected roe deer forelegs (808). *R. rupicapra* individuals

were found to be parasitized by *I. ricinus* only in GRI; at this site, from 10 individuals, a total of 30 ticks (5 larvae, 2 nymphs and 23 adults) were collected. In PIN, an *Ovis aries musimon* was controlled for ticks and 1 nymph and 3 adults were collected.

Table 4.6 Feeding ticks collected from large mammals (l=larvae; n=nymphs and a=adults).

Species	***Capreolus capreolus***					***Cervus elaphus***				
Site	Ind*	l	n	a	total	Ind*	l	n	a	total
CAD	25	124	18	20	162	3	0	0	10	10
CAV	16	244	67	20	331	1	5	6	14	25
LUN	9	107	27	0	134	3	5	9	1	15
GRI	13	13	6	21	40	5	0	0	16	16
CON	7	40	14	25	79	1	0	0	2	2
PIN	20	13	7	42	62	12	0	1	38	39
REV	16	118	25	30	173	9	8	4	32	44
PRI	5	7	3	9	19	6	16	4	23	43
GIO	11	111	27	20	158	2	15	5	28	48
MEZ[#]	7	31	11	4	46					0
Totals	129	808	205	191	1204	42	49	29	164	242

*number of animals checked (carcasses and/or forelegs) and parasitized by ticks; [#]*C. elaphus* not hunted in this site.

I sampled ticks from domestic sheep (*O. aries*) in a single flock in CAV; 16 *I. ricinus* females were collected from 16 animals.

Ticks were also collected from domestic dogs. In REV, 4 ticks adults were collected simultaneously from a dog living in a farm surrounded by patches of forest and apple orchards. In CON and PIN, ticks were collected over time from two dogs for a total of 11 and 3 adults, respectively; both were family pets frequently visiting forests.

4.2 DNA extraction

Here, I briefly report the most challenging problem we had to face during DNA extraction, i.e. obtaining good quality and highly concentrated gDNA for the application of RAD-Sequencing to *I. ricinus*.

As reported in section 3.3.2.2 for adult ticks, three different approaches were used. With the spin-column, Qiagen Blood & Tissue kit, we obtained good results both in terms of quality and quality of DNA when we applied this method to fresh ticks, or ticks that were frozen for only a few weeks (Fig. 4.2a); however, when the same kit was applied to samples preserved at -80 °C freezer for 2-3 months, insufficient quality and quantity gDNA was obtained (Fig. 4.2b). With the phenol-phenol protocol, we obtained good results even from ticks frozen for a longer period of time, but since the results of L. Cornetti (personal communication) suggested that samples extracted with this method did not produce good RAD sequencing data, we did not use this DNA protocol DNA for further extractions.

Subsequently, we tested the recently acquired magnetic-based DNA purification reagents and instrumentation. This method allowed gDNA extraction from frozen adult ticks, and remarkably, the quality and quantity of extracted DNA was significantly improved in comparison to the Qiagen kit used on fresh ticks (Tab. 4.7; Fig. 4.3), as the Mann-Whitney test for statistical difference between the two samples (i.e. different extraction methods) highly supports (Table 4.7).

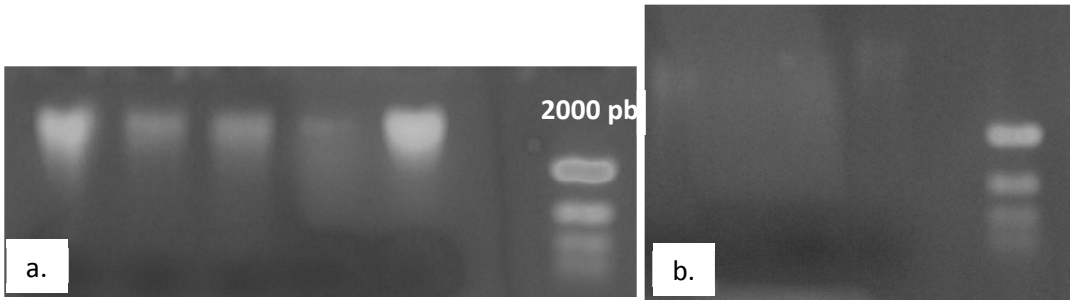


Fig. 4.2 Photographs of the gel electrophoresis of DNA extracted from questing adult ticks using the manual spin-column Qiagen Blood & Tissue Kit **a**. gDNA obtained from fresh ticks or from ticks frozen for a few weeks (-80°C); the compact bands with high molecular weight indicate good quality/quantity DNA. **b**. genomic DNA obtained from 3-4 months old ticks preserved at – 80°C; note that the gDNA is highly degraded.

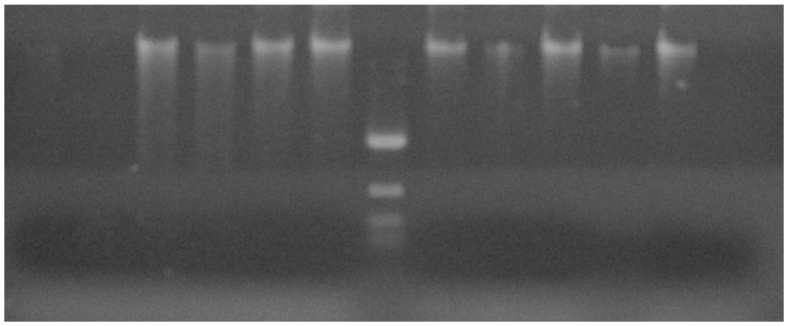


Fig. 4.3 Photographs of the gel electrophoresis of gDNA obtained from ticks preserved at – 80°C for 3-4 months using the KingFisher™ Cell and Tissue Kit with magnetic-beads.

Table 4.7 Comparison of gDNA extraction results from adult questing ticks (F= adult female; M= adult male) preserved at -80°C, with the two described methods. **p<0.05.

DNA extraction method	**Mean gDNA concentration (ng/µL ± S.E.)**	
	F	**M**
Spin-column, QIAgen Blood&Tissue	15.15 (±3.0)	4.9 (±1.7)
Magnetic-beads, KingFisher Cell and Tissue Kit	36.49 (±3.5)	12.02 (±1.2)
Mann-Whitney test	**	**

4.3 Protocol development for bloodmeal analysis in questing ticks

4.3.1 DNA sequencing: universal vertebrate *cytochrome b* primers

The method was applied to 33 questing nymphs with unknown bloodmeal source. As *Bos taurus* was amplified in the preliminary trials of this protocol, I excluded BSA from the reaction mix and *B. taurus* amplification no longer occurred. No potential host DNA was amplified in any sample. Human DNA was occasionally obtained. We attempted to design a new primer pair targeting a shorter fragment of *cytochrome b* that still allowed the amplification of a large number of vertebrate species. *Cytb* sequences available from GenBank and sequences we obtained during the optimization step from control samples, were aligned using Clustal X v. 2.0. A reference sequence of *Homo sapiens* was also added to avoid designing primers cross-reacting with contaminant human DNA. However, the high interspecific variability in this mtDNA region did not permit new primer design with the necessary features described above.

4.3.2 DNA sequencing: species-specific primers

Both primer pairs (*Apodemus* and *Capreolus*) proved to reliably amplify host DNA from control samples of target species and, therefore, the optimized protocols were applied to questing ticks (Table 4.8). Bloodmeal source was identified as *A. flavicollis* in 5 out of 115 ticks using the long amplicon primer set, and in 9 out of 109 using the short amplicon primer set. *C. capreolus* was amplified in 1 out of 97 questing ticks. There was no repeatability between amplifications obtained from the two *Apodemus* primer sets. In an attempt to improve repeatability of amplification results,

we used bloodmeal-positive questing tick DNA samples to test different settings in the thermal cycling and/or reaction mix. No improvements were achieved.

4.3.3 DNA sequencing: host group primers

As reported in Table 4.9, five group specific primer sets were designed; however, only use of degenerations allowed the primers to target all the selected species in *cytb* and *d-loop* mtDNA regions. Testing on control samples of target species showed that all primers, except Ruminants, resulted in reliable amplifications. Optimized protocols were tested on a large number (97 to 115) of questing ticks (mainly nymphs; Table 4.9). As primers were not always tested on the same tick, we could infer identification success rate only at primer set level and not globally for the 'host group' approach.

Table 4.8 Species-specific primer features; results of the testing on control samples (K+ampl) and testing on questing ticks.

	Primer name and sequence	Ta (°C)	length(bp)	Host species amplified	(n°) K+ ampl.	n° questing ticks	n° ticks with bloodmeal ID
Apodemus	*BM_cytb_Apodemus* (long amplicon) F_AATACACTATACATCAGACACA R_TACTGCGAATAGGAGAAT	53	214	*A. flavicollis; A. sylvaticus*	(4) good	115	5
	BM_cytb_Apodemus_S (short amplicon) F_AATACACTATACATCAGACACA R_s_GTCCTACGTGTAGAAATAAG	52	140	*A. flavicollis; A. sylvaticus*	(4) good	109	9
Capreolus	*BM_capreolus* (Garros et al. 2011) UNIV2_F_TGAGGACAAATATCATTYTGAGGRGC CAP_R_TTGTCCGCGTTTGATGGGATTCCTATC	56	240	*C. capreolus*	(3) good	97	1

Table 4.9 Host group primer features and target species; results of the testing on control samples (K+ampl) and testing on questing ticks.

Host group	Primer name and sequence	mtDNA target	Ta (°C)	length (bp)	Host species amplified	(n°) K+ ampl.	n° questing ticks	n° ticks with bloodmeal ID
Rodents	BM_cytb_Rodents F_GCTGTHATAGCMACWGCAT R_GTRGCTTTRTCWACTGAGAA	*cytb*	53	150	*A. flavicollis, A. sylvaticus, M. glareolus, S. vulgaris, M. avellanarius*	(2, -, 1, 1, 1) good	86	4
Soricomorpha	BM_d-loop_Inset F_GCRTATCAYCTCCAWTRGGTTAT R_GGGCGATTTTAGGTGAGAT	*d-loop*	55	173	*S. araneus, S. antinorii, C. russula, C. suaveolens, E. europaeus*	(-, 1, -, 1, 1) good	77	1
Ruminants	BM_d-loop_Rumin F_CCYCWTGCWTATAAGC R_GCAGGTSAWYAAGCTC	*d-loop*	53	144	*C. elaphus, R. rupicapra, C. capreolus*	(1, 1, 1) variable	70	1
Canidae	BM_cytb_Canidi F_CTGCCGAGACGTTAACTA R_CCAATRTTTCATGTTTCTATG	*cytb*		141	*C. l. domesticus, V. vulpes*	(2, 1) good	59	6
Passeriformes	BM_cytb_Passeriformi F_TACACAGCAGAYACBWCHCTAG R_GTTTCAGGTTTCTTTRTT	*cytb*	54	180	*T. merula, T. philomelus, E. rubecula, L. megarhyncos, P. major*	(1, 1, 1, 1, 1) good	83	0

4.3.4 Standardized semi-automated test for species-specific and host group primers

Complementary use of species-specific primers (for *Apodemus*, long and short amplicons) and host group primers (Rodents, Soricomorpha, Passeriformes) allowed bloodmeal source identification in 10 out of 94 analyzed nymphs (10.6%). Sequencing allowed species level host identification, as follows: three *C. lupus familiaris*, two *A. flavicollis*, and one each of *M. glareolus*, *M. musculus*, *C. suaveolens*, *B. taurus* and *L. europaeus*. However, we also report that: *C. l. familiaris* was amplified with the Rodents primers, *L. europaeus* with *Apodemus* (long amplicon) primers and *C. capreolus* with *Apodemus* (short amplicon). Additionally, only amplification of *L. europaeus* was obtained for each one of the three sample replicates. Considering the low specificity of these primers, we did not continue with this approach.

4.3.5 Real-time High Resolution Melting Analysis

4.3.5.1 Group-specific primers design and testing

Six group-specific primers, targeting the 20 chosen host species, were selected to allow identification of tick bloodmeal sources using HRMA (Table 4.10). For each of these primer pairs, conventional PCR resulted in amplification of the expected mtDNA target for all the control samples from both tissue and engorged ticks, as confirmed by BLASTn searches. None of the PCR negative controls or the human DNA samples were amplified. Primer dimers or short aspecific multiple amplicons were occasionally visible in the QIAxcel images (reported in Fig. 4.4 and Fig 4.5), as is typical of PCR reactions for which appropriate template is lacking. Since amplification of non-target species included in the testing of Muroidea, Soricidae, Passeriformes or Caprinae primers did not occur during conventional PCR (data not shown), only DNA control

samples from the target species were included in the testing of Real-time HRMA of each host group.

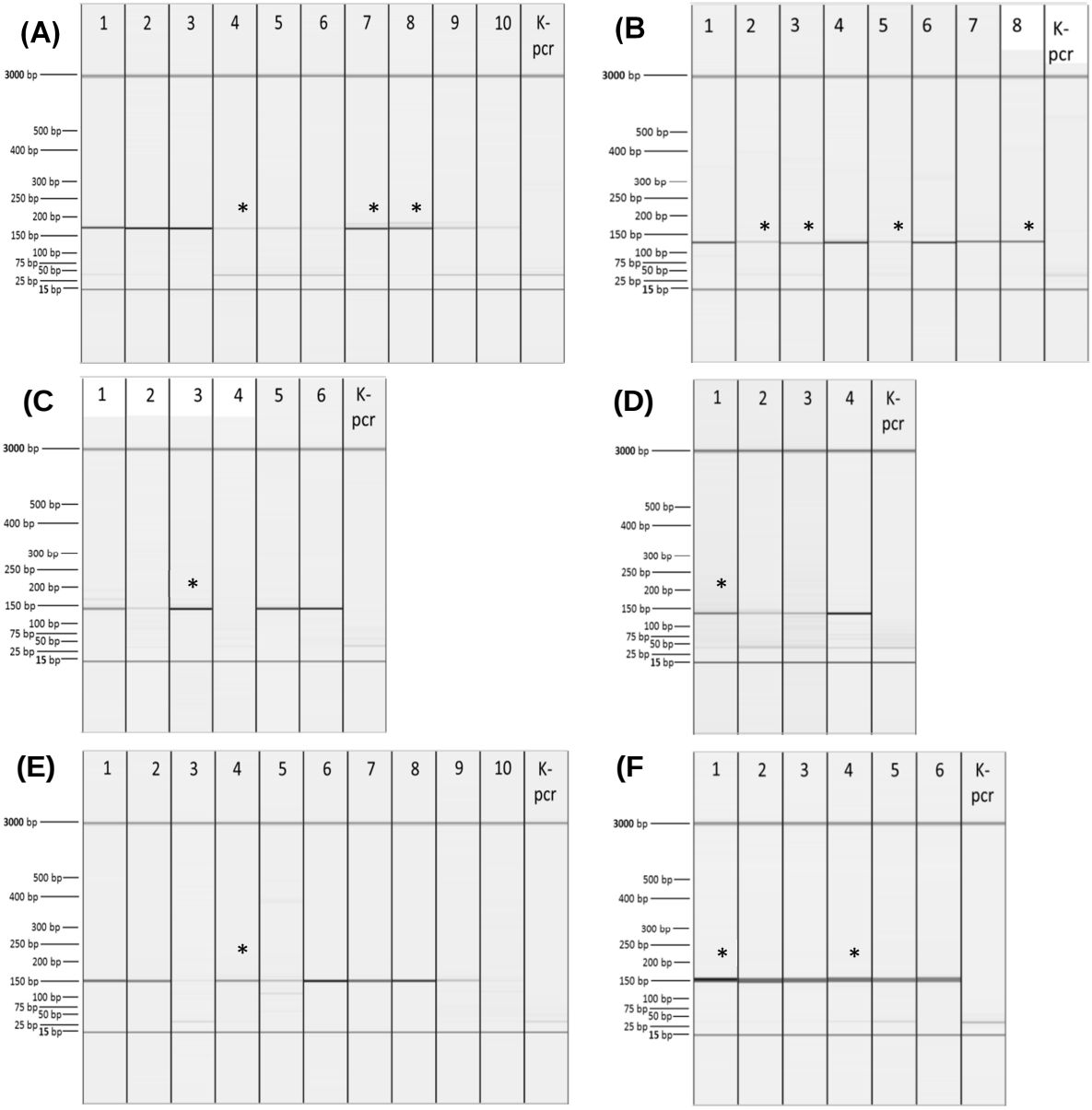


Fig. 4.4 QIAxcel capillary electrophoresis images of conventional PCR of a representative subset of DNA control samples obtained from host tissue or from engorged ticks directly collected from the host, and from human DNA, for each pair of HRMA group-specific primers. * sequenced PCR products. Size markers can be seen at 15 bp and 3000 bp. Bands appearing at about 40 bp are primer dimers. These bands disappeared when we increased the quantity of DNA template in the Real-time reaction. (A) Muroidea: *M. m. domesticus*: lane 1, tissue; *M. glareolus*: lanes 2-3, tissues; lanes 4-5, engorged *I. ricinus* larvae; *A. sylvaticus*: lane 6, 1:100 diluted tissue; lane 7, tissue; *A. flavicollis*: lane 8, tissue; lanes 9-10, engorged *I. ricinus* larvae; (B) Soricidae: *S. antinorii*: lane 1, tissue; lane 2, engorged *I. ricinus* larva; *S. minutus*: lane 3,

tissue; *C. leucodon*: lane 4, tissue; lane 5, engorged *I. ricinus* larva; *C. suaveolens*: lane 6, tissue; lanes 7-8, engorged *I. ricinus* larvae; (C) Passeriformes: *T. merula*: lane 1, tissue; lane 2, engorged *I. ricinus* nymphs; *T. philomelos*: lane 3, tissue; lane 4, engorged *I. ricinus* larva; *E. rubecula*: lane 5, tissue; lane 6, engorged *I. ricinus* larva; (D) Canidae: *C. l. familiaris*: lanes 1-3, engorged *I. ricinus* females; *V. vulpes*: lane 4, tissue; (E) Caprinae: *R. rupicapra*: lanes 1-2, tissues; lanes 3-5, engorged *I. ricinus* females; *C. hircus*: lanes 6-8, tissues; *O. aries*: lanes 9-10, engorged *I. ricinus* females; (F) Cervidae: *C. capreolus*: lanes 1-3, tissues; *C. elaphus*: lanes 4-6, tissues. K-pcr: PCR negative control.

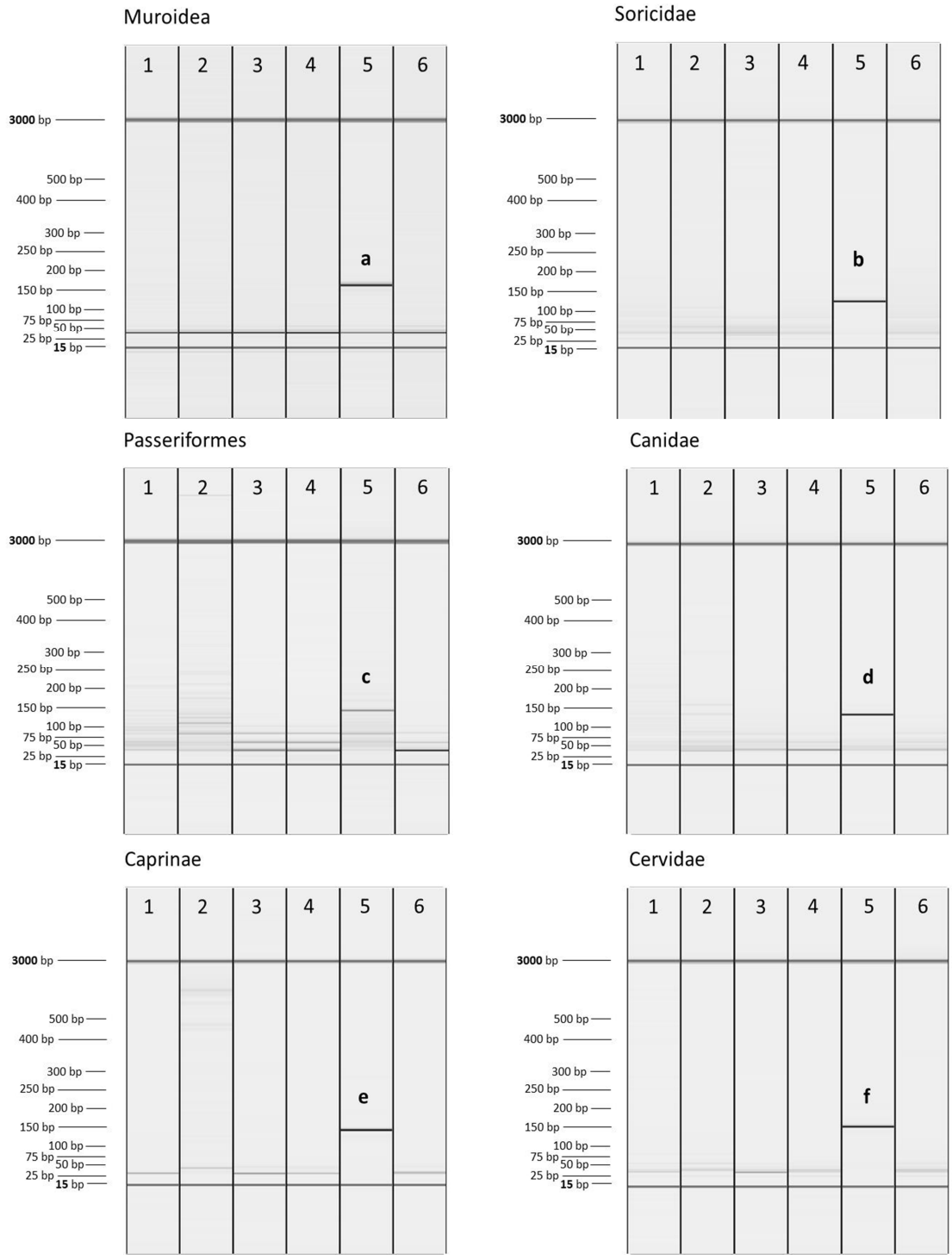


Fig. 4.5 (next page) QIAxcel capillary electrophoresis images of conventional PCR on human DNA for each HRMA group-specific primer set: lane 1, human DNA extracted from partially engorged *I. ricinus* nymph collected while feeding; lane 2, whole human blood; lane 3, human hair; lane 4, negative extraction control; lane 6, negative PCR control. Lane 5, target host species DNA control sample, as follows: a. *M. m. domesticus* (tissue); b. *C. suaveolens* (tissue); c. *T. merula* (engorged *I. ricinus* nymph); d. *V. vulpes* (tissue); e. *C. hircus* (tissue); f. *C. capreolus* (engorged *I. ricinus* female).

4.3.5.2 HRMA validation

Muroidea

Melting temperatures of *M. glareolus* (81.0-81.2 °C) and *M. m. domesticus* (80.7-80.9 °C) *d-loop* amplicons were different enough from each other and from the two *Apodemus sp*. to allow discrimination of these three genera; however, as shown by the normalized melting chart (Fig. 4.6A) and reported in Table 4.10, melting profiles of *A. sylvaticus* and *A. flavicollis* are fully overlapping having a T_m 81.3 °C and 81.4 °C, respectively, so discrimination of the two *Apodemus* species was not possible using these primers and HRMA. The amplified *d-loop* fragment produced a melting profile with a single peak for all four species tested (Fig. 4.6B).

Soricidae

Amplicons produced a melting profile with a single melting peak (Fig. 4.6D) with reliable results between replicates and, where tested, between samples of the same species. The melting temperature of *S. antinorii* (80.1-80.4 °C) and *S. minutus* (80.3-80.5 °C) amplicons were not sufficiently different from each other to be diagnostic for species identification (Fig. 4.6C and Tab. 4.10). However, the melting curves of *Sorex* species were well-separated from those of the *Crocidura* species tested. In addition, *C. suaveolens* (81.4-81.5 °C) can be distinguished from *C. leucodon* (82.1-82.3 °C).

Passeriformes

Melting profiles of the expected amplicons were consistent between replicates and samples of the same species. Furthermore, amplicons produced profiles with one melting peak with T_m diagnostic for the targeted species: *T. merula* 80.4-80.6 °C; *T. philomelos* 83.9-84.0 °C; *E. rubecula* 83.5-83.8 °C (Fig. 4.6E, F; Tab. 4.10). However, since only 0.2 °C separates the melting peaks of *T. philomelos* and *E. rubecula*, we

suggest amplicons be sequenced if unknown samples have melting peaks between the ranges of these two species.

Canidae

The melting profiles of these two species are easily discriminated by HRMA, both on the basis of T_m and by shape (Fig. 4.6G; Tab. 4.10). Amplicons from *C. l. familiaris* produced a profile with a single melting peak (80.3-81.3 °C), although some variability was recorded between samples. In fact, alignment of the sequences obtained from the two most differentiated *C. l. familiaris* HRM profiles (*a*, *b*) showed that sample *a* had one transversion (T->C) compared to sample *b* causing a + 0.5 °C shift in T_m. Instead amplicons of *V. vulpes* from the single DNA sample produced a profile with two melting peaks (81.3 °C and 83.5-83.6[e] °C) (Fig. 4.6H).

Caprinae

Melting profile analysis indicated that *C. hircus* and *R. rupicapra* can be discriminated thanks to their profile with two melting peaks (Fig. 4.6J): although the higher temperature peaks are completely overlapping (80.5 °C *C. hircus* and 80.3-80.6 °C *R. rupicapra*), the lower temperature peaks are separated by an average of 0.92 °C (i.e. T_m 76.5 °C for *C. hircus* and 75.3-75.7 °C for *R. rupicapra*). For *O. aries*, both melting temperature (74.5-75.0 °C and 81.7-82.0 °C) and melting profile shape easily permit identification (Fig. 4.6I and J).

Cervidae

The melting profile of the two species can be easily discriminated (Fig 4.6K and L; Tab. 4.10). Special feature of this *d-loop* amplicon is the presence of two melting regions, resulting in two peaks: 80.5-80.9 °C and 82.5-82.9 °C for *C. capreolus*, and 80.7-80.8

°C and 83.6-83.7 °C for *C. elaphus*. No intraspecific variability was observed between different samples of *C. elaphus*, while for *C. capreolus*, a clear shift of -0.3 °C in both melting peaks was observed for the sample from *C. capreolus* tissue compared to those obtained from engorged ticks (see Fig. 4.6K). To test if this shift could be attributed to variable DNA concentration, serial dilutions of the tissue DNA (starting concentration: 2.44 ng/µL) were subjected to HRMA. As shown in Fig. 4.7 (A, B), if the amplification curve rises early (C_T <25) above the threshold, there is a decrease in T_m (C_T 11: undiluted sample); conversely, the rising of amplification curves between 25 to 31 cycles (diluted samples) does not have a significant effect on melting temperature (Fig. 4.7). However, as Fig. 4.7C shows, even with this slight shift in melting temperature correlated with DNA content, all *C. capreolus* samples (diluted and undiluted) are easily differentiated from *C. elaphus*.

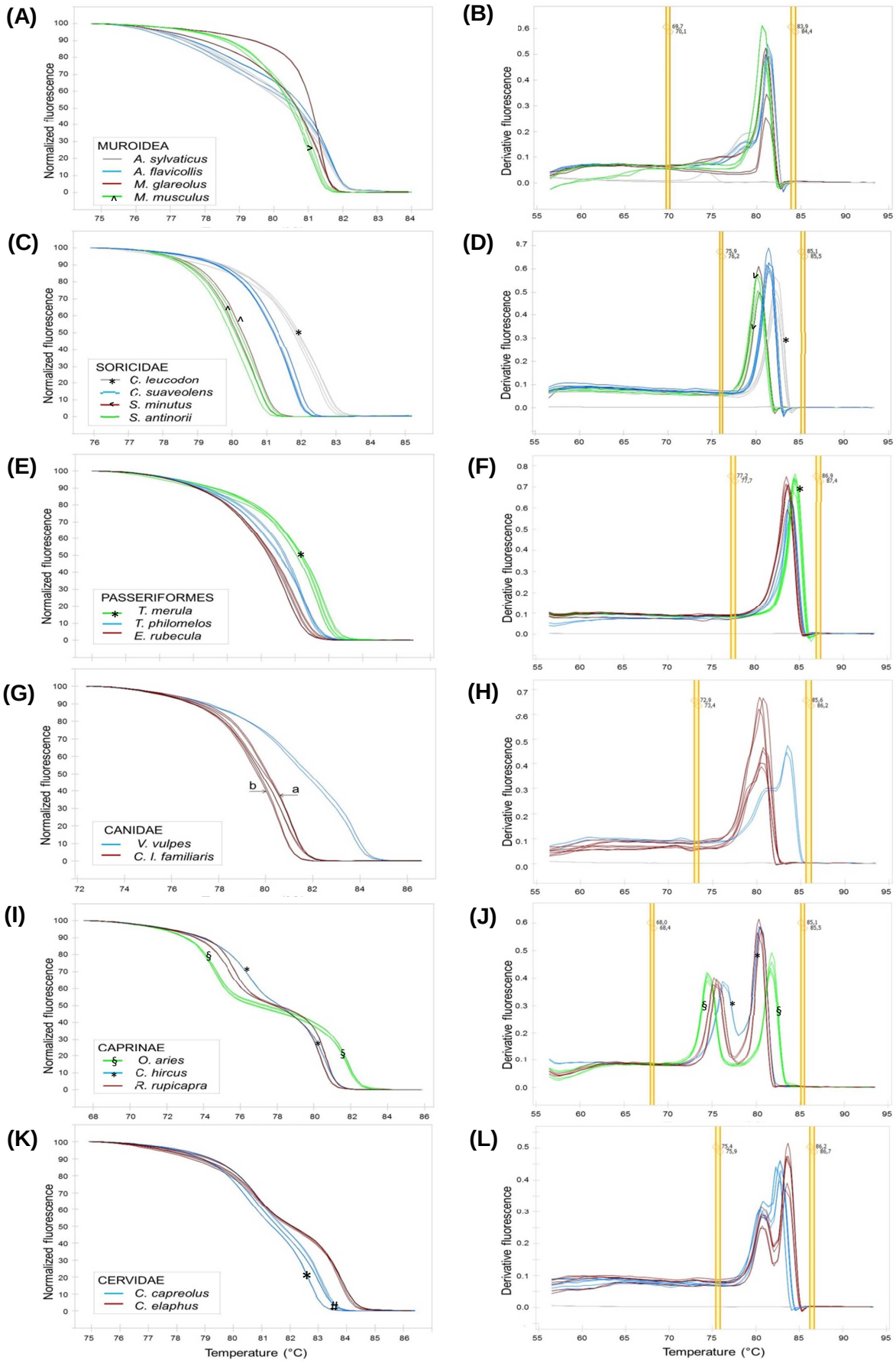
The results on control samples provide the proof-of-principle that the host group primers described here can reliably amplify host DNA and that selected amplicons permit to identify the targeted host species (or genera, in the case of *Sorex* and *Apodemus*) by HRMA.

Table 4.10 Target species within each host group, features of host group specific primers and amplicons and related HRMA parameters. All species within each host group can be distinguished from one another using the HRMA method described here unless otherwise indicated. T_a, annealing temperature; T_m sim., simulated melting temperature from uMELT^{TM} (Dwight et al., 2011); T_m1, melting temperature peak 1; T_m2, melting temperature peak 2 (if present); T_m obs, observed melting temperature expressed as *minimum-maximum* range; n.a., not available. Mean T_m obs was generally 2-6 °C lower than T_m sim. (Footnotes on the following page).

Targets and amplification parameters				**Amplicon features**				**HRMA**			
Host group	**Primer name mtDNA Target Primer (5' - 3')**	**T_a (°C)**	**Target species**	**Size (bp)**	**GC %**	**T_m sim (°C) T_m1**	**T_m2**	**Control samples[a]**	**T_m obs (°C) T_m1**	**T_m2**	**HRM[b] Fig. 4.6**
Muroidea	HRM_Rod *d-loop*	60	*A. flavicollis*	175	45	86.4	-	2E	81.4 [c]	-	A - B
	F_TCTGGTTCTTACTTCAGGGC		*A. sylvaticus*	175	44	86.1	-	2T	81.3 [c]	-	
	R_TTCATGCCTTGACGGCTATG		*M. glareolus*	176	44	85.9	-	2E	81.0-81.2	-	
			M. m. domesticus	175	43	85.5	-	2T	80.7-80.9	-	
Soricidae	HRM_Sor *d-loop*	62	*S. minutus*	137	44	84.2	-	1T	80.3-80.5[d-e]	-	C - D
	F_TCAGCCCATGCCGACACAT		*S. antinorii*	137	43	83.9	-	1T – 1E	80.1-80.4 [d]	-	
	R_GCCCCCATAGAGAATAAGCC		*S. araneus*	137	42	83.4	-	n.a.			
			C. leucodon	136	45	84.7	-	1T – 1E	82.1-82.3	-	
			C. suaveolens	137	47	86.0	-	1T – 2E	81.4-81.5	-	
			C. russula	136	46	85.0	-	n.a.			
Passeriformes	HRM_Pas *12S*	60	*T. merula*	155	52	90.2	-	1T – 3E	80.4-80.6	-	E - F
	F_ATCCACGATATTACCTGACCATT		*T. philomelos*	155	50	90.0	-	1T – 2E	83.9-84.0	-	
	R_TACCCCATTGCTTCCATTCC		*E. rubecula*	156	50	89.6	-	1T – 3E	83.5-83.8	-	
Canidae	HRM_Can *d-loop*	61	*C. l. familiaris*	147	44	82.7	85.9	4E	80.3-81.3[f]		G - H
	F_CCGCAACGGCACTAACTCTA		*V. vulpes*	146	49	84.5	88.6	1T	81.3	83.5-83.6[e]	
	R_CCATTGACTGAATAGCACCTTG										
Caprinae	HRM_Cap *12S*	57	*C. hircus*	158	37	80.1	86.3	1T	76.5	80.5	I - J
	F_TAAATCTCGTGCCAGCCA		*R. rupicapra*	158	37	79.6	86.6	2E	75.3-75.7	80.3-80.6	
	R_GTAGGGTTACTTTCGTCAT		*O. aries*	158	36	78.3	88.0	3E	74.5-75.0	81.7-82.0	
Cervidae	HRM_Cer *d-loop*	60	*C. capreolus*	168	45	84.5	88.6	1T – 2E	80.5-80.9	82.5-82.9	K - L
	F_CGATGGACTAATGACTAATCAG		*C. elaphus*	169	47	85.2	88.9	1T – 2E	80.7-80.8	83.6-83.7	
	R_TTATGGGGATGCTCAAGATG										

[a]Number of control DNA samples used in HRMA testing; T, extracted from host tissue; E, extracted from engorged ticks from different individuals.
[b]HRM normalized and derivative plots
[c]The melting temperature, as well as the melting profiles, for *A. sylvaticus* and *A. flavicollis d-loop* amplicons are fully overlapping, so discrimination of these two species is not possible using these primers.
[d]The melting temperature for *S. antinorii* and *S. minutus d-loop* amplicons are not sufficiently different from each other to be diagnostic for species identification.
[e]Reported range refers to the variation observed between the two replicates of the same single control sample.
[f]Wide range of T_m is related to mutations in the sequence of the used control samples (see Fig. 4.6G).

Fig. 4.6 (next page) HRMA of positive control samples of the species listed in the legends using group-specific primer sets. (A, B) Muroidea, (C, D) Soricidae, (E, F) Passeriformes, (G, H) Canidae, (I, J) Caprinae, (K, L) Cervidae; Figs. 1A, C, E, G, I and K are normalized melting plots, while Figs. 1B, D, F, H, J and L are derivative melting plots. Yellow bars delineate pre- and post- melting normalization regions. For Canidae (G, H), alignment of the sequences obtained from the two most differentiated *C. l. familiaris* HRMA profiles (*a*, *b*) showed that sample *a* had one transversion (T->C) compared to sample *b* causing a + 0.5 °C shift in T_m. In Fig.1K, L melting profiles from *C. capreolus* tissue (*) are notably different from melting profiles from engorged ticks (#), possibly related to a difference in DNA concentration.


(A)
(B)
(C)
(D)
(E)
(F)
(G)
(H)
(I)
(J)
(K)
(L)
Normalized fluorescence
Derivative fluorescence
Temperature (°C)
MUROIDEA
A. sylvaticus
A. flavicollis
M. glareolus
M. musculus
SORICIDAE
C. leucodon
C. suaveolens
S. minutus
S. antinorii
PASSERIFORMES
T. merula
T. philomelos
E. rubecula
CANIDAE
V. vulpes
C. l. familiaris
CAPRINAE
O. aries
C. hircus
R. rupicapra
CERVIDAE
C. capreolus
C. elaphus

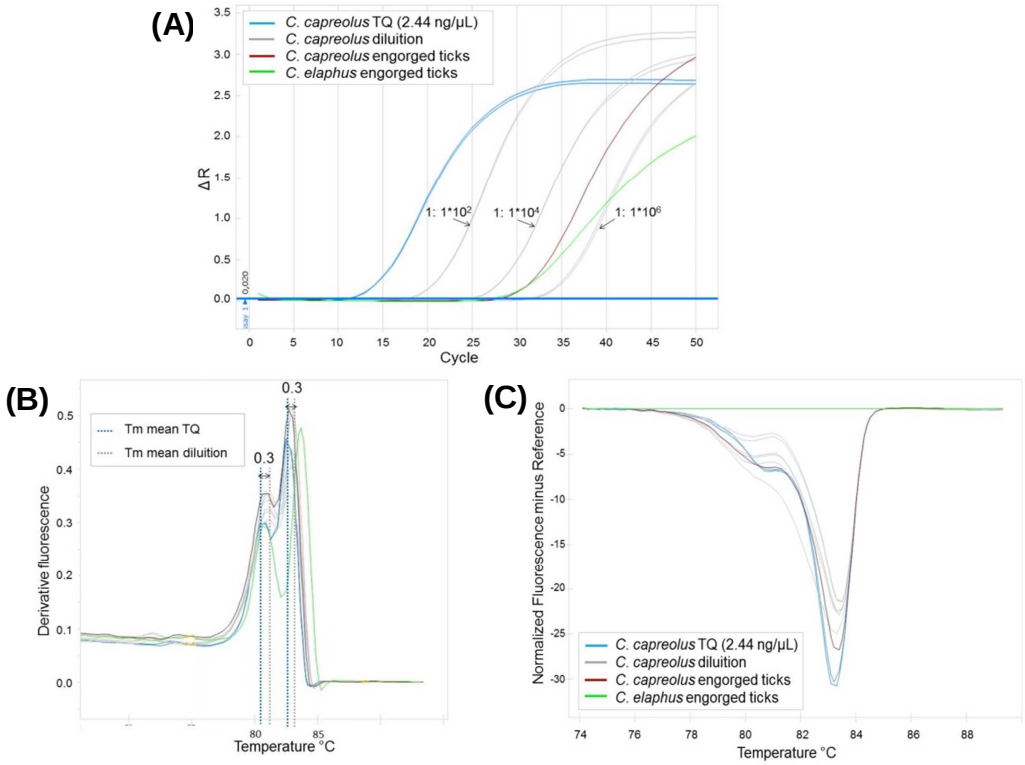


Fig. 4.7 Effect of DNA concentration on melting temperature of the targeted *d-loop* amplicon in Cervidae: (A) amplification graph of undiluted and diluted DNA from one tissue sample of *C. capreolus;* the start of amplification is directly correlated with the level of dilution; (B) derivative melting graph showing the shift in T_m (-0.3 °C) of both peaks in diluted compared to undiluted samples. Diluted samples, amplifying between 25 and 31 cycles all have comparable T_m; (C) difference graph showing the relative difference in melting profiles of all *C. capreolus* diluted and undiluted samples relative to the baseline (*C. elaphus* engorged tick). Despite the deviation in melting temperature of diluted *C. capreolus* samples, they can all be clearly differentiated from *C. elaphus*.

4.3.5.3 HRMA testing on questing ticks

Using our primers and the described HRMA protocol, bloodmeals were successfully identified in 34 out of 52 questing nymphs analyzed (i.e. sensitivity was 65.4 %; Table 4.11). However, as a result of mixed bloodmeals (see below) an additional eight amplicons were generated. Sequencing confirmed that HRMA allowed the correct identification of 35 out of 42 (83.3 %) bloodmeal sources to species level (including *Bos taurus*; see below), and five to genus level (*Sorex* sp. and *Apodemus* sp.; samples 14_CA and 13_PI: *S. antinorii*; samples 12_CO, 1_PI and 5_PI: *A. flavicollis*;

Appendix). Of the remaining two amplicons, sample 12_TR had an aspecific melting profile, within the target species range (see Appendix, Table A1) and was confirmed by sequencing as *Crocidura leucodon*. The last amplicon, 5_PI, was identified by HRMA as *M. glareolus*, whereas the BLASTn search of the sequenced amplicon suggested *A. flavicollis* (98% probability; see Appendix, Table A1). Alignment of sample 5_PI and the *A. flavicollis* control sample sequences revealed a 3 bp deletion (GTG) in sample 5_PI that caused a variation in melting temperature to match that of *M. glareolus.*

Identification of more than one host from the same tick (i.e. amplification by more than one host group primer set) occurred in 8 out of 34 (23.5 %) nymphs. *B. taurus* and *C. l. familiaris* were both found in 3 nymphs; for the other 5 nymphs, the pairs of hosts included: *Apodemus* sp. and *Crocidura suaveolens*; *Apodemus* sp. and *C. elaphus*; *C. leucodon* and *C. elaphus*; *C. leucodon* and *V. vulpes*; and *C. l. familiaris* and *C. elaphus.*

As reported in Table 4.11, amplicons from at least one questing tick tested positive for each of the targeted Canidae and Cervidae hosts; for Soricidae, *S. araneus* and *C. russula* were not found in any questing ticks, nor were *M. m. domesticus* and *M. glareolus* from Muroidea host group. No nymphs were positive for Caprinae or Passeriformes.

Table 4.11 Results of host identification with HRMA using field-collected questing nymphs.

Host DNA identification		**Site**[a]				
Host group	**Host genera/species**	**CO**	**TR**	**CA**	**PI**	**Total**
Muroidea	*Apodemus sp.*	1			2	3
	M. glareolus					0
	M. m. domesticus					0
Soricidae	*Sorex sp.*			1	1	2
	S. araneus					0
	C. leucodon		2	1	4	7
	C. suaveolens	1	1	1	2	5
	C. russula					0
Passeriformes	*T. merula*					0
	T. philomelos					0
	E. rubecula					0
Canidae	*C. l. familiaris*	6	3		1	10
	V. vulpes		2			2
'Caprinae'	*O. aries*					0
	R. rupicapra					0
	C. hircus					0
	B. taurus	3				3
Cervidae	*C. capreolus*	1		1		2
	C. elaphus	1	2	1	4	8
n ticks with identified bloodmeal/*n* ticks tested		9/13	8/12	5/14	12/13	34/52
% ticks with identified bloodmeal		69.2	66.7	35.7	92.3	65.4
n mixed bloodmeals		4	2	0	2	8
n different host species/genera identified		6	5	5	6	9

[a]CO, Condino; TR, Transacqua; CA, Cadine; PI, Pietramurata.

In most cases, identification of the bloodmeal was straightforward: i.e., the T_m and melting profile of questing tick (i.e. unknown) samples were clearly within the range of control samples (see example of Cervidae in Fig. 4.8).

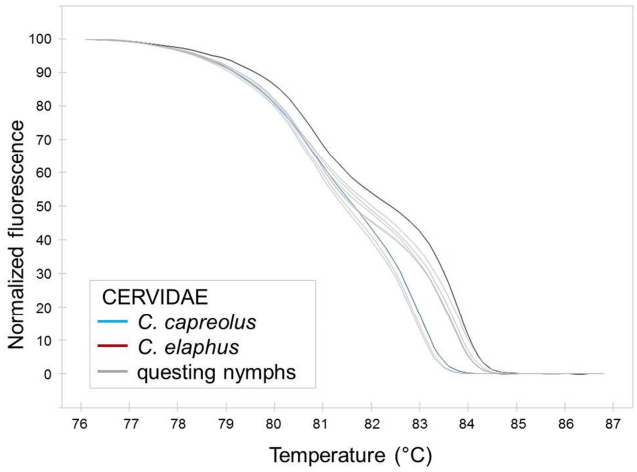


Fig. 4.8 Normalized melting plot of the Cervidae primer set showing the melting profile obtained from control samples and questing nymphs that provided amplification. Note that the melting curves of unknown samples are very similar to those of positive samples despite a slight deviation of T_m, ensuring correct species identification of unknown (i.e. questing tick) bloodmeals.

Very occasionally, amplicons gave T_m and/or melting curve profiles clearly divergent from those of control samples (see example in Fig. 4.9A). These amplicons were removed from the HRMA melting curve graphs to make examining the remaining curves easier (see example in Fig. 4.9B). The aspecific amplicons were then sequenced to verify their identity. A BLASTn search confirmed that these sequences were mainly derived from tick DNA or simply short primer dimer amplifications (see Appendix, Table A1). However, while testing the Caprinae primers on questing ticks, by sequencing three unusual amplicons, we confirmed that *Bos taurus* was also amplified (in samples 2_CO, 3_CO and 5_CO), and has an HRMA profile similar to but clearly distinguishable from those obtained for the tested Caprinae species, with double melting peaks and a T_m of 80.8-81.0 and 83.7-84.1 °C, respectively (Fig. 4.10).

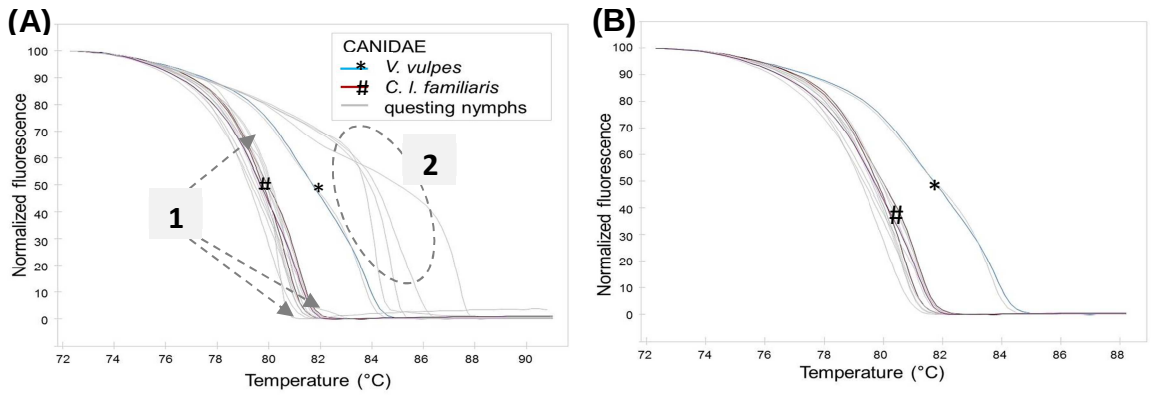


Fig. 4.9 Analysis of HRMA data by means of melting plots: an example using Canidae HRMA on questing nymphs. (A) Normalized melting graph of all amplified samples: note that the HRMA curves obtained for amplicons from questing ticks are very similar to (1) or clearly different from (2) control sample curves (in this case: *C. l. familiaris* and *V. vulpes*); (B) normalized melting graph reporting only those samples with melting profiles matching those of control samples.

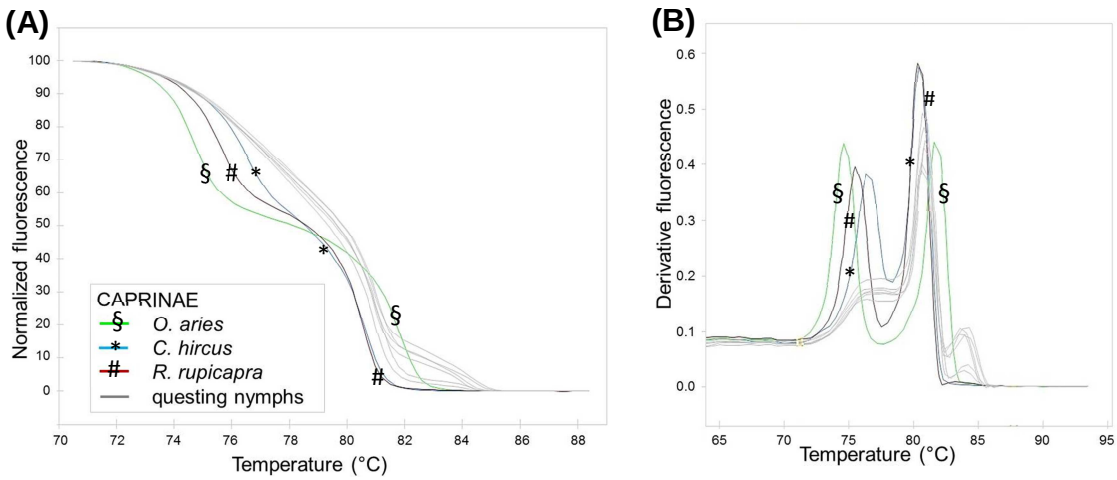


Fig. 4.10 HRMA plots for Caprinae. (A) Normalized; (B) derivative. Grey lines indicate unexpected HRMA profiles obtained from two replicates of samples 2_CO, 3_CO and 5_CO, later identified by sequencing as *Bos taurus*. Note that these curves are similar in shape to those of other hosts, and very different from the unusual curves in Fig. 4.9A..

4.3.6 Comprehensive Real-time HRMA bloodmeal analysis results for questing nymphs in Province of Trento

Results of the HRMA bloodmeal analysis of DNA extracted from questing both using magnetic-beads method (749) and manual spin-column method (99; of which 52 were also presented in the previous section, since they were part of the optimization procedure) are included so that about 30 nymphs for each of the 30 sampling sites were analysed (Table 4.12 for details).

Overall, larval bloodmeals were identified in 215 out of 848 nymphs (25.4% identification success; Table 4.13); DNA from multiple hosts (mixed bloodmeals) was recovered from 23 nymphs (10.7%). The linear model identified the DNA extraction method as the only factor that significantly affected the identification success ($p < 0.001$). Nymph DNA extraction (n=99) with QiaAmp® DNA Investigator resulted to be the most efficient, having a 55.1% identification success, compared to the 22.4% obtained from ticks extracted with KingFisher™ Cell and Tissue DNA kit (n=749) (Table 4.12). Identification success varied widely between sampling site (from 3.3% to 92.3%; Table 4.12), but no significant relationship between identification success and either sampling year, sampling month and or habitat type emerged from our data (Fig. 4.11). It should be noted that once DNA extracted with the applied to KingFisher™ Cell and Tissue DNA kit extracted nymphs, HRMA showed a constant consistent increase in the T_m of target species amplicons; sequencing gave proof of their identity and their melting temperatures were later used as reference. During this optimization step, a *Cricetus griseus* bloodmeal T_m (82.4 °C), amplified using HRM_Rod primers, was identified by sequencing and blasting (BLASTn identity score 99%); it possibly represent an alien individual deriving from escaped pets.

Table 4.12 DNA extraction methods, sampling time and number of analyzed nymphs for each site and relative identification success rate.

DNA extraction method	site	site code	sampling time	nymphs analyzed (n°)	identification success (%)
QiaAMP	Brentonico	BRE[a]	May-2012	15	46.7
	Transacqua	TRA	June-2012	12	66.7
	Grigno Valsugana	GRI[a]	April-2013	16	31.3
	Pietramurata	PIE	May-2012	13	92.3
	Cadine	CAD[a]	May-2012	14	35.7
	Giovo	GIO[a]	April-2013	16	43.8
	Condino	CON	April-2012	13	69.2
			Total	**99**	**55.1**
Thermo KingFisher	Caoria	CAO	April-2013	30	30.0
	Cavedine	CAV	April-2013	30	20.0
	Passo del Durone	PDU	May-2012	34	32.4
	Telve	TEL	June-2012	30	13.3
	Val Genova	VGE	May-2013	30	13.3
	Vervò	VER	May-2012	30	20.0
	Mezzocorona	MEZ	June-2012	30	16.7
	Molveno	MOL	May-2012	30	3.3
	Grigno Valsugana	GRI[a]	April-2013	16	43.8
	Laghi di Lamar	LAM	May-2012	31	9.7
	Tione di Trento	TIO	May-2012	30	10.0
	Volano	VOL	May-2012	30	50.0
	Lundo	LUN	May-2012	30	6.7
	Trento	TRE	May-2013	34	17.6
	Pinzolo	PIN	May-2012	30	20.0
	Cadine	CAD[a]	May-2012	8	25.0
	Tesino	TES	June-2012	30	20.0
	Pergine	PER	April-2012	30	20.0
	Revò	REV	May-2012	30	13.3
	Segonzano	SEG	May-2012	30	23.3
	Dimaro	DIM	May-2013	30	33.3
	Giovo	GIO[a]	April-2013	16	31.3
	Ala	ALA	May-2012	30	33.3
	Brentonico	BRE[a]	May-2012	17	29.4
	Levico	LEV	April-2012	30	26.7
	Cavalese	CVS	May-2012	23	21.7
	Ledro	LED	May-2012	30	20.0
			Total	**749**	**22.4**

[a]site having samples extracted either with QIAamp or Thermo King Fisher.

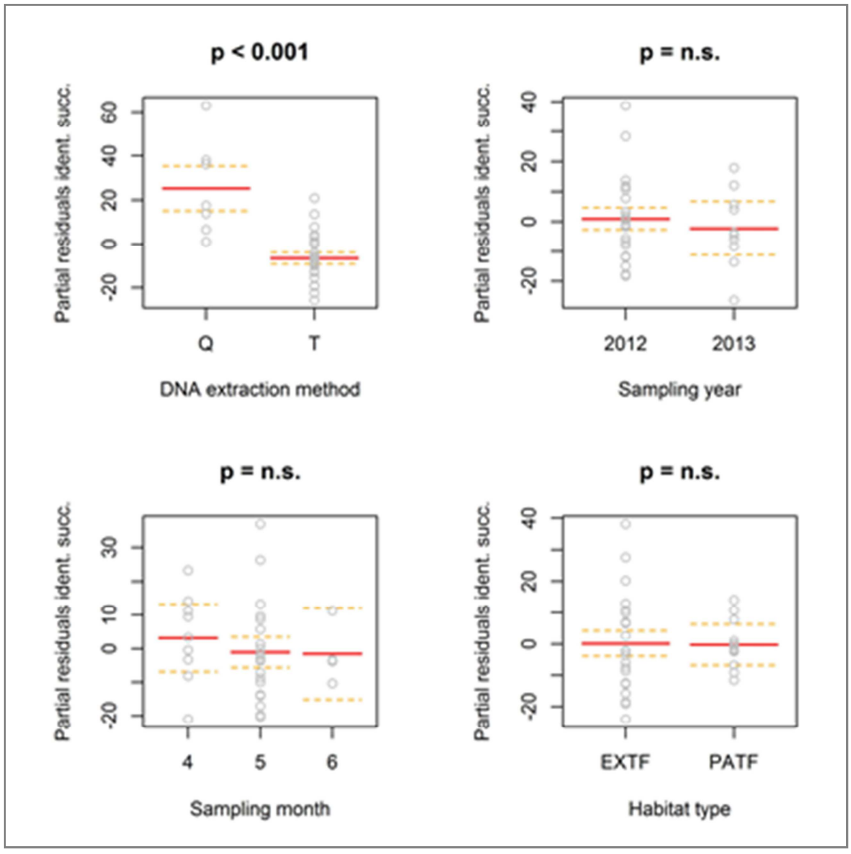


Fig. 4.11 Linear model, partial residuals for identification success in ticks (%) and explanatory variables: DNA extraction method (Q: QiaAMP® DNA Investigator; T: KingFisher™ Flex Magnetic Particle Processor), sampling year, sampling month (4: April; 5: May; 6: June) and habitat type (see text for detail). The only significant explanatory variable is DNA extraction method; Q is more efficient than T, as underlined by the residuals plot. n.s.= not significant.

Identification of bloodmeal source at species level was possible for 137 amplicons, while genus level was reached, as already described in the previous chapter, for *Apodemus* (N=67) and *Sorex* (N=9), and additionally for *Turdus* (N=19) and *Ovis* (N=7) (Table 4.13). In fact, first HRMA could not reliably discriminate between *T. philomelos* and *E. rubecula* control samples and, in addition, sequencing of amplicons having a slightly deviating T_m (84.10 °C, 84.14°C and 84.24 °C) from *T. philomelos/E. rubecula* (T_m of amplicons from questing nymphs 84.44-84.50°C) revealed amplifications of two *Turdus* spp. DNA, not recognizable at species level by BLASTn search, and also a *T. philomelos* (100% identity score by BLASTn); therefore amplicons

having a T_m in the range 84.00 to 84.50 °C were all classified as *Turdus* spp. / *E. rubecula* (Table 4.13). Moreover, the application of the protocol on a larger sample set showed that the Cervidae primers amplify a larger range of Cetartiodactyla hosts than predicted, because they also amplify some species targeted by Caprinae primers (HRM_Cap) (*Ovis* spp. and *R. rupicapra*), as well as tick hosts not previously considered, such as *Dama dama.* Because T_m of several species overlapped, identification by HRMA only became complex for Cervidae amplicons and therefore I had to resort to sequencing more often than expected in HRMA. Sequence analysis of *Ovis* amplicons obtained with HRM_Cer primers, both from an engorged tick collected while feeding on a mouflon and from questing nymphs, showed the inability of the chosen mitochondrial control region fragment to confidently discern domestic (*Ovis aries*) and wild (*Ovis aries musimon*) sheep. BLASTn reports 99-100% identity scores for *Ovis aries, O. a. musimon* and *O. orientalis* (Alignment and BLASTn results in Appendix 2); consequently, all sequences for which BLASTn reported various sheep species and subspecies at similar identity score, were classified here as *Ovis* spp. In five cases, HRMA led to misidentification of *Apodemus* spp bloodmeal as *M. glareolus*, because of intraspecific mutations (Appendix 3). The most common larval hosts were Rodentia (28.9%), mainly *Apodemus* spp (28.0%). The second most frequent host group was Carnivora (28.4%), with *C. l. familiaris* accounting for 21.3% and *V. vulpes* for 7.1%. Cetartiodactyla species fed 17.2% of larvae, *C. elaphus* and *C. capreolus* being the most common hosts (6.3% and 4.2%, respectively). 14.6% of the identified bloodmeals belongs to Passeriformes order and, lastly, 10.9% of bloodmeals were from Soricomorpha. Of the entire list of target hosts (Table 4.13) *C. russula*, *S. araneus* and *M. m. domesticus* did not appear to be larval bloodmeal sources in the study area.

Larval host identification success was similar in EXTF (130/506; 25.7%) and PATF (85/342; 24.9%) sites (Fig. 4.12; Table 4.13). The proportion of Soricomorpha bloodmeals was higher in EXTF than in PATF (15.2% and 4.3%; $p(\chi^2) < 0.05$), while the opposite was true for Passeriformes (PATF: 22.3% and EXTF: 9.6%; $p(\chi^2) < 0.05$) and Carnivora (PATF: 37.2% and EXTF: 22.8%, respectively; $p(\chi^2) = 0.06$). No significant differences were observed between the proportions of Rodentia (EXTF = 31.7%; PATF = 24.5%; $p(\chi^2) = 0.34$) or Cetartiodactyla (EXTF = 20.7%; PATF = 11.7%; $p(\chi^2) = 0.11$ acting as larval hosts in the two forest habitats (Fig. 4.12).

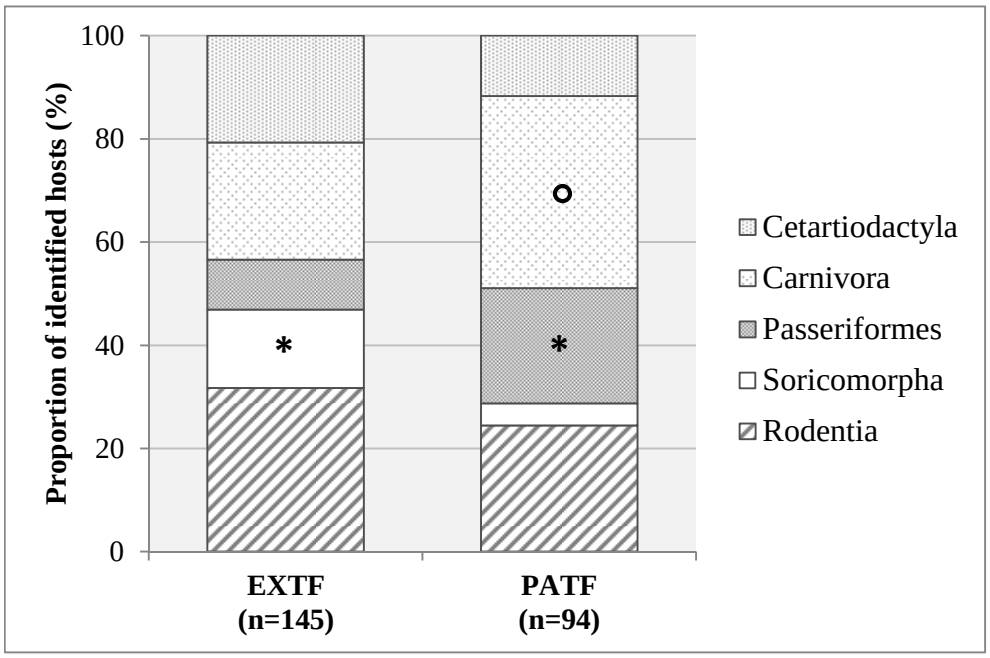


Fig. 4.12 Proportion of host identification in *I. ricinus* nymphs collected in the two forest habitat types: extensive forest (EXTF) and forest patches (PATF). * ($p<0.05$) and O ($p<0.10$) as indicated by the chi-square test. A significantly higher number of bloodmeals from Soricomorpha were identified in EXTF sites than in PATF, while Passeriformes were significantly more represented in PATF sites. Carnivores tend to be more abundant in PATF sites.

Table 4.13 Larval host identifications in questing nymphs collected in EXTF and PATF sites in Province of Trento

Target hosts groups (primer used)	**Tot EXTF[a] (%)**	**Tot PATF[a] (%)**	**TOTALS (%)**
Rodentia (HRM_Rod)	**46 (31.7)**	**23 (24.5)**	**69 (28.9)**
Apodemus spp.	44 (30.3)	23 (24.5)	67 (28.0)
M. glareolus	1 (0.69)	0	1 (0.4)
M. musculus	0	0	0
C. grigeus[b]	1 (0.69)	0	1 (0.4)
Soricomorpha (HRM_Sor)	**22 (15.2)**	**4 (4.3)**	**26 (10.9)**
Sorex spp.	9 (6.2)	0	9 (3.8)
S. araneus	0	0	0
C. leucodon	7 (4.8)	2 (2.1)	9 (3.8)
C. suaveolens	6 (4.1)	2 (2.1)	8 (3.3)
C. russula	0	0	0
Passeriformes (HRM_Pas)	**14 (9.6)**	**21 (22.3)**	**35 (14.6)**
T. merula	8 (5.5)	8 (8.5)	16 (6.7)
Turdus spp.[b] / *E. rubecula*[c]	6 (4.1)	13 (13.8)	19 (7.9)
Carnivora (HRM_Can)	**33 (22.7)**	**35 (37.2)**	**68 (28.4)**
C. l. familiaris	28 (19.3)	23 (24.5)	51 (21.3)
V. vulpes	5 (3.4)	12 (12.7)	17 (7.1)
Cetartiodactyla	**30 (20.7)**	**11 (11.7)**	**41 (17.2)**
Ovis spp[e, d]	5 (3.4)	2 (2.1)	7 (2.9)
R. rupicapra[d]	3 (2.1)	0	3 (1.3)
C. hircus[d]	2 (1.4)	0	2 (0.8)
B. taurus[d]	0	3 (3.2)	3 (1.3)
C. capreolus[e]	8 (5.5)	2 (2.1)	10 (4.2)
C. elaphus[e]	12 (8.3)	3 (3.2)	15 (6.3)
D. dama[b, e]	0	1 (1.1)	1 (0.4)
n° nymphs analyzed	506	342	848
n° nymphs with blood meal	130	85	215
n° total hosts	145	94	239
% identification success	25.7	24.9	25.4
n° mixed bloodmeal	14	9	23

[a] EXTF: extensive forest; PATF patchy forest (see text for details)

[b] host not originally considered as target for primer set, but identified after HRMA, by sequencing and BLASTing

[c] HRMA of Passeriformes amplification did not allowed reliable discrimination between *Turdus* spp. and *E. rubecula*

[d] host amplified with HRM_Cap (Caprinae primer set, HRM_Cap)

[e] host amplified with HRM_Cer (Cervidae primer set, HRM_Cer)

4.4 Population genetics

4.4.1 Microsatellites amplification

STR genotyping at the 10 selected loci proved to be unsatisfactory from the outset. As reported in Table 4.14, issues both molecular and technical in nature, such as mutations in repeat sequences other than repeats, lack of amplifications, and stuttering were observed.

PCR artifacts, like stuttering and incomplete terminal adenylation (+/-A'), increased the complexity of STR scoring, but could be easily solved by an accurate electropherogram check. More concerning was the presence of 'extra alleles' (Fig. 4.13), the lack of amplification and the high number of homozygotes, the presence of alleles with 'not-repeat' length (i.e. in dinucleotide, presence of both even and odd alleles; Fig. 4.14), and alleles outside of the expected range (Fig. 4.15). Sequencing of some of the most difficult STRs (IRN-4; IRN-12; IRN-17; IRN-37; IR-32 AND IR 39; Table 4.14) and their alignment against GenBank deposited reference sequences, provided evidence that mutations (in/del), other than expected variation in the number of repeats, were occurring, both in the region flanking the repeats and in the repeat itself. Null allele and unexpected allele lengths could be consequences of these mutations. 'Extra alleles' could be a consequence of locus duplication in the *I. ricinus* genome.

Testing of the 4 STR loci, developed for *I. scapularis* (Fagerberg et al., 2001), on *I. ricinus* ticks, did not give any positive PCR amplification.

For all of the above reasons, STR genotyping of *I. ricinus* was abandoned as unreliable, also because current statistical methods cannot deal with fragments of mixed mutation types and rates of mutation.

Table 4.14 Microsatellite (STR) loci amplification testing results and issues observed. (STR seq= locus sequencing; yes= done and sequence readable; not readable=presence of multiple peaks; no=not performed).

STR locus	n° samples	STR seq	Results (n° samples/n° samples tested)
IRN-4	17	yes	# some amplifications failed # high number of homozygotes (11/17) # mutation in the STR flanking region and in the STR region
IRN-7	41	not readable	# odd and even alleles # anomalous electrophoretic profile
IRN-8	25	no	# no amplification (5/25) # high number of homozygote (15/25) # stutters
IRN-12	38	yes	# odd and even alleles in heterozygotes # mutation in the STR flanking region # in/del in the STR repeat region
IRN-17	31	yes	# mutation in the STR flanking region # in/del in the STR repeat region
IRN-28	31	no readable	# stutters # more than 2 alleles/individual (contamination or aspecific DNA amplification)
IRN-30	38	no	# more than 2 alleles/individual (contamination or aspecific DNA amplification)
IRN-37	27	yes	# stutters # mutation in the STR flanking region # in/del in the STR repeat region
IR-32	37	yes	# mutation in the STR flanking region and in the STR region # no amplification (2/26) # high number of homozygote (11/26)
IR-39	19	yes	# no amplification (3/19) # high number of homozygote (11/19) # mutation in the STR flanking region and in the STR region

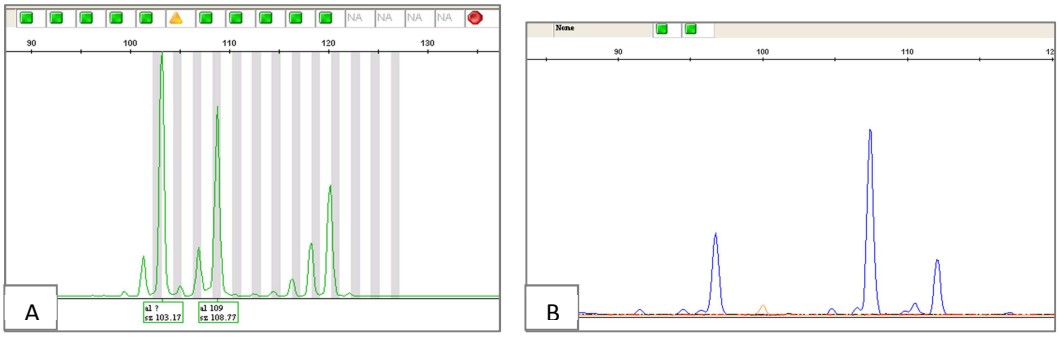


Fig. 4.13 GeneMapper screenshot of the perfect IRN-28 (A) and the compound IRN 30 (B) dinucleotide loci genotypes for two different ticks. Both presented an ‘extra allele’.

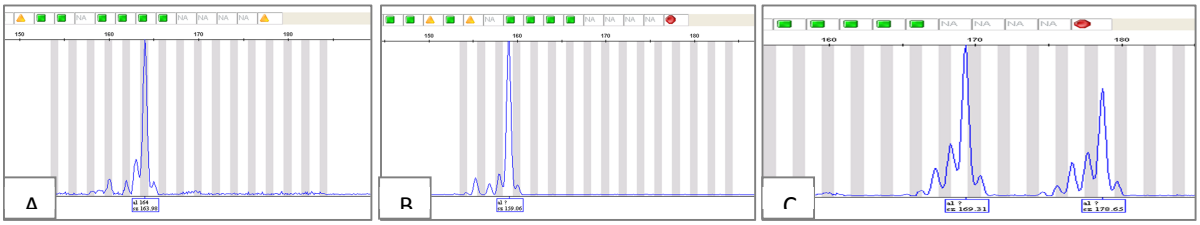


Fig. 4.14 GeneMapper screenshot of IRN-12 dinucleotide locus genotypes of three different individuals; A. homozygote for an odd allele (164); B. homozygote for an even allele (159), and C. heterozygote individual having an even (169) and an odd (178) allele.

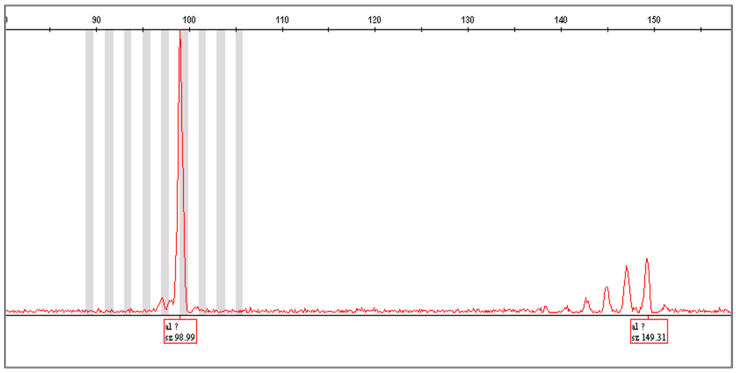


Fig. 4.15 GeneMapper screenshot of the perfect IRN-7 dinucleotide locus presenting one allele far outside the expected locus range (85-101 bp). Additionally, the first allele (the shorter) did not present stuttering, while the second did.

4.4.2 RAD-Sequencing

Illumina sequencing of the library created during protocol optimization gave 191,312,456 for both Reads 1 and Reads 2 sequences (46 Mb of data); the Quality Score, expressed in Phred 33 coding, was on average high for all sequence length (Fig. 4.16).

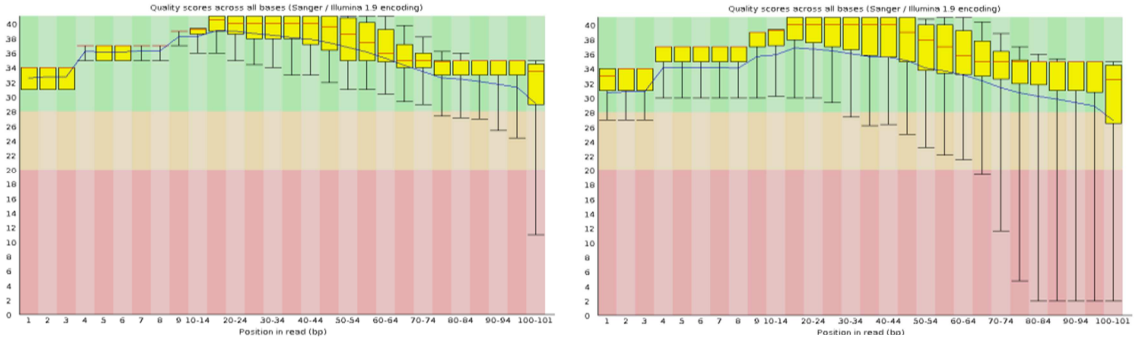


Fig. 4.16 Graphic representation of Quality Score (Phred 33) for bp position for the 2012 RAD-Seq library; Read1 left, Reads2 right.

After quality filtering, 76.3% of the total sequences (Reads1 and Reads2) were retained, 16.4% were discarded because of low quality; 15.7% of the raw Reads 1 were discarded because of ambiguous nucleotides in the RAD-Tag site and 7.7% because of ambiguous nucleotides in the barcode. Retained read numbers varied considerably between the methods used for gDNA extraction: phenol protocol gave a mean of only 586431 reads, while Qiagen spin-column and Thermo Scientific magnetic-beads methods gave, on average, 8115099 and 9004833 reads respectively, with Qiagen having an higher variation between samples (±S.D. 6636828 reads; Table 4.15). As a consequence, the *de novo* loci identification gave nearly zero results for phenol extracted samples, while Qiagen and Thermo both gave good results (Table 4.15). Therefore, the phenol protocol was subsequently abandoned, also given that

ThermoScientific magnetic-beads purification provides high quality DNA even from frozen ticks stored for more than a few weeks (see DNA extraction results).

Table 4.15 Quality filtering and *de novo* processing results for the first sequenced library (2012).

DNA extraction method	Retained reads		RAD-loci	
	mean	±S.D.	mean	±S.D.
Phenol protocol	586431	150498	70	63
Spin-columns, Qiagen	8115099	6636828	30971	11746
Magnetic-beads, ThermoScientific	9004833	3785041	40518	10698

Overall, Illumina sequencing of the 11 paired-end libraries yielded about 2 376 million 100 bp reads for 246 samples (Table 4.16 and Table 4.17). Here I report the genotyping results obtained from the different approaches and the relative results obtained from population structure analysis.

Table 4.16 Raw RAD-Seq paired-end sequencing results.

Sequencing lane	Raw paired-end sequencing results
Library 2012	382624912
Lane1-2014	458303582
Lane2-2014	430250288
Lane3-2014	341141432
Lane4-2014	411335750
Lane5-2014	352676018

Table 4.17 Final sample numbers and sex of RAD-Seq adults *I. ricinus* for each site are reported.

SAMPLES SUMMARY FOR RAD-SEQ			
	Female	Male	TOT
Ala	5	0	5
Brentonico	8	1	9
Cadine	7	1	8
Caoria	7	2	9
Cavedine	8	2	10
Cavalese	7	1	8
Condino	9	2	11
Dimaro	4	3	7
Finland	2	0	2
Giovo	7	2	9
Grigno	7	1	8
Lamar	8	2	9
Ledro	5	2	7
Levico	7	1	9
Lundo	4	3	7
Mezzocorona	7	1	8
Molveno	7	0	7
Passo Durone	8	2	10
Pergine	4	4	8
Pietramurata	7	1	8
Pinzolo	7	2	9
Revò	5	2	7
Segonzano	4	1	5
Telve	7	1	8
Tesino	7	1	8
Tione	7	1	8
Transacqua	7	3	10
Trento	7	1	8
Val Genova	3	2	5
Vervò	7	2	9
Volano	8	2	10
TOT	**197**	**49**	**246**

4.4.2.1 De novo *SNP calling without PCR duplicate removal*

Preliminary trimming of raw data to 80 bp and successive quality filtering with minimum quality score of 10 retained between 57.36 and 76.26% of sequences per lane (Table 4.18).

Preceeding analysis of the whole dataset, an investigative analysis regarding sequencing results was conducted. Examining the number of reads retained for unique barcode, each one been used 2 to 6 times in different lanes, showed for barcodes

ACACG (used 6 times) and AGGAC (used 5 times) a clear deficiency in providing sequencing results, having respectively a median number of retained reads of 1522477 and 878894 in comparison to a median of 6186485 (Fig. 4.17).

No bias related to sampling site could be identified, as in each site different numbers of retained reads were retrieved for each sample. PER is the site with the lowest mean number of retained reads (3128341), CON the one with the highest (9008863) (Fig. 4.18).

On the other hand, females ticks (N=197) gave a 30% higher median number of retained reads (6916706) than males (N=59; 4869346), but the variability inside the two samples is high (Fig. 4.19).

Table 4.18 Quality filtering of raw data results according to sequencing lane. * Nine individuals for which gDNA was extracted using a phenol protocol did not give any RAD-Seq results.

	LANE 2012	**LANE 1**	**LANE 2**	**LANE 3**	**LANE 4**	**LANE 5**
Number of individuals	42*	42	42	43	44	43
Ambiguous Barcodes	31407784	37240680	40775142	50475630	40 616592	50232018
% R1	16.42	16.25	18.95	29.59	19.75	28.49
Ambiguous RAD-Tag	30011675	39932747	55385298	67858032	49823692	60323208
% R1	15.69	17.43	25.75	39.78	24.23	34.21
Low Quality	29403378	46445918	45638096	27282401	33901437	26212172
%	7.68	10.13	10.61	8.00	8.24	7.43
Retained Reads (R1+R2)	291802075	334684237	288451752	195525369	286994029	215908620
%	76.26	73.03	67.04	57.32	69.77	61.22

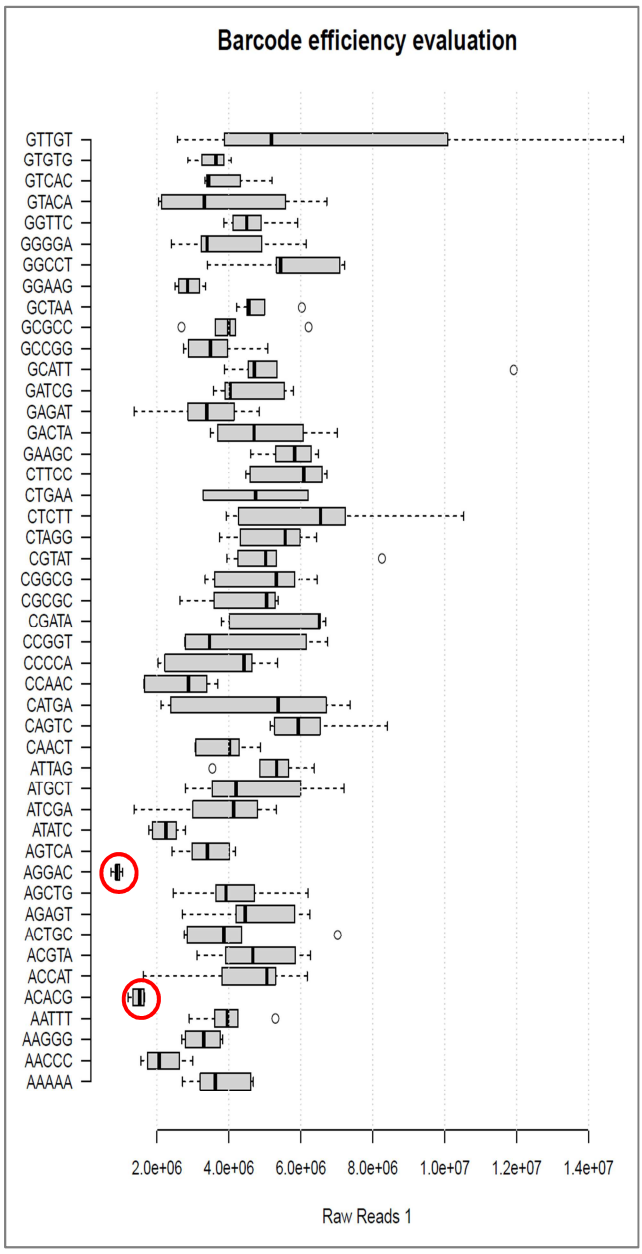


Fig. 4.17 Boxplot presenting median values of retained reads for each unique barcode. Note that barcodes AGGAC and ACACG having an extremely low median number of reads (red circles).

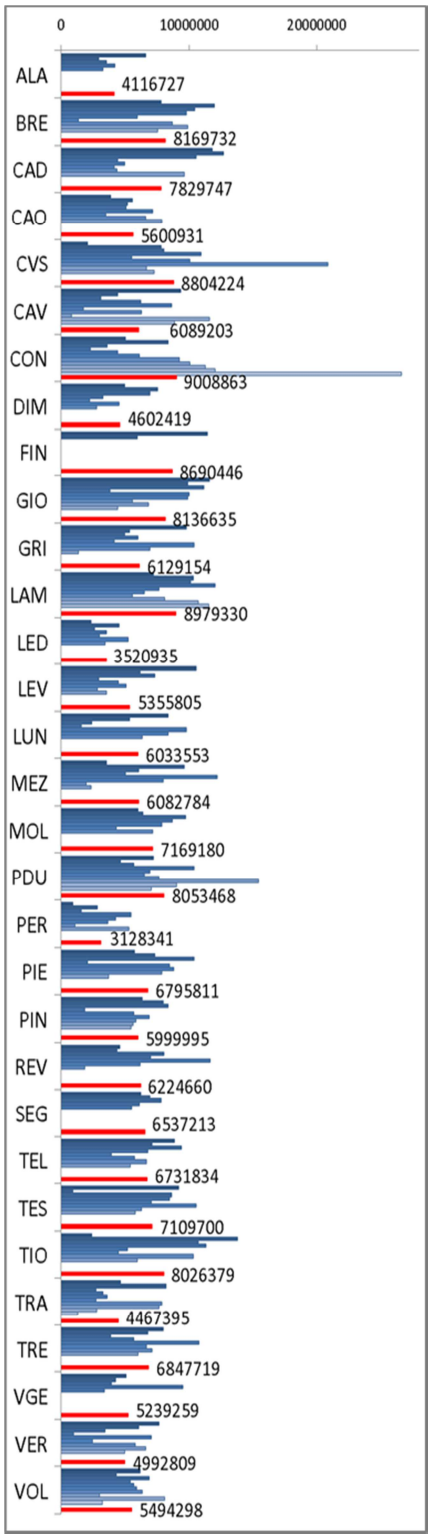


Fig. 4.18 Number of retained reads for each sample according to sampling site (blue bars) and mean values of retained reads per site (numbers and red bars).

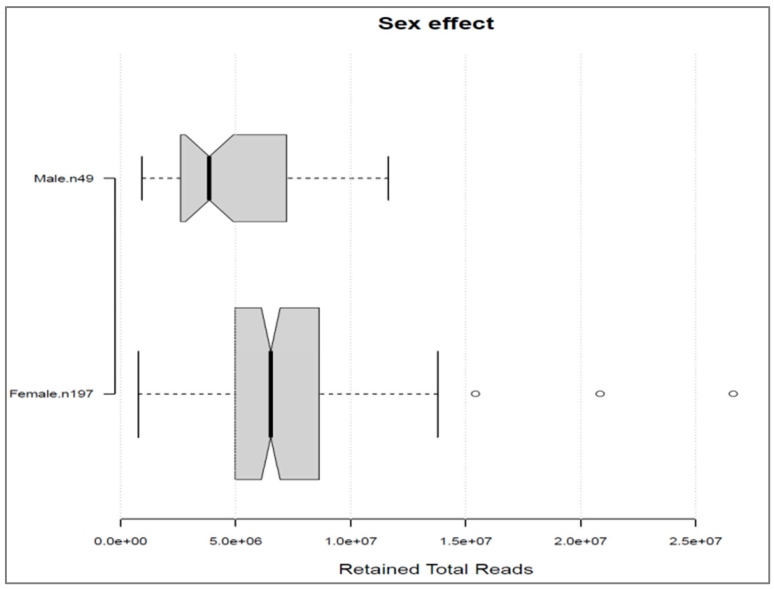


Fig. 4.19 Boxplot presenting median values of retained reads for males and females ticks.

Results of *de novo* loci identification and SNP calling were not those expected: about 45 000 loci (unique stack) were predicted using *radcounter v4* according to the estimate of genome size and GC content of *I. scapularis*, but the number of unique stacks (or loci) actually identified ranged from 357 to 56253 (mean 14265), and only 35 individuals had more than 30000 stacks. Mean merged coverage of identified loci ranged from 55 to 810 reads (mean 300). Analysis for the presence of PCR duplicates using *clone_filter* on raw sequences reported a mean number of clone reads of 75% (Lane2012 76.47%; Lane 1 –Lib1-12 72.19%; Lane 3 –Lib4-5 71.76%). The same analysis performed on two individual samples of Library 2012 and Lane 1, after demultiplexing and quality filtering, showed a clone reads for individuals ranged between 89.90 - 94.00% and 89.06 - 91.94%, respectively. The reason for which few loci for individual are identified and overrepresented, in terms of mean merged coverage, can most likely be attributed to an adapter concentration lower than the one declared by the supplier, leading to an inefficient ligation of the DNA fragment obtained

through the *SbfI* enzymatic digestion. However, the results could be also related to imprecise gDNA quantification and quality assessment or other errors in wet lab procedures or the specific feature of the investigated genome. The technical 'errors' meant that few adapter-ligated DNA fragments were then duplicated through PCR multiple times, increasing the % of PCR duplicates over the limit usually expected for RAD-Seq experiments. As a consequences of the few sequenced loci, final SNPs dataset is characterized by a high level of missing data and, consequently, statistical method had to be applied carefully to obtain robust results. By selecting only individuals with more than 30 000 identified loci and non-deleveraged loci (non-repetitive loci, according to algorithm identification process), with one or maximum two SNPs, removing loci for which at least one SNP position showed more than two alleles per individual, as they may represent paralogs, called in at least 50% of the selected individuals, a 'high coverage' dataset was produced and a *smart_pca* analysis was performed with default parameters and the *lsq project* algorithm enabled. Grouping individuals by site (Fig. 4.20) or geographical area (Fig. 4.21) did not suggest any particular genetic pattern between individuals.

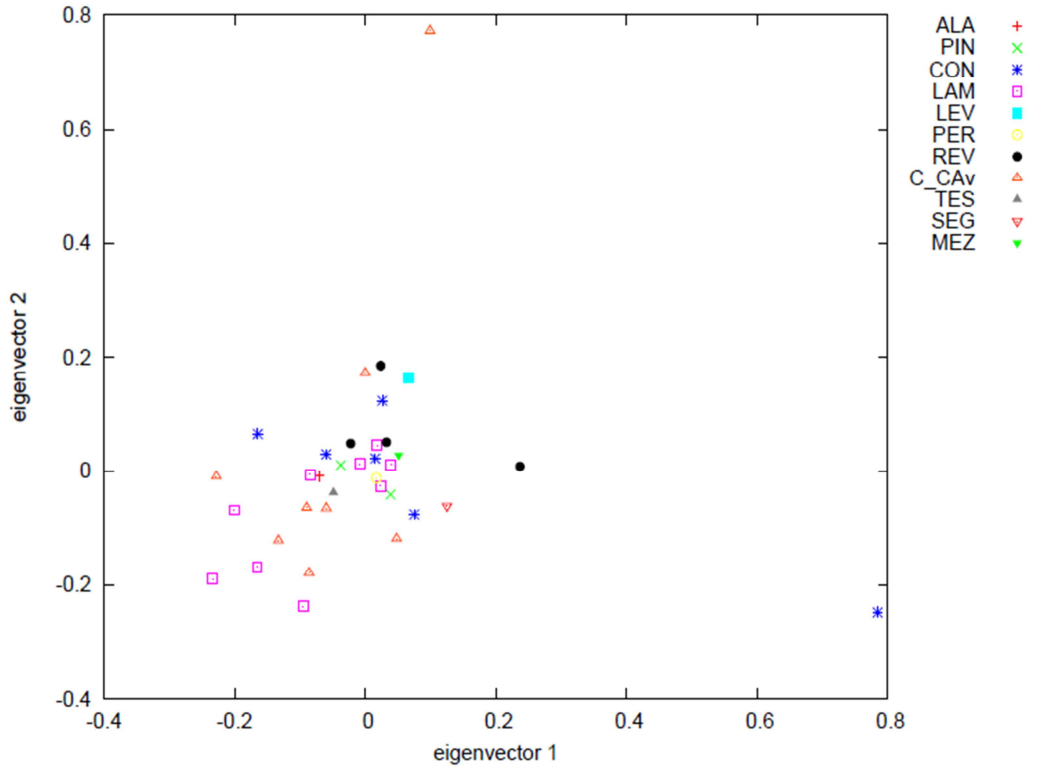


Fig. 4.20 *Smart_pca* plot of the 35 individuals presenting more than 30000 RAD loci; 1219 SNP were used in this analysis, being the ones with less than 50% missing rate; individuals are represented by means of the sampling site (C_CAv=CVS). Individuals coming from different sites are completely admixed.

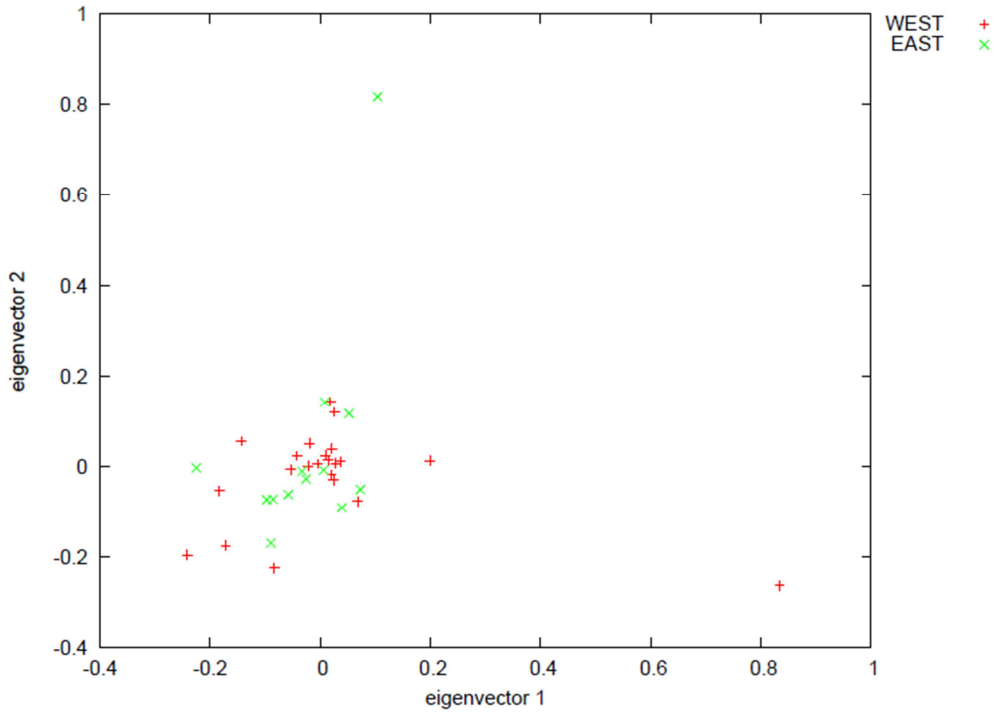


Fig. 4.21(previous page) *Smart_pca* plot of the 35 individuals presenting more than 30000 RAD loci; 1219 SNP were used in this analysis, being the ones with less than 50% missing rate; PAT I. ricinus individuals are classified by area (east or west of the Adige Valley). Some individuals lie out from the main central cluster; however no patterns of clustering could be identified.

The same *smart_pca* analysis was carried on a less conservative dataset. In this case, non-deleveraged loci with one to five SNPs and genotyped in at least 50 individuals were selected. Biallelic SNP, with a MAF (Minor Allele Frequency) higher than 1.5% (this is because a lower MAF could be a signature of SNPs derived by sequencing error) and missing call rate smaller than 80%, individuals with a missing call rate smaller than 99%, were retained producing a dataset of 228 individuals and 7967 SNPs (filtering performed with *vcftools*). Given specific algorithms applied by the *smart_pca* and the *lsqproject* algorithm, only 117 individuals and 6095 SNPs were analyzed. Again, no genetic structure can be identified among samples, as the PCA plot in Figure 4.22 shows.

Finally, a total of 214 109 biallelic SNPs were derived from the raw SNP dataset, selecting non deleveraged loci, having 1 to 5 SNPS each and retaining all individuals. This large SNP dataset was used to compute a MultiDimensional Scaling (MDS) analysis. All SNP loci and individual were retained without considering the missing rate. The resulting plot in Figure 4.23, showing individuals in a space of genetic distances, grouped by color according to sampling area and the two Finnish ticks in blue, displays a homogenous clouds of points, without any indication of spatial clustering.

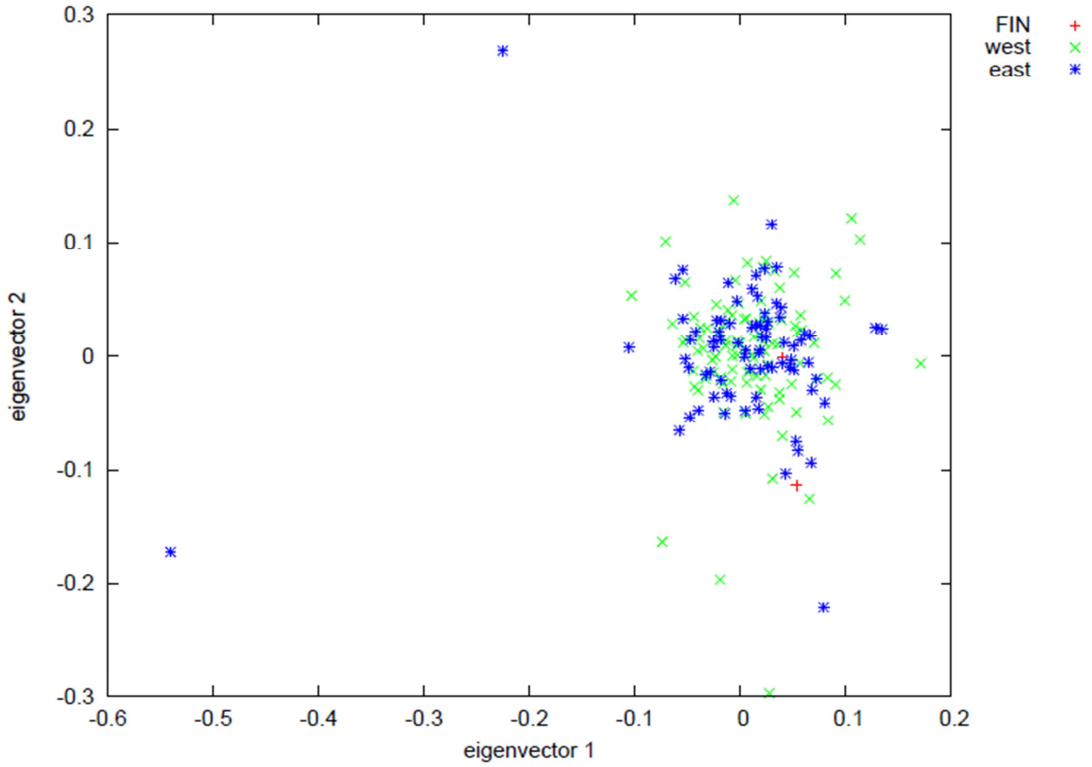


Fig. 4.22 *Smart_pca* plot obtained from the dataset of 7967 SNP (only SNP loci with <80% missing rate were retained) for 117 individuals; PAT individuals are classified by area (east or west of the Adige Valley). FIN: Finnish samples. Some individuals lie out from the main central cluster; however no patterns of clustering could be identified.

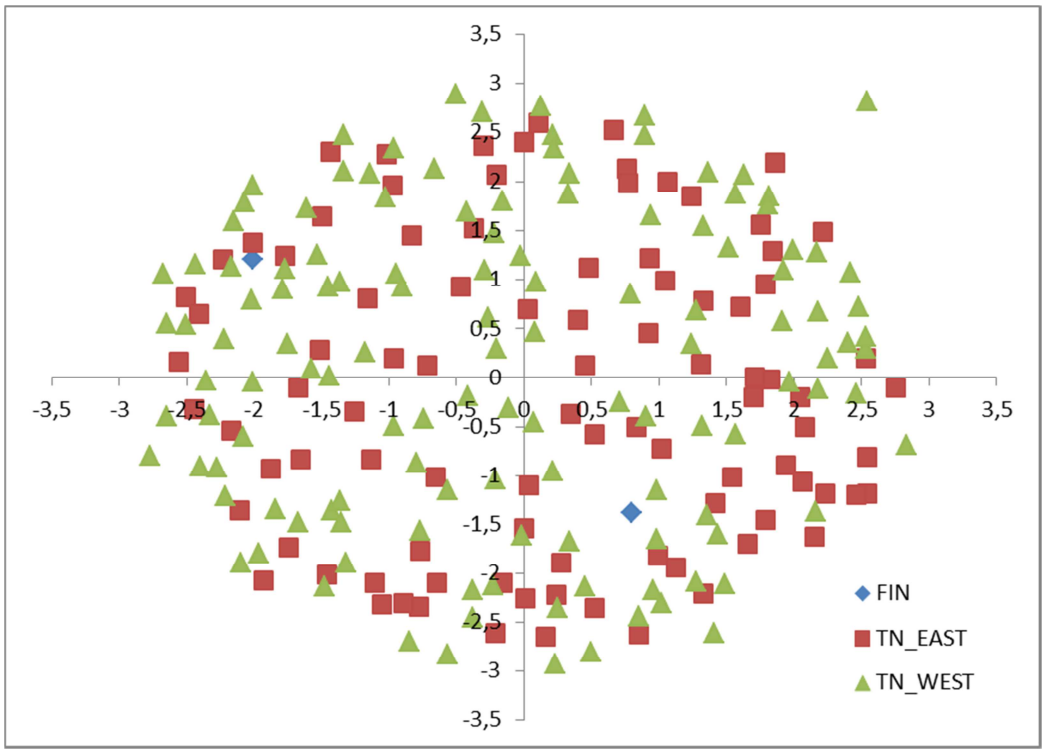


Fig. 4.23 Spatial distribution of 244 *I. ricinus* adults from PAT populations and two from Finland (FIN), based on MDS of genetic distances calculated from 214 109 biallelic SNPs. No genetic clustering emerges based on sampling area.

4.4.2.2 De novo *SNP calling with Stacks after PCR duplicate removal*

Trimming raw sequences to 90 bp resulted in a slightly lower sequence retention during the cleaning process, because of the generally lower Phred quality score in the final part of the sequences (Tab. 4.19).

After demultiplexing and quality filtering, sequences were cleaned from PCR duplicates, and consequently R1 sequences were reduced to a mean of 239153 per individual, (minimum 6168 to maximum 938819). According to that, minimum stack depth in *denovo_map.pl* was reduced to 2 (-m 2, see Materials and Method 3.3.6.3 *B.*).

Table 4.19 Quality filtering of raw data results according to sequencing lane.

	Lib2012	**Lib1-12**	**lib2-3**	**lib4-5**	**lib7-8**	**lib10-11**
Total Sequences	382624912	458303582	430250288	341141432	411335750	352676018
Ambiguous Barcodes	31407784	37240680	40775142	50475630	40616592	50232018
Low Quality	32073365	51204198	49496692	29098397	37013382	28177307
Ambiguous RAD-Tag	30011675	39932747	55385298	67858032	49823692	60323208
Retained Reads	289132088	329925957	284593156	193709373	283882084	213943485
% retained reads	**75.57**	**71.99**	**66.15**	**56.78**	**69.01**	**60.66**

De novo genotyping produced a minimum of 267 loci to a maximum of 57846 loci per sample, with a mean of 12792 loci per sample. Mean merged coverage per locus also varied per sample, ranging from 4X to 38X. A total of 219566 SNPs were identified, in non-deleveraged RAD loci with 1 to 5 SNPs each; missing call rate in individuals was high, varying between 72 to 99%. Before starting any population genetic structure analysis, all SNPs identified in positions higher than 79 bp were removed; in addition, according to results obtained from mini-contigs blasting (see 4.4.2.4), all SNPs identified in loci giving significant alignment with tick vectored pathogen (rickettsioses) and endoparasites (*C. m. mitochondrii*) were removed, as well as all loci for which at least one SNP position showed more than two alleles per individual, as they may represent paralogs. This left a dataset of 163444 SNPs. Using this dataset, analyses on genetic structuring were conducted on individual genotypes and on consensus genotypes at site level. *Smart_pca* and MDS, were used to analyse filtered SNPs datasets, but results of *smart_pca* clearly showed the bias caused by missing data for each individual, especially when used with default parameters (i.e. normalization of allele frequencies and out-group removal; see Fig. 4.24 for an example). Therefore, MDS was preferred to *smart_pca*, because it showed no bias related to sample missing rate.

MDS analysis did not suggest any obvious genetic clustering of the individuals; for example, Fig. 4.25 shows the MDS plot obtained from a dataset of 6824 SNPs, having a maximum missing call rate of 80%, for 215 individuals, as the 31 worst individuals (having more than 99% of missing genotypes) were removed. The same lack of clustering was obtained when MDS was applied to more conservative SNP dataset (1568 SNP for 215 individuals) or to only 'high quality samples' (Fig. 4.26). Even classification of samples by means of sampling site or sex (female vs males) did not reveal any genetic clustering.

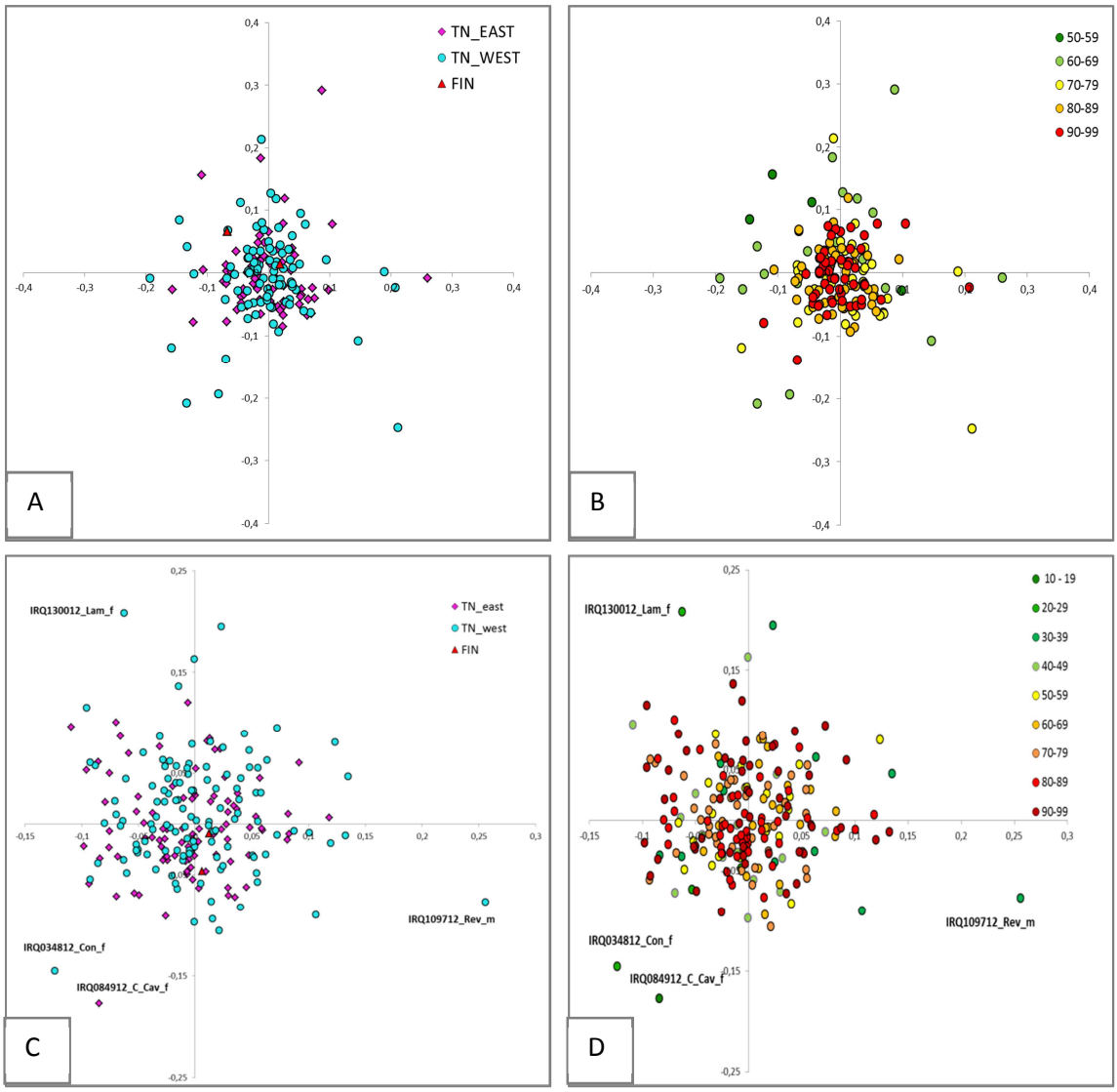


Fig. 4.24 (previous page) *Smart_pca* plots obtained by applying different parameters from a dataset of 6824 SNP for 215 individuals (only loci with <80% missing rate were selected). Left hand plots (A and C) present individuals pooled by sample area (east or west of the Adige River), while righthand plots (B and D) by means of their missing rate class (as defined in the legends). In A and B PCA was performed using default parameters. In C and D PCA was performed disabling normalization of allele frequencies and the outlier removal algorithm. In all cases, *lsq project* algorithm was enabled and no significant eigenvalues, by means of Tracy-Widom statistics, were identified. From A and B clearly emerge as the PCA analysis is biased by missing rate in the samples: high missing rate ones are concentrated in the center of the plot. Additionally, as emerge from C and D, removed outlier were the ones with the lowest rate of missing genotypes. Even disabling normalization, individual missing rate biased the output of the analysis.

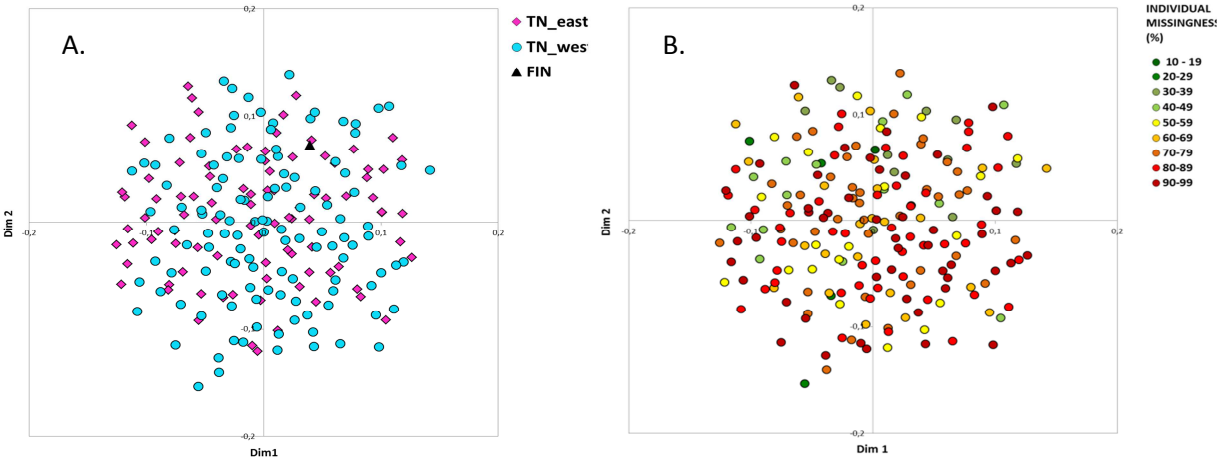


Fig. 4.25 MDS obtained from a dataset of 6824 SNPs for 215 individuals (only loci with <80% missing rate were selected). A. Individuals pooled by area of origin within PAT (east or west of the Adige River),; B. individual represented by means of their missingness class, as reported in the plot legend.

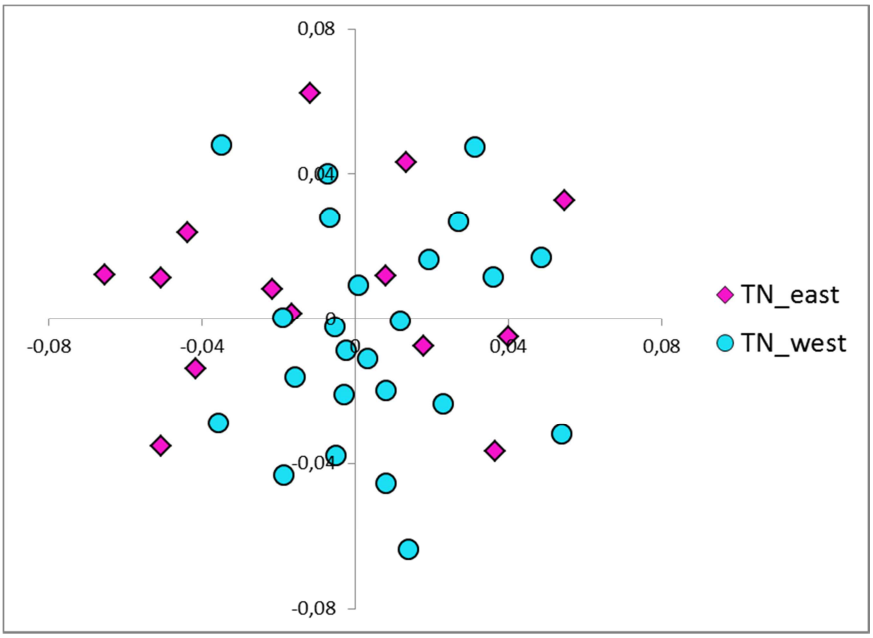


Fig. 4.26 MDS obtained from a dataset of 6977 SNPs, only loci with less than 50% of missing rate were allowed, for the 39 individuals for which > 25 000 stacks were identified.

In the same way, several SNP datasets, with varying stringencies in terms of locus/individual missingness, were used as input in *MDS* for site consensus genotypes, but in every case no genetic clustering was identified (example in Fig. 4.27).

The removal of PCR duplicates from the raw RAD-seq dataset meant that a *de novo* SNP calling and genotyping could be performed with higher confidence. All analyses confirm the absence of genetic clustering the sampled individuals, regardless of site origin. In contrast to population genetic studies of vertebrate species in PAT, no genetic pattern between east and west PAT could be detected.

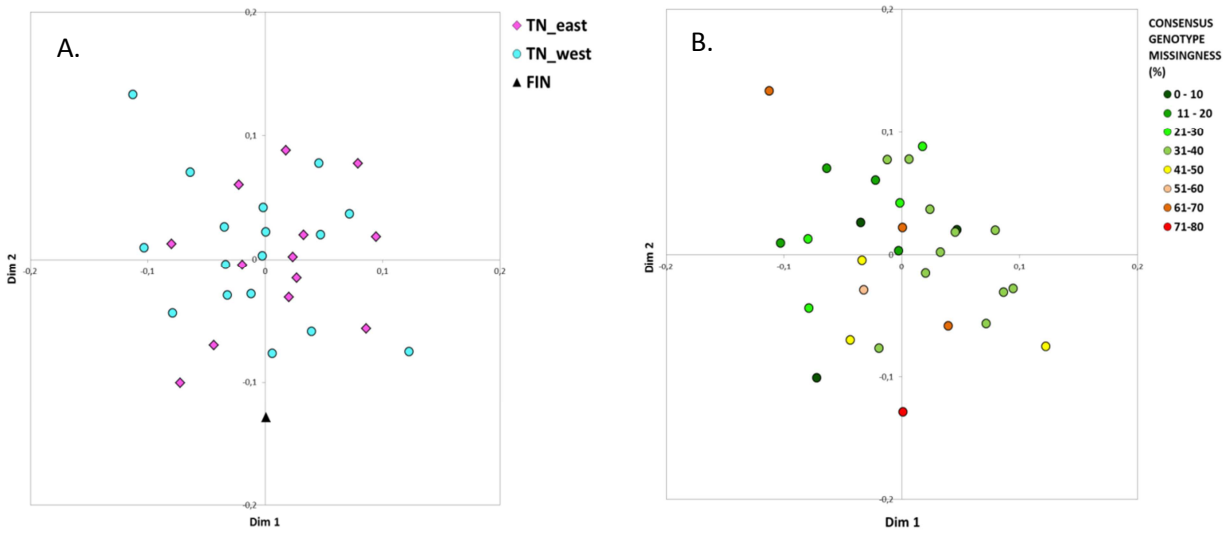


Fig. 4.27 MDS analysis of 22138 SNPs for 31 sites (consensus genotypes); loci with a missing rate less than 50% were retained. A. Individuals are pooled by area of origin within PAT (east or west of the Adige River),; B. individuals represented by missingness class, as reported in the legend.

4.4.2.3 SNP calling with use of a reference genome

Rather surprisingly, only 0.5% to 0.1% of *I. ricinus* individual paired-end sequences aligned to *I. scapularis* reduced reference genome and only 36 SNPs were obtained as a final output.

The PCA obtained from genotype probabilities (i.e. individual datasets pooled by PAT area), again showed an absence of structure, with a central cluster and three outliers individuals (Fig. 4.28). These three (IRQ084912_C_Cav, IRQ034812_Con and IRQ076112_Lam) are characterized by a high number of reads that aligned well with the reference genome. This result suggests that the *I. ricinus* population is well admixed and no genetic structure is present, but could be also a consequence of missing data, so that all individuals have the same genotype probabilities and therefore, cannot be distinguished one each other.

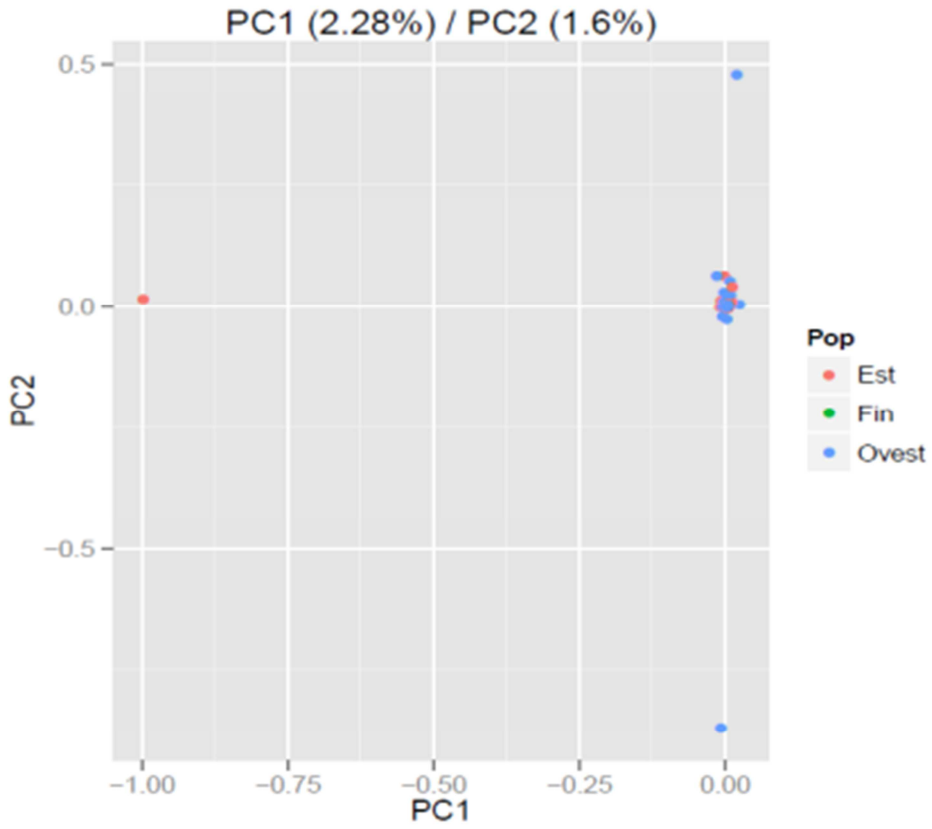


Fig. 4.28 PCA from genotype probabilities computed with ANGSD; individuals are represented by means of PAT area (Est or Ovest=East or West, as before used) and the two Finnish individuals (Fin) are in green.

4.4.2.4 Mini-contigs blasting results

A total of 71664 contigs, with a minimum length of 200 bp, were obtained. Of this, 12426 (17%) gave significant alignment with BLASTn. As reported in Fig. 4.29, most contigs (10762/12426; 86.6%) aligned with *I. scapularis*, 11 with *I. ricinus*, and 13 with *R. appendiculatus*. Interestingly, 13 hits aligned with sequences of the tick endosymbiont *Candidatus midichloria mitochondrii*, while 5 aligned with the Rickettsia spotted fever group.

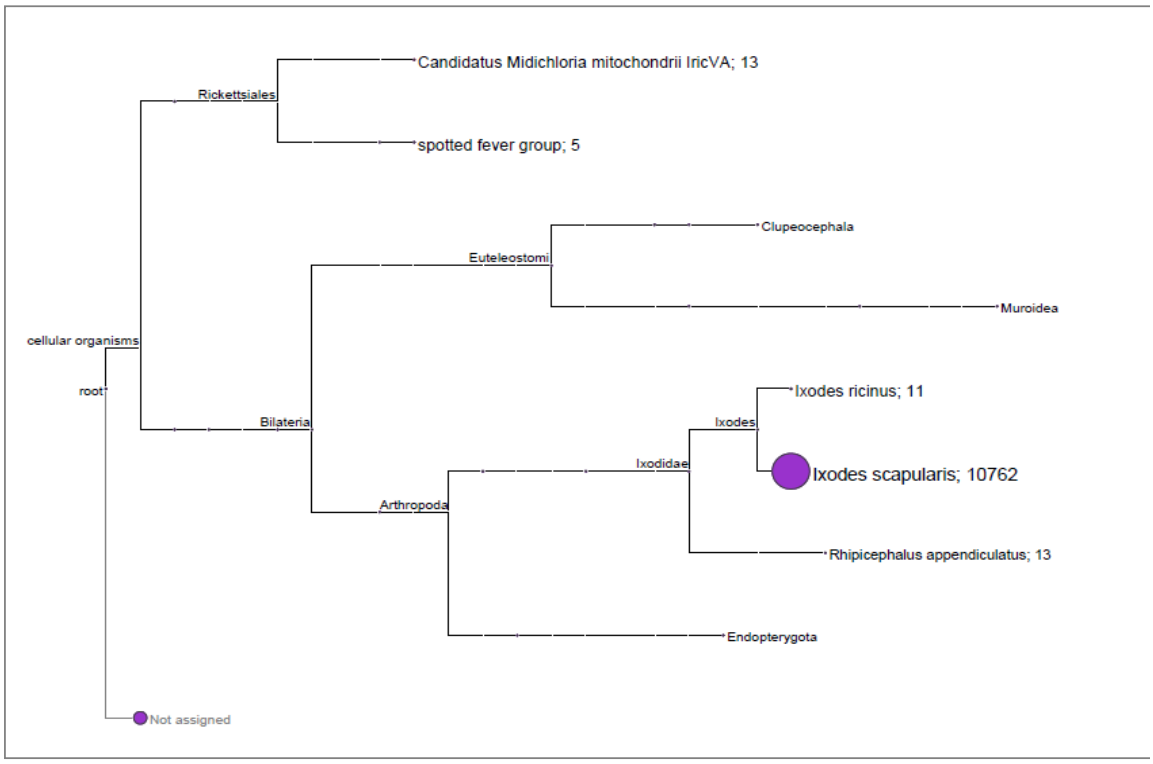


Fig. 4.29 MEGAN contigs BLASTn results visualization.