UNIVERSITÀ DEGLI STUDI DI MILANO

Graduate School of Animal Health and Production:
Science, Technology and Biotechnologies

PhD Course in Biotechnologies Applied to Veterinary and
Animal Husbandry Sciences
(Cycle XXVII)

Doctoral Thesis

New molecular tools for the study of the sheep tick (*I. ricinus* L.): development, application and epidemiological implications

(VET/06)

Margherita COLLINI
Nr. R09500

Tutors: Prof. Michele MORTARINO and Dr Heidi C. HAUFFE

Coordinator: Prof. Fulvio GANDOLFI

Academic Year 2013-2014

ACKNOWLEDGEMENTS

I would first especially thanks both my supervisors, Prof. Michele Mortarino and Dr. Heidi C. Hauffe. This work would not have been possible to complete without the outstanding support that I received from them, by sharing expertise, giving me valuable research guidance and as being two important leading figures. It is thanks to the scientific creativity of Prof. Mortarino that the bloodmeal analysis nightmare found a solution. Dr. Heidi C. Hauffe, despite her various commitment as Department leader, always found time to share concerns and to help in finding solutions; but above all, she always believed in my abilities (more than myself!) and continuously encouraged me. Together with Dr. Annapaola Rizzoli, thanks also to Annapaola, enthusiastically pushed me towards results, making me understand their importance in public health protections.

It will be a long list of thanks, but the work carried out needed complex interactions of people and expertise to get to the presented results. So, …

Thanks to Valentina Tagliapietra, Daniele Arnoldi, Ivana Baràková and everyone else helping in ticks collections from all around Trentino and from any catchable vertebrate animal. I am especially grateful to Franco Rizzolli, because without his expertize bird ticks collection would not have been possible; thanks him to introduce me to the fascinating birds world. Thanks to Associazione Cacciatori del Trentino and its wardens and hunters for the collaboration and for their interest in this project.

A huge thanks is deserved to Matteo Girardi, Barbara Crestanello, Francesca Albonico, Luca Cornetti, Emiliano Trucchi, Cristiano Vernesi, Fausta Rosso, Chiara Rossi, Lorenza Conterno, Gaurav Sablok for their help in establishing the protocols for the molecular approaches used here and for their incomparable moral support. Thanks

to the students Ben, Kieran and Alessandra that joining me during endless afternoons at identifying and cutting ticks, made these tasks far more entertaining (and faster!) than if I had been by my own. Thanks to all PhD students I meet, because sharing doubts and anxiety about life and science, made me feeling less lonely.

A special thanks to Prof. Jeremy Searle, that made me feeling actually part of his research team at Searle lab at the Cornell University, and that made a dream come true; but also to Prof. Giorgio Bertorelle and his research group at the Università di Ferrara for their help in bioinformatics analysis.

Finally, thanks to my wonderful family that encouraged and supported me in any possible way, by sharing doubts, happy moments and ticks collection, and pushing me to strive for, and to wish, all the best from the life.

At last, but not least, a deep thanks to my husband Ilario, that being half of my soul, suffered and enjoyed together with me in this path of growth.

PUBLICATIONS, PRESENTATIONS AND OTHER SCIENTIFIC ACTIVITIES

PEER REVIEWED

Collini M., Albonico F., Hauffe H.C. and Mortarino M., 2015 (in press), Identifying the last bloodmeal of questing sheep tick nymphs (*Ixodes ricinus* L.) using high resolution melting analysis, *Veterinary Parasitology* DOI: 10.1016/j.vetpar.2015.04.007 [IF 2.7]

> All authors designed the research. Under FA supervision, MC optimized the primers and performed the laboratory analyses. All authors contributed to the paper, but MC wrote the first draft, prepared all Tables Figures and Supplementary files, and made the final submission. The developmental procedure is reported here in section 3.3.5.3 and results in 4.3.5.

Baráková I., Derdáková M., Carpi G., Rosso F., Collini M., Tagliapietra V., Ramponi C., Hauffe H.C., Rizzoli A., 2014, Genetic and ecologic variability among strains of *Anaplasma phagocytophilum*, northern Italy [letter], *Emerging Infectious Diseases*, 20: 1082–1085 [IF 7.3]

> MC organized the collection of deer legs and collected many of the feeding ticks from different hosts, especially deer, birds and rodents (section 3.2.2). Feeding and questing ticks were identified morphologically and by molecular analyses by using the newly designed 16SrRNA primers and protocol that MC optimized (see section 3.3.2).

In preparation:

Collini M., Albonico F., Rosà R., Tagliapietra V., Rossi C., Rizzoli A., Mortarino M., Hauffe H.C., Hosts use by *Ixodes ricinus* peri-urban and extensive forests by Real-time HRMA bloodmeal analysis: epidemiological implications *OR* Domestic dogs are important larval wood tick hosts, especially in peri-urban forests [target journal *Parasites and Vectors*; IF 3.25]

Collini M., Trucchi E., Panziera A., Bertorelle G., Tagliapietra V., Arnoldi D., Rossi C., Rizzoli A., Hauffe H.C., High gene flow between I. ricinus populations in the northeastern Alps detected using RAD-Seq analysis

SCIENTIFIC POSTER

Collini M., Hauffe H.C., Masséglia S., Albonico F., Arnoldi D., Bailly X., Bard E., Galan M., Rossi C., Tagliapietra V., Vourc'h G., Rizzoli A., Mortarino M., 2015, Identifying the last bloodmeal of questing wood tick nymphs (*Ixodes ricinus* L.) by DNA amplification: three approaches tested, Heraklion, Crete, Greece, "Genes, Ecosystems and Risk of Infections", 21-23 April 2015.

AWARDED one of two POSTER PRIZES at the International Conference "Genes, Ecosystems and Risk of Infections" (GERI), Heraklion, Greece, in April 2015, for being "very efficiente in addressing a complex issue".

SCIENTIFIC CONFERENCE PRESENTATIONS (speaker is in **bold**)

Collini M., Hauffe H.C., Rizzoli A., Vernesi C., Arnoldi D., Baráková I., Cornetti L., Crestanello B., Girardi G, Konecny A, Rossi C, Rosso F, Tagliapietra V, Albonico F, Mortarino M., Population genetic of the wood tick *Ixodes ricinus* (acari: Ixodidae): new insight into dispersal capacity in relation to host movements - Bloodmeal analysis in questing ticks: application of High Resolution Melting (HRM), Bellaterra, EDENext Annual Meeting, PhD Day, 18 March 2013

Crestanello B., **Collini M.,** Il contributo della genetica della conservazione alla conoscenza e alla gestione di alcune specie di Galliformi alpini: pernice bianca *Lagopus muta*, fagiano di monte *Tetrao tetrix* e gallo cedrone *Tetrao urogallus*, VII Convegno Italiano di Ornitologia, Trento, 13 September 2013

Plantard O., Quillery E., Collini M., Panziera A., Trucchi E., Rizzoli A., Hauffe H.C., Using population genetics to assess tick dispersal, from the mainland to the landscape scale: a review of current knowledge and its utility to design tick-control methods, Heraklion, Crete, Greece, "Genes, Ecosystems and Risk of Infections", 21-23 April 2015 [as coauthor, I provided 4 slides on my research on RAD-Seq for the speaker]

SCIENTIFIC MEETINGS PRESENTATIONS (speaker is in **bold**)

Collini M., **Hauffe H.C.,** Population Genetics of the wood tick Ixodes ricinus (Acari: Ixodidae): new insight into dispersal capacity and host preference. FP7 TBD EDENext Meeting. Bratislava, Slovakia. 25 January 2012. [I could not make the presentation because I was ill and could not attend at the last minute; however, I prepared the talk myself]

Hauffe H.C., Collini M., Population Genetics of the wood tick Ixodes ricinus (Acari: Ixodidae): new insight into dispersal capacity and host preference. FP7 EDENext

Annual Meeting. Izmir, Turkey. 28 March 2012. [I prepared the talk together with my supervisor, H.C. Hauffe]

Hauffe H.C., Collini M., Rizzoli A., Biology of the wood tick Ixodes ricinus. EDENext Annual Meeting, Rovaniemi, Finland. 25 March 2014. [I prepared the talk together with my supervisor, H.C. Hauffe]

Collini M., Presentazione attività PhD project, D-BEM PhD Day, Fondazione Edmund Mach, 6 June 2013.

INVITED PRESENTATIONS

Collini M., Population genetics of the wood tick *Ixodes ricinus* (acari: Ixodidae): new insight into dispersal capacity in relation to host movements, 21 November 2012, Centre for Ecological and Evolutionary Synthesis, University of Oslo, Norway.

Collini M., Population genetics of the wood tick *Ixodes ricinus* (acari: Ixodidae): new insight into dispersal capacity in relation to host movements, 8 April 2014, Searle Lab meeting, Ecology and Evolutionary Biology Dept., Cornell University NY, USA.

PROPOSAL FOR PATENT AT FEM

H.C. Hauffe and M. Collini presented a new technique of possible patent interest to the CRI IP Committee.

Collini M., Hauffe H.C., Mortarino M., Albonico F., 2013, Patent Proposal: Application of High Resolution Melting to identification of host and disease-causing pathogens in wood ticks.

OUTREACH

Collini M., La Notte dei Ricercatori. Participation as high school student guide; Trento, 26 September 2012.

Hauffe H.C., Collini M., Ossi F., Obertegger U., “Ecofiera” di Tione di Trento (TN): Fondazione E Mach. La biodiversità dietro casa... Impariamo a ri-conoscerla! 28.100 visitors, 4 newpaper articles. 3-6 October 2013.

STUDENTS SUPERVISED

Ben Fowle, Kieran David O'Malley, Alessandra Fauri, Francesca Michelon.

COURSES

Gene Expression and Pathway Reconstruction, Phylogeny & Evolution Day, International PhD Programme in Genomics and Molecular Physiology of Fruits (gmpf) (FEM), San Michele a/A, Italy. 26-29 June 2012.

Population and Quantitative Genetics, International PhD Programme in Genomics and Molecular Physiology of Fruits (gmpf), FEM, Viote del Monte Bondone, Italy. 2-6 July 2012.

RADseq experimental design and analysis course, GENECO, University of Lund, Lund, Sweden, 15- 19 ottobre 2012.

International Environment 'Omics Synthesis 2013, Univ. Of Cardiff, UK, 9-11 September 2013.

RAD-Seq library preparation for *Ixodes ricinus* population genetic study. RAD-TAG platform, Centre for Ecological and Evolutionary Synthesis, University of Oslo, Norway, November 2013.

Visiting fellow at Searle Lab, Ecology and Evolutionary Biology Department, Cornell University (NY- USA), 2 April - 29 September 2014.

NEWSPAPER ARTICLES

Collini M., 2011, La Pernice bianca alpina: una sottospecie endemica in declino. *Il Cacciatore Trentino*, 83, 14-15

Collini M., 2012, I piccoli abitanti del bosco, *La Gus dai Buiac* - Comune di Giustino, 15

ABBREVIATIONS

bp: base pairs

DBEM: Department of Biodiversity and Molecular Ecology, Research and Innovation Centre, FEM

dsDNA: double strand DNA

EXTF: extensive forest

FEM: Fondazione Edmund Mach

gDNA: genomic DNA

HRMA: High Resolution Melting Analysis

LB: Lyme borreliosis

MDS: Multi-Dimensional Scaling

mtDNA: mitochondrial DNA

NGS: Next Generation Sequencing

PAT: Provincia Autonoma di Trento/ Province of Trento

PATF: forest patches near to villages

PCA: Principal Component Analysis

PCR: Polymerase Chain Reaction

RAD-Seq: Restriction-site Associated DNA Sequencing

RFLP: Restriction Fragment Length Polymorphism

RLBH: Reverse Line Blot Hybridization

SNP: Single Nucleotide Polymorphism

TBD: Tick-borne diseases

TBEv: Tick-Borne Encephalitis virus

T_m: melting temperature

T_a: annealing temperature

VBD: Vector-borne disease

WP: work package

ABSTRACT

The sheep tick, *Ixodes ricinus*, is the most important zoonotic vector in Europe; its dispersal potential and the relative importance of various vertebrate hosts it exploits, both essential to understand emergence of tick-borne diseases are virtually unknown.

I applied two molecular approaches to 30 *I. ricinus* populations in the Province of Trento, Italy. A novel bloodmeal analysis Real-time HRMA protocol was developed and tested on questing nymphs. RAD-Seq was used for the first time on *I. ricinus* to individually genotype SNP loci in adult ticks.

Bloodmeal analysis confirmed that rodents feed about 30% of tick larvae, but also illustrate that large mammals play a central role in feeding larval ticks. Since birds also feed about 15% ticks, the results of this analysis imply that larval ticks are carried long distances by their hosts; hence dispersal potential is high. In fact, population genetics support that investigated alpine populations are genetically admixed, confirming other phylogenetic studies showing that panmixia of *I. ricinus* population is a general phenomenon. This is the first study showing that dogs are important sheep tick hosts. As dogs feed a higher proportion of ticks in peri-urban forest, they may enhance the contact rate between human and infected ticks, by bringing them into human habitations and urban parks.

My results have added new knowledge to tick dispersal and host use, which will be used to improve models of the spread of *I. ricinus*, and related pathogens, in new climatically suitable areas. My novel bloodmeal analysis protocol, which eliminates previous problems of contamination, could help to identify and explain local TBD dynamics in other areas of the EU.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Chapter 3

Chapter 4

LIST OF TABLES

1. GENERAL INTRODUCTION

Ticks (Acari: Ixodidae) are obligate hematophagous ectoparasites that exploit every class of vertebrate worldwide (Hoogstraal and Aeschlimann, 1982). Importantly, they are among the most important vectors for many pathogens affecting humans, as well as wild and domestic animals. Zoonotic spillover to humans has become a key public health concern in the last decades, especially since the incidence of Lyme borrelliosis (LB) (which emerged in the early 1980s) as well as tick-borne encephalitis (TBEv, known from the 1930s in Europe), have risen dramatically in the Northern Hemisphere (Parola and Raoult, 2001; Gray et al., 2009; Randolph, 2009). At the same time, the tick-borne diseases (TBDs) anaplasmosis and rickettsioses are also frequently reported in Europe (Heyman et al., 2010). In a global context, where climate change together with anthropogenic factors (trade, land use, etc.) are affecting TBD epidemiology by modifying tick population dynamics, habitat, hosts, pathogen reservoirs and human contact rate with infected ticks, these diseases are expected to become an increasingly important public health issue (Parola & Raoult, 2001; Heyman et al., 2010; Rizzoli et al., 2014).

In this thesis, I focus on the sheep (or wood) tick *Ixodes ricinus* L., defined as "the most important multi-potent vector in Europe" (Randolph, 2009). Many studies of this species have investigated the ecology of *I. ricinus* habitat, such as the impact of abiotic (e.g. climate) and biotic (e.g. vegetation) variables on its population, and local TDB dynamics and the interconnection between hosts, vector and pathogen (e.g.: Estrada-Peña et al., 2006; Estrada-Peña, 2009; Randolph et al., 2002; Cagnacci et al., 2012; Bolzoni et al., 2012; Rizzoli et al., 2007; Rizzoli et al., 2009; Rosà and Pugliese, 2007; Carpi et al., 2008; Krasnov et al., 2007; Hudson et al., 2001). In contrast, there is

very little knowledge about tick dispersal capacity or host preferences and host-specialization, information that would help the improvement in the modelling of TBD risk and spread (Bolzoni et al., 2012; Medlock et al., 2013; Léger et al., 2013; Gray et al., 2009; Madhav et al., 2004).

As highlighted by De Meeûs and colleagues (2007), the study of zoonotic disease vectors is challenging, since these organisms are generally small, and certain features of their life cycle, usually including diapause, make direct observation of them almost impossible. A molecular genetics approach could provide indirect measures of population parameters (effective size, substructure, dispersal rate, and host-race formation) of the arthropod populations, based on the analysis of patterns of genetic variability with appropriate markers.

Here, feeding behavior and dispersal of *I. ricinus* were investigated using two *state-of-the-art* molecular approaches: high resolution melting analysis (HRMA) and Restriction-site Associated DNA Next Generation Sequencing (RAD-Seq NGS) technology. Since *I. ricinus* movement, and therefore gene flow, is correlated with host vagility, host use and dispersal capacity can be inferred from the comparative analysis of these two molecular techniques. Population genetics patterns can also tell us about population isolation and dynamics, which have epidemiological implications with regards disease emergence and spread.

The PhD project was designed and funded within the framework of the large collaborative FP7-HEALTH.2010.2.3.3-1 EDENext project 'Biology and control of vector-borne infections in Europe' (EDENext) at the Department of Biodiversity and Molecular Ecology, Research and Innovation Centre, Fondazione Edmund Mach (Italy)

in collaboration with the Dipartimento di Scienze veterinarie e sanità pubblica, Università degli Studi di Milano.

1.1 EDENext project

Initiated in 2011, EDENext brings together 46 international partners to investigate the biological, ecological and epidemiological components of vector-borne disease (VBD) introduction, emergence and spread and, using the newly acquired knowledge, to create new tools to control them (Fig. 1). EDENext takes advantage of the results, as well as concepts, methods and tools, of the earlier FP6 project *EDEN Vector-borne disease in a changing European environment*. While the latter focused on the effects of environmental changes on the emergence of VBDs, EDENext aims to explain and to model the processes leading to the introduction, spreading and establishment of VBD and, most importantly, to define the possible control strategies to break the epidemiological cycles of VBDs. The project addresses five groups of vectors and associated zoonoses: rodents and insectivores (RBD), hard ticks (TBD; Acari, Ixodidae), mosquitoes (MBD; Diptera, Culicidae), sand flies (PhBD; Diptera, Psychodidae), and biting midges (CBD; Diptera, Ceratopogonidae). Research activities are organized vertically into Work Packages (WPs) according to the five vector groups, while horizontal themes (Modelling, Data management and Public Health) provide technical input to the WPs and integration of the datasets and results (Fig. 1.1).

My PhD research project, as part of this framework, was designed to address the objectives of the WP 1.1.1 *Emergence and Spread of bacterial and protozoan tick-borne pathogens* for the sheep tick *I. ricinus*, closing gaps in the knowledge of its ecology and dispersal (Group 4). This WP brings together the expertise in population genetics and TBDs of the Fondazione Edmund Mach (FEM, Italy: headed by A.

Rizzoli, in collaboration with H.C. Hauffe) and the Institut National de la Recherche Agronomique (INRA, France: O. Plantard, G. Vourc'h).

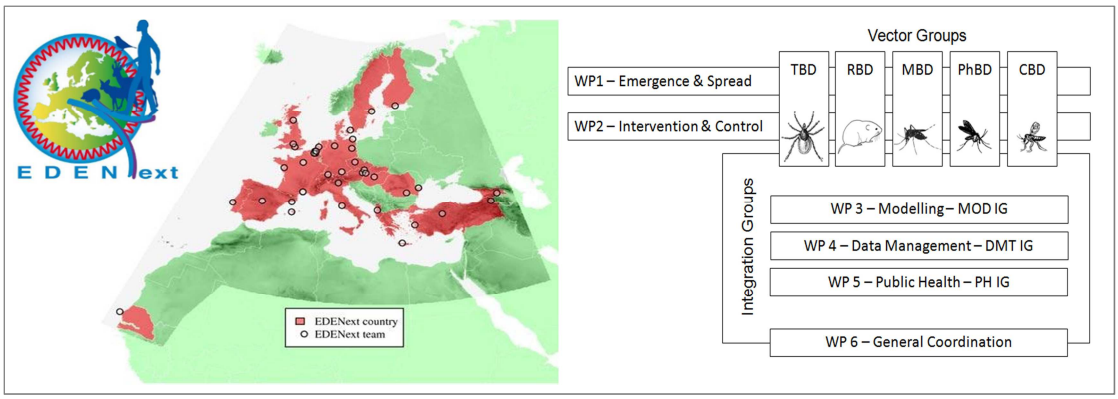

Fig. 1.1 The EDENext project: logo, participating countries and schematic representation of WP organization (reprinted from the EDENext Project).

1.2 The biology of *Ixodes ricinus* L.

Ticks (Suborder: Ixodida) belongs to the Phylum Arthropoda and, as part of the Subclass Acari, Order Parasitiformes, they are closely related to mites (Suborders: Holothyrida, Mesostigmata and Opilioacarida; Nava et al., 2009). The 896 known tick species are distributed worldwide and classified into three Families: the monotypic Nutalliellidae, containing a single species *Nutalliella namaqua*; Argasidae, consisting of 193 species; and Ixodidae comprising 702 species in 14 genera, the most numerous being *Ixodes* (243 species), *Haemaphysalis* (166) and *Amblyomma* (130) (Guglielmone et al., 2010). A recently published study of Mans et al. (2011), together with molecular clock estimates (Jeyaprakash and Hoy, 2009; Dunlop and Selden, 2009), dates the origin of ancestral tick lineages back to the middle Permian (260-270 Mya) in the Karoo-basin (South Africa), and suggests that they parasitised therapsids, based on the

phylogenetic evidence that *N. namaqua*, which feeds on lizards, is at a basal position in relation to the major tick families (Fig. 1.2).

Argasidae, called soft ticks, lack a sclerotized dorsal scutum; nymphs and adults present a ventral capitulum and a highly sculptured integument, and in their life cycles multiple nymphal instars are present. In contrast, Ixodidae, or hard ticks, are characterized by an anterior capitulum, a simple striate integument and a sclerotized scutum, and only one nymphal instar is required before molting to the adult stage. Ixodidae species are divided in two morphological groups according to the position of the anal groove: the Prostriata have the anal groove anterior to the anus, and the Metastriata have the anal groove posterior to the anus. Prostriata are represented only by the genus *Ixodes* and are considered the basal lineage to all the other genera included in the Metastriata (Nava et al., 2009).

I. ricinus L. has been placed, together with 13 other *Ixodes* species, in the *I. ricinus* complex (Xu et al., 2003), a paraphyletic group of closely related and morphologically similar species that are all vectors for Lyme diseases spirochetes. However, the paraphyly of the group suggests that acquisition of the ability to transmit borreliosis agents, distributed worldwide, may have multiple origins (Fig. 1.3).

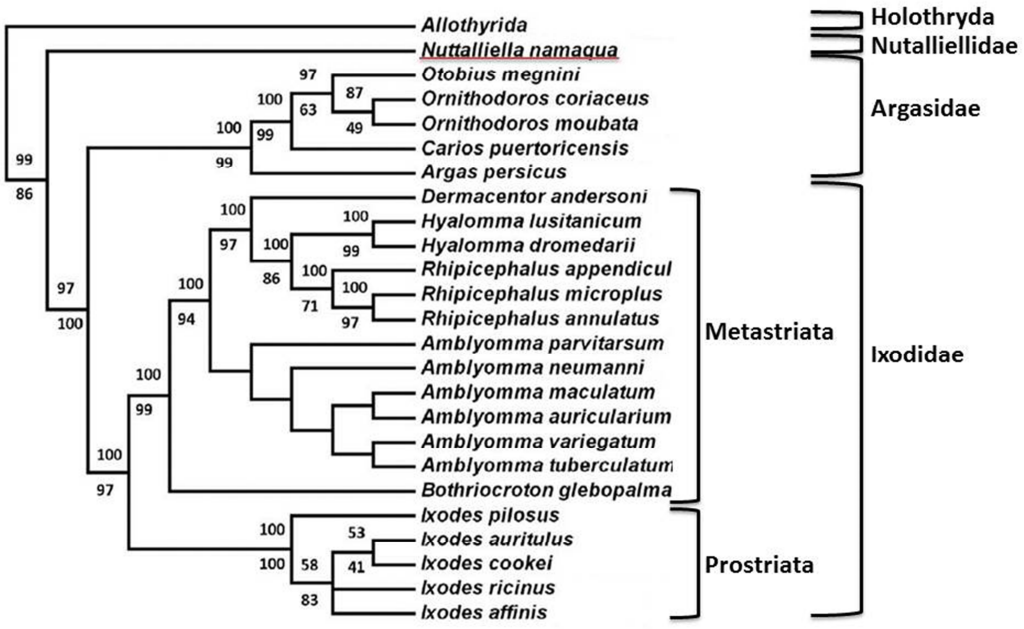

Fig. 1.2 The phylogenetic tree of Ixodida (18S-16S rRNA genes), obtained with Bayesian as well as maximum parsimony analysis. Posterior probability and bootstrap support values are indicated above and below the nodes, respectively (reprinted from Mans et al., 2011).

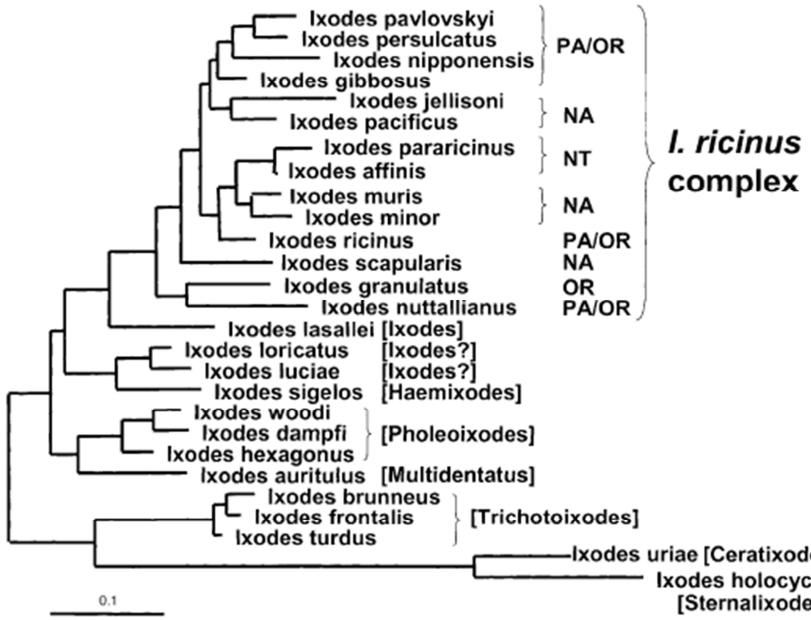

Fig. 1.3 16S Bayesian tree of *Ixodes* species, with special reference to the *I. ricinus* complex; it indicate that members of the *I. ricinus* species complex are closely related despite the fact that they are distributed in different geographic regions of the world (OR=oriental region; NA= Neartic region; NT= Neotropical region). However, the complex is not a monophyletic group unless 3 more species, *I. muris*, *I. minor*, and *I. granulatus*, are also included in it (reprinted from Xu et al. 2003).

The *I. ricinus* life cycle consists of three instars: larva, nymph and adult, commencing from the hatching of the clutch of eggs deposited on the ground by the female (Fig. 1.4). Sexual dimorphism only appears at the adult stage. This species is defined as a 'three-host tick', meaning that each instar takes a single bloodmeal (but see Results and Gray et al., 1999), before molting to the next stage. *I. ricinus* is an exophilic ticks, i.e. actively questing on the vegetation in order to latch on to passing hosts. The feeding phase usually takes several days, and varies between life stages: larva, 2–3 days; nymph, 4–5 days; adult female, 7–9 days. The adult male rarely feeds and never becomes fully engorged. As the majority of their lifespan is spent independently from hosts, questing on the vegetation or developing to the next instar in the ground litter, ticks are very vulnerable to desiccation. Hence, in both the questing and developing phases, ticks can obtain water from sub-saturated air by secreting and then re-ingesting hygroscopic fluid that is produced by the salivary glands. However, this process promotes tick survival only if the relative humidity in their microclimate does not fall below 80%. Accordingly, preferential habitat of the sheep tick includes areas with good vegetation cover and a mat of decaying vegetation on the ground, such that even during the driest periods of the year, a damp microclimate is maintained near the ground. Deciduous woodland, particularly those containing oak and beech, harboring good numbers of large mammals are preferred to conifer forest and cultivated lands, however ticks may also be found in abandoned farmland with dense underbrush, and in patchy vegetation in sub-urban and urban environments (Gray et al., 1998, 1999; Lindgren and Jaenson, 2006; Cagnacci et al., 2012; Rizzoli et al., 2009). Large vertebrate hosts such as wild ungulates, carnivores, cattle and sheep, are the main determinants of tick abundance in preferred habitat, as the female needs a large

bloodmeal to produce eggs, whereas larvae and nymphs are less demanding in their host preference, feeding in addition to large mammals, on a variety of birds, rodents, insectivores, and reptiles (Hoogstraal and Aeschlimann, 1982). Humans are considered accidental hosts (Fig 1.4; Gray et al., 1998; Lindgren and Jaenson, 2006).

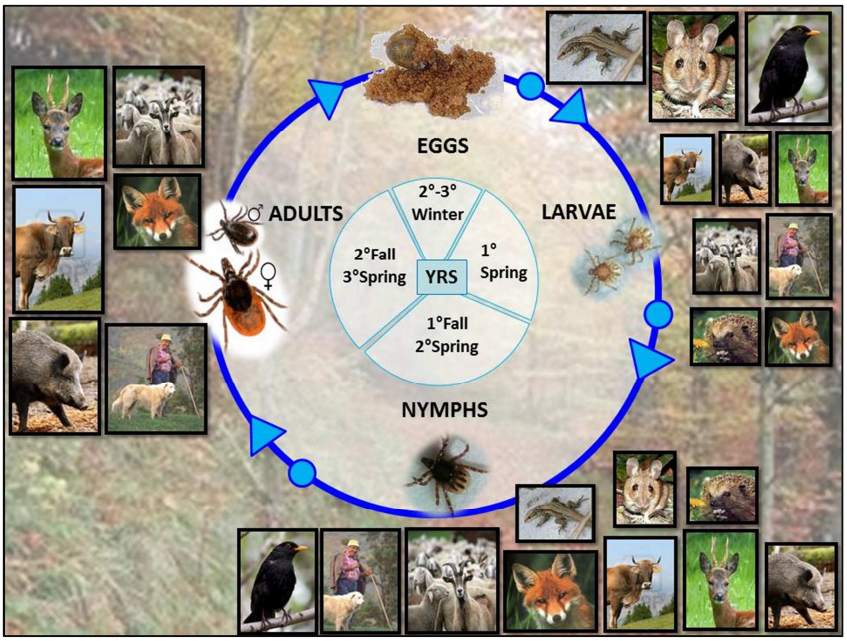

Fig. 1.4 *Ixodes ricinus* life cycle; between each of the life stages (larva to nymph, nymph to adult, adult for producing eggs), the sheep tick needs a bloodmeal. As a generalist ectoparasite, *I. ricinus* uses 350 different vertebrate species as hosts, including, occasionally, humans (Hoogstraal and Aeschlimann, 1982).

The phenology of *I. ricinus*, i.e. the variation in abundance of the three developmental stages over time and space, is known to be greatly influenced by climatic factors, but also host populations seasonal abundance (Tagliapietra et al., 2011; Randolph et al., 2002; Estrada-Peña et al., 2006; Medlock et al., 2013; Randolph, 2008, 2009; Gray et al., 2009; Gilbert, 2010). Ticks finely regulate their questing activity and developmental phases according to these variables, thanks to two diapause mechanisms:

the developmental diapause, involving arrested development of the engorged stages or of eggs, and the behavioral diapause, involving a form of quiescence of the unfed ticks at times when environmental conditions are unsuitable for seeking hosts (Gray et al., 1998). These mechanisms enable the tick to avoid entering host-seeking phases at unfavorable times of the year, such as high summer and winter. Since tick population phenology may be highly variable according to local climatic conditions (Randolph et al., 2002; Tagliapietra et al., 2011), completion of tick development to the adult stage may take from 2 to 6 years.

Ticks are responsible for the transmission of a variety of microorganisms including bacteria, viruses and protozoa, and toxin associated diseases, to wild and domesticated animals, as well to humans. Ticks are recognized as second only to mosquitoes in importance for the transmission of zoonoses (Parola and Raoult, 2001). The significance of *I. ricinus* as a vector is the result of various peculiarities in its biology; in fact, its wide host range permits interaction with potentially infected reservoir hosts in vertebrate communities, and allows different stages to feed on the same hosts and the trans-stadial maintenance of infection, greatly improves its performance as vector (Randolph et al., 2004). Therefore, it is epidemiologically relevant to discriminate, within the host community, between 'tick maintenance hosts' (or amplifying hosts, mainly large bodied mammals) involved only in feeding tick populations and regulating density; and 'reservoir hosts' (rodents and birds for Borrelia spp.; rodents for TBEv), that maintain and amplify the pathogens involved in the zoonotic cycle and are responsible for infecting the ticks, except in the case of co-feeding, where an infected tick could transfer the pathogen to nearly feeding ticks on a non-infected hosts (Gray et al., 1998; Randolph et al., 1997). As vector potential of a

pathogen in a certain area is a function of the vector-reservoir host contact rate, the specific composition of the host community could have a great impact on the epidemiological cycle of TBD. In fact, some studies (Ostfeld and Keesing, 2000; Schmidt and Ostfeld, 2001; LoGiudice et al., 2003) have underlined that the presence of incompetent host species for a specific pathogen could decrease the transmission of the pathogen itself by diverting tick bite (i.e. infection of new ticks) from more competent hosts. This is called the 'diluition effect' and seems to be likely supported by high levels of biodiversity (Johnson and Thieltges, 2013). However, the relationship between biodiversity and zoonotic diseases appears to be case-specific rather than a general dynamic and still needs careful evaluation in disease epidemiology (Pfäffle et al., 2013).

Rizzoli et al. (2014) reports an updated list of the main *I. ricinus* associated pathogens and hosts (Fig. 1.5). Even if the list refers to host-pathogen association retrieved in the European peri-urban and urban environment, it is also valid for zoonotic cycles taking place in natural systems, where, in terms of human health, TBD risk is more meaningful.

Fig. 1.5 (next page) List of the most important host group and species involved in *I. ricinus*-borne diseases systems in peri-urban and urban habitats. Pathogens in bold indicate those for which reservoir competence has been experimentally proven. (reprinted from Rizzoli et al., 2014).

Order	Species	Associated *I. ricinus* stage	Associated pathogens
Rodentia	*Apodemus flavicollis*	L, N	**TBEV** ***Borrelia afzelii*** ***Borrelia burgdorferi* s.s.** ***Borrelia spielmanii*** ***Borrelia miyamotoi*** ***Cand.* N. mikurensis** *Anaplasma phagocytophilum* *Babesia microti*
	Apodemus sylvaticus	L, N	**TBEV** ***Borrelia afzelii*** ***Borrelia burgdorferi* s.s.** ***Borrelia spielmanii*** ***Cand.* N. mikurensis** *Anaplasma phagocytophilum* *Babesia microti*
	Apodemus agrarius	L, N	***Borrelia afzelii*** ***Cand.* N. mikurensis** *Anaplasma phagocytophilum* ***Babesia microti***
	Myodes glareolus	L, N	**TBEV** ***Borrelia afzelii*** ***Borrelia burgdorferi* s.s.** ***Borrelia miyamotoi*** ***Cand.* N. mikurensis** *Anaplasma phagocytophilum* *Babesia microti*
	Microtus agrestris	L, N	**TBEV** *Borrelia afzelii* *Babesia microti* *Cand.* N. mikurensis *Anaplasma phagocytophilum*
	Microtus arvalis	L, N	*Cand.* N. mikurensis *Anaplasma phagocytophilum* *Babesia microti*
	Rattus norvegicus	L, N	***Borrelia afzelii*** ***Borrelia spielmanii***
	Rattus rattus	L, N	*Borrelia afzelii* *Anaplasma phagocytophilum*
	Eliomys quercinus	L, N	***Borrelia spielmanii***
	Muscardinus avellanarius	L, N	***Borrelia spielmanii***
	Glis glis	L, N	**TBEV** *Borrelia afzelii*
	Sciurus carolinensis	L, N	***Borrelia afzelii*** *Borrelia burgdorferi* s.s.
	Sciurus vulgaris	L, N	**TBEV** *Borrelia burgdorferi* s.s. *Borrelia afzelii* *Borrelia garinii*
	Eutamias sibiricus	L, N	*Borrelia burgdorferi* s.s. *Borrelia afzelii* *Borrelia garinii*

Order	Species	Associated *I. ricinus* stage	Associated pathogens
Lagomorpha	*Lepus europaeus*	L, N, A	*Borrelia burgdorferi* s.l. *Anaplasma phagocytophilum*
	Lepus timidus	L, N, A	*Borrelia burgdorferi* s.l.
Soricomorpha	*Sorex araneus*	L, N	**TBEV** *Borrelia burgdorferi* s.l. *Anaplasma phagocytophilum* *Babesia microti*
	Sorex minutus	L, N	*Borrelia burgdorferi* s.l.
Erinaceomorpha	*Erinaceus europaeus*	L, N, A	*Borrelia afzelii* *Borrelia spielmanii* *Borrelia bavariensis* *Anaplasma phagocytophilum*
	Erinaceus roumanicus	L, N, A	**TBEV** *Borrelia afzelii* *Borrelia bavariensis* *Anaplasma phagocytophilum* *Cand.* N. mikurensis
Artiodactyla	*Capreolus capreolus*	L, N, A	*Anaplasma phagocytophilum* *Babesia venatorum*
	Cervus elaphus	L, N, A	*Anaplasma phagocytophilum*
	Dama dama	L, N, A	*Anaplasma phagocytophilum*
Carnivora	*Vulpes vulpes*	L, N, A	*Borrelia burgdorferi* s.l. *Anaplasma phagocytophilum*
	Meles meles	L, N, A	*Borrelia afzelii* *Borrelia valaisiana*

Fig. 1.5 Continued

The two most important TBDs in Europe are Lyme disease (LB) and Tick Borne Encephalitis (TBE). LB is a multi-systemic inflammatory disorder caused by an immune response to the pathogenic genspecies of *Borrelia burgdorferi* sensu lato, and is the most prevalent arthropod-borne disease in the temperate regions of the northern hemisphere, causing approximately 65,500 patients yearly in Europe. As is evident from Fig. 1.5, experimentally confirmed competent reservoir hosts include many common

species of small and medium-sized rodents (mice, rats, squirrels, hares and rabbits), as well as several bird species (especially passerines), reptiles and insectivores. In contrast, large wild and domesticated mammals are considered non-competent hosts for the pathogen, but important vector-maintaining hosts (reviewed in Rizzoli et al., 2011). TBE, instead is caused by the TBE virus of the genus *Flavivirus* within the Flaviviridae family. It is present only on the Eurasian continent with three subtypes (European subtype, the Siberian sub-type and the Far Eastern subtype) which are associated with varying degrees of disease severity. The TBEv cycle involves permanently infected ticks and small mammals, especially rodents; transmission occurs horizontally between tick and vertebrate hosts, but co-feeding of infected and non-infected ticks on the same hosts and trans-stadial and trans-ovarial transmission of the virus, also play a major role in virus transmission (reviewed in Mantke et al., 2011).

Given the low active mobility of ticks in general (Falco and Fish, 1991), *I. ricinus* transportation on hosts while blood feeding is the only means of tick movement and introduction into new regions, along with vectored pathogens. Their survival in the new area is related to the climatic condition, but also on the presence of a suitable hosts community (reviewed in Léger et al., 2013; Gray et al., 2009; Semenza and Menne, 2009).

Concern for the impact of TBDs on human health has arisen from the current altitudinal and latitudinal expansion of *I. ricinus* populations, already widely distributed across Europe (Fig. 1.6; Léger et al., 2013; Medlock et al., 2012), but most importantly from the higher potential contact rate between ticks and humans as a result of increased human exploitation of tick habitat for recreational activities and by reports of infected ticks population in peri-urban and urban green areas of Europe (reviewed in Rizzoli et

al., 2014). Several strongly interlinked factors are favoring the establishment of ticks populations in new environments: climate change (i.e. increase in the mean annual temperature and milder winters) is positively affecting both tick survival at higher altitudes and latitudes, as well as hosts community density; changes in land use (i.e. reforestation; fragmentation) and wildlife management are affecting hosts spatial distribution (Gray et al., 2009; Cagnacci et al., 2012; Medlock et al., 2013).

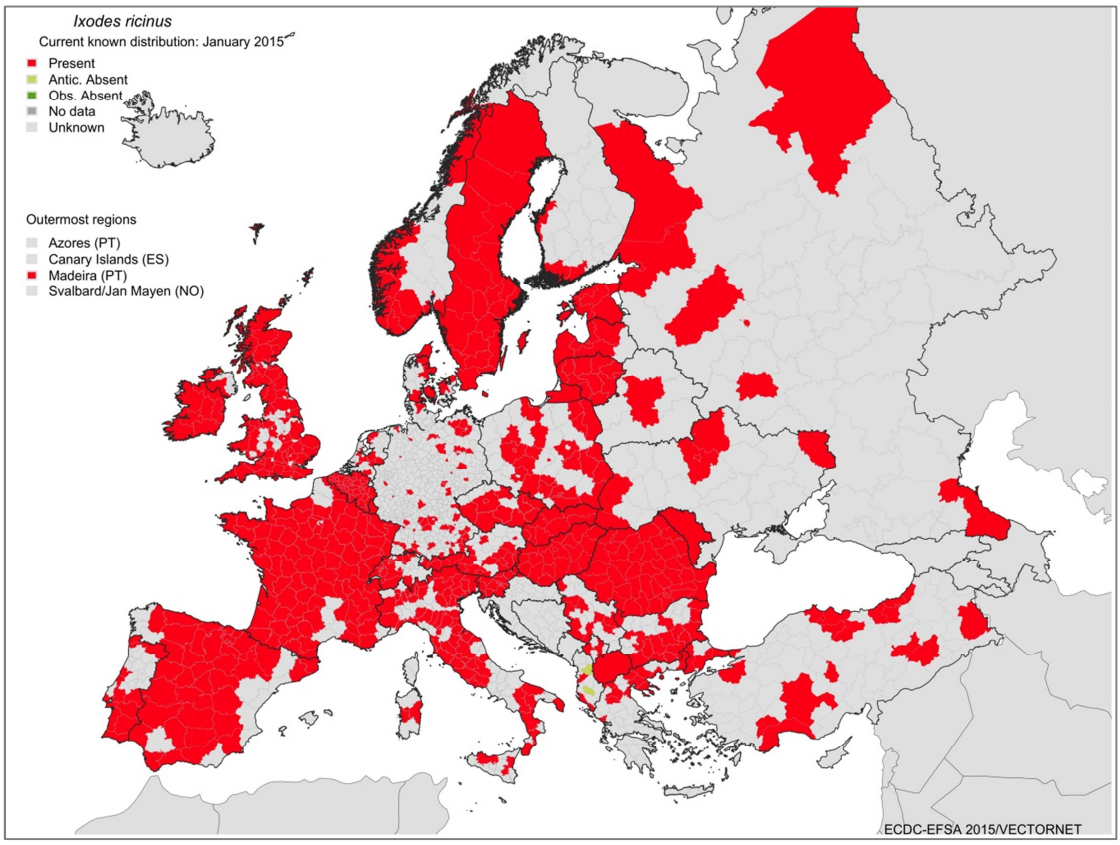

Fig. 1.6 *I. ricinus* distribution updated to January 2015, from www.ecdc.europa.eu.

1.3 Bloodmeal analysis in questing ticks

TBD systems are characterized by a complex network of interactions. Multiple vertebrate hosts are involved, many of them are competent reservoirs for etiological agents of infections (Fig. 1.5). In addition, transmission can occur vertically (e.g. trans-ovarial and trans-stadial), by co-feeeding, as well as by vector transmission between competent infected hosts and non-infected individuals. In this scenario, the knowledge of vector feeding ecology, host community composition and host status (competent reservoirs or amplification host) is of critical importance for understanding the epidemiology of VBD, in order to improve disease control strategies and to model and predict disease risk for public health (Mukabana, et al., 2002; Kent, 2009; Gómez-Díaz and Figuerola, 2010; Bolzoni et al., 2012).

In a generalist tick such as *I. ricinus*, estimation of host exploitation by field observations (i.e. host capture and tick counting) is generally difficult, expensive and may provide an unrealistic assessment, given the fact that *I. ricinus* spend only few days per life stage on the host (Kirstein and Gray, 1996, Estrada-Peña et al., 2005). Serological methods opened the way for indirect study of feeding patterns of hematophagous arthropods by means of analysis of their gut content, simply named *bloodmeal analysis*. But the advent of the polymerase chain reaction (PCR) and molecular genetic markers, and their ability to amplify DNA even in degraded and damaged samples (aDNA, Pääbo et al., 2004), has meant that DNA-based methods have been increasingly used for bloodmeal analysis (Mukabana et al., 2002; Kent, 2009; Gómez-Díaz and Figuerola, 2010), especially in combination with molecular screening for vectored pathogens. A number of methods have been applied to different hematophagous vectors, as summarized in Table 1.1 and reviewed in Kent (2009); these

usually rely on amplification of a target locus of the host DNA and its identification by means of various post-PCR techniques. Sequencing would be a more straightforward and specific method to identify the host, but the cost of this technique limits its application in high-throughput processing for large sample sets; therefore, more cost-effective post-PCR identification methods are usually employed, such as heteroduplex analysis, restriction length fragment polymorphisms (RLFP), and reverse line blotting hybridization (RLBH; Kent, 2009).

In contrast to mosquito vectors, where recently engorged females can be easily captured by using baited oviposition traps, field collection of engorged ticks following detachment from the host is not possible (Sonenshine, 1991). Only questing ticks can be collected easily by the conventional blanket dragging method; therefore, bloodmeal analysis can only applied to questing ticks. Because field-collected questing ticks may have had their previous bloodmeal up to one year before, in the previous instar (Randolph et al., 2002), specific issues must be addressed: i) only a few intact copies of host DNA are likely to be present, stored in the endosome of the tick midgut (Sonenshine, 1991); ii) remnant host DNA quality will be compromised by digestive and hemolytic process (Kirstein ans Gray, 1996; Sojka et al., 2013); iii) the presence of a high concentration of heme molecules inhibits polymerase activity; iv) environmental DNA or human DNA contamination may mask the true DNA bloodmeal signature. Using PCR-based methods, points i) and ii) have been addressed by using short amplicons of mitochondrial DNA. The mitochondrial genome may be present in hundreds to thousands copies in a single cell, making mitochondrial markers an ideal option to robustly amplify vertebrate DNA from the tiny bloodmeal remnants; in addition, vertebrate mtDNA has an evolutionary rate 5 to 10 times faster than the

nuclear genome and a rapid mutation fixation between species, which makes this molecule valuable for species identification, even when short fragment are targeted (Kirstein and Gray, 1996; Kent, 2009). Mitochondrial DNA can be retrieved in all blood cells of birds, while for mammals, having anucleated red blood cells, leukocytes and epithelial tissue still contribute in significant numbers. To cope with PCR inhibitions and contamination problems (points iii) and iv)), a careful protocol design is required, from primers selection to reagent and reaction optimizations. All sample handling procedures must be carried out in sterile conditions and the use of negative controls in DNA extractions and PCR provide a control for contamination.

Kirstein and Gray (1996) were the first to report bloodmeal analysis in questing ticks by means of host DNA amplification with degenerate vertebrate *cytb* primers and host identification with RLFP and RLBH. They proved the ability of molecular methods to reliably amplify host DNA up to 200 days post-engorgement, and defined RLBH has the most feasible method for host DNA identification. Since then, a variety of PCR-RLBH methods have been tested, gradually extending the array of probes for host identification. Pichon and colleagues (2003, 2005, 2006) targeted the multicopy 18S rRNA gene, while Humair et al. (2007) and Morán Cadenas et al. (2007), applying Humair's protocol, targeted the 12S rDNA vertebrate gene. Although partially successful, these methods show a wide variability in bloodmeal identification success (from 26.4% Bown et al., 2009 to 49.4% Pichon et al., 2005 and 62.8% Allan et al., 2010 on *Amblyomma americanum*) according to questing ticks collection time (i.e. time passed since the last bloodmeal), collection site microclimate (affecting speed in digestion) and life stage (quantity of ingested blood) (Morán Cadenas et al., 2007; Pichon et al., 2006). In addition, the taxonomic level of host identification reached in

these studies varies according to the set of probes developed: in the majority of the publications cited above, most hosts were only identified to group or genus level, and rarely to species.

Recently, a PCR-RLFP protocol, targeting the *12S* rDNA gene has been optimized and applied to a large sample dataset (Wodecka et al., 2014). This method appears to provide a higher mean identification success than reported for RLBH (62.8%). However, although the article claims that RFLP patterns are available for about 60 host species, only 19 host species were actually identified in the study.

Alternative methods to DNA-based ones have been proposed, such as protein analysis (Wickramasekara et al., 2008; Laskay et al., 2012), proteomics-based, spectral-matching (Önder et al., 2013), and stable isotope analysis (SIA; Rasgon, 2008), but none of them is well-established or have been applied to large sample of questing ticks.

The large body of tick bloodmeal analysis studies and the range of methods proposed underline the importance of this topic in the epidemiological study of TBDs, but there is still no reliable method free from contaminations and sensitivity problems (Gómez-Díaz and Figuerola, 2010; Estrada-Peña et al., 2013; Estrada-Peña et al., 2005). Within the present thesis, I aim to define a new protocol for bloodmeal analysis, including the application of this new method to the study of feeding patterns of *I. ricinus* larvae in the Province of Trento using a large collection of questing ticks. As reported in the Materials and Methods chapter, we began by testing if there was still room to improve the current DNA sequencing approaches, by designing new primers and by performing a careful optimizations. However, the unsatisfactory results led us to move to a more recent and promising approach: HRMA.

Table 1.1 Comparative overview of current molecular methods for arthropod bloodmeal identification (reprinted from Kent 2009).

Marker	Technique	Benefits	Drawbacks	Applications
COI	DNA sequencing	*Can get specific identity of the host *Sequence database excellent *Can be used to confirm results of another method	*Requires expensive sequencing equipment or contracting services *Sequence data may be missing for some species, potentially leading to bloodmeal mis-identification	*Mosquitoes*: Kent *et al.* (unpublished data)
cytb	DNA sequencing	*Can get specific identity of the host *Sequence database good *Can be used to confirm results of another method	*Requires expensive sequencing equipment or contracting services *Sequence data may be missing for some species, potentially leading to bloodmeal mis-identification	*Mosquitoes*: Kent *et al.* (unpublished data); Hamer *et al.* (2008); Savage *et al.* (2007); Molaei *et al.* (2006, 2007); Kilpatrick *et al.* (2006a, 2007); Richards *et al.* (2006); Cupp *et al.* (2004); Apperson *et al.* (2002, 2004); Hassan *et al.* (2003) *Ticks*: Tobolewski *et al.* (1992); *Blackflies*: Malmqvist *et al.* (2004); Hellegren *et al.* (2008)
	Group-specific primers	*Can obtain broad classification for bloodmeal *Can potentially detect mixed bloodmeals *Only requires PCR and gel electrophoresis	*Conserved primers may be cross-reactive (see Table 1 footnote) *Amplicons often sequenced anyway for more specific results *Multiplexing can sometimes lead to problems with primer interference	*Mosquitoes*: Kent & Norris (2005); Ngo & Kramer (2003); Temu *et al.* (2007); Molaei *et al.* (2006, 2007)
	Heteroduplex analysis	*Sensitive *Many samples can be analysed at once	*Technique difficult to master *Results may be ambuigous and/or need sequence confirmation *Only samples on the same gel can be reliably compared	*Mosquitoes*: Lee *et al.* (2002); Hassan *et al.* (2003); Apperson *et al.* (2002, 2004); Richards *et al.* (2006); Savage *et al.* (2007); *Black flies*: Boake *et al.* (1999); *Tsetse flies*: Boake *et al.* (1999); Njiokou *et al.* (2004); Simo *et al.* (2008)
	PCR–Restriction Fragment Length Polymorphism	*Can target nucleotide substitions or minor sequence differences between organisms	*Requires prior knowledge of polymorphic restriction sites *Requires RFLP profile library to match unknown samples	*Mosquitoes*: Ngo & Kramer (2003); Oshaghi *et al.* (2006a); *Tsetse flies*: Steuber *et al.* (2005); *Ticks*: Kirsten & Gray (1996)
	Terminal RFLP	*Good, user-friendly database for searching results	*Complex, labor-intensive procedure *Sequencing equipment necessary	*Mosquitoes*: Meece *et al.* (2005)
	Real-time PCR	*Highly sensitive and specific *Added ability to quantify starting template	*Need fluorescent probes, real-time thermal cycler and software *Limited number of flurophores available	*Mosquitoes*: Van Den Hurk *et al.* (2007) *Fleas*: Woods *et al.* (unpublished data)
18S, 12S rDNA	Reverse line-blot hybridization	*Can screen many samples for many hosts simultaneously *Potential to identify mixed bloodmeals *Less expensive than similar microarray technology	*May require designing and optimizing novel primers and probes if not already published	*Ticks*: Pichon *et al.* (2003, 2005); Humair *et al.* (2007); Morán Cadenas *et al.* (2007)
Microsatellites	DNA profiling	*Can match bloodmeals to individual hosts	*Dependent on the characterization of microsatellite loci from the host being studied *Requires profiles be generated for all individuals potentially fed upon	*Mosquitoes and people*: Chow-Shaffer *et al.* (2000); Michael *et al.* (2001); De Benedictis *et al.* (2003); Soremekun *et al.* (2004); Scott *et al.* (2006); *Mosquitoes and birds*: Darbro *et al.* (2007); *Tsetse flies and cattle*: Torr *et al.* (2001)

1.3.1 High Resolution Melting Analysis

Invented in 2002 by University of Utah and Idaho Technology, HRMA is a simple, rapid post-PCR method, widely applied for genotyping, mutation scanning and sequence matching, as well as for pathogen screening and identification (Gundry et al., 2003; Wittwer et al., 2003, 2009; Reed et al., 2007). HRMA exploits the fundamental property of double stranded DNA (dsDNA) to melt, that is, to separate into two strands with heat. The quantity of heat needed for the melting to occur strongly depends on the GC content, length and nucleotide arrangement of the amplified dsDNA fragment. Introduced in the 1960's, when melting was first monitored by UV-absorbance, melting analysis became popular from 1997 with the advent of the Real-time PCR LightCycler® and the use of fluorescent intercalating dye (Ririe et al., 1997). Thanks to new instruments and dyes, dsDNA melting can now be monitored at high resolution. The acquisition of high density data, i.e. fluorescence measures per unit time, and precise temperature control, allows detection of small variations in DNA sequences, down to single nucleotide polymorphisms (SNPs).

In HRMA, amplification of the dsDNA region of interest, using conventional or Real time PCR, is performed in the presence of a specialized dsDNA binding dye. The ability of this dye to be highly fluorescent when ligated into the dsDNA, while poorly fluorescent when DNA in the unbound state, is exploited for amplification control in Real time PCR, but in HRMA, it is used essentially to monitor the melting step: after PCR, the amplicons are gradually denatured by increasing the temperature in small increments (0.008 - 0.2 °C). The dsDNA fragments denature gradually, releasing the dye, which results in a drop of fluorescence (Fig. 1.7).

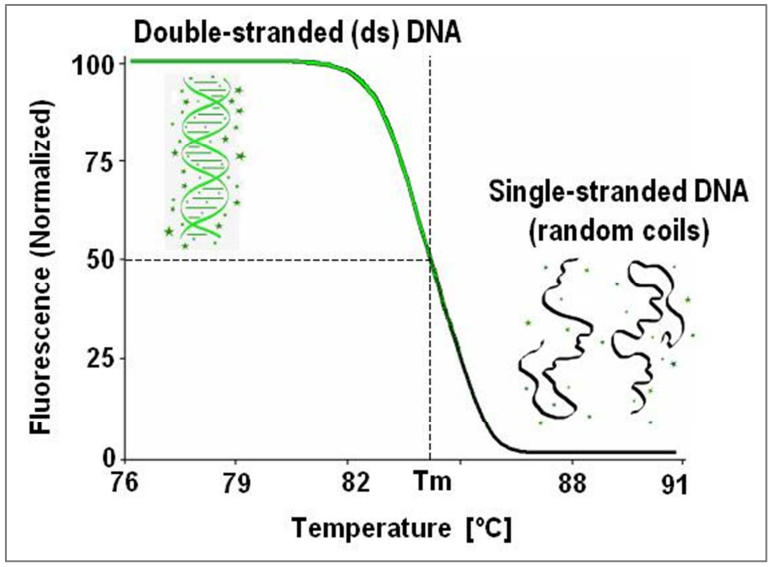

Fig. 1.7 Example of HRMA principle with a normalized fluorescence plot. (https://dna.utah.edu/Hi-Res/TOP_Hi-Res%20Melting.html)

Fluorescence readings and temperature changes collected during the HRM step are compared with specific software. First, visualization of the 'raw melting curves plot' is carried out (see example in Fig. 1.8a). The highest rate of fluorescence decrease is generally at the melting temperature of the DNA amplicon of interest (called T_m), more specifically defined as the temperature at which 50% of the DNA sample is double-stranded and 50% is single-stranded. T_m could be derived by plotting the derivative of fluorescence vs. temperature (-dF/dT against T), or the 'derivative melting plot' (Fig. 1.8b). Since initial fluorescence values on the raw melting profile can be variable between samples making analysis difficult, it is standard to perform a simple normalization of pre- and post-melting regions in order to align data and magnify the differences in melting properties of genetically different samples, visualized as a 'normalized melting plot' (Fig. 1.8c). Another way to visualize melting data is the

‘difference melting plot’, where normalized melting data of a user-defined genotype (reference) are subtracted from the normalized data of all other samples; the reference genotype is visualized as the baseline and the position of the other samples is plotted against the temperature. This process aids the visualization of normalized data (Fig. 1.8d).

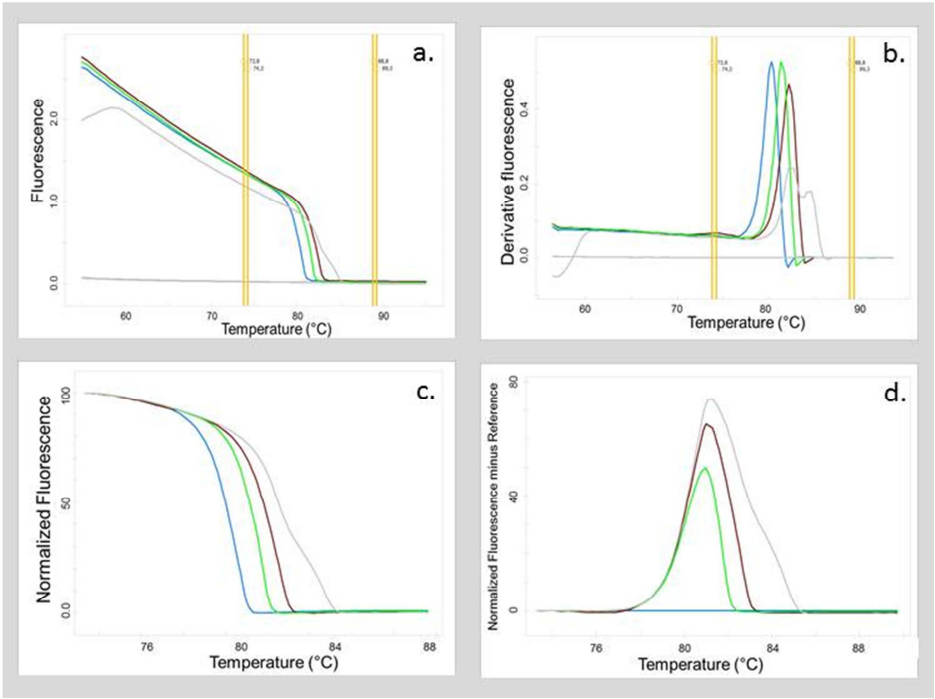

Fig. 1.8 Process of software analysis of high resolution melting data; a.) raw melting curves plot, yellow bars define the normalization area ;b.) derivative melting plot; c.) normalized melting plot; and d.) difference melting plot. (plots from the Real-time HRMA of control samples in bloodmeal analysis).

In the design of a new HRMA assay, the choice of target DNA and primer design are the most critical steps. Several factors should be carefully considered, based on the fact that melting properties are a function of amplicon length, nucleotide compositions and arrangement. Generally, small amplicons are required for maximum sensitivity (optimal 50 to 300 bp) and with known sequence variation. Under these conditions, the number of potential DNA regions for primer design is delimited. In

addition, primers should be highly specific to the target organism, possibly without degenerations; and have a high melting temperature and low probability of dimers formation (www.kapabiosystems.com/assets/Introduction_to_High_Resolution_Melt_Analysis_Guide.pdf).

As mentioned above, HRMA is widely used for genotyping (Erali and Wittwer, 2010) and mutation scanning; however, most importantly for the purpose of this research, HRMA has been applied successfully to sequence matching (Reed et al., 2007), pathogen screening and identification (Do et al., 2008; Lin et al., 2008), and species identification (Winder et al., 2011; Kang and Sim, 2013; McCarthy et al., 2013). Additionally, HRMA has proved to be a sensitive and reliable method even when DNA is degraded and of low quantity, especially when coupled with Real-time PCR (Do et al., 2008). HRMA has already been applied to bloodmeal analysis in the Chagas diseases vector (*Trypanosoma cruzi*; Peña et al., 2012), with bloodmeals identifiable to species 30 days after the bloodmeal. In contrast to RLBH and other currently used methods, the lack of sample processing after the amplification step and the possibility for amplicons to be run on agarose gel after the HRMA and to be sequenced, are additional features of HRMA that are promising for its application to bloodmeal analysis in questing ticks.

1.4 *Ixodes ricinus* phylogeography and population genetics

The analysis of the patterns of genetic variability in molecular markers using appropriate statistical methods can indirectly provide key information about the biology of vectors (i.e. mating system, dispersal and gene flow), demography (effective population size, past demographic events) and evolutionary potential (adaptation, host-race formation and speciation). These features of the molecular ecology approach are particularly important in *I. ricinus*, characterized by an extremely complex life history, living for a few days a year in contact with the host, and of limited physical dimensions (reviewed in De Meeûs et al., 2007).

I. ricinus population structure was firstly investigated by means of allozyme variability by Delaye et al. (1997); a limited number of polymorphisms and absence of differentiation was observed among *I. ricinus* collected from neighbouring populations in Switzerland. Other studies have been carried out using microsatellites markers. Population genetic substructure was suggested by a significant heterozygosite deficit in several analyzed populations: as geographical based genetic structuring or isolation by distance was not significant, the heterozygote deficit was attributed to sex-biased dispersal (De Meeûs et al., 2002; Kempf et al., 2010), and host-parasite association (De Meeûs et al., 2002; Kempf et al., 2011). Using the same microsatellite markers, non-random pairing in males and females (e.g. assortative mating) was proposed in *I. ricinus* (Kempf et al., 2009b). For other tick species (*I. uriae*, McCoy et al., 2003 and Kempf et al., 2009a; *Rhipicephalus micropilus*, De Meeûs et al., 2010), host-race evolution has been hypothesized.

At the European range of the species, phylogeography studies using mitochondrial (Casati et al., 2008), and a combination of mitochondrial and nuclear

markers (Noureddine et al., 2011; Porretta et al., 2013) concluded that *I. ricinus* present one panmictic population at the European level. Noureddine et al. (2011) noted that European *I. ricinus* populations were genetically differentiated from north-African populations, but its species status was not resolved. Further genetic discontinuity was noted between the British and Latvian tick populations using a mitochondrial genes multilocus sequence typing (Dinnis et al., 2014). These results support the hypothesis that there is a high level of gene flow between European populations of *I. ricinus*, presumably as a result of host movements (Noureddine et al., 2011).

As part of the EDENext project, population genetic patterns were to be compared at the European, regional and local levels. While INRA was assigned to study gene flow at the EU and local levels (with some of those results cited above in Nourredine et al., 2011), we were to investigate the level and distribution of genetic variation at the regional level (PAT). Given the number of previously published papers using microsatellite markers for this purpose, we decided to genotype both questing and feeding ticks from across PAT in order to evaluate both the influence of geographical features and host species in shaping genetic variation in the overall population.

However, as later presented in Results, our findings proved that microsatellites were not suitable markers for population genetics studies of *I. ricinus*, as confirmed by Quillery et al. (2013). Given that microsatellites are not abundant in the arthropod genome (Fagerberg et al., 2001), development of new STR markers was not feasible in the time available. On the other hand, Single Nucleotide Polymorphisms (SNPs) have been successfully used in the investigation of genetic variation in populations of non-model species (Helyar et al., 2011). SNPs are the most abundant and uniformly distributed

markers in the genome (Schlötterer, 2004), allowing informative patterns of genetic variation in populations to be analysed (Lao et al., 2006; Paschou et al., 2007).

A SNP dataset was not available at that time for *I. ricinus* (later Quillery et al., 2013) or similar ixodid species and, therefore, we decided to apply the recently developed NGS RAD-Seq technology to obtain one.

1.4.1 Next Generation Sequencing: RAD-Seq

NGS technologies have revolutionized genomic and transcriptomic approaches to biology, but also the study of genetic variation in populations. At the base of this success lies the ability of NGS platforms, such as Illumina, Roche 454 and ABi SOLiD, to produce giga-bases of DNA sequences at minimal cost (Seeb et al., 2011; Davey and Baxter, 2010). It is now possible to produce whole genome sequences of several individuals of a target species and, from their comparison, design new markers for genetic variation analysis, such as microsatellites and SNPs. However, for eukaryotes with large genome sizes lacking an assembled reference genome, whole genome sequencing was still a challenging task, until the introduction of Reduced Representation Libraries (RRL; Seeb et al., 2011; Etter et al., 2011; Davey and Baxter, 2010). RAD-Seq falls into this category (Miller et al., 2007), and is particularly suitable for our purpose as it allows the simultaneous *de novo* discovery and genotyping of tens of thousands of SNPs throughout the genome at limited cost, and requires no prior development of genomic resources (Miller et al., 2007; Davey and Baxter, 2010). The first application of this approach to population genomics (Hohenlohe et al., 2010), SNP discovery and genome mapping (Baird et al., 2008), and phylogeography (Emerson et al., 2010), illustrated the versatility of the method, even in non-model organisms,

without reference genomes and characterized by complex evolutionary histories (Rowe et al., 2011). As presented in Figure 1.9, RAD-Seq combines two molecular biology techniques: DNA fragmentation with a restriction enzyme (as used in restriction fragment length polymorphisms and amplified fragment length polymorphisms methods), and association of each individual/population to a specific molecular identifier (MID), allowing pooling of individuals into a single library to be sequenced in one Illumina sequencing lane. The Illumina HiSeq sequencing technology used here allowed me to read 100 bp of the genomic region flanking the restriction site; paired-end sequencing allows me to extend the genome investigation by 300- 400 bp region downstream of the restriction site (Etter et al., 2011; Fig. 1.9G). In this way, the entire genome is randomly sampled and interrogated, leading to a reduction in the analysis complexity usually associated with whole genome sequencing (Davey and Baxter 2010; Rowe et al., 2011).

Genetic variation in terms of SNPs, insertion/deletions, and microsatellites, can be investigated in the 100 bp reads generated from the restriction site (R1) (Etter et al., 2011). The software *Stacks* is a modular pipeline specifically designed for efficiently processing and assembly of the large numbers of short-read sequences originating from multiple samples generated by RAD-Seq and by other RRL protocols (Catchen et al., 2011, 2013). It incorporates a maximum likelihood statistical model to identify sequence polymorphisms and distinguish them from sequencing errors, either *de novo* or with sequences aligned to a reference genome. The basic *Stacks* workflow and the programs involved, are presented in Figure 1.10. According to the RAD-Seq features, this method promised a dataset that could be analysed to answer our biological questions.

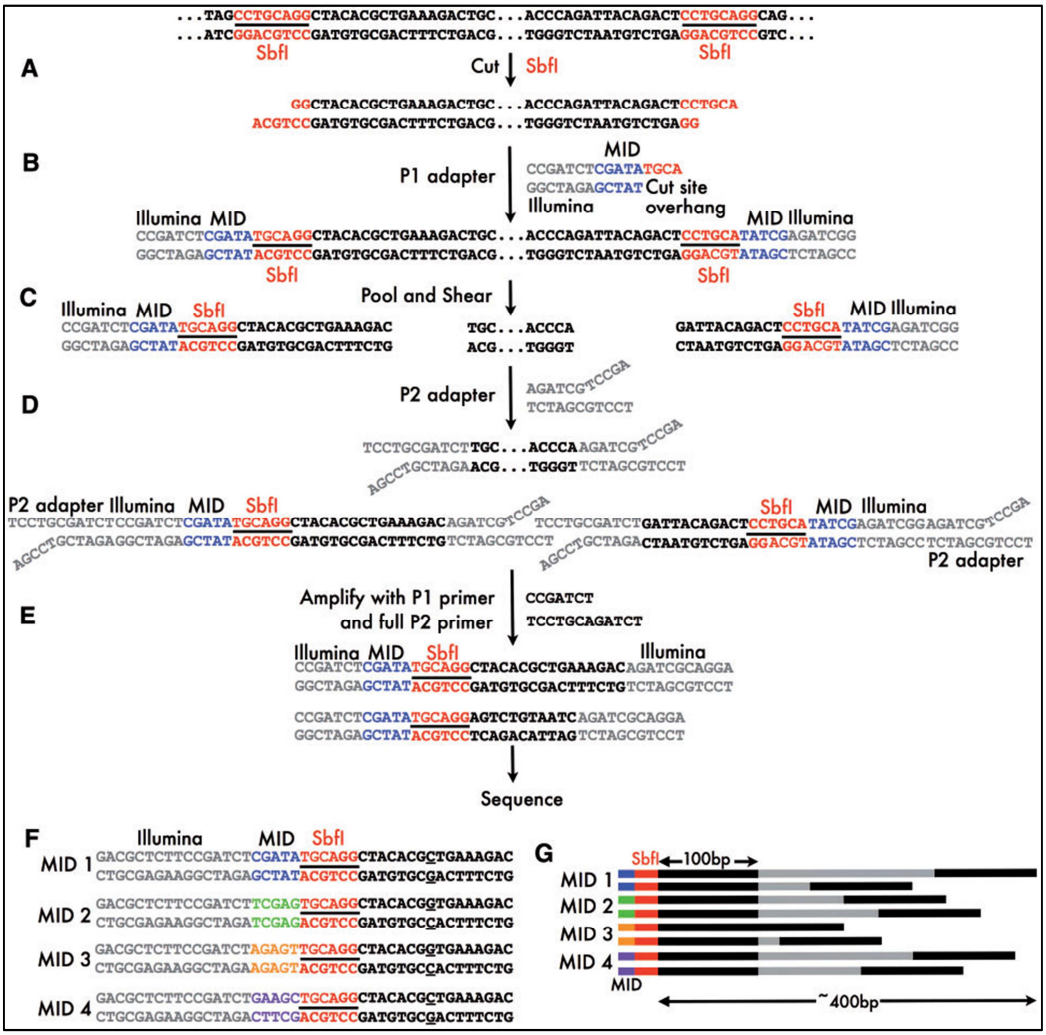

Fig. 1.9 The RADSeq method. (A) Genomic DNA is sheared with a restriction enzyme of choice (SbfI in this example). (B) P1 adapter is ligated to SbfI-cut fragments. The P1 adapter is adapted from the Illumina sequencing adapter (full sequence not shown here), with a molecular identifier (MID; CGATA in this example) and a cut site overhang at the end (TGCA in this example). (C) samples from multiple individuals are pooled together and all fragments are randomly sheared. Only a subset of the resulting fragments contains restriction sites and P1 adapters. (D) P2 adapter is ligated to all fragments. The P2 adapter has a divergent end. (E) PCR amplification with P1 and P2 primers. The P2 adapter will be completed only in the fragments ligated with P1 adapter, and so only these fragments will be fully amplified. (F) Pooled samples with different MIDs are separated bioinformatically and SNPs called (C/G SNP underlined). (G) As fragments are sheared randomly, paired-end sequences from each sequenced fragment will cover a 300- 400 bp region downstream of the restriction site. (Reprinted from Davey and Baxter, 2010).

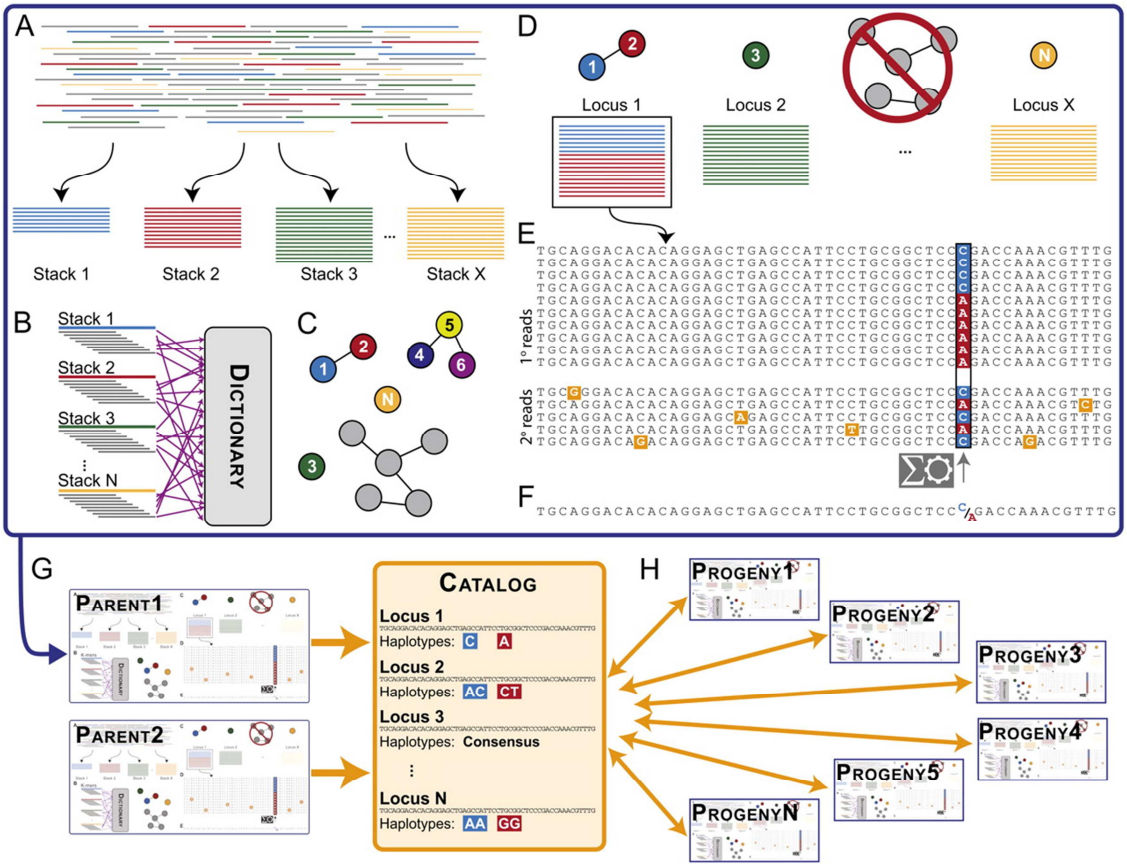

Fig. 1.10 Schematic *Stacks* workflow. (A) The *ustacks* program forms stacks in an individual from short sequencing R1 reads (cleaned by *process_radtags.pl*) that match exactly. (B) The *ustacks* program breaks down the sequence of each stack into k-mers and loads them into a dictionary. The *ustacks* program breaks down each stack again into k-mers and queries the k-mer Dictionary to create a list of potentially matching stacks, which can be visualized as nodes in a graph connected by the nucleotide distance between them. (C) *ustacks* merges matched stacks to form putative loci. (D) *ustacks* matches secondary reads that were not initially placed in a stack against putative loci to increase stack depth. An SNP model in *ustacks* checks each locus at each nucleotide position for polymorphisms. (E) *ustacks* calls a consensus sequence and records SNP and haplotype data. (F) The *cstacks* program loads stacks from the parents of a genetic cross into a Catalog to create a set of all possible loci in a mapping cross. (G) *sstacks* matches map cross progeny against the Catalog to determine the haplotypes at each locus in every individual in the cross. (Reprinted from Catchen et al., 2011)

2. OBJECTIVES

The TBD spread and emergence facing Europe urgently calls for improved knowledge of *I. ricinus* biology, especially dispersal and host-association patterns. The present study focuses on a relatively small geographical area (Province of Trento, Italy), where 30 *I. ricinus* populations were sampled intensively both while feeding and while questing. The main objective of this thesis was to apply *state-of-the-art* molecular approaches to fill this knowledge gap, specifically:

1. Population genetics will be applied to investigate the genetic structure of sheep tick populations, in order to understand if geographical barriers (i.e. mountain chains; rivers) and/or host exploitation affects the rate of gene flow (and therefore, dispersal) on this scale;
2. Bloodmeal analysis of questing tick nymphs will be optimized to define the relative importance of various wild and domestic vertebrate species as larval hosts in two important tick habitats (extensive forest and forest patches near urban settlements).

Since feeding ecology and dispersal are interrelated in sheep ticks (because they move as they feed), I will also meet the main objective by merging the results of the two approaches, such that the epidemiological implications of the research regarding TBD circulation, spread and maintenance, can be discussed.

3. MATERIALS AND METHODS

For reasons that will become obvious below, field and laboratory protocols are listed in the order they were developed and applied. Therefore, because the list of Materials and Methods is rather extensive, a chart is provided below to aid the reader in following the workflow. Red X's indicate methods that were tested, but subsequently abandoned because reliable, repeatable results were not obtainable.

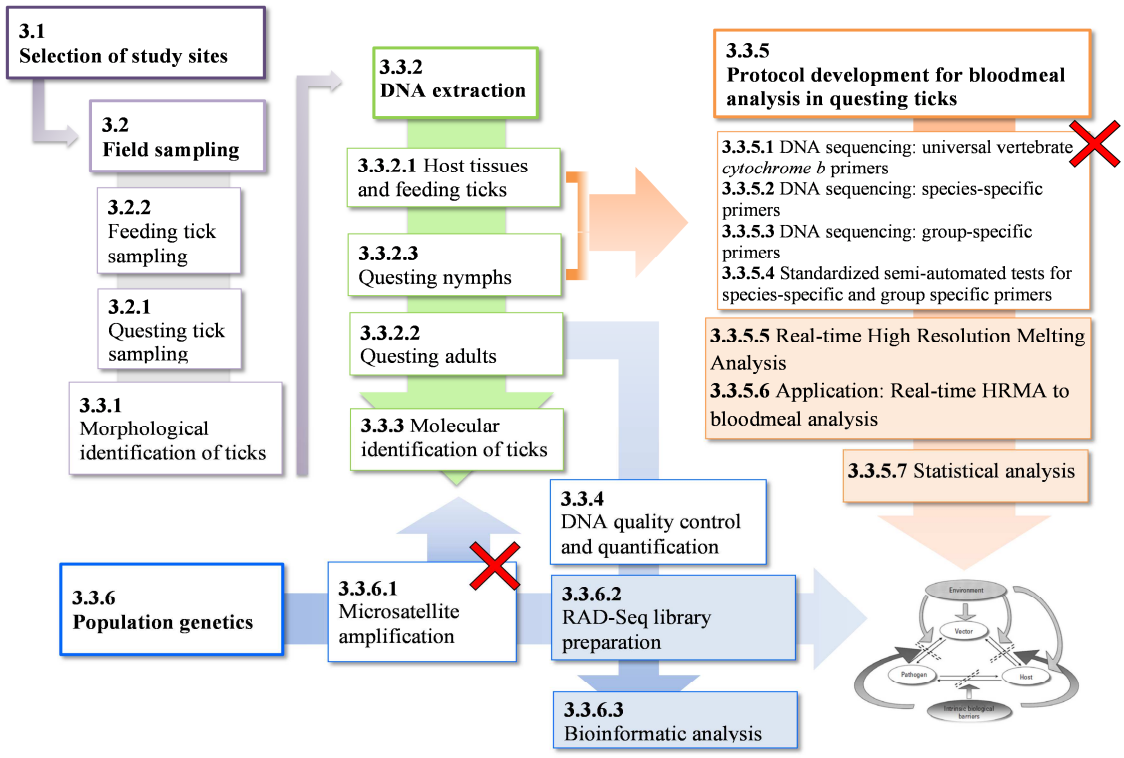

Fig. 3.1 Schematic workflow of material and methods protocol used and their interactions.

3.1 Selection of study sites

The Province of Trento (PAT; 6206 km^2) lies in the heart of the Dolomitic Alps (Fig. 4.1); 77% of the territory is above 1000 m a.s.l.. Its complex geomorphology, consisting of a network of river valleys of various orientations and extent, surrounded by rugged mountain chains, and dotted with lakes, and hills, results in a wide diversity

of local climatic conditions, ranging from sub-mediterranean to continental and alpine, that vary according to exposure, altitude, and prevailing winds. Forest covers about 55% of the region. In the lowest reaches of the southern valleys, thermophile deciduous woods and coppices with hophornbeam (*Ostrya carpinifolia*) and flowering ash (*Fraxinus ornus* L.) prevail; in sub-continental and continental valleys, beech forests (*Fagus sylvatica* L.) are common at the lowest altitudes, replaced by mountain and subalpine Norway (*Picea abies* L. Kasten) and European silver spruce forests (*Abies alba* Miller) along the altitudinal gradient. Much of the coniferous forest is heavily managed; in addition, widespread anthropic disturbance is present across the study area. In fact, although towns and larger urban areas are mainly concentrated in the valley floors, many villages are at higher altitudes, embedded in patchy agro-ecosystems, representing the natural bridge with the described forests. Cultivated crops (especially grape and apple) are a strong feature of the landscape, but agricultural activities at higher altitudes are less profitable and hay meadows and pasture somewhat abandoned. Additionally, tourism, is one of the most important economic driving forces in the area, and increases the exploitation of natural areas for recreational activities. The fauna is characteristic of natural Alpine habitats. Importantly for this thesis, many of the vertebrate species are also recognized as important hosts maintaining sheep tick (*I. ricinus*) populations, such as small mammals (*Apodemus* spp., *Myodes glareolus*, *Sorex* spp. and *Crocidura* spp.), passerine birds (*Turdus merula, T. philomelos, Erithacus rubecula*), wild ruminants (*Capreolus capreolus*, *Cervus elaphus*, *Ovis musimon*, *Rupicapra rupicapra*) and Carnivores (*V. vulpes*; Carpi et al., 2008; Bolzoni et al., 2012; Cagnacci et al., 2012). In addition, some are competent reservoirs for tick-borne pathogens such as *Borrelia burgdorferi* s.l., Tick Borne Encephalitis virus and

Anaplasma phagocytophila (Rizzoli et al., 2004; Mantelli et al., 2006; Pecchioli et al., 2007; Carpi et al., 2008; see General Introduction).

3.2 Field sampling

In order to carry out an in-depth investigation of the host use and genetic structure of tick populations at a local scale, 30 sites were selected in typical tick habitat across PAT. At least one site was selected from each of the main valleys of the Province. Questing ticks were sampled in all 30 sites, while sampling of all stages of feeding ticks from the main host species was carried out in 10 of these (Fig. 4.1 in Results).

3.2.1 Questing tick sampling

Material:

- 1 m^2 white felt blanket;
- measuring tape;
- forceps;
- sterile 2 mL sterile vials with plug seal cap (Sigma);
- 10% bleach;
- gloves.

Questing nymphs and adults were collected by conventional blanket-dragging (Sonenshine, 1993). A 1m x 1 m white felt blanket, attached by one side to a wooden pole, was dragged over leaf-litter, grass and low understory. Every 5 m the blanket was checked for ticks. Ticks were removed from the blanket by researchers wearing gloves and using forceps sterilized with diluted bleach (10%), placed individually in 2 mL

vials, transported live at room temperature and subsequently frozen at -80°C, until DNA extraction. Initially, a 100 m transect was dragged (as part of a long-term monitoring program at FEM), then dragging continued at random in the surrounding area until enough ticks were collected (see below).

Sampling was carried out on dry days during the spring activity peaks of tick populations, i.e. middle of April to June (2012 and 2013) (Tagliapietra et al., 2011). At each site we aimed to collect at least 30 questing nymphs and 12 questing adults, half males and half females. Altitude, exposure and geographical coordinates, as well as predominant vegetation type were noted; tick density was expressed as the number of ticks (adults and nymphs) collected in 100 m^2.

3.2.2 Feeding tick sampling

At the outset of the project, I planned to genotype both questing and feeding ticks from various host species using microsatellite loci in order to compare the genetic patterns of ticks found on certain host species, to determine if different tick genotypes were associated with certain vertebrate hosts ('host races'). Therefore, to collect feeding ticks of the dominant tick-host species in PAT, small mammals and birds were trapped, and large mammals were sampled at hunting inspection stations.

In the 10 sites selected for feeding tick sampling, the aim was to collect at least 30 ticks of any stage from the main tick host species, with a maximum of 5 ticks from the same individual.

Target host species:

- ✓ small mammals: Muroidea: yellow-necked mouse (*Apodemus flavicollis*), wood mouse (*Apodemus sylvaticus*) and bank vole (*Myodes glareolus*), the most

numerous and widespread forest-dwelling rodents in PAT, and competent hosts for several tick-borne diseases; *A. flavicollis* inhabits forested areas from the valley floor to the vegetation limit, *M. glareolus* is also fairly widely distributed, but prefers coniferous woodland up to 1800 m; *A. sylvaticus* is a less common rodent species in PAT;

- ✓ birds: Passeriformes: blackbird (*Turdus merula*), European thrush (*T. philomelos)* and European robin (*Erithacus rubecula*), the most important host for *I. ricinus* in forested and peri-urban habitats (Marsot et al., 2012; reviewed in Rizzoli et al., 2014). These species forage low in the vegetation or on the ground and, therefore, are more prone to *I. ricinus* infestations (Marsot et al., 2012). It is worth noting that *E. rubecula* is predominantly a forest species, preferring dense coniferous or deciduous tree clusters; however, it is commonly found in urban garden and parks; *T. merula* and *T. philomelos*, are also ubiquitous species, inhabiting forest as well as forest edge and more urbanized areas;
- ✓ wild large mammals: Cetartiodactyla: roe deer (*Capreolus capreolus*), red deer (*Cervus elaphus*) and chamois (*Rupicapra rupicapra*). These deer species are widely distributed in the Province, while chamois prefer higher altitudes and more open habitat, where ticks are often absent.
- ✓ domesticated animals: Canidae: domestic dog (*Canis lupus familiaris*); Caprinae: domestic sheep (*Ovis aries*).

3.2.2.1 Trapping of small mammals

Material:

- Ugglan live traps (8x9x23 cm);
- Potatoes cut into chunks;
- whole sunflower seeds;
- polyethylene transparent bags;
- forceps;
- 10% bleach;
- 70% ethanol (Sigma) in 2 mL sterile vials with plug seal cap (Sigma);
- gloves and facial mask with virus filter.

In sites CON, PIN, REV, MEZ and TRA (Table 4.1, Fig. 4.1) trapping was carried out by FEM personnel and myself, after health and safety training in small mammals trapping and handling. At these sites, the only aim of small mammal trapping was the collection of feeding ticks. Trapping sessions took place in spring-summer 2012 and 2013, with traps set on the first day then checked daily for two or three consecutive nights and repeated twice or three times for each site until enough ticks were collected. Live traps were placed in linear transects, when possible, ca. 10 m apart, next to fallen logs, tree roots, rocks and where rodent burrows were visible. A variable number of traps was used: between 42 and 47 for each session.

CAV is a long-term sampling site of DBEM with 4 permanent grids of traps. A single grid is composed of 8 transects with 8 traps each; both transects and traps are ca. 10 m apart. Traps are activated monthly for three nights. Sites LUN and CAD were established as part of the PAT-funded project ROCOALPS (Rodent communities in a changing environment: implications for human health in the Alps; PI: Konečny Adam,

2010-2013) with two 8x8 grids each, similar to those in CAV; here, traps were activated monthly for two nights. Site GRI was trapped as part of a collaborative project between DBEM and MUSE (Trento) for the definition of small mammal biodiversity in PAT; in this case, transects were used and a variable number of traps (37 – 40) was used for each trapping session.

In all cases, Ugglan live-traps were baited with sunflower seeds and pieces of potato to provide a moisture source. Traps were activated in the late morning or early afternoon and checked the subsequent morning. Small mammals were released from the trap into a clear polyethylene bag and, once confined in the corner of the bag, were taken firmly by the scruff of the neck and tail. Species identification by morphological traits did not allow us to discriminate between *A. flavicollis* and *A. sylvaticus*; therefore sampled individuals are later identified at genus level (*Apodemus* spp.; Michaux et al., 2001). All ticks were gently removed with sterile forceps and placed individually in 2 mL Eppendorfs with 1 ml of 70% ethanol. Ticks coming from the same individual rodent were placed in the same tube, unless damaged during removal.

Small mammal and trap handling were carried out under strict health and safety rules for field work (e.g. using gloves and face mask). Permission to carry out the trapping in the defined areas and all animal handling procedures were authorized by the Comitato Faunistico Provinciale della Provincia di Trento prot. n. 595 04.05.2011. Ethical guidelines concerning animal welfare were followed as defined by the European Commission and detailed in the EDENext project.

3.2.2.2 Bird netting

Material:

- 12x2.5 m mist-nets with 5 shelves, mesh size 16 mm, normal filament;
- 12x2 m mist-nets with 5 shelves, mesh size 16 mm and thin filament;
- 4 m telescopic poles for mist-net support;
- forceps;
- 10% bleach:
- 70% ethanol (Sigma) in 2 mL sterile vials with plug seal cap (Sigma);
- birding morphological measurement instruments;
- gloves.

In order to analyse feeding ticks acquired by hosts in the area of interest (and not outside the Province), our netting regime aimed at capturing resident birds, rather than migrating ones. Therefore, live-trapping was carried out in all but CAV sites during the reproductive period for the bird species of interest. Netting was carried out with the assistance of ornithologist Dott. Franco Rizzolli, a licensed bird ringer. At each site, a minimum of 120 m of mist-nets, divided into two or more transects, were set and maintained for one or two days per trapping session. Trapping efficiency drastically dropped after just one day and consequently, we had to move the nets to a different position daily. In general, ecotonal areas, like meadow-forest borders, or wetlands, as well as areas near food sources (like orchards and bushes with berries) were selected. Mist-nets were activated before sunrise and monitored for trapped birds every hour; they were closed during heavy rain and during the hottest hours in the middle of the day. Only the licensed ornithologist handled the birds; species and sex were identified, standard morphological measurements taken, reproductive and moult condition noted,

and ectoparasites were collected. Ticks were mainly removed from the region surrounding the eyes, ear holes and beak. Once removed with sterile forceps, ticks were placed in 70% ethanol. Ticks from the same individual were placed in the same 2 mL vial, unless damaged. Each bird was ringed before release at the trapping site. At the CAV site there is a permanent ringing station, managed by Mauro Segata (forest warden and ornithologist), who kindly collected ticks from birds netted at this site during the same period.

3.2.2.3 Collection of ticks feeding on large mammals

Material:

- polyethylene transparent bags;
- forceps;
- 10% bleach;
- gloves;
- magnifier lamp;
- 70% ethanol (Sigma) in 2 mL sterile vials with plug seal cap (Sigma).

Ticks were collected from the most common large wild mammals present in PAT (see list above), bagged during the autumn hunting season, when these ungulates are still territorial and consequently, feeding ticks are representative of the animal's area of origin (Carpi et al., 2008). In collaboration with the Trentino Hunter's Association, the wardens of each game reserve encompassing our sampling site, and/or local hunting guards were contacted. The Hunter's Association is responsible for ensuring that the species, sex and age of each bagged animal matches those of the hunter's license. However, such controls are done in two ways: either the hunter is required to bring the

animal to a central checkpoint on the evening of the kill, or the hunter contacts the warden who subsequently goes to the hunter's house to make his report. Therefore, to collect ticks, either:

- ✓ carcasses were checked for ticks in the presence of wardens and/or hunting guards, within 24 h of the kill, by myself or trained DBEM personnel; or,
- ✓ the warden of the game reserve or local hunting guards were trained to remove the forelegs (part of the leg distal to the carpal joint, which mainly hosts tick larva and nymphs), as soon as possible after the kill, and to place them in a sealed plastic bag, which I provided; if the hunting personnel were willing, I also provided sterile vials filled with high grade ethanol for adult ticks. These samples were conserved at -20°C until delivery to FEM. Forelegs were stored at -20°C until tick removal, carried out under a magnifying lamp, following Carpi et al. (2008).

Species, sex and age, as well as altitude, exposure, and location of the kill were registered for each sample.

For PIN and CON sites, ticks were occasionally collected from acquaintances' dogs. Domestic sheep (*Ovis aries*) ticks were collected by contacting local sheep breeders.

Ticks coming from the same individual were placed in the same 2 mL vial filled with 70% high grade ethanol, unless damaged.

3.3 Laboratory methods

3.3.1 Morphological identification of ticks

Material:

- identification key (Cringoli et al., 2005; Estrada-Peña et al., 2004);
- dissecting microscope (Zeiss);
- forceps;
- petri dishes;
- 10% bleach and ethanol for cleaning instruments;
- gloves.

In PAT, the most widespread tick is *I. ricinus*, but on wildlife it is also possible to retrieve *I. hexagonus* and *I. trianguliceps*. Given the specialist, nidicolous life-style of these tick species, I could assume that the vast majority of questing ticks collected by blanket dragging would be *I. ricinus*; however, feeding ticks collected from parasitized wildlife could be both nidicolous and generalist ticks. Therefore, for both questing and feeding ticks, I attempted to identify collected ticks morphologically using a recognized identification keys (Cringoli et al., 2005; Estrada-Peña et al., 2004), at 40X magnification under a dissecting microscope.

Although this procedure was straightforward for most of the questing ticks, morphological identification was sometimes difficult for engorged specimens in the larval and nymphal stages, since some morphological features were damaged or missing as a result of collection, or deformed because of engorgement. In this case I confirmed species identity by molecular analysis (see section 3.3.3).

3.3.2 DNA extraction

DNA isolation from ticks, in general, is a challenging task due to their size (especially larval and nymphal stages) and the hard chitinous exoskeleton; furthermore previous studies have shown that tick DNA is susceptible to degradation (Hubbard et al., 1995; Hill and Gutierrez, 2003; Halos et al., 2004). Additionally, according to the use of DNA from collected ticks (i.e. pathogen detection, bloodmeal analysis, NGS application) with the EDENext project, various DNA qualities and quantities were required. Therefore, I used a number of different protocols and commercial and non-commercial kits in order to meet the post-extraction applications requirements in terms of quantity, quality of DNA, but also with the aim of improving the efficiency of the protocols, in terms of time and cost.

3.3.2.1 Host tissues and feeding ticks

DNA was extracted from engorged ticks collected from the host while feeding (see *Feeding tick collection*) and from tissue samples available from previous or ongoing projects at FEM, using the Qiagen Dneasy Blood and Tissue kit (Qiagen, Valencia, CA, USA). DNA from a large variety of host species was needed to validate bloodmeal protocols (as control samples; see section 3.3.5). Specifically, for Real-time HRMA, host DNA extracted from tissue and engorged ticks was needed i) to test that the primer pairs amplified the target fragments correctly and reliably using conventional PCR; ii) to validate the species-discriminating power of HRMA and; iii) as positive samples in Real-time HRMA of questing ticks with unknown bloodmeal sources. Feeding ticks were also employed for microsatellite genotyping validation.

For engorged ticks, a pre-lysis step, following morphological identification,

included a physical disruption step. A single tick was placed in 2 mL autoclaved Eppendorf tube with a 5mm stainless steel bead and 100 µL PBS (Phosphate-buffered saline) solution. Using a mixer mill, tubes placed in the proper sampler holder, were shaken for 3 min at 30 Hz to completely crush the tick. The homogenate was then transferred to a fresh autoclaved Eppendorf tube, 20 µL Proteinase-K 10 mM/mL (Sigma-Aldrich, Saint Louis, USA) and 180 µL Qiagen ATL tissue lysis buffer were added, and left overnight to digest on a rotary tube mixer placed in an incubator at 56 °C. Spin-column DNA purification was performed following manufacturer instructions for QiaAmp® DNA Investigator Kit (Qiagen, Valencia, CA, USA), *Purification of Total DNA from Nail Clipping and Hair*.

3.3.2.2 Questing adults

To avoid contamination from environmental DNA, human DNA and cross-contamination among samples, all recommended precautions were used: forceps were sterilized with bleach and ethanol; washing and DNA extraction were performed under a biological (UV sterilized) hood; sterile and DNA-free consumables and reagents were used for all methods.

Sample lysis

The sample lysis procedure was common to the different DNA extraction methods used, as follows. After morphological identification, each tick was washed twice in DNA-free distilled water to rehydrate it and to eliminate possible surface contaminants. Briefly, ticks were immerse in 200 µL of RNase DNase free water (Sigma-Aldrich, Saint Louis, USA) in 0.5 mL sterile tubes for at least 20 s; after

agitation of the tube by tipping it up and down, the tick was transferred, using a fresh sterile 200 µL tip, to another 0.5 mL tube and the procedure repeated. The tick was then transferred in the same way to a fresh 0.5 mL tubes, still containing 200 µL RNase DNase free water (Sigma-Aldrich, Saint Louis, USA), until processed for lysis. In 1.5 mL sterile vial containing 180 µL ATL Buffer, each tick was cut in small pieces using a sterile disposable scalped with pointed blade (WVR, Radnor, PA, USA). 20 µL Proteinase K 10 mM/mL (Sigma-Aldrich, Saint Louis, USA) for Qiagen column extraction or 20 µL Proteinase K (KingFisher™ Cell and Tissue DNA Kit; Thermo Fisher Scientific, Vantaa, Finland) for Thermo magnetic-beads extraction, 30 µL DTT 1 M (Sigma-Aldrich, Saint Louis, USA) were added. After vortexing for reagent-sample mixing, tubes were sealed with parafilm, and placed in an incubator for overnight digestion at 56°C.

Before purification with one of the following methods, digested samples were treated with RNase A 100 mg/ml (Qiagen, Valencia, CA, USA). 1 µL for each 50 µL digested sample was added to the lysate and incubated at room temperature for 30 min.

Spin-column protocol

Following indications for obtaining high quality and quantity genomic DNA from the RAD-Seq library generation protocol of Etter et al. (2011), DNeasy® Blood & Tissue Kit (Qiagen, Valencia, CA, USA) using the manual protocol *Purification of Total DNA from Animal Tissues* was used to purify DNA from lysed samples. Eleven samples plus a negative extraction control were manually processed at a time. Double elution of genomic DNA in two separate 1.5 mL sterile tubes was performed with 50 µL of the kit elution buffer. To improve DNA release from the column, elution buffer was

heated at 56°C, the sample was added to the silica membrane, and incubated for 5 min. Eluted gDNA was stored at 4°C until DNA quality control and quantification.

Phenol protocol

After lysis and RNAse treatment, 1 volume of phenol (pH 8) was added to the lysate sample, and placed for 15 min on a rotary tube roller, and centrifuged for 15 min at 5000 g. The aqueous phase was transferred to a fresh 1.5 mL Eppendorf tube. These steps were repeated twice. After phenol extraction, 0.4 volume of sodium acetate (3 M, pH 4.6) and 2 volumes of chilled 100% ethanol were added to the sample to precipitate the DNA. Samples were vortexed to aggregate DNA and briefly centrifuged to remove solution drops from the lid, before placing them at -80 °C for 10 min. DNA was pelleted by centrifuging for 14000 g for 15 min. Ethanol was removed from the sample. 200 ml of 70% ethanol was added and the tube gently vortexed. After sample centrifugation at 14000 g for 15 min, as much of the 70% ethanol as possible was removed. To eliminate residual ethanol, tubes were placed in a heated centrifuge under vacuum. DNA was resuspended in 50 µL of ATE buffer (Qiagen, Inc., Valencia, CA, USA). Samples were gently rocked to enhance DNA resuspension. gDNA was then stored at 4°C until DNA quality control and quantification.

Magnetic-beads protocol

The recently-developed magnetic-beads DNA purification technology promises high quality DNA, free of protein, nucleases, and other contaminants or inhibitors. The KingFisher Cell and Tissue DNA Kit with KingFisher™ Flex Magnetic Particle Processors (Thermo Fisher Scientific, Vantaa, Finland) combines the high quality DNA

and efficient extraction with an additional decrease in time of extraction, by employing a robotic handling of samples. KingFisher™ Flex Magnetic Particle Processors allow simultaneous processing of 96 samples, and the purification protocol can be easily controlled and modified with the Thermo Scientific™ BindIt™ Software.

We used the KingFisher Cell and Tissue DNA Kit (Thermo Fisher Scientific, Vantaa, Finland) following the manufacturer's instructions, although the lysis step was performed as described above. Plates and consumables for samples and reagents were UV-sterilized in a biological hood before loading into the instrument under a UV-sterilized biological hood. Final elution volume was adjusted to 80 µL for females and to 60 µL for males to fulfill final gDNA concentration requirements. In order to check for contamination, three negative controls were added in each 96-well plate. Eluted gDNA and negative controls were transferred from the elution plate to 1.5 mL sterile tubes and conserved at 4°C until DNA quality control and quantification.

3.3.2.3 Questing nymphs

Spin-column protocol

After morphological identification, each questing nymph was washed as described above (see *Sample lysis*). The protocol *Purification of Total DNA from Nail Clipping and Hair* of QiaAmp® DNA Investigator Kit (Qiagen, Valencia, CA, USA) was used with minor modifications. Briefly, each nymph was placed in a sterile vial containing 230 µL ATL Buffer (Qiagen) and cut into small pieces with a sterile scalpel; 40 µL Proteinase-K 10 mM/mL (Sigma-Aldrich, Saint Louis, USA) and 30 µL DTT 1 M (Sigma-Aldrich, Saint Louis, USA) were added to the solution. Overnight digestion was performed on a rotary tube mixer placed in an incubator at 56 °C. Prior to

processing, 1 µL Carrier RNA (1 µg/µL in ATE (Qiagen) was added to the lysate, as suggested in the kit manual for low quantity samples to increase DNA binding to the silica membrane. Purification was performed manually following the standard protocol, except that: ATE elution buffer (Qiagen) was heated at 56°C and elution volume was adjusted to 60 µL; incubation time after ATE addition to the column was increased to 5 min. In some cases, as specified below, the same purification protocol was performed using the QIAcube (Qiagen, Inc., Valencia, CA, USA), a robotic workstation for automated purification of DNA, RNA and proteins. To avoid contamination from environmental DNA, all DNA extraction procedures were carried under a laminar flow hood (UV-sterilized); 11 nymphs were processed at a time with one negative control, for cross-contamination control. DNA was stored at -20 °C until use.

Magnetic-beads protocol

Since I obtained good quantity and quality of gDNA from adult questing ticks with the magnetic-beads method (see Results), I decided to apply it to questing nymphs for Real-time HRMA bloodmeal analysis, decreasing the time and funds necessary for identifying bloodmeals the new HRMA protocol. Sample lysis and DNA purification protocols and instruments are the same as those described in section 3.3.3 for adult ticks, except that final elution volume for nymphs was 80 µL. Samples and negative controls were transferred in 96 x 0.2 mL-well PCR plate to allow automated robotic PCR reaction set-up, and stored at -20°C until HRMA.

3.3.3 Molecular identification of ticks

According to Cringoli et al. (2005), the most important tick species found in the Alps, in addition to *I. ricinus*, are *I. hexagonus, I. trianguliceps, I. canisuga, I. frontalis, Riphicephalus sanguineus e Haemaphisalis punctata*. Sequences of the mitochondrial 16S gene for these species were downloaded from the database Taxonomy of NCBI (http://www.ncbi.nlm.nih.gov/taxonomy). Sequence alignment was done with Clustal X v. 2.0. PRIMER3 (www.frodo.wi.mit.edu/primer3), a new primer set targeting a 470 bp region of the 16S gene, allowing species discrimination through BLASTn (http://blast.ncbi.nlm.nih.gov/Blast.cgi) search, was designed. Primer sequences are as follows:

Ir_16S_681-F CAAAAACATTTCATTTTGG

Ir_16S_1159-R GAACTCAGATCATGTAGGAA

Conventional PCR was performed at a final volume of 20 µL, containing 0.5 µM of each primer, 0.25 mM of each dNTP, 1x HotMaster Taq Buffer, 1.25 U HotMaster Taq (5-Prime), and 1 µL of template DNA. The thermal cycling consisted of 94 °C for 2 min, 30 to 35 (in case of larvae) cycles at 94 °C for 30 s, 51°C for 1 min, 65 °C for 1 min and 40 s, and final elongation at 65 °C for 10 min, performed in a Veriti® Thermal Cycler (Applied Biosystems, Foster City, CA, USA). Amplification results were checked via capillary electrophoresis on a QIAxcel system (Qiagen, Valencia, CA, USA) using a DNA High Resolution Cartridge and QX 15 bp-3 Kb size marker and by applying the method OM500, and were analyzed using QIAxcel ScreenGel 1.0.2.0. PCR products were purified with Exo-SAP-ITTM (GE Healthcare, Little Chalfont, England) and both forward and reverse strands were sequenced on an ABI 3130 XL using Big Dye Terminator v3.1 (Applied Biosystems, Foster City, CA, USA). After

electropherogram check and creation of a consensus sequence using Sequencher v. 5.1 (Gene Codes corp.), a BLASTn search was carried out for species identification.

3.3.4 DNA quality control and quantification

Genomic DNA concentration was measured with the fluorometer Qubit® 2.0 (InvitrogenTM, Carlsbad, USA), as suggested by Etter et al. (2011). Fluorometric quantification is based on the properties of the Molecular Probes® Dyes that emit a fluorescent signal only when ligated to the target molecules. In our case, using the double stranded DNA kit (dsDNA), only double stranded DNA was quantified, even in presence of free nucleotides and degraded nucleic acids. DsDNA HS (high sensivity) Assay kit was used following the manufacturer's instructions; 1 µL of DNA was quantified in a dilution ratio of 1:200 with the kit working solution. The DNA in a sample was concentrated using a vacuum/pressure kit (Concentrator 5301, Eppendorf AG, Hamburg, Germany) when gDNA concentration lower than 10 ng/µL.

The quality of gDNA was assessed by running 3 µL of each sample in a 0.7% agarose gel stained with ethidium bromide, prepared with 10x TAE buffer. Electrophoresis was carried out in a 10x TAE buffer solution at 80 V for 40 minutes and viewed using a UV-transilluminator. High quality sample would present a compact DNA band at high molecular weight and absence of DNA smearing.

3.3.5 Protocol development for bloodmeal analysis in questing ticks

3.3.5.1 DNA sequencing: universal vertebrate cytochrome b primers

Universal primer set targeting about 350 bp of the vertebrate *cytochrome b* mitochondrial gene, described first by Kocher et al. (1989), was first tested on DNA extracted from engorged ticks (see 3.3.2.1 Host tissue and engorged ticks) collected while feeding from different host species (*A. flavicollis*, *T. merula, T. philomelos, E. rubecula, C. l. familiaris*, *O. aries*, *M. glareolus, E. europaeus, S. vulgaris, P. muralis*).

To optimize the conventional PCR amplification protocol, in order to amplify DNA from a number of different tick host species and avoid amplification of contaminant human DNA, different annealing temperature (52 – 62 °C), magnesium concentrations (1.0 – 2.0 mM) and touch-down protocols were tested. The following selected protocol was applied to questing ticks with unknown bloodmeal source. Reaction was performed with AmpliTaq Gold® DNA Polymerase (Applied Biosystem, Foster City, CA, USA) in 50 µL final volume mix containing: 1x AmpliTaq Gold® 360 Buffer 10x, 2 mM $MgCl_2$ 25 mM, 5 µL 360 GC Enhancer, 0.25 mg/mL of Bovine Serum Albumine (BSA) 0.4 µM of each primer, 0.25 mM of each dNTP, 1.25 U AmpliTaq Gold® 360 DNA Polymerase and 4 µL of template DNA. Thermal cycling, performed in a Veriti® Thermal Cycler (Applied Biosystem, Foster City, CA, USA), consisted of an initiation step at 95 °C for 10 min; denaturation at 95 °C for 30 sec, annealing at 54 °C for 30 sec and elongation for 1 min and 10 sec at 72 °C, 50 times; final elongation at 72 °C for 7 min. Negative controls for both DNA extraction and PCR amplifications were included in all amplification reactions. Amplification results were checked via capillary electrophoresis on a QIAxcel system (Qiagen, Valencia, CA, USA) using a DNA High Resolution Cartridge and QX 15 bp-3 Kb size marker with the

OM500 method, and were analyzed using QIAxcel ScreenGel 1.0.2.0. PCR products were purified with the Exo-SAP-IT™ kit (GE Healthcare, Little Chalfont, England) and both forward and reverse strands were sequenced on an ABI 3130 XL using Big Dye Terminator v3.1 (Applied Biosystems, Foster City, CA, USA). Following the electropherogram check and the creation of a consensus sequence using Sequencher v. 5.1, a BLASTn search was carried out to identify the species represented by the amplified fragment.

3.3.5.2 DNA sequencing: species-specific primers

A. flavicollis e *C. capreolus* were selected to test the feasibility of a bloodmeal identification approach that uses species-specific primer and sequencing since they are both bloodmeal sources for *I. ricinus* larvae and nymphs, and reservoir hosts for several TBDs.

For *A. flavicollis*, I created an alignment of *cytochrome b* sequences retrieved from Genbank. New primers were then designed using the online software PRIMER3 (www.frodo.wi.mit.edu/primer3) (see Table 4.8). For *C. capreolus* we used the primer F_UNIV2 and R_CAPREOLUS following Garros et al. (2011), to amplify 240 bp of *cytochrome b* in this species.

To validate the primers and optimize reaction conditions, PCR amplification was first performed on control DNA samples of *A. flavicollis* and *C. capreolus*. Optimized protocols (which were then applied to amplification of bloodmeal DNA in questing ticks) are as follows.

- ✓ *Apodemus*: the PCR reaction was performed with AmpliTaq Gold® DNA Polymerase (Applied Biosystem, Foster City, CA, USA) in 20 µL final volume

mix containing: 1x AmpliTaq Gold® 360 Buffer 10x, 2 mM $MgCl_2$ 25 mM, 2 µL 360 GC Enhancer, 0.5 µM of each primer, 0.25 mM of each dNTP, 1.25 U AmpliTaq Gold® 360 DNA Polymerase and 2 µL of template DNA. Thermal cycling, performed in a Veriti® Thermal Cycler (Applied Biosystem, Foster City, CA, USA), consisted of an initiation step at 95 °C for 10 min; denaturation at 95 °C for 30 sec, annealing at primer set specific T_a °C (see Table 4.8) for 20 sec and elongation for 30 sec at 72 °C, 50 times; final elongation at 72 °C for 7 min.

- ✓ *Capreolus*: the PCR reaction was performed with AmpliTaq Gold® DNA Polymerase (Applied Biosystem, Foster City, CA, USA) in 20 µL final volume mix containing: 1x AmpliTaq Gold® 360 Buffer 10x, 2 mM $MgCl_2$ 25 mM, 2 µL 360 GC Enhancer, 0.375 µM of each primer, 0.25 mM of each dNTP, 1.25 U AmpliTaq Gold® 360 DNA Polymerase and 2 µL of template DNA. Thermal cycling, performed in a Veriti® Thermal Cycler (Applied Biosystem, Foster City, CA, USA), consisted of an initiation step at 95 °C for 10 min; denaturation at 95 °C for 30 sec, annealing at 56 °C for 20 sec and elongation for 30 sec at 72 °C, 50 times; final elongation at 72 °C for 7 min.

Amplification results were checked using capillary electrophoresis with QIAxcel system (Qiagen, Valencia, CA, USA) using a DNA High Resolution Cartridge and QX 15 bp-3 Kb size marker with the OM500 method, and were analyzed using QIAxcel ScreenGel 1.0.2.0. PCR products were purified with Exo-SAP-ITTM (GE Healthcare, Little Chalfont, England) and both forward and reverse strands were sequenced on an ABI 3130 XL using Big Dye Terminator v3.1 (Applied Biosystems, Foster City, CA, USA).

If species-specific primers prove to give reliable results, additional primers

would be designed and cost effectiveness of the protocol ameliorate by multiplexing primers and by species identification by means of amplicon length, thus avoiding sequencing.

3.3.5.3 DNA sequencing: group-specific primers

With this approach, we targeted tick host DNA by means of host group specific primers. A first selection of the most important ticks hosts in the investigated area was done, taking into account the availability of genetic resources in public databases, such as GenBank. Clustal X v. 2.0 was used to create sequences alignment of different mtDNA regions of the chosen host species groups (*cytochrome b*, *d-loop*, *cytochrome oxidase I*, *16S rDNA*). Primers design was performed using PRIMER3 (www.frodo.wi.mit.edu/primer3).

Each primer set was tested on control samples from engorged ticks or from host tissue of the target species of the primer set. Reactions was performed with AmpliTaq Gold® DNA Polymerase (Applied Biosystem, Foster City, CA, USA) in 20 µL final volume mix containing reagents according to the optimized conditions for each primer set reported in Table 3.1.

Thermal cycling was performed on a Veriti® Thermal Cycler (Applied Biosystem, Foster City, CA, USA), and consisted of an initiation step at 95 °C for 10 min; denaturation at 95 °C for 30 sec, annealing at primer set specific T_a °C (see Table 4.9) for 20 sec and elongation for 30 sec at 72 °C, 50 times; final elongation at 72 °C for 7 min. Optimized protocols were then applied to DNA extracted from questing nymphs in order to identify larval bloodmeal source. Amplification results were checked using capillary electrophoresis with the QIAxcel system (Qiagen, Valencia, CA, USA) using a

DNA High Resolution Cartridge and QX 15 bp-3 Kb size marker with the OM500 method and analyzed using QIAxcel ScreenGel 1.0.2.0. PCR products were purified with Exo-SAP-ITTM (GE Healthcare, Little Chalfont, England), and both forward and reverse strands were sequenced on an ABI 3130 XL using Big Dye Terminator v3.1 (Applied Biosystems, Foster City, CA, USA).

Table 3.1 Reaction mix for each host group primer set.

Reagents	**ROD**	**SOR**	**PAS**	**CAN**	**RUM**
AmpliTaq GOLD 360 Buffer 10x (X)	1	1	1	1	1
25 mM $MgCl_2$ (mM)	2.00	1.50	1.75	1.51	2.00
360 GC Enhancer (μL)	2	2	2	2	2
Primer F 10 ρmol/μL (μM)	0.5	0.5	0.5	0.5	0.75
Primer R 10 ρmol/μL	0.5	0.5	0.5	0.5	0.75
dNTP 10 mM (mM)	0.25	0.25	0.25	0.25	0.25
AmpliTaq GOLD 360 Polymerase (U)	1.25	1.25	1.25	1.25	1.25
DNA control samples (μL)	1	1	1	1	1
DNA questing ticks (μL)	3	3	3	3	3
total reaction mix (μL)	20	20	20	20	20

ROD= Rodents; SOR= Soricomorpha; PAS= Passeriformes; CAN= Canidae; RUM= Ruminants

3.3.5.4 Standardized semi-automated test for species-specific and group-specific primers

In order to test the newly designed primers described above using semi-automated procedures, the following protocols were performed as follows:

1. a DNA extraction was performed using spin-columns (see protocol 3.3.2.3 Questing nymphs) and the QIAcube robotic workstation from 94 questing nymphs collected in

the same day from a single site;

2. a PCR reaction was carried out in a 96 well-plate using the QIAgility robotic workstation (Qiagen, Inc., Valencia, CA, USA), with the Rodentia, Soricomorpha, Passeriformes, *Apodemus* (short and long amplicons) primers (see Tables 4.8 and 4.9); each sample was tested three times for each primer set and a control sample and negative PCR control was included in each reaction;

3. PCR amplification results were controlled with QIAxcel® (Qiagen, Inc., Valencia, CA, USA), DNA High Resolution Kit;

4. amplicons in the expected bp range were sequenced and BLASTn search confirmed species identity.

3.3.5.5 Real-time High Resolution Melting Analysis

In High Resolution Melting Analysis (HRMA), the choice of molecular markers is the most fundamental and challenging step. The use of universal primers amplifying a short fragment (about 110 bp) of vertebrate 12S mtDNA would have represented the best solution in terms of number of *I. ricinus* hosts targeted, possibly allowing the discovery of neglected host species in the investigated area. However, *in silico* analysis demonstrated that only the design of degenerate primers would allow us to target the wide range of tick hosts; moreover, HRMA was not able to reliably identify host species of the generated amplicons, because melting temperatures for the species-specific amplicons overlapped. Therefore, we decided to focus our attention to a restricted set of tick hosts and to group hosts such that each Real time HRMA reaction targeted a few species at a time. *In silico* trials allowed me to confirm that this approach could reliably identify host species. As presented in Fig. 3.2 (HRMA optimization workflow), primer

selection was organized in several steps, described in the following paragraphs. Every step was repeated multiple times in order to identify the highest performing primer sets in terms of number of targeted species and taxonomic level of host identifications for each one of the host groups. Then, the Real Time HRMA protocol was tested on field collected questing nymphs.

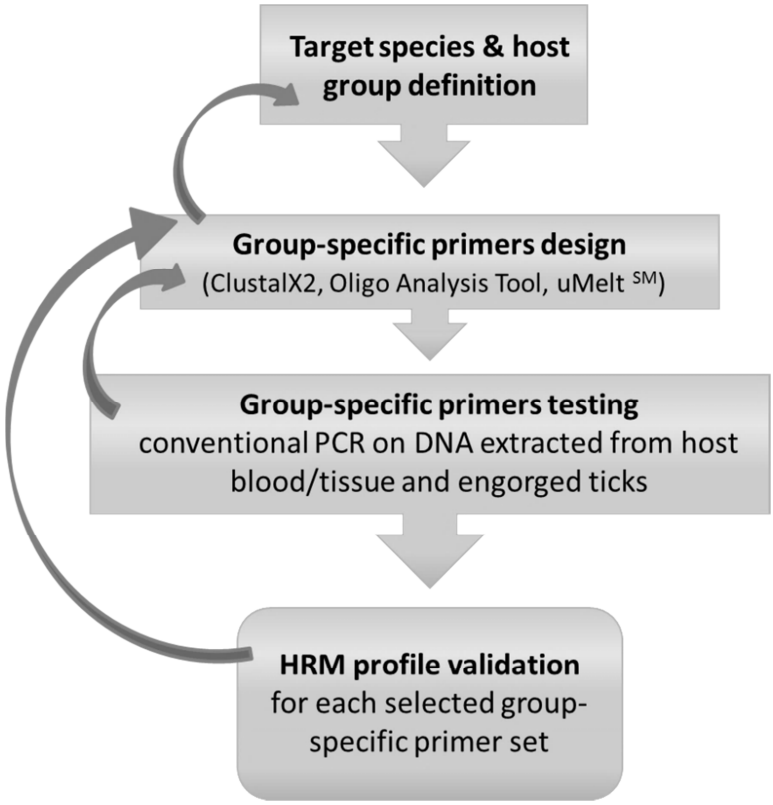

Fig. 3.2 Schematic optimization workflow followed during HRMA primers selection for bloodmeal analysis in questing ticks.

Target species and host group definition

Twenty of the most important vertebrate hosts of *I. ricinus* in Europe were chosen on the basis of their role as maintenance hosts, reservoir competence, occurrence in the Alps, and relevance as livestock and companion animals, as well as availability of control samples and GenBank sequences (Gray, 1998; Morán Cadenas et al., 2007; De

la Fuente et al., 2008; Bown et al., 2011; Marsot et al., 2012; Wodecka et al., 2014). Selected hosts were grouped in the following taxonomic categories:

1. Muroidae (Superfamily): *A. flavicollis*, *A. sylvaticus*, *M. glareolus*, *M. musculus*;

2. Soricidae (Family): *S. minutus*, *S. antinorii*, *S. araneus*, *C. leucodon*, *C. suaveolens*, *C. russula*;

3. Passeriformes (Order): *T. merula*, *T. philomelos*, *E. rubecula*;

4. Canidae (Family): *C. l. familiaris*, *V. vulpes*;

5. Caprinae (Subfamily): *O. aries*, *R. rupicapra*, *C. hircus*;

6. Cervidae (Family): *C. capreolus*, *C. elaphus*.

Humans were not considered a main tick host as suggested from previous studies (e.g. Humair et al., 2007; Pichon et al., 2003; Morán Cadenas et al., 2007; Wodecka et al., 2014). In any case, the risk of human contamination during collection was fairly high (a common problem when working with low quality/quantity DNA), and discriminating contaminant human DNA from that of the larval bloodmeal is not possible with the approach chosen (even after tick washing).

Group-specific primer design

Clustal X v. 2.0 was used to create a sequence alignment of the different mtDNA regions of the chosen host species. Sequences were mainly retrieved from GenBank where possible, but for some species for which no or insufficient sequences were available, sequences were generated by us from host tissues or engorged ticks, using universal 16S and 12S mtDNA primers (GenBank accession numbers: KJ676686 *T. merula*; KJ676687 *T. philomelos*; KJ676688 *E. rubecula*). In order to design primers that did not cross-react with DNA other than the target hosts, sequences of non-target

species and a *Homo sapiens* reference sequence were included in the alignment. In fact, avoiding the amplification of contaminant human DNA when working with extremely low quality/quantity DNA using highly sensitive methods is not easily achieved (Pääbo et al., 2004), even if both field sampling and subsequent handling of ticks are carried out using all possible precautions to avoid contamination. Alignments were checked visually to identify DNA regions that would optimize the discriminating power of HRMA; i.e. highly conserved intraspecific mtDNA regions of about 200 bp with well-defined interspecific variation (at least two single nucleotide polymorphisms, SNPs; see alignment in Appendix 1). Non-degenerate group-specific primers were designed to have melting temperatures (T_m) of about 60 °C and a low probability of dimer formation as predicted by the Oligo Analysis Tool (www.operon.com/tools/oligo-analysis-tool.aspx). Identified mtDNA regions were tested for their species identification potential in HRMA with uMELT^{TM} (https://dna.utah.edu/umelt/um.php; Dwight et al., 2011) using standard parameters and the thermodynamic parameter set of Unified-SantaLucia (SantaLucia, 1998). Finally, each selected sequence was blasted (BLASTn; http://blast.ncbi.nlm.nih.gov/Blast.cgi) to test if the mtDNA fragment would allow unequivocal species identification by sequencing.

Group-specific primers testing

In order to optimize thermal cycling conditions, each primer set was tested on control DNA samples of the target species (see 3.3.2.1, and Table 4.10). In order to verify that non-target DNA amplification did not occur, control samples of some non-target species were included in the conventional PCR of Muroidea (*C. capreolus, C. l. familiaris, S. antinorii*), Soricidae (*C. capreolus, C. l. familiaris, M. glareolus*),

Passeriformes (*C. capreolus, A. flavicollis*) and Caprinae (*C. capreolus, A. flavicollis*). Moreover, each group-specific primer set was tested using conventional PCR on three human DNA templates extracted from a partially engorged nymph collected while feeding, whole human blood, and human hair, including a negative control of the extraction and positive controls for each primer set (see Fig. 4.4 and 4.5 in Results for additional details).

Conventional PCR was performed at a final volume of 20 µL, containing 0.5 µM of each primer, 0.25 mM of each dNTP, 1x HotMaster Taq Buffer, 1.25 U HotMaster Taq (5-Prime), and 1 µL of template DNA. The thermal cycling consisted of 94 °C for 2 min; 40 cycles at 94 °C for 30 s, T_a (°C) of the group-specific primer set (Table 4.10) for 30 s, 65 °C for 1 min; 65 °C for 10 min, and was performed in a Veriti® Thermal Cycler (Applied Biosystems, Foster City, CA, USA). Amplification results were checked via capillary electrophoresis on a QIAxcel system (Qiagen, Valencia, CA, USA) with a DNA High Resolution Cartridge and QX 15 bp-3 Kb size marker using the OM500 method, and were analyzed using QIAxcel ScreenGel 1.0.2.0. At least one PCR product per group-specific primer set was purified with Exo-SAP-ITTM (GE Healthcare, Little Chalfont, England) and both forward and reverse strands were sequenced on an ABI 3130 XL using Big Dye Terminator v3.1 (Applied Biosystems, Foster City, CA, USA); after the electropherogram check and the creation of a consensus sequence using Sequencher v. 5.1, a BLASTn search was carried out to verify the amplification target.

HRMA validation

Real-time PCR coupled with HRMA was conducted on an ECO™ Real-Time PCR machine (Illumina®, San Diego, USA) twice for each sample at a final volume of 15 µL, containing 0.3 µM of each primer, 1x SsoFast™ EvaGreen® Supermix (Bio-Rad, Hercules, CA), and 3 µL of genomic DNA. Thermal cycling conditions were 95 °C for 5 min, 50 cycles at 95 °C for 15 s and T_a (°C) of the group-specific primer set (Table 4.10) for 15 s, directly followed by HRMA carried out at 95 °C for 15 s, 55 °C for 15 s, then by an increase of temperature from 55 °C to 95 °C, and 95 °C for 15 s; fluorescence data was collected every 0.1 °C. HRMA was performed using ECOTM v. 4.0 (Illumina®, San Diego, USA). Raw fluorescence plots were normalized by setting pre- and post-melting regions to 100% and 0%, respectively as in standard HRMA. Both normalized and derivative graphs were analysed for melting temperatures (T_m °C) and melting profile shapes. HRMA of control samples was used to define the minimum and maximum T_m for each host species. Note that the observed T_m was generally 2-6 °C lower than that predicted by uMELT™ (Table 4.10).

HRMA testing on questing ticks

The selected group-specific primers and the Real time HRMA protocol were tested on DNA from questing nymphs (see section 3.3.2.3 Questing nymphs - *Spin-column protocol*). Each tick was tested twice for each one of the group-specific primer sets, with minor modifications (i.e. 55 cycles of amplification was needed for the Muroidea and Soricidae primers). In each Real-time HRMA reaction, one positive control for each target species and one negative control were included. Normalized and derivative HRMA plots were obtained using ECOTM v. 4.0 (Illumina®, San Diego,

USA). Amplicons from questing ticks were assigned to species or genera by visually matching their melting patterns (T_m, melting curve shape, number of melting peaks) to those of control samples (see Results for examples). To verify the accuracy of HRMA in identifying the host species providing the larval bloodmeal, the amplicons of all amplified samples were sequenced, both with reverse and forward primers. Amplicons with T_m and melting curve profiles divergent from those of control samples were also sequenced. Using Sequencher v. 5.1 (Gene Codes Corporation, Ann Arbor, USA), consensus sequences were created, visually checked and then aligned in the corresponding group-specific alignment; in addition, a BLASTn search was carried out to confirm species identity.

3.3.5.6 Application: Real-time HRMA to bloodmeal analysis

Once the sensitivity and specificity of the newly designed Real time HRMA for bloodmeal analysis in questing nymphs was proven (see Results), the protocol was applied to questing nymphs collected in the remaining PAT sampling sites, at the same time introducing more automation into the protocol in order to try to save time and lower cost.

Nymph DNA extraction was performed as described above (see 3.3.2.3 Questing nymphs - *Magnetic-beads protocol*). A QIAgility robotic workstation (Qiagen, Valencia, CA, USA) was used for the automated high precision reaction setup. Real time HRMA was performed in a Rotor-Gene™ 6000 real time rotary analyzer (Corbett Life Science) with a 72-well rotor, as reported above, with the following modifications: initiation step 95 °C for 5 min, 55 annealing and elongation cycles at 95 °C for 15 s and T_a (°C) of the group-specific primer set for 15 s, directly followed by HRM with a pre-

melt conditioning step of 70 °C for 90 s, then increasing the temperature from 70 °C to 95 °C, by 0.2 °C steps for 2 seconds each. The Rotor-Gene 6000 Series Software 1.7 was used to analyse amplifications and HRM results by means of both normalized and derivative melting profile shapes and melting temperatures (T_m °C). Initially, amplifications were sequenced to verify their identity; then, following the procedure described in Collini et al. (2015 in press), only amplicons with aspecific melting properties were additionally investigated via capillary electrophoresis and/or sequence analysis. Capillary electrophoresis was carried out using the QIAxcel system (Qiagen, Valencia, CA, USA) with a DNA High Resolution Cartridge and the QX 15 bp-3 Kb size marker, method OM500; results were analysed with QIAxcel ScreenGel 1.0.2.0. For sequencing, Real time PCR products were purified with Exo-SAP-IT™ (GE Healthcare, Little Chalfont, England); both forward and reverse strands were sequenced on an ABI 3130 XL using Big Dye Terminator v3.1 (Applied Biosystems, Foster City, CA, USA). Raw sequences were checked and a consensus sequence created using the software Sequencher v 5.1; a BLASTn search (http://blast.ncbi.nlm.nih.gov/Blast.cgi) was carried out to verify species identity of each amplicon.

3.3.5.7 Statistical analysis

The results of bloodmeal analysis of questing nymphs obtained from Real-time HRMA using DNA extracted and amplified during both the optimization and application phases of protocol development were pooled in order to have a comprehensive analysis of the feeding behavior of *I. ricinus* larvae in PAT. Statistical analyses were performed with R v 3.1.0 (The R Foundation for Statistical Computing, 2014) in collaboration with Roberto Rosà (DBEM, Animal Ecology research group).

Identified hosts were grouped according to taxonomic orders (Rodentia, Soricomorpha, Passeriformes, Carnivora, Cetartiodactyla). A chi-square test was used to compare EXTF and PATF sites by means of the proportion of identified bloodmeals in the different host groups. A Linear Model was used to assess variation of the number of ticks with identified bloodmeal (identification success) in relation to the explanatory variables DNA extraction method, sampling year, sampling month and habitat type.

3.3.6 Population genetics

3.3.6.1 Microsatellite amplification

At the outset of the project, microsatellites or short tandem repeats (STR) were believed to be the appropriate molecular markers for the genotyping of feeding and questing ticks, because of their wide use and apparently good resolution in population genetics (Beaumont and Brufford, 1999); in addition, previous studies on *I. ricinus* and other Ixodidae, made use of microsatellites (Delaye et al., 1998; McCoy and Tirard, 2000; Fagerberg et al., 2001; Røed et al., 2006; De Meeûs et al., 2002; Kempf et al., 2009b), as presented in the General introduction. Therefore, ten microsatellite (STR) markers for *I. ricinus* were selected among those described by Røed et al. (2006) and Delaye et al. (1998) (see Table 3.2). According to reported data, they lack Linkage Disequilibrium and deviation from Hardy-Weimberg equilibrium. Subsequently, 4 STR markers described for *I. scapularis* by Fagerberg et al. (2001) were tested. Each locus was tested singly and optimization of PCR conditions was performed on template DNA from three questing adult ticks from different sampling sites. Finally, PCR was performed at a final volume of 20 µL, containing 0.1 µM of each primer, 0.25 mM of each dNTP, 1x HotMaster Taq Buffer, 1.25 U HotMaster Taq (5-Prime) and 1 µL of template DNA. Amplification was performed in a Veriti® Thermal Cycler (Applied Biosystems, Foster City, CA, USA); thermal cycling consisted of 94 °C for 2 min; 30-40 cycles at 94 °C for 15 s, Ta (°C) of the STR marker (Table 3.2) for 15 s, 65 °C for 45 sec; 65 °C for 10 min. 1 µL of PCR product was diluted in 18 µL formamide (Hi-Ti Applied Biosystem) with 0.6% GS500LIZ size standard (Applied Biosystems, Foster City, CA, USA) to be visualized on a ABI3130 sequencer (Applied Biosystems, Foster City, CA, USA); microsatellites scoring was performed using the software GeneMapper

v 3.7 (Applied Biosystem, Foster City, CA, USA).

Table 3.2 Features of selected microsatellites loci.

STR	Repeat motif	Primer sequences and fluorescence label	T_a (°C)	Size range
IRN-4[a]	$(CA)_{14}$	F: NED_GCCATTTTATGTGCCGTTTT R: CTTTGAGTGCGTGCGTGT	54	148-168
IRN-7[a]	$(CA)_{13}$	F: PET_CGGATGATCAATAGTCGATTCC R: CCTAGTCACAAACTCTACCAAGTTA	52	85-101
IRN-8[a]	$(CA)_{15}$	F: PET_CGCTTCGAAGACGACTAAACA R: TGCGAACAATGACAAACAGA	50	169-179
IRN-12[a]	$(GT)_{14}$	F: FAM_GACAAAGGCTGTCAAAGGCTGCATCATA R: CGAGGAAGCCACGACTTGCAGAACTATT	50	151-225
IRN-17[a]	$(GT)_7AT(GT)_{12}$	F: VIC_CATGAGTGTTATATTCGCATTT R: GCTATTACGTCGACGATTTT	55	180-212
IRN-28[a]	$(CA)_{23}$	F: VIC_AGCCACGCTAGTTCTGAGA R: CCTGTTGTGTTTTGTTGGTC	56	100-128
IRN-30[a]	$(GT)_6(GC)_4(GT)_3\ldots(GT)_6$	F: FAM_GCAATTGCTATTCTTTGT R: AGTCTACTAAATCGTCACCA	45	101-109
IRN-37[a]	$(GT)_{19}TT(GT)_2CT(GT)_2$	F: NED_CGGGGCGTTTTTCTTTATTCT R: GAAGCGTCAGACTCCGTAACAG	49	96-122
IR32[b]	$(AG)_{12}$	F: FAM_TCGACAAGTGCAGTGGAGAC R: GTTTCCTACCACAGATTCTCC	61	233-250
IR39[b]	$(AG)_9$	F: PET_ATACCCGTAGAACGAGAG R: GTTTTTCAAGATTTCCGCC	59	121-149
IsAC4[c]	AC	F: AAGCGTATCCGATTTGCCCTTCAT R: GGGTCCCAACGATTGCTAAACCAG	var	n.a.
IsAC8[c]	AC	F: GAGCTACCCCTTTCATCGTCTTCG R: TCTTCCCGCTGCTGTCTCGTATTC	var	n.a.
IsAG25[c]	AG	F: AAATGTCCGAACAGCCTTAT R: GCCCTTGAGTCTACCCACTA	var	n.a.
IsGATA4[c]	GATA	F: CAGACAATGTCATTCAATCGCA R: CGCACAATGCAAAACAAATCTA	var	n.a.

[a]Røed et al., 2006.

[b]Delaye et al., 1998.

[c]Fagerberg et al., 2001; this primer has been tested unlabelled.

3.3.6.2. RAD-Seq library preparation

As reported in Results, STR proved not to be reliable markers for the study of *I. ricinus* population genetics and consequently we moved to SNPs genotyping. As no reference SNPs database is available for *I. ricinus*, we decided to use the recently developed Restriction site Associated DNA sequencing, NGS approach, to *de novo* identify SNPs loci and, at the same time, genotype *I. ricinus* ticks of PAT population (see General introduction).

Restriction enzyme and RAD-Seq strategy choice

To choose the most appropriate RAD-sequencing strategy in terms of restriction enzyme, coverage and number of samples per lane, we used *radcounter v4* (GenePool, Edinburgh). Our target was to paired-end sequencing about 10 individual for each of the 30 sampling sites in PAT (Figure 4.1), for a total of 300 genotyped ticks. Since there is no reference genome currently available for *I. ricinus*, genome data of *Ixodes scapularis* (Geraci et al., 2007) was used with the software, with following sequencing parameters:

- Genome size: 2300 Mbp
- Mean GC%= 45%
- Sequencing technology: HiqSeq2000, 130 million reads per lane;
- plexity (individuals per lane): 46
- coverage: 40x.

R*adcounter v4* analysis suggested *SbfI* as the most appropriate restriction enzyme. It is a non-frequent cutter that identifies the palindromic sequence CCTGCA*GG as the restriction site; it was estimated to be present about 10 times every Mb in the *I. scapularis* genome and therefore it was expected to produce 45136 tags. In such a

RAD-Seq setting, the maximum number of pooled individuals could be 72.

Library preparation

I optimized the RAD-Seq library generation protocol of Etter et al. (2011) for *I. ricinus* at the RAD-TAG platform of the Centre for Ecological and Evolutionary Synthesis (CEES), Oslo University, in collaboration with Emiliano Trucchi in 2012. It was then transferred to the Animal Genetics Laboratory of DBEM-FEM to complete *I. ricinus* genotyping.

Feeding ticks were not used for the RAD-Seq experiment since host DNA would have been over-represented in comparison to *I. ricinus* DNA, reducing RAD-loci coverage; furthermore, bioinformatics analysis would have been extremely complex, since few reference genomes are present for wild host species and none for *I. ricinus*, so that non-tick sequences could not be removed from the dataset. Only adults questing ticks provided to have enough high quality genomic DNA for the application of individual RAD sequencing. However, long trials have been conducted with the different DNA extraction methods described in section 3.3.2.2 in order to select the most efficient, as briefly described in Results. After quality assessment, samples showing high molecular weight and highly concentrated DNA were employed in NGS of RAD-tags.

Briefly, 250 ng of genomic DNA per sample were used, instead of the suggested 1 µg, in the enzymatic *SbfI* digestion (SbfI-HF™, New England BioLabs®). Decreasing gDNA input allowed to use ticks samples for which the DNA concentration was as low as 10 ng/µL. Forty-six Illumina P1 adapters and the P2 adapter were provided us by the RAD-TAG Platform (CEES, University of Oslo). Illumina P1 barcoded adapters were

ligated to digested samples. The 5-nucleotide barcode allows sample identification in bioinformatics data processing and each one differs by at least 2 bp, to limit erroneous sample assignment due to sequencing error. Once barcoded, samples were pooled in libraries of multiple individuals. Pooled samples were sheared by sonication (Bioruptor® Plus, Diagenode) to a mid-size range of 300-600 bp and, after concentration to 25 µL using the QIAquick PCR purification Kit (Qiagen, Valencia, CA, USA), they were size selected on a 1% agarose gel stained with ethidium bromide. Purification steps after the P2 adapter ligation were performed using DNA capture on magnetic beads following the manufacturer's instructions (Agencourt AMPure XP, Beckman Coulter, MA, USA) instead of the standard Qiagen column based purification, at the following ratio *beads solution : DNA = 0.8 : 1*, in an attempt to improve the quantity of retrieved DNA and reduce the carry-over of un-ligated adapters. PCR amplification was performed in 8 x 12.5 µL aliquots, using 16 µL of library template, and setting 21 to 23 replication cycles. To obtain the enriched library concentration requested for sequencing (10 µL library at 10 ng/µL), it was necessary to perform multiple distinct amplifications for each library and combine them before the final concentration step. Amplification product was quantified using the fluorometric-method (Qubit® 2.0), after concentration with magnetic beads, as above, the 260/280 ratio was measured with Nanodrop 8000 UV-Vis Spectrophotometers (Thermo Scientific).

A preliminary library was prepared while optimizing the protocol at the CEES, Oslo and sequenced in November 2012; forty-two adult questing ticks, coming from five different sampling sites and extracted with three different methods (spin-column; phenol protocol and magnetic-beads), were sequenced in a single lane (details in Table 3.3) on a Illumina HiSeq2000 at the Norwegian Sequencing Center, University of Oslo.

In 2014 at FEM, twelve libraries were prepared from 256 adult questing ticks from PAT questing adults, two females from Finland were added to the analysis; males and females, randomly chosen by sampling site and gDNA concentration, were processed in separate libraries, as for the two sex gDNA concentration was significantly different. As two libraries of male individuals did not give good amplification results, probably as a result of low quality/quantity DNA (erroneous gDNA quantification), only 10 libraries were sent for sequencing (2 males and 8 females libraries). Those were pooled two by two in equimolar ratio prior to sequencing, in order to maintain proportional RAD-loci representation for each individual (details in Table 3.4; Fig. 3.3). Paired-end sequencing was performed on a Illumina HiSeq2000 at the Norwegian Sequencing Center, University of Oslo.

Table 3.3 Samples sequenced in the preliminary 2012 RAD-Seq library. Sampling site, sex (F=female; M=male), DNA extraction method used (A=spin-column; B=phenol protocol; and C=magnetic-beads) are reported, as well as the final concentration (ng/ µl) of the enriched library sent for sequencing at the Norwegian Sequencing Center, University of Oslo.

	Samples		**DNA extraction**				**Concentration**
Site	**F**	**M**	**A**	**B**	**C**	**Total**	**(ng/µL)**
CAV	5	3		8		8	
LAM	7	2	6		3	9	
CON	7	2	5		4	9	
REV	6	1		1	6	7	
CVS	7	2	1		8	9	
			Final library			42	14.0

Table 3.4 Wet lab results and sample composition (number of individuals and sex) for the 2014 RAD-Seq libraries. Sex (F=female; M=male), final concentration (ng/ µl) and 260/280 ratios are reported.

Lane	Library name	n° ind	sex	concentration (ng/µl)	ratio 260/280
1	LIB-1	21	F	10.3	-
	LIB-12	22	F	21.8	1.78
2	LIB-2	21	F	9.6	-
	LIB-3	21	F	12.6	-
3	LIB-4	22	F	22.0	1.87
	LIB-5	21	M	10.0	1.75
4	LIB-7	22	F	10.1	-
	LIB-8	22	F	10.9	1.74
5	LIB-10	21	M	10.2	-
	LIB-11	22	F	12.2	1.85

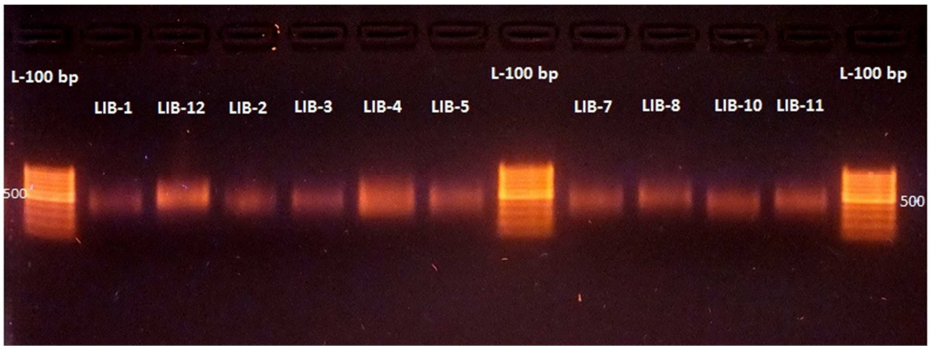

Fig. 3.3 Photograph of the gel electrophoresis of the final 2014 RAD-Seq libraries after amplification and concentration. Medium library size is 500 bp.

3.3.6.3 Bioinformatic analysis

Preliminary de novo genotyping analysis on first sequenced lane

Stacks v. 0.99997 (Catchen et al., 2011, 2013) was used for analysis of the preliminary library of 42 individuals, to investigate the efficiency of the different methods used for DNA extraction. Using the *process_radtags* program, raw data were demultiplexed, according to assigned individual barcodes and quality filtered setting a

minimum Phred quality score of 10 (90% of base call accuracy) in a sliding window of 15% of the sequence. No trimming of the sequences was set at this point. Sequences with no scored nucleotides (N), ambiguous nucleotides in the barcode region or in the restriction site were automatically discarded. As a standard in *de novo* mapping and genotyping RAD-Seq experiment, only first reads (the ones originated at the restriction site and containing the barcode R1), were used and fed in the pipeline *denovo_map.pl;* SNP identification and definition of diploid individual genotypes at each nucleotide position as performed in a maximum likelihood statistical framework (Hohenlohe et al., 2010). Briefly, the pipeline first identifies "*stacks*", identical sequences from a single individual, and then searches for the same stack in all other individuals and merges them to form a locus. In this putative locus, genetic variation is examined and genotype called (see also General Introduction Fig. 1.10). Parameters for the assembly were set as follows: minimum stack depth 5 (-m), maximum distance between stack 3 (-M), maximum distance to align secondary reads 5 (-N) and maximum number of stacks allowed for *de novo* locus formation 3 (default). The genetic variability model type was set to SNP and the alpha significance level to call a SNP was left as default (0.05). Deleveraging algorithm (-t), allowing identification of loci deriving from repeat region and PCR artifact, was enabled. For parameter definitions and effects on *de novo* formation of *stacks* and loci, please refer to http://creskolab.uoregon.edu/stacks/param_tut.php.

Full data set de novo *genotyping analysis*

As no reference genome for *I. ricinus* is available, *de novo* RAD-Seq loci building and SNP calling was performed; at first following the standard procedure (A.)

described in (http://creskolab.uoregon.edu/stacks/) and then, once we had identified some anomalies in the RAD-Seq sequence dataset (i.e. PRC clone; low number of RAD-loci; see Results), more conservative approaches were used (B.)

A. De novo *SNP calling with Stacks; no PCR duplicates removal*

Raw sequences were trimmed to 80 bp using *fastx_trimmer* (FASTX-Toolkit v 0.0.13.2) and then demultiplexed and quality filtered, setting a minimum Phred quality score of 20 (99% base call accuracy), using the *process_radtags* command of Stacks v. 0.99997 (Catchen et al., 2011); sequences with uncalled nucleotides and more than one ambiguous nucleotide in the barcode or in the restriction site sequence were discarded (Catchen et al., 2011, 2013). Simultaneous identification of SNPs and individual genotyping was performed without a reference genome using the *denovo_map.pl* pipeline of Stacks v 1.12. Parameters for the *denovo_map.pl* pipeline were set as follows: minimum stack depth 5 (–m), maximum distance between stack or within individual distance 2 (-M), maximum number of mismatches between loci when building the catalog 2 (–n) and maximum number of stacks allowed for locus 2 (-X "ustacks:--max_locus_stacks"). SNP calling from secondary reads was disabled (-H). The genetic variability model type was set to SNP and the alpha significance level to call a SNP was left as default (0.05). Deleveraging algorithm, allowing the identification of loci deriving from the repeat region and PCR artifact was enabled (-t). The pipelines *export.pl* or *population.pl* were used to export the genotyping results in different file formats.

B. De novo *SNP calling with Stacks; PCR duplicates removal*

According to investigative analysis showing an extremely high rate of PCR clones per individual in the sequence data (>85%, expected 20-50%; see Results), we

decided to remove these duplicate as they may introduce bias in the SNP calling process.

Raw reads were trimmed to 90 bp using *fastx_trimmer* command line of FASTX-Toolkit v 0.0.13.2 and then demultiplexed and cleaned using *process_radtags.pl* program of *Stacks* (v 0.99997, Catchen et al., 2011); sequences with a medium Phred Quality score under 20 (99% of base call accuracy) in a sliding windows of 15% of the sequence itself, uncalled nucleotides and more than one ambiguous nucleotide in the barcode or in the restriction site sequence were discarded (Catchen et al., 2011). Individual paired-end reads were cleaned of PCR-duplicates with c*lone_filter* program. It identify a PCR clone (PCR duplicate) as a pair of reads that match exactly, because a paired-end reads from two different DNA molecules will nearly always be slightly different in length. Individual loci and SNPs were called *de novo* with a maximum-likelihood function by feeding R1 reads in the *denovo_map.pl* pipeline of *Stacks* v 1.12. Parameters for the assembly and SNP calling were set as following: minimum stack depth 2 (–m), maximum distance between stack or within individual distance 2 (–M), maximum number of mismatches between loci when building the catalog 2 (–n) and maximum number of stacks allowed for locus 2. Calling of SNP from secondary reads was disabled (-H). The genetic variability model type was set to SNP and the alpha significance level to call a SNP was left as default to 0.05. A deleveraging algorithm, allowing the identification of loci deriving from repeat regions and PCR artifacts, was enabled. A table including all loci having at least 1 SNP was built using *export_sql.pl* in *Stacks*. Genotype outputs were generated in *.plink* and *.vcf* formats using *population* in Stacks.

As a preliminary analysis showed a steep increase in the number of SNPs

identified from the 80th bp on (Fig. 3.4), only SNPs identified up to the 79th bp were retained in the following analyses.

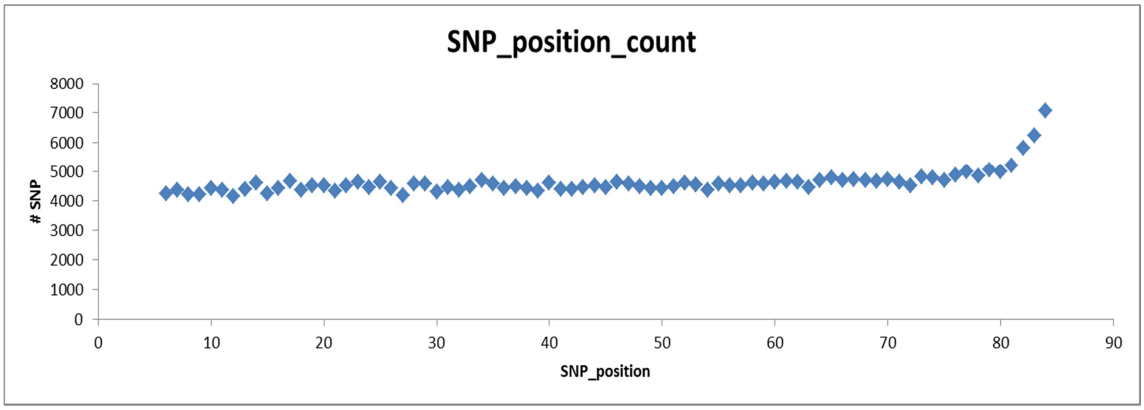

Fig. 3.4 Number of SNP per nucleotide position. Note that there is a significant increase in number of SNPs in the last five nucleotides, suggestive of sequencing errors.

C. SNP calling with alignment to reference genome

The following analyses were carried out in collaboration with the Department of Life Sciences and Biotechnology, University of Ferrara, in the research group led by Giorgio Bertorelle, with the FEM-UNIFE co-funded PhD student Alex Panziera (who is developing and testing analyses of RAD-tag data) and the bioinformatician Andrea Benazzo.

Given the peculiarities of the starting raw data (i.e. high level of PCR duplicates; low number of identified stacks; low genome coverage; see Results), a SNP calling approach with alignment against a reference genome was attempted, in order to obtain more robust SNPs and a higher quality genotyping dataset. Since there is no reference genome for *I. ricinus*, I aligned *I. ricinus* sequences against that of *I. scapularis* (even though *I. scapularis* is not considered a sister species of *I. ricinus*, it is the most similar genome available), thanks to software specifically developed to deal with genome alignment of distantly related species. The IscaW1 (GenBank accession number:

ABJB000000000.1; assembly version: 2008) reference genome is composed of 369492 scaffolds, totalling 1.76 Gb, with a supercontig N50 size of 72 Kb. In the alignment process, repeat regions were masked and only contigs with a length higher than 10k bp were used (17 000 contigs retained out of 369492), by filtering them with a specifically written script. For each *I. ricinus* individual, paired-end sequences were demultiplexed and cleaned (trimmed to 80 bp) and, both R1 and R2 were aligned against the filtered set of scaffolds of IscaW1 using *Stampy* (Lunter and Goodson, 2011), setting a mutation rate of 10%, estimated by aligning nuclear gene sequences from *I. scapularis* and *I. ricinus* retrieved from GenBank, and an insert size of 300 bp. Only sequences aligning with a quality score of 20 were retained for the following steps. After the indexing step of alignment files, performed with SAMtools (Li et al., 2009), PCR duplicates were filtered out of the dataset using *Picard* (http://broadinstitute.github.io/picard/). Finally, GATK (McKenna et al., 2010) was used to check alignment in correspondence of *indel* polymorphisms. For each individual only 0.01 – 0.50 % of reads were *not* PCR duplicates and resulted to be properly aligned. GATK was used to call and identify SNPs; only SNPs having a coverage of at least 4 reads and an alignment quality score higher than 20 were considered. Only a few usable SNPs were retrieved using this approach; therefore, a PCA was performed using genotype probabilities, instead of actual genotypes, computed with the software *ANGSD* (Korneliussen et al., 2014).

Population genetics analysis

VCFtools (Danecek et al., 2011) was used to filter full *.vcf* SNPs dataset. By means of individual missing rate, site missing rate and to convert *.vcf* files to *.plink*. PGDSpider 2.0.5.2 (Lischer and Excoffier, 2012) was used in file format conversions.

The PCA was computed with *smart_pca* of the EIGENSOFT package (Patterson et al., 2006). The *lsqproject* algorithm was always enabled, as suggested for datasets with a high proportion of missing data. Plots of the top two principal components were built with *ploteig* or within the Excel environment, according to different grouping strategies. Significance of the computed eigenvalues was assessed by running *twstat* that calculate Tracy-Widom statistics (Patterson et al., 2006). MultiDimensional Scaling (MDS) analysis was computed with R (The R Foundation for Statistical Computing, 2014), using a custom script wrote by Andrea Benazzo (University of Ferrara); plots were created in Excel. MDS is a robust analysis for genetic structure investigation, even for a dataset with a high rate of missing data.

PCA and MDS were computed on individual genotypes, but also on consensus genotypes built for each of the 30 PAT and the Finland one (which were added for comparison). The consensus genotypes were built from individual genotypes of the defined site; the following rule were followed: at SNP locus level, the more frequent genotype will be chosen, in case of missing genotype in all individuals from a site will also be missing also in the consensus in other sites, otherwise, the more frequent in the other individuals of the site will be chosen. Consensus genotypes were created with a custom script provided by Alex Panziera.

Building mini-contigs from paired-end sequences

Using the command lines *sort_read_pairs.pl* and *exec_velvet.pl* implemented in Stacks v 1.12 and following the procedure described at http://creskolab.uoregon.edu/stacks/pe_tut.php, mini-contigs were built from individual paired-end sequences (90 bp trimmed). Resulting contigs were blasted with default

parameters (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and MEGAN v 5.9.1 (MEtaGenome Analyzer, D. Huson, University of Tübingen) was used to visualize results; only sequences having a BLASTn e-value of 10^{-5} and a minimum score of 80 were retained in the analysis.