Population structure, adaptation and divergence of the meadow spittlebug, *Philaenus spumarius* (Hemiptera, Aphrophoridae), revealed by genomic and morphological data

Sofia Seabra [Corresp., 1], Ana SB Rodrigues [1], Sara E Silva [1], Ana Carina Neto [2], Francisco Pina-Martins [1], Eduardo Marabuto [1], Vinton Thompson [3], Michael R Wilson [4], Selçuk Yurtsever [5], Antti Halkka [6], Maria Teresa Rebelo [2], Paulo AV Borges [7], José A Quartau [1], Chris D Jiggins [8], Octávio S Paulo [1]

[1] cE3c - Centre for Ecology, Evolution and Environmental Changes, Departamento de Biologia Animal, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal

[2] Centre for Environmental and Marine Studies (CESAM), Departamento de Biologia Animal, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal

[3] American Museum of Natural History, New York, USA

[4] Department of Natural Sciences, National Museum of Wales, Cardiff, United Kingdom

[5] Biology Department, Science Faculty, Trakya University, Edirne, Turkey

[6] Department of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland

[7] cE3c - Centre for Ecology, Evolution and Environmental Changes/Azorean Biodiversity Group, Faculty of Agriculture and Environment, Department of Environmental Sciences and Engineering, Universidade dos Açores, Angra do Heroísmo, Açores, Portugal

[8] Department of Zoology, University of Cambridge, Cambridge, United Kingdom

Corresponding Author: Sofia Seabra
Email address: sgseabra@gmail.com

Understanding patterns of population differentiation and gene flow in insect vectors of plant diseases is crucial for the implementation of management programs of disease. We investigated morphological and genome-wide variation across the distribution range of the spittlebug *Philaenus spumarius* (Linnaeus, 1758) (Hemiptera, Auchenorrhyncha Aphrophoridae), a recently known vector of the plant pathogenic bacterium *Xylella fastidiosa* in Europe. We found genome-wide divergence between *P. spumarius* and a very closely related species, *P. tesselatus* Melichar, 1899, at RAD sequencing markers. The two species may be identified by the morphology of male genitalia but are not differentiated at mitochondrial COI, making DNA barcoding with this gene ineffective. We detected admixture between *P. tesselatus* from Morocco and *P. spumarius* from the Iberian Peninsula, suggesting gene-flow between them. Within *P. spumarius*, we found a pattern of isolation-by-distance in European populations, likely acting alongside other factors restricting gene flow. Varying levels of co-occurence of different lineages, showing heterogeneous levels of admixture, suggest other isolation mechanisms. The transatlantic populations of North America and Azores were genetically closer to the British population analysed here, suggesting an origin from North-Western Europe, as already detected with mitochondrial DNA. Nevertheless, these may have been produced through different

colonization events. We detected SNPs with signatures of positive selection associated with environmental variables, especially related to extremes and range variation in temperature and precipitation. The population genomics approach provided new insights into the patterns of divergence, gene flow and adaptation in these spittlebugs and led to several hypotheses that require further local investigation.

Population structure, adaptation and divergence of the meadow spittlebug, *Philaenus spumarius* (Hemiptera, Aphrophoridae), revealed by genomic and morphological data

Sofia G Seabra[1*], Ana SB Rodrigues[1], Sara E Silva[1]; Ana Carina Neto[2]; Francisco Pina-Martins[1]; Eduardo Marabuto[1]; Vinton Thompson[3]; Michael R. Wilson[4]; Selcuk Yurtsever[5]; Antti Halkka[6]; Maria Teresa Rebelo[2], Paulo AV Borges[7]; José A Quartau[1]; Chris D Jiggins[8]; Octávio S Paulo[1]

[1] cE3c - Centre for Ecology, Evolution and Environmental Changes, Departamento de Biologia Animal, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, Edifício C2, 1749-016 Lisboa, Portugal.

[2] Centre for Environmental and Marine Studies (CESAM), Departamento de Biologia Animal, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal.

[3] American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024, USA

[4] Department of Natural Sciences, National Museum of Wales, Cardiff, United Kingdom

[5] Biology Department, Science Faculty, Trakya University, Edirne, Turkey

[6] Department of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland

[7] cE3c - Centre for Ecology, Evolution and Environmental Changes/Azorean Biodiversity Group, Faculty of Agriculture and Environment, Department of Environmental Sciences and Engineering, Universidade dos Açores, 9700-042 Angra do Heroísmo, Açores, Portugal

[8] Department of Zoology, University of Cambridge, Cambridge, UK

* Present address: Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Rua da Junqueira nº 100, 1349-008 Lisboa, Portugal

Corresponding Author:
Sofia G. Seabra[1]
cE3c - Centre for Ecology, Evolution and Environmental Changes, Departamento de Biologia Animal, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, Edifício C2, 1749-016 Lisboa, Portugal.
Email address: sgseabra@gmail.com

Abstract

42 Understanding patterns of population differentiation and gene flow in insect vectors of plant
43 diseases is crucial for the implementation of management programs of disease. We investigated
44 morphological and genome-wide variation across the distribution range of the spittlebug
45 *Philaenus spumarius* (Linnaeus, 1758) (Hemiptera, Auchenorrhyncha Aphrophoridae), a
46 recently known vector of the plant pathogenic bacterium *Xylella fastidiosa* in Europe. We found
47 genome-wide divergence between *P. spumarius* and a very closely related species, *P. tesselatus*
48 Melichar, 1899, at RAD sequencing markers. The two species may be identified by the
49 morphology of male genitalia but are not differentiated at mitochondrial COI, making DNA
50 barcoding with this gene ineffective. We detected admixture between *P. tesselatus* from
51 Morocco and *P. spumarius* from the Iberian Peninsula, suggesting gene-flow between them.
52 Within *P. spumarius*, we found a pattern of isolation-by-distance in European populations, likely
53 acting alongside other factors restricting gene flow. Varying levels of co-occurence of different
54 lineages, showing heterogeneous levels of admixture, suggest other isolation mechanisms. The
55 transatlantic populations of North America and Azores were genetically closer to the British
56 population analysed here, suggesting an origin from North-Western Europe, as already detected
57 with mitochondrial DNA. Nevertheless, these may have been produced through different
58 colonization events. We detected SNPs with signatures of positive selection associated with
59 environmental variables, especially related to extremes and range variation in temperature and
60 precipitation. The population genomics approach provided new insights into the patterns of
61 divergence, gene flow and adaptation in these spittlebugs and led to several hypotheses that
62 require further local investigation.

63

Introduction

65 Speciation involves the evolution of reproductive isolation and the buildup of genetic
66 differentiation through selection and drift, but gene flow can counteract such divergence by
67 homogenizing allelic variation and also by allowing recombination to oppose or break
68 associations between loci underlying isolating traits (Smadja & Butlin, 2011; Sousa & Hey,
69 2013). However, several mechanisms may favour divergence in the face of gene flow, such as
70 ecologically driven selection or sexual selection (Smadja & Butlin, 2011; Nosil, 2012).
71 According to the genic model of speciation, at the start of the speciation process, and in the
72 presence of gene flow, a few localized regions in the genome being subject to divergent selection
73 will differentiate, while the remaining genome continues to be freely exchanged between
74 populations (Wu, 2001). Genome-wide analyses have allowed the detection of these "genomic
75 islands" of differentiation in several systems (e.g., Malinsky et al., 2015; Vijay et al., 2017),
76 although other processes not related to speciation or reproductive isolation may also be
77 responsible for them, such as linked selection, variable recombination rates and/or density of
78 targets of selection (Wolf & Ellegren, 2017). As populations diverge through the action of

79 selection and drift, a genome-wide differentiation will emerge and eventually lead to full
80 reproductive isolation and diversification. Designated species may thus lie somewhere in this
81 'speciation continuum' with different levels of divergence and gene flow (Hendry, Bolnick,
82 Berner, & Peichel, 2009; Peccoud, Ollivier, Plantegenest, & Simon, 2009; Renaut et al., 2012;
83 Riesch et al., 2017).

84

85 Distinguishing taxa and understanding the patterns of gene flow and local adaptation in insect
86 species that transmit diseases is crucial for a better management of those diseases (Busvine,
87 1980; Pélissié, Crossley, Cohen, & Schoville, 2018; Bahrndorff et al., 2020). *Philaenus*
88 *spumarius* (Linnaeus, 1758) (Insecta, Hemiptera, Auchenorrhyncha, Aphrophoridae), the
89 meadow spittlebug is a xylem-feeding vector of *Xylella fastidiosa* Wells et al. 1987, a plant
90 pathogenic bacterium of South American origin that is emergent in Europe (Saponari et al., 2014).
91 The olive quick decline syndrome (OQDS), caused by *X. fastidiosa*, was first detected in Apulia,
92 southern Italy in 2013 (Saponari et al., 2014; Cornara et al., 2017). Since then, *X. fastidiosa* has
93 already been detected in several other European countries and is a cause of major concern (EFSA
94 et al., 2019). *X. fastidiosa* is native to the Americas, where it causes important diseases such as
95 Pierce's disease of grapevine, citrus variegated chlorosis, almond leaf scorch and several others
96 in perennial crops and ornamental plants (Baldi & La Porta, 2017). There the main vectors are
97 sharpshooters (another xylem-feeding Auchenorrhyncha group, the Cicadellidae Cicadellinae),
98 while spittlebugs have minimal epidemiological importance (Almeida et al., 2019; Cornara et al.,
99 2019). One of the main vectors of Pierce's disease of grapevines in California is the glassy-
100 winged sharpshooter *Homalodisca coagulata* (Say). It is native to the southern United States and
101 it became established in late 1990´s in California, being a costly invasive species to agriculture.
102 Population genetic structure studies based on DNA fingerprinting and mitochondrial DNA on
103 this species have revealed highly differentiated geographic groups in the natural range and
104 indicated the likely sources of the California insects were in Texas (Léon, Jones & Morgan,
105 2004; Smith, 2005). This result has led, for example, to further work in Texas for a better
106 understanding of the natural population dynamics (Yoon et al., 2014).

107

108 Potential movements and dispersal of *P. spumarius* should be included in models of risk
109 assessment for *X. fastidiosa* spread in Europe (EFSA Panel on Plant Health, 2015) for which data
110 on population genetic structure of this insect will be of paramount importance. Since the vectors
111 are the only means of natural dissemination of *X. fastidiosa* (Sicard et al., 2018), this information
112 is crucial for the successful management of this pathogen.

113

114 *P. spumarius* is a polyphagous xylem-feeding insect, widespread in the Holarctic, whose nymphs
115 produce a protective foam (spittle masses) from a liquid excretion. Humidity and temperature
116 are particularly limiting in the earlier nymphal stages (Weaver & King, 1954). This species is

117   thought to have a Palaearctic origin having more recently colonised North America, Azorean
118   islands, Hawaii and New Zealand, probably through accidental man-mediated introductions
119   (Halkka & Halkka, 1990; Rodrigues et al., 2014), although natural colonisation cannot be
120   excluded for the S. Miguel island in the Azores (Borges et al., 2018; Rodrigues et al., 2014) as
121   the populations in this island are restricted to native vegetation. In North America it was reported
122   as a crop pest (Weaver & King, 1954), but surveys of the spittle masses across coastal California
123   have revealed a recent population decline of this species, very accentuated in some places
124   (Karban & Strauss, 2004; Karban & Huntzinger, 2018), including the Wonalancet, new
125   Hampshire population sampled for this report (VT unpublished observations).

126

127   Previous studies based on mitochondrial and nuclear DNA genes have revealed the major
128   phylogeographic patterns in *Philaenus spumarius* (Maryańska-Nadachowska, Kajtoch, &
129   Lachowska, 2011; Rodrigues et al., 2014). Two main mitochondrial lineages have initially
130   diverged during the Pleistocene: the "Western", currently found in the Mediterranean region and
131   also in Central and Northern Europe, and the "North-Eastern", currently found from Eastern Asia
132   to Central and Northern Europe. The "Western" lineage is further differentiated into sublineages:
133   the ''Western'', predominant in the Iberian Peninsula but also in western parts of Central and
134   Northern Europe; and the ''Eastern-Mediterranean'' present in the Balkans and Middle East and
135   others around Black Sea and in the Caucasus (Maryańska-Nadachowska et al., 2011; Rodrigues
136   et al., 2014). Throughout the wide distribution area where the species occurs, there are contact
137   zones where these lineages co-occur (Lis, Maryańska-Nadachowska, Lachowska-Cierlik, &
138   Kajtoch, 2014; Rodrigues et al., 2014). The occurrence of different infection rates of the
139   maternally inherited endosymbiont *Wolbachia* in the different lineages of *P. spumarius* has
140   pointed to a possible mechanism to explain the maintenance of genetic differentiation in the
141   Carpathians contact zone (Lis et al., 2015). Previous studies have also revealed close
142   relationship and even shared haplotypes between samples from North America, Azores, New
143   Zealand and those from Great Britain (Rodrigues et al., 2014), indicating a recent human-assisted
144   colonization, as previously suggested for North America and New Zealand (Hamilton, 1979;
145   Yurtsever, 2002).

146

147   Seven other species of the genus *Philaenus* occur in the Mediterranean area, having a much more
148   restricted distribution range, and partially overlapping that of *P. spumarius* (Drosopoulos, 2003;
149   Maryańska-Nadachowska, Drosopoulos, Lachowska, Kajtoch, & Kuznetsova, 2010). One such
150   species is *P. tesselatus* Melichar, 1889, which was originally described from Tunisia and was
151   later synonymized with *P. spumarius* (Nast, 1972), being considered a geographic subspecies.
152   Later the synonymy was re-assessed based on morphological evaluation, with the best diagnostic
153   characters being the size and shape of the appendages of the male aedeagus (Drosopoulos &
154   Quartau, 2002). However, geographic variation in the curvature of the aedeagal apical
155   appendages in *P. spumarius* has been reported in both Europe and North America (Wagner,

156 1955; 1959; Hamilton, 1979). Such variation in aedeagus structure within *P. tesselatus* is still
157 largely unexplored (Drosopoulos & Quartau, 2002). Recent genetics studies based on
158 mitochondrial *cytochrome c oxidase I* (COI) and *cytochrome b* (cytB), as well as on nuclear
159 *internal transcribed spacer 2* (ITS2) and *elongation factor 1-alpha* (EF-1alpha) DNA sequence
160 analysis have questioned the species status of *P. tesselatus*, since individuals showing *P.*
161 *tesselatus*-like male genitalia have the same or very similar sequences to *P. spumarius*
162 (Maryańska-Nadachowska et al., 2011; Rodrigues et al., 2014). It is expected that genome-wide
163 markers will provide greater resolution to understand the divergence between these cryptic
164 species. Delimitation of species boundaries is a difficult taxonomic endeavour but it is now
165 widely recognised that an integrative taxonomic approach should include phenotypic, genetic
166 (with a large number of nuclear and mitochondrial markers) and ecological data (Edwards &
167 Knowles, 2014; Tonzo, Papadopoulou & Ortego, 2019).

168

169 In this study, we applied restriction site-associated DNA sequencing (RAD-seq), a reduced-
170 representation sequencing approach that simultaneously discovers and genotypes thousands of
171 single nucleotide polymorphisms for a large number of individuals (Baird et al., 2008; Andrews
172 et al., 2016). We had three main objectives: i) to characterize the morphological (appendages of
173 male aedeagus) and genome-wide divergence between *P. spumarius* and *P. tesselatus;* ii)
174 characterise the patterns of genome-wide differentiation of *P. spumarius* populations across the
175 distribution range of the species; and iii) detect local adaptation by finding genomic regions
176 under selection and associated with environmental variation. The information on gene flow
177 between populations and on the environmental factors associated with local adaptation, as well
178 as on the most appropriate diagnostic methods for the identification of the closely related *P.*
179 *spumarius* and *P. tesselatus,* will be important for future risk assessment of *X. fastidiosa* spread
180 in Europe.

181

182

183 **Materials & Methods**

184 *Sampling*

185 Adult and nymphs of *Philaenus spumarius* were collected in 2010 and 2011 from eight
186 populations (Figure 1) across the distribution range of the species: Cerkes, Anatolia, Turkey
187 (TUR); Mount Parnassus, Greece (GRE); Haapanaki-Keuruu, Finland (FIN); Fitou, South of
188 France (FRAN); Gouveia/Fontanelas, Sintra, Portugal (POR); Aberdare, South Wales, United
189 Kingdom (UK); S. Miguel island, Azores (AZO); Wonalancet, New Hampshire, United States of
190 America (USA). *P. tesselatus* was sampled in Morocco (MOR), from three main localities, near
191 Azrou, near Rabat and near Ceuta. A total of 170 specimens were used for sequencing, up to 22
192 from each sampling site, except from Morocco where only 7 specimens were included
193 (Supplementary Table S1).

194

195   *Morphological characters*

196   *Philaenus* species distinction based on the morphology of male genitalia was accomplished for a
197   subset of 38 males from across all populations, except the Azores and France, for which only
198   immature individuals or females were collected. Nine additional males from MOR, POR and
199   TUR were included to increase morphological sample size, but were not used for genetic
200   analyses (Supplementary Table S1). Preparation and measurements of male genitalia were done
201   as detailed in Supplementary Material – Supplementary Information 1.

202

203   Five variables calculated from the nine previously taken measurements (Figure 2) were used in
204   the morphometric analysis: total length of aedeagus (TotLen), mean length of lower appendages
205   (LowLen), mean length of middle appendages (MidLen), mean length of upper appendages
206   (UpLen), mean curvature of upper appendages (UpCur). The mean value of measurements of
207   paired structures was considered instead of both left and right measurements individually to
208   reduce some of the variability and the number of specimens to be dropped out of the analysis due
209   to missing values related to appendages that were occasionally broken or tilted during aedeagus
210   removal.

211

212   A Principal Component Analysis (PCA) was used to evaluate if morphological characters of the
213   aedeagus could separate *Philaenus* species and/or populations. PCA was applied to standardised
214   variables (centred by the mean and scaled by the variance), since they were measured in different
215   units. Three specimens were left out of the analysis due to missing values. The analysis was
216   performed in R version 3.4.1 (R Core Team, 2017) using the "prcomp" function and figures were
217   produced using the package "ggplot2" version 3.2.1 (Wickham, 2016).

218

219   *DNA extraction and mitochondrial DNA analyses*
220   DNA was extracted from the head and thorax of each specimen using DNeasy Blood & Tissue
221   kit (Qiagen) following the manufacturer's instructions and including RNase A treatment step.
222   Wings and abdomen were not used for DNA extraction to avoid extracting DNA of
223   endosymbionts and parasites. The obtained DNA was assessed for the presence of a high
224   molecular weight band on the agarose gel after electrophoresis, and it was quantified in Qubit 2.0
225   (Invitrogen), using Qubit dsDNA HS Assay kit.

226

227   A subset of 48 specimens from the nine areas was sequenced for mitochondrial DNA
228   (Supplementary Table S1) from which we amplified an 800 bp fragment of the 3'-end of the
229   mitochondrial gene cytochrome *c* oxidase subunit I (COI) by polymerase chain reaction (PCR).
230   Primers used were: C1-J-2195 (5'–TTGATTTTTTGGTCATCCAGAAGT–3') and TL2-N-3014
231   (5'–TCCAATGCACTAATCTGCCATATTA–3') (Simon et al., 1994). PCR was performed in a
232   12.5 uL reaction volume containing: 1x Colorless GoTaq Flexi Buffer, 2 mM $MgCl_2$, 0.2 mM

233 dNTPs, 0.6 mg/ml of BSA, 0.5 µM of each primer, 0.0375 U GoTaq DNA Polimerase
234 (Promega) and approximately 30 ng of DNA. PCR conditions were: an initial denaturation step
235 at 95ºC for 5 min, followed by 35 cycles of denaturation at 95ºC for 45 s, annealing at 50ºC for
236 35 s and extension at 72ºC for 2 min, with a final extension period at 72ºC for 10 min.

237

238 Chromatograms were verified and edited using SEQUENCHER v. 4.0.5 (Gene Codes
239 Corporation), they were aligned using CLUSTAL W on BIOEDIT v. 7.0.9 (Thompson, Higgins,
240 & Gibson, 1994; Hall, 1999) and subsequently trimmed to the same length. We followed the
241 designation of haplotypes of Rodrigues et al. (2014). A median-joining haplotype network was
242 constructed using PopART version 1.7 (Bandelt, Forster, & Rohl, 1999; Leigh & Bryant, 2015).

243

244 *RAD libraries preparation and sequencing*

245 RAD libraries were prepared using a protocol by Etter et al. (2011) and available at
246 https://www.wiki.ed.ac.uk/display/RADSequencing/Home, with modifications as in Rodrigues et
247 al. (2016). The restriction enzyme used was *SbfI* (New England Biolabs). Six libraries were
248 prepared, with ⬚ to 31 individually barcoded samples multiplexed. ⬚ ed-end sequencing used
249 three lanes of Illumina HiSeq 2000 at Genepool (Ashworth Laboratories)
250 (http://genepool.bio.ed.ac.uk/). The individuals from each population were distributed over the
251 different libraries and lanes to avoid library or lane-specific biases.

252

253 *Assembly and SNP calling*

254 The sequence reads from each run were examined by *process_radtags* from STACKS version
255 1.45 (Catchen et al., 2013), to remove those with uncalled bases and low quality scores (phred
256 score lower than 10), to check that the barcode and restriction site were intact in each read and to
257 demultiplex the samples based on the barcode identification. Reads were trimmed at the 3' end,
258 using TRIMMOMATIC v. 0.38, to keep only 87 bases, since preliminary analyses using the
259 entire read revealed a high number of (possibly false) SNPs at the 3' end after this number of
260 bases (data not shown). This may be due to higher sequencing errors towards the end of the reads
261 characteristic of Illumina sequencing (Dohm et al., 2008). These reads were r⬚ on the Stacks
262 module *ustacks* to first assemble *de novo* the reads into 'stacks' (identical sets of reads, called
263 loci) for one individual. The minimum depth of coverage to build a stack (-m) was set to 10, the
264 maximum number of nucleotide differences allowed between stacks to form a locus (-M) was set
265 to 2. Then the Stacks module *cstacks* was used to build a catalog by merging stacks (loci) from
266 multiple individuals, using the default options. Finally, the module *sstacks* was used to match
267 loci from an individual against the catalog. Stacks with very high coverage were removed since
268 they may represent highly repetitive regions and that may include non-orthologous sequences.

269

270   The St ks module *populations* was used to create a Variant Call Format (VCF) file with the bi-
271   allelic genotypes of each individual for each variable nucleotide position. The minimum number
272   of populations a locus must be present in to process a locus (-p) was set to the number of
273   populations analysed (eight or nine, excluding or including Morocco, respectively – see below),
274   the minimum percentage of individuals in a population required to process a locus for that
275   population (-r) was set to 0.5 (50%). Only one SNP per locus was kept, using the option –
276   write_random_snp. Other parameters (-M 2, 3 and 6, -m 5 and 10, -n 1 and 4, --min_maf 0.05 or
277   with no min maf) were tested and the differences were assessed by general patterns in the
278   Principal Component Analysis.

279

280   The VCF file was then filtered using VCFtools (v 0.1.14) (Danecek et al., 2011), excluding sites
281   with less than 75% of individuals with genotype (--max-missing 0.75) and/or with minor allele
282   count of 2 (--mac 2), to exclude singletons (Linck & Battey, 2019). In order to exclude
283   overclustered loci, we filtered out those with a mean depth (across individuals) higher than 200x
284   (--max-meanDP 200).

285

286   The filtered VCF file with the SNP genotypes was converted into the file formats needed for the
287   different analysis programs using PGDSpider 2.0.4.0 (Lischer & Excoffier, 2012). For
288   conversion of geste format to baypass format, we used the script
289   https://github.com/CoBiG2/RAD_Tools/blob/master/geste2baypass.py as of commit b99636e.

290

291   VCFtools was used to calculate summary statistics of coverage and percentage of missing data.
292   Genetix v. 4.05.2 was used to obtain expected and observed heterozygosity, as well as F_{IS} in each
293   population. Pairwise differentiation between populations (F_{ST}) were calculated in Arlequin
294   3.5.1.3, and the significance of F_{ST} was obtained from permutation tests with 10000 repetitions.
295   Mantel tests between $F_{ST}/(1-F_{ST})$ matrices and the natural logarithm of the geographical distance
296   (Rousset, 1997) were performed with ape package version 5.0 (http://ape-package.ird.fr/) in R
297   version 3.4.0, using 9999 permutations.

298

299   *Population structure*

300   Principal Components Analysis (PCA) was used as an exploratory tool of the population
301   structure (Novembre & Stephens, 2008). Computations of PCA were performed in R using
302   package SNPRelate version 1.12.0 (Zheng et al., 2012). Population structure was further
303   examined with model-based clustering algorithm using STRUCTURE v. 2.3.4 (Falush, Stephens,
304   & Pritchard, 2003; Pritchard, Stephens, & Donnelly, 2000). We obtained the coefficients of

305   ancestry using the admixture model and assuming correlated allele frequencies among
306   populations, and K from 1 to 9, with 10 replicate runs of each, applying 50,000 steps of burnin
307   and 1,000,000 MCMC steps after burnin. Structure_threader version 1.2.2 (Pina-Martins et al.,
308   2017) was used to parallelize the runs and find the K best explaining the data (Evanno, Regnaut,
309   & Goudet, 2005; Earl & vonHoldt, 2012). Clumpp version 1.1.2 (Jakobsson & Rosenberg, 2007)
310   was then used to obtain the optimal alignment of ancestry proportions, by permuting the 10
311   replicate runs of STRUCTURE for each value of K.

312

313   The complete dataset consisted of 9 populations and 133 individuals. We also analysed a dataset
314   excluding *P. tesselatus* individuals from Morocco, which consisted of 8 populations and 127
315   individuals of *P. spumarius*. In order to compare morphological variation and genetic variation
316   in *P. spumarius,* we used a dataset of the 32 individuals for which we had data for both
317   morphometry and RAD-seq. We performed PCA for both types of data, as above, and we
318   calculated Spearman correlation between the Principal Component scores obtained from both
319   PCAs using R.

320

321   *Detection of selection – outlier analyses and environmental associations*

322   In order to detect loci with signs of selection for the *P. spumarius* RAD-seq dataset (without
323   Morocco), two approaches were taken: one that detects outlier loci departing from expectation
324   under neutral demographic models (Foll & Gaggiotti, 2008; Vitalis et al., 2014), and another that
325   detects loci associated with environmental variation between populations (Coop et al., 2010;
326   Gautier, 2015).

327

328   Outlier analyses were carried out using Bayescan v. 2.1 (Foll & Gaggiotti, 2008) and SelEstim
329   v1.1.4 (Vitalis et al., 2014). Bayescan uses a Bayesian approach to estimate the posterior
330   probability of two alternative models for each locus, with or without selection. Posterior odds are
331   then obtained and False Discovery Rate calculated to control for multiple testing. The parameters
332   of the chain and of the model were set to the default values. Outlier SNPs were defined to be
333   those with *q*-values lower than 5%. SelEstim v1.1.4 (Vitalis et al., 2014) estimates the intensity
334   of selection at each locus and the posterior distributions of the locus-specific coefficients of
335   selection are compared with a distribution derived from the genome-wide effect of selection
336   using Kullback-Leibler divergence (KLD). KLD is calibrated with simulations from posterior
337   predictive distribution based on observed data (Vitalis et al., 2014). A total of 50 pilot runs of
338   length 1,000 were followed by a run of 1,000,000 with burnin of 10,000. The criterion for a
339   candidate SNP for selection was defined to be the 99% quantile of the KLD distribution.

340

341 Environmental and geospatial variables used in the association analysis were 19 bioclimatic
342 variables, as well as longitude and latitude. Bioclimatic variables were mined from WorldClim
343 version 1.4 (release 3) (http://www.worldclim.org/) and the data was extracted for each location
344 using DIVA-GIS 7.5.0 (http://www.diva-gis.org). Associations between SNP allele frequency
345 differences and the environmental variables were assessed with BayPass v. 2.1 (Gautier, 2015),
346 using the script Baypass_workflow.R
347 (https://gitlab.com/StuntsPT/pyRona/blob/master/pyRona/R/Baypass_workflow.R) as of
348 pyRONA v0.3.7 (Pina-Martins et al., 2019). Significant associations were assessed with Bayes
349 Factor (BF) obtained with the auxiliary covariate model, considering a threshold for BF of 15.
350 We did not exclude any variable at the start of the study based on their correlations, but we tested
351 for correlation the ones that were significantly associated. Spearman correlations between
352 variables were calculated using R.

353

354 After finding the candidate SNPs for selection and environmental association, two datasets were
355 created: a "neutral" dataset, for which we excluded the candidate SNPs, and a "candidate"
356 dataset, that contained only the candidate SNPs. STRUCTURE analyses were also performed on
357 these two datasets.

358

359 RAD tags with candidate SNPs were queried against the available *P. spumarius* partial draft
360 genome and transcriptome (Rodrigues et al., 2016), using blastn with an e-value threshold of 1E-
361 30. From the genome alignment, a longer sequence, bp extended from each end of the RAD
362 tag, was obtained and queried against the NCBI nucleotide database (nr/nt) using BLASTN
363 version 2.9.0 (Altschul et al., 1997), setting a threshold e-value of 1E-5.

364

365 Scripts used in this analysis are available at
366 https://github.com/seabrasg/popgenom_Philaenus.git.

367

368
369
370
371
372
373
374
375
376

377 **Results**

378 *Morphology of male aedeagus*

379 The analysis of male genitalia revealed strong differentiation of the three Morocco samples,
380 which showed a characteristic *P. tesselatus* aedeagus, as originally described in Drosopoulos and
381 Quartau (2002): with the upper appendages longer and weakly curved, extending beyond the
382 lateral appendages and the lower appendages longer and more regularly curved than *P.*
383 *spumarius* (Supplementary Figure S1). All the remaining samples showed *P. spumarius*-like
384 aedeagus (Supplementary Figure S1). Morphometric analysis confirmed this distinction, with the
385 segregation of Moroccan samples along the first Principal Component in the PCA (Figure 3A
386 and 3B), mainly because of variables mean lengths of upper and lower appendages of the
387 aedeagus (PCA loadings in Supplementary Table S2a; boxplots in Supplementary Figure S2).
388 The longer appendages in *P. tesselatus* are expected to be related to the longer body size in
389 general in this species (Drosopoulos & Quartau, 2002) but, when controlling for the total length
390 of the aedeagus, the relative size of the lower appendages remained largest in Moroccan samples
391 (Supplementary Figure S2).

392

393 Within *P. spumarius* we also noted variation across samples, mainly due to the morphometric
394 variables of length and curvature of the upper appendages (Supplementary Figure S2; PCA
395 loadings in Supplementary Table S2b). In particular, there was some geographical structure, for
396 example UK and Finland lying on one extreme of PC2 and Greece lying on the other,
397 corresponding also to the extremes of latitude in this study. A less accentuated curvature and also
398 smaller length of the upper appendages in the Finnish and British than in the Greek samples may
399 be behind this differentiation (boxplots in Supplementary Figure S2).

400

401 *Mitochondrial DNA*

402 The fragment of COI spanned 540 bp and was analysed for 48 specimens, revealing 25
403 haplotypes, 21 of which had already been described in Rodrigues et al. (2014). The remaining
404 four (haplotypes UK15, UK18, GR18_13 and FIN9) differed from previously known haplotypes
405 by 1 or 2 substitutions (Supplementary Figure S3). Two of these new haplotypes (GR18_13 and
406 FIN9) lay in an intermediate position in the haplotype network between the previously defined
407 "North-Eastern" and "Western" haplogroups. In fact, the three haplogroups are not completely
408 distinct but we maintain their designation in order to more easily describe and visualize the
409 mitochondrial variation in relation to the RAD-seq variation: "Eastern-Mediterranean" (EM) in
410 red, "Western" (W) in green and "North-Eastern" (NE) in blue (Supplementary Figure S3. We
411 also attributed similar colours to the groups resulting from RAD-seq for ease of visualization
412 (see below Figure 6C).

413

414   All seven specimens from Morocco (MOR) sequenced for mtDNA either showed the most
415   common haplotype of the "W" haplogroup (H29) or a haplotype differing by only one
416   substitution (H28 and H37) (Supplementary Figure S3). All haplotypes from the Azores (AZO)
417   and continental Portugal (POR) belonged to "W". France (FRAN) haplotypes belonged in "W"
418   or in between "W" and "EM" (haplotype H49). Haplotypes from Greece (GRE) belonged to
419   "EM" or in between "W" and "NE" (haplotype GR18_13). Haplotypes from Turkey (TUR)
420   belonged to "NE". In Finland (FIN), there were haplotypes from "EM", "NE" and also one
421   between "W" and "NE". The USA population comprised haplotypes from "NE" and the UK
422   population from "NE" and "W" (Supplementary Figure S3). The four new haplotypes were
423   submitted to GenBank under accession numbers MT025773-MT025776.

424

425   *RAD sequencing*

426   A total of 838,730,936 reads was obtained from the Illumina sequencing. The *process_radtags*
427   step in Stacks retained 647,870,180 reads. This corresponds to an average of 3,811,001 ±
428   3,524,799 (standard deviation) reads per individual. Thirty-seven individuals with lower
429   numbers of reads (<500,000 reads) or large amounts of missing data (>60%) were excluded from
430   the analysis (Supplementary Table S1), leaving a total of 133 individuals, for which the number
431   of reads ranged from 736,248 to 23,798,148 (average 4,507,853 ± 3,668,879 sd). Raw reads after
432   demultiplexing were deposited in SRA database with accession PRJNA606428. The *population*
433   Stacks module, followed by filtering, produced 1691 SNPs, with a mean coverage of 105.5 reads
434   per locus per individual (Supplementary Figure S4) and mean percentage of missing data per
435   individual of 12.3 % (Supplementary Table S1). We applied relatively stringent filtering criteria
436   to avoid having large amounts of missing data per individual resulting from the large genome
437   size in this species. This has produced a relatively small number of SNPs but that have a good
438   representation across individuals and that are expected to be scattered across the genome.

439

440   *Population structure*

441   Principal Component 1 in the PCA clusters Morocco individuals away from the others (Figure
442   3C and D). When testing other assembly and filter parameters we obtained similar patterns in the
443   groupings of samples (Supplementary Figure S5). Also, Structure analysis gave support to a
444   genetic group solely comprising Moroccan samples (Figure The average F_{ST} of Morocco
445   *versus* other populations was 0.4, much higher than average F_{ST} of other populations'
446   comparisons (0.13) (Table 1). In all population-pairwise F_{ST} calculations involving Morocco,
447   there were a considerable number of SNPs that were fixed or nearly fixed for one allele in
448   Morocco and for the other allele in all the other populations, as seen in the relatively high
449   frequency of high F_{ST} values on the histograms in all comparisons and in the high correlations
450   between F_{ST} values among population pairs (Supplementary Figure S6a). There was neither such
451   a high number of fixed SNPs nor such high correlations between F_{ST} values when considering

452   the other pairs of populations (Supplementary Figure S6b). Moroccan samples are thus clearly
453   differentiated, at the genome-wide markers, from the remaining eight populations here analysed
454   in contrast with mtDNA results that showed no differentiation between Morocco and the Iberian
455   Peninsula (Supplementary Figure S3).

456

457   The relationship between geographical and genetic distances was significant when considering
458   European populations (excluding from the dataset USA, Azores and Morocco) (Mantel test,
459   z=19.49793, p=0.0177; Figure 5). When considering the comparisons involving Morocco, a
460   positive correlation is seen between genetic and geographical distances, but mainly because of
461   the lower F_{ST} value obtained between Morocco and Portugal (F_{ST} = 0.34), than between Morocco
462   and the other populations (F_{ST} > 0.4; Figure 5). This lower differentiation may be the result of
463   some level of admixture, which was detected in the STRUCTURE analysis, where all individuals
464   from the Portuguese population show a small contribution from the genetic group present in
465   Morocco (Figure 4).

466

467   Mean diversity (expected heterozygosity, H_E) ranged from 0.0373 (in Morocco) to 0.0808 (in
468   Greece) and mean observed heterozygosity (H_O) from 0.0258 (in Morocco) to 0.0537 (UK). H_O
469   values were generally lower than expected under Hardy-Weinberg equilibrium (HWE) in all
470   populations (average F_{IS} of 0.346) (Table 1). For Morocco, interpretation of H_E should be carried
471   out carefully, since individuals come from three different locations and thus are not necessarily
472   expected to be in HWE. Additionally, we found a positive and significant correlation between
473   observed heterozygosity and sequence read depth (r_S=0.686, p=0.0412; Supplementary Figure
474   S7). This suggests that lower read depths may have led in some cases to allele dropout,
475   contributing towards false homozygotes. However, in the case of Morocco, the mean read depths
476   were not the lowest in this dataset, being similar to others (Supplementary Figure S4) and thus
477   this should not be the main factor contributing to the low observed heterozygosity.

478

479   The dataset without Moroccan samples consisted of 127 individuals and 2083 SNPs. For this
480   dataset, PCA revealed two distinct clusters along PC1, generally separating Greece and Turkey
481   from the remaining populations (Figure 6A). The latter were separated along PC2 in three
482   groups, one including mainly Portugal and France, another including mainly Finland. A third
483   one, in between these two, including USA, UK and the Azores (Figure 6A). This structure had
484   already been detected in PC3 of the analysis of the dataset that included Morocco (Figure 3D).
485   For this dataset, when excluding the USA and the Azores populations, there was again
486   significant isolation-by-distance for the European populations (Mantel test, z=19.49793,
487   p=0.0229). However, there were a few individuals that were genetically more similar to others
488   more distant geographically, which is also seen in the STRUCTURE analysis (Figure 6). In
489   Greece and Finland, there were no admixed individuals between the two main clusters ("Eastern-

490 Mediterranean" in red, and "North-Eastern" in blue) – they were either from one or the other
491 group, with a few exceptions (Figure 6B). An analysis of the Turkish population revealed the
492 presence of possibly admixed individuals from these two groups (with a smaller contribution
493 from the "North-Eastern" cluster). In the USA and UK populations, all individuals showed some
494 level of admixture between the "North-Eastern" (blue) and the "Western" (green) clusters. The
495 Azores allies a small contribution from the "Western" group to a major one belonging to the
496 "North-Eastern" group and one individual was admixed between "North-Eastern" and "Eastern-
497 Mediterranean" (also seen in PCA). The admixture in USA, UK and the Azores is also apparent
498 from their intermediate position between the Portugal+France group and the
499 Finland+Greece+Turkey group in the PCA (Figure 6A).

500

501 The majority of specimens for which COI sequence was available, had a correspondence
502 between the mtDNA and the genomic cluster. However, there were some specimens showing a
503 mismatch consisting of a mtDNA haplotype belonging in a different genomic cluster (Figure 6B
504 and C; Supplementary Figure S3). For example, one individual from UK (UK6) bearing a
505 mtDNA haplotype (H24) belonging to "Western" haplogroup (green), turned up "North-Eastern"
506 (blue) in the genome analysis. Two individuals from France, bearing mtDNA H49 haplotype
507 (intermediate between "Western" and "Eastern-Mediterranean"), came up as differentiated at
508 genomic markers, one "Western" (green) and the other intermediate "Western" / "North-
509 Eastern". In Greece and Finland, COI sequenced individuals show both a mtDNA and genomic
510 makeup belonging to either "Eastern-Mediterranean" or "North-Eastern", except two individuals
511 assessed as "North-Eastern" in genomic markers but with a mtDNA haplotype in intermediate
512 position in the network, between "Western" and "North-Eastern" haplogroups. The four
513 individuals from Turkey sequenced for COI belonged in "North-Eastern" haplotypes. While one
514 of them had full ancestry from "North-Eastern" group, the other three had their largest
515 proportion of ancestry from "Eastern-Mediterranean", based on the genome-wide markers.

516

517 For samples for which both morphometric and RAD-seq data was available (N=32), we
518 computed Principal Components Analysis (Supplementary Figure S8) and calculated the
519 correlation between PC1 and PC2 scores for both analyses. There was a significant correlation
520 between PC1 from morphometry and PC2 from RAD-seq ($r_S=0.63$, p=1E-04), while all the
521 remaining were low and non significant ($r_S=-0.22$, p=0.2194 between PC1 from each; $r_S=0.1$,
522 p=0.5927 between PC2 from each; $r_S=0.29$, p=0.102 between PC2 from morphometry and PC1
523 from RAD-seq).

524

525

526 *Detection of selection – outlier analysis and environmental associations*

527 Candidate SNPs for positive selection were identified by detection of highly differentiated
528 outliers: eight were detected by Bayescan (Supplementary Figure S9); and 25 by KLD (quantile
529 99% KLD 2.037087) in SelEstim (Supplementary Figure S10; Supplementary Table S3). Six
530 outlier SNPs were common to both analyses. No outlier SNPs for balancing selection were
531 detected in the Bayescan analysis (Supplementary Figure S9).

532

533 The BayPass analysis detected 163 SNPs associated with environmental variables (BF > 15)
534 (Supplementary Table S3 and S4; Supplementary Figure S11), one of them common to the
535 candidate SNPs detected with Bayescan. Variables showing association were: Longitude,
536 Temperature Annual Range, Precipitation of Driest Quarter, Precipitation of Wettest Quarter,
537 Mean Temperature of Warmest Quarter, Mean Diurnal Range (Mean of monthly (max temp –
538 min temp)). Spearman correlations between these 6 variables were generally low (absolute
539 values below 0.6), with only two values above 0.6 (Supplementary Table S4c).

540

541 When excluding these candidate SNPs (188 in total) from the "full" dataset, creating a "neutral"
542 dataset, the main pattern of structuring was maintained, differing only in admixture proportions
543 at higher values of K (4 and 5) (PCA in Supplementary Figure S12 and Structure in
544 Supplementary Figure S13). When analysing the "candidates" dataset, the PCA showed a
545 separation that corresponded generally to longitude variation along PC1 and to latitude along
546 PC2 (Supplementary Figure S12). The STRUCTURE analysis, although artificial for the dataset
547 of loci under selection, revealed similar structuring when compared to the other datasets, but
548 with less admixture (Supplementary Figure S13). This is an expected outcome considering that
549 these candidate loci have similar allelic variation within each population and different allelic
550 variation between populations. The fact that there are still differentiated individuals within
551 populations in this dataset, consistently assigned to the same groups as in the other datasets, is a
552 reflection of the methods for detecting selection, based on population allelic variation.

553

554 Seventy-three candidate SNPs had hits (threshold evalue of 1E-30) with the draft genome of *P.*
555 *spumarius* and seven with the transcriptome (Rodrigues et al 2016). From these, nine had hits
556 (threshold evalue of 1E-5) with predicted genes in the NCBI nucleotide database (Supplementary
557 Table S3).

558

559 **Discussion**
560 RAD sequencing analysis revealed the genetic distinction of North-African relative to other
561 samples here analysed, which matched the morphological differences at the male genitalia,

562 identifying these as *Philaenus tesselatus*. This genetic differentiation was however not detected
563 at the mitochondrial DNA level, since *P. tesselatus* and *P. spumarius* share mtDNA haplotypes,
564 as already described in previous studies (Maryańska-Nadachowska et al., 2011; Rodrigues et al.,
565 2014). These results thus reinforce the importance of taking an integrative approach when
566 studying the taxonomy of a group of species, especially cryptic ones (Edwards & Knowles,
567 2014; Dejaco et al., 2016; Borges et al., 2017; Tonzo et al., 2019).

568

569 The fact that there are mitochondrial DNA haplotypes shared between *P. tesselatus* and *P.*
570 *spumarius*, while the nuclear genome is differentiated, may indicate selection on mtDNA
571 following introgressive hybridization (Gompert et al., 2008). One possible mechanism for
572 selection on mtDNA described in several insects, is the occurrence of maternally inherited
573 endosymbionts associated with certain haplotypes, like *Wolbachia*. These endosymbionts are
574 known to manipulate reproductive output, mainly through cytoplasmic incompatibility: no viable
575 offspring are produced when an infected male fertilizes an uninfected female, or a female
576 infected with a different strain (Werren, Baldo, & Clark, 2008). Haplotypes associated with the
577 *Wolbachia* infection could thus spread, hitchhiking through the population. Mitochondrial
578 introgression between closely related species caused by *Wolbachia* has been described in several
579 species of Diptera and Lepidoptera (Jiggins, 2003; Narita et al., 2006; Rousset & Soulignac,
580 1995; Whitworth et al., 2007). In these cases, different species share the same mitochondrial
581 haplotypes, making DNA barcoding ineffective. Such a scenario is plausible as *Wolbachia*
582 infection has already been detected in *P. spumarius* across Europe by Lis et al. (2015) (see
583 below).

584

585 The admixture from the Moroccan genetic group detected in all the individuals from the
586 Portuguese population (located in the Central-West part of the Iberian Peninsula) suggests some
587 level of recent or ongoing gene-flow between *P. spumarius* and *P. tesselatus*. Despite previous
588 doubts about the taxonomic status of these two taxa, our data point towards them being closely
589 related but independent gene-pools, probably early in the speciation "continuum" (Seehausen et
590 al., 2014). Both taxa co-occur in some locations in southern Iberian Peninsula (personal
591 observation by J. A. Quartau and A. C. Neto, based on identification by male aedeagus
592 morphology) and it will be important to study these sympatric areas. Genital traits are relevant
593 since they may contribute to reproductive isolation, either structural or sensory, if differences in
594 genital morphology between species prevent or reduce the success of copulation and
595 insemination (Masly, 2012). Structural isolation was shown for example in the species pair
596 *Drosophila yakuba* and *D. santomea* (Kamimura & Mitsumoto, 2012), but in many species no
597 convincing evidence for such isolation has been found so far (Masly, 2012). Morphological
598 variation in female genitalia, as well as behavioural and physiological responses during mating
599 may also aid in understanding potential mechanisms of reproductive isolation, particularly in
600 sympatry. We recognise that species identification based on male genitalia characteristics may be

601 insufficient when there is intraspecific variation with some overlap between species. Despite our
602 limited *P. tesselatus* samples to fully understand the range of its variation, genomic data allow
603 higher resolution in detecting genetic differentiation, but this is not enough to infer species status
604 (Tonzo et al., 2019). A more comprehensive study on morphology, mtDNA and genome-wide
605 variation of a wider sample from the Mediterranean region of both *P. spumarius* and *P.*
606 *tesselatus* is required.

607

608 The morphometric geographical variation detected in *P. spumarius* showed some correlation
609 with genetic variation, although this was not fully clear. Clinal latitudinal variation, as well as
610 elevation variation, in the shape of male genitalia had already been described in European
611 populations of species (Wagner, 1955; 1959). Shorter and less curved upper appendages were
612 found in the north compared to the south, and in higher than in lower altitudes at the same
613 geographical regions (Wagner, 1955; 1959). RAD sequencing data permitted detection of finer
614 population genetic structure within *P. spumarius* than previously known with mtDNA and a
615 limited number of nuclear genes. Although there was a pattern of isolation-by-distance in
616 European populations, there were clear distinctions between groups in the PCA and
617 STRUCTURE analyses not related to geographical distance. The most likely K of three in
618 STRUCTURE corresponded loosely to the three mitochondrial haplogroups already described in
619 Rodrigues et al. (2014) and Maryańska-Nadachowska et al. (2011) but we detected some degree
620 of admixture along contact zones. We found admixed individuals in France, Turkey and Finland,
621 and several other individuals belonging to a different genetic group, with no admixture. This may
622 suggest recent migration or the maintenance of reproductive barriers. In particular, there was
623 almost no admixture between the "Eastern-Mediterranean" and the "North-Eastern" groups.
624 Maryańska-Nadachowska et al. (2011) described a contact zone in the Carpathians between
625 North-Eastern and South-Western haplogroups (this last group corresponds to our "Western" and
626 "Eastern-Mediterranean" together) and detected heteroplasmic mitochondrial DNA, likely
627 caused by paternal leakage from hybridization between members of these two clades.
628 Interestingly, Lis, Maryańska-Nadachowska and Kajtoch (2015) have found different
629 prevalences of *Wolbachia* infection between the different mitochondrial lineages of *P.*
630 *spumarius*. The North-Eastern clade showed a higher proportion of infected individuals than the
631 South-Western. In the Carpathian contact zone, infection was more prevalent in both groups,
632 which however harboured different supergroups of *Wolbachia*. The authors suggest that there
633 may be limited gene-flow between genetically distinct populations through a mechanism of
634 cytoplasmic incompatibility. This could be an explanation for the low level of admixture
635 detected in our study between the two genetic groups. A genome-wide survey in a wider
636 sampling of both genetic groups will allow testing these hypotheses.

637

638 The intermediate position of UK, USA and Azores individuals in the PCA analysis, as well as
639 the admixture detected in STRUCTURE, suggest they are the result of mixed gene pools. It

640 further corroborates the mtDNA results of Rodrigues et al. (2014) that showed the occurrence of
641 different mitochondrial lineages in the UK and USA and that Azores and some USA samples
642 were genetically similar to those from the UK. Across North America, variation in the
643 morphology of male aedeagus in *P. spumarius* was reported by Hamilton (1979) and different
644 mtDNA haplogroups were detected by Rodrigues et al. (2014), leading to the suggestion of
645 multiple colonization events. The analysed population from New Hampshire, USA, showed very
646 low genome-wide differentiation from the UK population (mean F_{ST}=0.042, compared to other
647 pairwise comparisons in this study, and also a large number of COI haplotypes belonging to the
648 "North-Eastern" haplogroup (5 in the 7 samples analysed for mtDNA). This supports a likely
649 origin of the North American *P. spumarius* from Northern Europe, perhaps with multiple
650 colonisation events and with large effective population sizes. In S. Miguel island (Azores), only
651 two COI haplotypes, differing by one substitution, were found so far (in 6 samples, Rodrigues et
652 al., 2014 and this study), which are closely related to the UK haplotypes from the "Western"
653 haplogroup. From the genomic results, this population was more differentiated from UK and
654 USA (mean F_{ST}=0.098 between AZO and UK, and mean F_{ST}=0.078 between AZO and USA)
655 than these two were from one another (mean F_{ST}=0.042 between UK and USA), showing the
656 lowest genetic diversity (expected heterozygosity) of all *P. spumarius* populations here analysed.
657 These results suggest a likely origin of the Azores colonization from Northern Europe, not from
658 Southern Europe, and that this colonization is likely to have involved a bottleneck event leading
659 to reduced genetic diversity observed. The low number of colour morphs found in S. Miguel
660 (Borges et al., 2018) in this highly polymorphic species further supports this hypothesis,
661 although selective processes may also be involved. Expanding the sampling and analyses will
662 allow more precise determination of the origin and eventually the timing of these transatlantic
663 colonisations.

664

665 Population genomics approaches provide genome-wide information that is expected to reflect a
666 baseline of neutral processes and, at the same time, allow detection of loci with signatures of
667 selection deviating from this baseline (Hohenlohe, Phillips, & Cresko, 2010). We focused on
668 detecting local adaptation, by finding those loci that are more differentiated (F_{ST}) between
669 populations than expected from the neutral background, and also by finding loci that have allelic
670 variation correlated with environmental variation. When discarding such candidate loci for
671 selection from our dataset, the population structure patterns remained very similar to the neutral
672 dataset, which means that these 9% of loci are not affecting the genome-wide neutral pattern of
673 population structure. Despite the usefulness of RAD sequencing for detection of selection in
674 natural populations (Catchen et al., 2017), this analysis is limited by the number of SNPs
675 analysed and also by the fact that RAD tags are usually distributed non-uniformly across the
676 genome (Lowry et al., 2016). The large genome size of *P. spumarius* (Rodrigues et al., 2016)
677 makes it more difficult to have a good genomic representation with these scattered markers.
678 Also, genetic signatures of selection in individual loci can be weak and not easily detected in
679 cases of soft selective sweeps (adaptation from standing genetic variation), epistatic interactions

680 among loci or genotype-by-environment interactions (Hohenlohe et al., 2010). Whole-genome
681 analyses, by analysing patterns of diversity, differentiation and linkage disequilibrium along the
682 genome, will be essential to better understand the evolutionary forces of selection and
683 recombination shaping the genomic variation (Ellegren, 2014). The fact that we did not detect
684 loci under balancing selection in the Bayescan analysis may also be related to low marker
685 density. *P. spumarius* is particularly known for its balanced polymorphism in the dorsal colour
686 form and the assessment of population variation in the colour-associated loci (Rodrigues et al.,
687 2016) remains to be done.

688

689 In the environmental association analysis, the associated variables were primarily longitude and
690 those related to the extreme values and range variation in temperature and precipitation. This
691 analysis is tentative, since we have a low number of populations from a wide geographic range.
692 The low number of hits of the candidate loci with the *P. spumarius* transcriptome may indicate
693 their location was mostly in non-coding regions, while the low number of hits with the partial
694 genome denotes its incompleteness (Rodrigues et al., 2016). A more complete draft genome is
695 now available (Biello et al., 2020), and new genomic and transcriptomic resources will be soon
696 generated for *P. spumarius* which will provide important tools to further explore the molecular
697 basis of adaptation in this species.

698

699 Understanding species divergence and population genetic structure of these insect vectors is
700 relevant to the management of the eventual progression of the plant pathogenic bacterium *X.*
701 *fastidiosa*, since their dispersal patterns might aid or constrain disease transmission. Also, the
702 ecological requirements of different taxa or local populations may be different, for example in
703 host plant preference or performance as vectors (acquisition of *X. fastidiosa* and transmission
704 efficiency). Understanding the specific ecology of the vector has been shown to be crucial in the
705 management of *X. fastidiosa* diseases in America (Redak, Purcell, Blua, Mizell, & Andersen,
706 2004). All thi pe of information is important for epidemiological models of *X. fastidiosa* in
707 Europe and other Mediterranean countries. The risk of *X. fastidiosa* transmission and disease
708 progression is generally expected to be related to long-range human-assisted movements of
709 infected plants and with shorter-range natural dispersal by vectors (EFSA et al., 2019). The
710 genetic studies of *P. spumarius* have shown that it does not constitute a panmictic population,
711 and geographical distance is not the only factor restricting gene flow. Other factors have to be
712 taken into account, namely habitat fragmentation, barriers to gene exchange such as
713 endosymbionts or behavioural differences, rapid climate changes that may cause major shifts in
714 distribution ranges, as well as unpredictable adaptive responses (Kellermann & Van
715 Heerwaarden, 2019). Even without detectable gene flow, adults of *P. spumarius* may be able to
716 migrate occasionally, or consistently but without reproductive outcome, and spread the
717 bacterium. More ecological studies on the abundance and distribution of the insect's vectors
718 through the seasons and across years are needed to understand the dispersal patterns across

719 geographical regions and the potential for disease spread. In particular, it will be important to
720 understand the dispersal patterns from South Italy, which is a potential source of contamination
721 by *Xylella fastidiosa* (subspecies *pauca*). In North and Central Italy, as well as in Sicily,
722 Rodrigues et al. (2014) detected haplotypes from both "Eastern-Mediterranean" and "Western"
723 haplogroups, unveiling the pivotal role of that region in the dispersal patterns of the species
724 between the Mediterranean peninsulas.

725

726 **Conclusions**

727 In this study, morphological and genomic analysis allowed a more detailed view of the
728 divergence between *P. spumarius* and *P. tesselatus*, as well as of the population structure and
729 adaptation in *P. spumarius*. We found genome-wide divergence between these two species,
730 despite the lack of mitochondrial DNA differentiation between them. The population genomics
731 approach taken here showed admixture but also co-occurrence of non-admixed individuals in
732 contact zones of diverging mitochondrial lineages of *P. spumarius*. The potential role of
733 *Wolbachia* in shaping these patterns of divergence and introgression should be further explored.

734

735 The findings on species divergence and population structure described here point to the need for
736 elucidating the dispersal and ecological requirements of the different taxa and local populations
737 of these vectors for a better management of *X. fastidiosa* progression.

738

739

740

741

742 **Acknowledgements**

743 We thank Carla Pereira, Carla Rego, Rui Nunes and Luis Ferreira for help in field-work; Karim
744 Gharbi and Mark Blaxter for help with the RAD sequencing; Vera Nunes and Juan Farinas for
745 help in the laboratory; the Faculty of Sciences of the University of Lisbon's Microscopy Facility,
746 a node of the Portuguese Platform for Bioimaging, reference PPBI-POCI-01-0145-FEDER-
747 022122, where images from insect specimens were acquired.

748

749

750

751

752

753 **References**

754 Almeida, R. P. P., Fuente, L. D. La, Koebnik, R., Lopes, J. R. S., Parnell, S., & Scherm, H. (2019).
755 Addressing the New Global Threat of *Xylella fastidiosa*. *Phytopathology*, *109*, 172–174.
756 https://doi.org/10.1094/PHYTO-12-18-0488-FI
757
758 Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J.
759 (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.
760 *Nucleic Acids Research, 25,* 3389–3402. doi: 10.1093/nar/25.17.3389.
761
762 Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of
763 RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, *17*(2), 81–92. doi:
764 10.1038/nrg.2015.28.
765
766 Baldi, P., & La Porta, N. (2017). *Xylella fastidiosa*: Host Range and Advance in Molecular
767 Identification Techniques. *Frontiers in Plant Science*, *8*, 944.
768 https://doi.org/10.3389/fpls.2017.00944
769
770 Bahrndorff, S., Ruiz-González, A., Jonge, N. De, Nielsen, J. L., Skovgård, H., & Pertoldi, C.
771 (2020). Integrated genome-wide investigations of the housefly, a global vector of diseases reveal
772 unique dispersal patterns and bacterial communities across farms. *BMC Genomics*, *21*, 66. doi.
773 10.1186/s12864-020-6445-z.
774
775 Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., … Johnson,
776 E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS*
777 *One, 3*(10), e3376. doi: 10.1371/journal.pone.0003376.
778
779 Bandelt, H. J., Forster, P., & Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies.
780 *Molecular Biology and Evolution*, *16*(1), 37–48. doi: 10.1093/oxfordjournals.molbev.a026036.
781
782 Biello, R., Mathers, T. C., Mugford, S. T., Liu, Q., Rodrigues, A. S. B., Neto, A. C. …
783 Hogenhout, S. A. (2020). Draft genome assembly version 1 of the meadow spittlebug *Philaenus*
784 *spumarius* (Linnaeus, 1758) (Hemiptera, Aphrophoridae) (Version version 1) [Data set]. *Zenodo*.
785 doi: 10.5281/zenodo.3368385.
786
787 Borges, P. A. V., Amorim, I. R., Terzopoulou, S., Rigal, F., Emerson, B., & Serrano, A. R. M. (2017).
788 Cryptic diversity in Azorean beetle genus *Tarphius* Erichson, 1845 (Coleoptera: Zopheridae): An
789 integrative taxonomic approach with description of four new species. *Zootaxa,* 4236, 401–449. doi:
790 10.11646/zootaxa.4236.3.1.
791
792 Borges, P. A. V, Rodrigues, A. S. B., Silva, S. E., Seabra, S. G., Paulo, O. S., & Quartau, J. A.
793 (2018). New data on polymorphism of the meadow spittlebug *Philaenus spumarius* (L.)
794 (Hemiptera: Aphrophoridae) from the island of São Miguel (Azores), *Zootaxa*, *4369*(1), 144–
795 150. doi: 10.11646/zootaxa.4369.1.9.

796

797   Busvine, J. R. (1980). Cryptic species of insect disease vectors and their importance. *Endeavour*,
798   *4*(3), 108–112.

799

800   Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an
801   analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140. doi:
802   10.1111/mec.12354.

803

804   Catchen, J. M., Hohenlohe, P. A., Bernatchez, L., Funk, W. C., Andrews, K. R., & Allendorf, F.
805   W. (2017). Unbroken : RADseq remains a powerful tool for understanding the genetics of
806   adaptation in natural populations. *Molecular Ecology Resources*, *17*, 362–365. doi:
807   10.1111/1755-0998.12669.

808

809   Coop, G., Witonsky, D., Di Rienzo, A., & Pritchard, J. K. (2010). Using environmental
810   correlations to identify loci underlying local adaptation. *Genetics*, *185*(4), 1411–1423. doi:
811   10.1534/genetics.110.114819.

812

813   Cornara, D., Saponari, M., Zeilinger, A. R., Stradis, A. De, Boscia, D., Loconsole, G., …
814   Martelli, G. P. (2017). Spittlebugs as vectors of *Xylella fastidiosa* in olive orchards in Italy.
815   *Journal of Pest Science*, *90*(2), 521–530. doi: 10.1007/s10340-016-0793-0.

816

817   Cornara, D., Morente, M., Markheiser, A., Bodino, N., Tsai, C.-W., Fereres, A., … Lopes, J. R. S.
818   (2019). An overview on the worldwide vectors of *Xylella fastidiosa*. *Entomologia Generalis*, *39*(3–
819   4), 157–181. https://doi.org/10.1127/entomologia/2019/0811

820

821   Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … Group, 1000
822   Genomes Project Analysis. (2011). The variant call format and VCFtools. *Bioinformatics*,
823   *27*(15), 2156–2158. doi: 10.1093/bioinformatics/btr330.

824

825   Dejaco, T., Gassner, M., Arthofer, W., Schlick-Steiner, B. C., & Steiner, F. M. (2016).
826   Taxonomist ' s nightmare … Evolutionist ' s delight: an integrative approach resolves species
827   limits in jumping bristletails despite widespread hybridization and parthenogenesis. *Systematic*
828   *Biology*, *65*(6), 947–974. doi: 10.1093/sysbio/syw003.

829

830   Dohm, J. C., Lottaz, C., Borodina, T., & Himmelbauer, H. (2008). Substantial biases in ultra-
831   short read data sets from high-throughput DNA sequencing. *Nucleic Acids Research, 36,* e105.
832   doi: 10.1093/nar/gkn425.

833

834   Drosopoulos, S. (2003). New data on the nature and origin of colour polymorphism in the
835   spittlebug genus *Philaenus* (Hemiptera: Aphorophoridae). *Annales de La Société Entomologique*
836   *de France*, *39*(1), 31–42. doi: 10.1080/00379271.2003.10697360.

837

838 Drosopoulos, S., & Quartau, J.A. (2002). The spittlebug *Philaenus tesselatus* Melichar, 1899
839 (Hemiptera, Auchenorrhyncha, Cercopidae) is a distinct species. *Zootaxa* 68: 1–8. doi:
840 10.11646/zootaxa.68.1.1.

841

842 Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for
843 visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics*
844 *Resources*, *4*(2), 359–361. doi: 10.1007/s12686-011-9548-7.

845

846 Edwards, D. L., & Knowles, L. L. (2014). Species detection and individual assignment in species
847 delimitation: can integrative data increase efficacy? *Proceedings of the Royal Society B:*
848 *Biological Sciences*, *281*, 20132765. doi: 10.1098/rspb.2013.2765.

849

850 EFSA Panel on Plant Health (PLH). (2015). Scientific Opinion on the risks to plant health posed
851 by *Xylella fastidiosa* in the EU territory, with the identification and evaluation of risk reduction
852 options. *EFSA Journal*, *13*(1), 3989. doi: 10.2903/j.efsa.2015.3989.

853

854 EFSA (European Food Safety Authority), Vos, S., Camilleri, M., Diakaki, M., Lázaro, E.,
855 Parnell, S., … Vicent, A. (2019). *Pest survey card on Xylella fastidiosa. EFSA supporting*
856 *publication* (Vol. 2019:EN–16). doi: 10.2903/sp.efsa.2019.EN-1667.

857

858 Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms.
859 *Trends in Ecology & Evolution*, *29*(1), 51–63. doi: 10.1016/j.tree.2013.09.008.

860

861 Etter, P. D., Bassham, S., Hohenlohe, P.A., Johnson, E., & Cresko, W.A, (2011) SNP discovery
862 and genotyping for evolutionary genetics using RAD sequencing. In: *Molecular Methods for*
863 *Evolutionary Genetics* (eds Orgogozo V, Rockman MV), pp. 157– 178. Humana Press, New
864 York. Doi: 10.1007/978-1-61779-228-1_9.

865

866 Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals
867 using the software STRUCTURE: a simulation study. *Molecular Ecology*, *14*, 2611–2620. doi:
868 10.1111/j.1365-294X.2005.02553.x.

869

870 Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using
871 multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, *164*, 1567–
872 1587.

873

874 Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for
875 both dominant and codominant markers: a bayesian perspective. *Genetics*, *180*, 977–993. doi:
876 10.1534/genetics.108.092221.

877

878 Gautier, M. (2015). Genome-wide scan for adaptive divergence and association with population-
879 specific covariates. *Genetics*, *201*(4), 1555–1579. doi: 10.1534/genetics.115.181453.
880

881 Gompert, Z., Forister, M. L., Fordyce, J. A., & Nice, C. C. (2008). Widespread mito-nuclear
882 discordance with evidence for introgressive hybridization and selective sweeps in Lycaeides.
883 *Molecular Ecology*, *17*, 5231–5244. doi: 10.1111/j.1365-294X.2008.03988.x.
884

885 Halkka, O., & Halkka, L. (1990). Population genetics of the polymorphic meadow spittlebug,
886 *Philaenus spumarius* (L.). *Evolutionary Biology (USA), 24,* 149–191.
887

888 Hall, T.A. (1999). BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis
889 Program for Windows 95/98/NT. *Nucleic Acids Symposium Series, 41*, 95–98.
890

891 Hamilton, K. G. A. (1979). Synopsis of North American Philaenini (Rhynchota: Homoptera:
892 Cercopidae) with a new genus and four new species. *The Canadian Entomologist*, *111*, 127–141.
893

894 Hendry, A. P., Bolnick, D. I., Berner, D., & Peichel, C. L. (2009) Along the speciation
895 continuum in sticklebacks. *Journal of Fish Biology*, 75, 2000–2036. doi: 10.1111/j.1095-
896 8649.2009.02419.x.
897

898 Hohenlohe, P. A., Phillips, P. C., & Cresko, W. A. (2010). Using population genomics to detect
899 selection in natural populations: key concepts and methodological considerations. *International
900 Journal of Plant Science*, *171*(9), 1059–1071. doi: 10.1086/656306.
901

902 Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation
903 program for dealing with label switching and multimodality in analysis of population structure.
904 *Bioinformatics*, *23*(14), 1801–1806. doi: 10.1093/bioinformatics/btm233.
905

906 Jiggins, F. M. (2003). Male-killing Wolbachia and mitochondrial DNA : selective sweeps, hybrid
907 introgression and parasite population dynamics. *Genetics*, *164*, 5–12.
908

909 Kamimura, Y., & Mitsumoto, H. (2012). Lock-and-key structural isolation between sibling
910 *Drosophila* species. *Entomological Science*, *15*, 197–201. doi: 10.1111/j.1479-
911 8298.2011.00490.x.
912

913 Karban, R., & Huntzinger, M. (2018). Decline of meadow spittlebugs, a previously abundant
914 insect, along the California coast. *Ecology*, *99*, 2614–2616. doi: 10.1002/ecy.2456.
915

916   Karban, R., & Strauss, S. (2004). Physiological tolerance, climate change, and a northward range
917   shift in the spittlebug, *Philaenus spumarius*. *Ecological Entomology*, *29*, 251–254. doi:
918   10.1111/j.1365-2311.2004.00576.x.
919
920   Kellermann, V., & Van Heerwaarden, B. (2019). Terrestrial insects and climate change: adaptive
921   responses in key traits. *Physiological Entomology, 44,* 99–115. doi: 10.1111/phen.12282.
922
923   Leigh, J. W., & Bryant, D. (2015). PopART: Full-feature software for haplotype network
924   construction. *Methods in Ecology and Evolution, 6*(9), 1110–1116. doi: 10.1111/2041-
925   210X.12410.
926
927   León, J., Jones, W. & Morgan, D. (2009). Population genetic structure of *Homalodisca*
928   *coagulata* (Homoptera: Cicadellidae), the vector of the bacterium *Xylella fastidiosa* causing
929   Pierce's disease in grapevines. *Annals of the Entomological Society of America, 97*(4), 809-818.
930
931   Linck, E., & Battey, C. J. (2019). Minor allele frequency thresholds strongly affect population
932   structure inference with genomic data sets, *Molecular Ecology Resources, 19*(3), 639–647. doi:
933   10.1111/1755-0998.12995.
934
935   Lis, A., Maryańska-Nadachowska, A., & Kajtoch, Ł. (2015). Relations of *Wolbachia* infection
936   with phylogeography of *Philaenus spumarius* (Hemiptera: Aphrophoridae) populations within
937   and beyond the Carpathian contact zone. *Microbial Ecology*, *70*(2), 509–521. doi:
938   10.1007/s00248-015-0570-2.
939
940   Lis, A., Maryańska-Nadachowska, A., Lachowska-Cierlik, D., & Kajtoch, L. (2014). The
941   secondary contact zone of phylogenetic lineages of the *Philaenus spumarius* (Hemiptera:
942   Aphrophoridae): an example of incomplete allopatric speciation. *Journal of Insect Science*, *14*,
943   227. doi: 10.1093/jisesa/ieu089.
944
945   Lischer, H. E. L., & Excoffier, L. (2012) PGDSpider: An automated data conversion tool for
946   connecting population genetics and genomics programs. *Bioinformatics, 28*, 298–299. doi:
947   10.1093/bioinformatics/btr642.
948
949   Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer,
950   A. (2016). Breaking RAD: An evaluation of the utility of restriction site associated DNA
951   sequencing for genome scans of adaptation. *Molecular Ecology Resources*, *17*, 142–152. doi:
952   10.1111/1755-0998.12596
953
954   Malinsky, M., Challis, R. J., Tyers, A. M., Schiffels, S., Terai, Y., Ngatunga, B. P., Miska, E. A.,
955   Durbin, R., Genner, M. J., & Turner, G. F. (2015). Genomic islands of speciation separate cichlid

956 ecomorphs in an East African crater lake. *Science (New York, N.Y.)*, *350*(6267), 1493–1498.
957 https://doi.org/10.1126/science.aac9927
958
959
960 Maryańska-Nadachowska, A., Drosopoulos, S., Lachowska, D., Kajtoch, L., & Kuznetsova, V.
961 G. (2010). Molecular phylogeny of the Mediterranean species of *Philaenus* (Hemiptera:
962 Auchenorrhyncha: Aphrophoridae) using mitochondrial and nuclear DNA sequences. *Systematic*
963 *Entomology*, *35*, 318–328. doi: 10.1111/j.1365-3113.2009.00510.x.
964
965 Maryańska-Nadachowska, A., Kajtoch, Ł., & Lachowska, D. (2011). Genetic diversity of
966 *Philaenus spumarius* and *P. tesselatus* (Hemiptera, Aphrophoridae): implications for evolution
967 and taxonomy. *Systematic Entomology*, *37*(1), 55–64. doi: 10.1111/j.1365-3113.2011.00592.x.
968
969 Masly, J. P. (2012). 170 Years of "Lock-and-Key": genital morphology and reproductive
970 isolation. *International Journal of Evolutionary Biology*, *2012*, 247352. doi:
971 10.1155/2012/247352.
972
973 Nadeau, N. J., Whibley, A., Jones, R. T., Davey, J. W., Dasmahapatra, K. K., Baxter, S. W., …
974 Jiggins, C. D. (2012). Genomic islands of divergence in hybridizing *Heliconius* butterflies
975 identified by large-scale targeted sequencing. *Philosophical Transactions of the Royal Society B:*
976 *Biological Sciences*, *367*, 343–353. doi: 10.1098/rstb.2011.0198.
977
978 Narita, S., Nomura, M., Kato, Y., & Fukatsu, T. (2006). Genetic structure of sibling butterfly
979 species affected by *Wolbachia* infection sweep : evolutionary and biogeographical implications.
980 *Molecular Ecology*, *15*, 1095–1108. doi: 10.1111/j.1365-294X.2006.02857.x.
981
982 Nast, J. (1972). *Palaearctic Auchenorrhyncha (Homoptera): An Annotated Check List*. Institute
983 of Zoology, Polish Academy of Sciences, Warsaw: Polish Scientific Publisher, 550 pp.
984
985 Nosil, P. (2012). *Ecological Speciation*. Oxford Series in Ecology and Evolution. 300 pp.
986
987 Novembre, J., & Stephens, M. (2008). Interpreting principal component analyses of spatial
988 population genetic variation. *Nature Genetics*, *40*(5), 646–649. doi: 10.1038/ng.139.Interpreting.
989
990 Peccoud, J., Ollivier, A., Plantegenest, M., & Simon, J.-C. (2009). A continuum of genetic
991 divergence from sympatric host races to species in the pea aphid complex. *Proceedings of the*
992 *National Academy of Sciences of the United States of America*, *106*(18), 7495–7500. doi:
993 10.1073/pnas.0811117106.
994

995   Pélissié, B., Crossley, M. S., Cohen, Z. P., & Schoville, S. D. (2018). Rapid evolution in insect
996   pests: the importance of space and time in population genomics studies *Current Opinion in*
997   *Insect Science*, *26*, 8–16. doi: 10.1016/j.cois.2017.12.008.
998

999   Pina-Martins, F., Silva, D. N., Fino, J., & Paulo, O. S. (2017). Structure_threader: An improved
1000   method for automation and parallelization of programs structure, fastStructure and MavericK on
1001   multicore CPU systems. *Molecular Ecology Resources*, *17*(6), e268–e274. Doi: 10.1111/1755-
1002   0998.12702.
1003

1004   Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using
1005   multilocus genotype data. *Genetics*, *155*, 945–959.
1006

1007   R Core Team (2017). R: A language and environment for statistical computing. R Foundation for
1008   Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.
1009

1010   Redak, R. A., Purcell, A. H., Blua, M. J., Mizell, R. F., & Andersen, P. C. (2004). The biology of
1011   xylem fluid-feeding insect vectors of *Xylella fastidiosa* and their relation to disease
1012   epidemiology. *Annual Review of Entomology*, *49*, 243–270. doi:
1013   10.1146/annurev.ento.49.061802.123403.
1014

1015   Renaut, S., Maillet, N., Normandeau, E., Sauvage, C., Derome, N., Rogers, S. M., & Bernatchez,
1016   L. (2012). Genome-wide patterns of divergence during speciation: the lake whitefish case study.
1017   *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*, 354–363. doi:
1018   10.1098/rstb.2011.0197.
1019

1020   Riesch, R., Muschick, M., Villoutreix, R., Comeault, A. A., Farkas, T. E., Lucek, K., … Nosil, P.
1021   (2017). Transitions between phases of genomic differentiation during stick-insect speciation.
1022   *Nature Ecology & Evolution*, *1*, 0082. doi: 10.1038/s41559-017-0082.
1023

1024   Rodrigues, A. S. B., Silva, S. E., Marabuto, E., Silva, D. N., Wilson, M. R., Thompson, V., …
1025   Seabra, S. G. (2014). New mitochondrial and nuclear evidences support recent demographic
1026   expansion and an atypical phylogeographic pattern in the spittlebug *Philaenus spumarius*
1027   (Hemiptera, Aphrophoridae). *PLoS ONE*, *9*(6), e98375. doi: 10.1371/journal.pone.0098375.
1028

1029   Rodrigues, A. S. B., Silva, S. E., Pina-Martins, F., Loureiro, J., Castro, M., Gharbi, K., …
1030   Seabra, S. G. (2016). Assessing genotype-phenotype associations in three dorsal colour morphs
1031   in the meadow spittlebug Philaenus spumarius (L.) (Hemiptera: Aphrophoridae) using genomic
1032   and transcriptomic resources. *BMC Genetics*, *17*(1), 144. doi: 10.1186/s12863-016-0455-5.
1033

1034 Rousset, F. (1997). Genetic differentiation and estimation of gene flow from F-Statistics under
1035 isolation by distance. *Genetics*, *145*, 1219–1228.
1036
1037 Rousset, F., & Solignac, M. (1995). Evolution of single and double *Wolbachia* symbioses during
1038 speciation in the *Drosophila simulans* complex. *Proceedings of the National Academy of*
1039 *Sciences of the United States of America*, *92*, 6389–6393.
1040
1041 Saponari, M., Loconsole, G., Cornara, D., Yokomi, R. K., Stradis, A. D. E., Boscia, D., …
1042 Porcelli, F. (2014). Infectivity and transmission of *Xylella fastidiosa* by *Philaenus spumarius*
1043 (Hemiptera: Aphrophoridae) in Apulia, Italy. *Journal of Economic Entomology*, *107*(4), 1–4. doi:
1044 10.1603/ec14142.
1045
1046 Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., …
1047 Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews Genetics*, *15*(3), 176–
1048 192. doi: 10.1038/nrg3644.
1049
1050 Sicard, A., Zeilinger, A. R., Vanhove, M., Schartel, T. E., Beal, D. J., Daugherty, M. P., &
1051 Almeida, R. P. P. (2018). *Xylella fastidiosa*: insights into an emerging plant pathogen. *Annual*
1052 *Review of Phytopathology*, *56*, 181–202. doi: 10.1146/annurev-phyto-080417-045849.
1053
1054 Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., & Flook, P. (1994). Evolution,
1055 weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of
1056 conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*,
1057 *87*, 651–701.
1058
1059 Smadja, C. M., & Butlin, R. K. (2011). A framework for comparing processes of speciation in
1060 the presence of gene flow. *Molecular Ecology*, *20*, 5123–5140. doi: 10.1111/j.1365-
1061 294X.2011.05350.x.
1062
1063 Smith P. T. (2005). Mitochondrial DNA variation among populations of the glassy-winged
1064 sharpshooter, *Homalodisca coagulata*. *Journal of insect science (Online)*, *5*, 41.
1065 https://doi.org/10.1093/jis/5.1.41
1066
1067 Sousa, V., & Hey, J. (2013). Understanding the origin of species with genome-scale data:
1068 modelling gene flow. *Nature Reviews. Genetics*, *14*(6), 404–14. doi: 10.1038/nrg3446.
1069
1070 Thompson, J. D., Higgins, D. G., Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity
1071 of progressive multiple sequence alignment through sequence weighting, position-specific gap
1072 penalties and weight matrix choice. *Nucleic Acids Research*, *22,* 4673–4680. doi:
1073 10.1093/nar/22.22.4673.
1074

1075 Tonzo, V., Papadopoulou, A., & Ortego, J. (2019). Genomic data reveal deep genetic structure
1076 but no support for current taxonomic designation in a grasshopper species complex. *Molecular*
1077 *Ecology*, *28*(17), 3869–3886. doi: 10.1111/mec.15189.
1078
1079 Turner, T. L., Hahn, M. W., & Nuzhdin, S. V. (2005). Genomic islands of speciation in
1080 *Anopheles gambiae. PLoS Biology*, *3*(9), e285. doi: 10.1371/journal.pbio.0030285.
1081
1082 Vijay, N., Bossu, C. M., Poelstra, J. W., Weissensteiner, M. H., Suh, A., Kryukov, A. P., &
1083 Wolf, J. B. W. (2016). Evolution of heterogeneous genome differentiation across multiple
1084 contact zones in a crow species complex. *Nature Communications*, *7*, 1–10.
1085 https://doi.org/10.1038/ncomms13195
1086
1087 Vitalis, R., Gautier, M., Dawson, K. J., & Beaumont, M. A. (2014). Detecting and measuring
1088 selection from gene frequency data. *Genetics*, *196*(3), 799–817. doi:
1089 10.1534/genetics.113.152991.
1090
1091 Wagner, W. (1955). Die Bewertung morphologischer Merkmale in den unteren taxonomischen
1092 Kategorien, aufgezeigt an Beispielen aus der Taxionomie der Zikaden. Mitt. Hamburg. Zool.
1093 Mus. Inst. 53, 75–108.
1094
1095 Wagner, W. (1959). Homoptera. In *Zoologische Studien in Westgriechenland. IX* (pp. 583–605).
1096 Wien.
1097
1098 Weaver, C., & King, D., (1954). *Meadow Spittlebug*. Research Bulletin, Ohio Agricultural
1099 Experiment Station, 741, pp. 99
1100
1101 Werren, J. H., Baldo, L., & Clark, M. E. (2008). *Wolbachia*: master manipulators of invertebrate
1102 biology. *Nature Reviews. Microbiology*, *6*, 741–751. doi: 10.1038/nrmicro1969.
1103
1104 Whitworth, T. L., Dawson, R. D., Magalon, H., & Baudry, E. (2007). DNA barcoding cannot
1105 reliably identify species of the blowfly genus *Protocalliphora* (Diptera: Calliphoridae).
1106 *Proceedings of the Royal Society B: Biological Sciences*, *274*, 1731–1739. doi:
1107 10.1098/rspb.2007.0062.
1108
1109 Wickham, H. (2016). *ggplot2: elegant graphics for data analysis*. Springer: Verlag, New York.
1110
1111 Wolf, J. B. W., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in
1112 light of speciation. *Nature Reviews Genetics*, *18*, 87–100. doi: 10.1038/nrg.2016.133.
1113

1114   Wu, C.-I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology*,
1115   *14*, 851–865.
1116
1117   Yoon, J.-.I., Hrynkiv, V., Morano, L., Nguyen, A. T., Wilder, S., Mitchell, F. (2014).
1118   Mathematical modelling of glassy-winged sharpshooter population. *Mathematical Biosciences*
1119   *and Engineering,* 11, 667-677.
1120
1121   Yurtsever, S. (2002). Hybrid crosses of the meadow spittlebug *Philaenus spumarius* (L.)
1122   (Homoptera: Cercopidae) between New Zealand and Welsh populations. *New Zealand Journal*
1123   *of Zoology*, *29*, 245–251.
1124
1125   Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-
1126   performance computing toolset for relatedness and principal component analysis of SNP data.
1127   *Bioinformatics*, *28*(24), 3326–3328. doi: 10.1093/bioinformatics/bts606.
1128
1129
1130

Figure 1

Map with the sampling locations of *Philaenus.*

The points indicate the sampling locations of *Philaenus spumarius* in Turkey (TUR), Greece (GRE), Finland (FIN), France (FRAN), Portugal (POR), United Kingdom (UK), Azores (AZO) and United States of America (USA), and of *Philaenus tesselatus* in Morocco (MOR). In Morocco, three locations were sampled (details in Supplementary Table S1).

Figure 2

Schematic representation of the aedeagus of Philaenus with morphometric characteres measured

(A) Morphometric characters measured on the aedeagus of *Philaenus* spittlebugs: *TotLen* – total length of aedeagus; *LowLen* – length of lower appendages (left and right); *MidLen* – length of middle appendages (left and right); *UpLen* – length of upper appendages (left and right); *UpCur* – curvature of upper appendages (left and right). **(B)** Diagram of geometric measurements of the curvature of the aedeagus upper appendages.

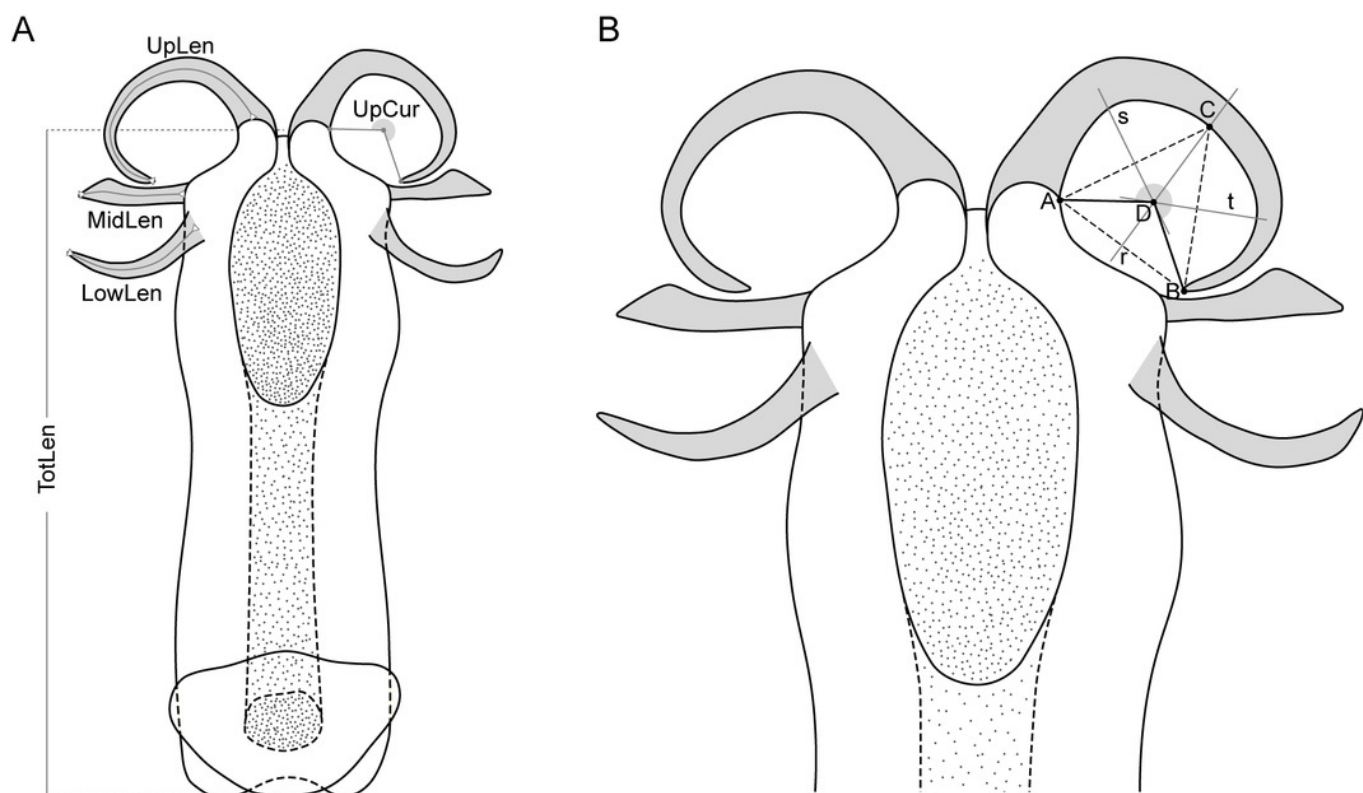



Figure 3

Principal component analysis of morphometric data of male aedeagus and of RAD-seq data.

Scatterplots of the three first principal components (PC1, PC2 nd PC3) from the Principal Component Analysis (PCA) of morphometric data of male aedeagus (A and B) and of RAD-seq data (C and D).

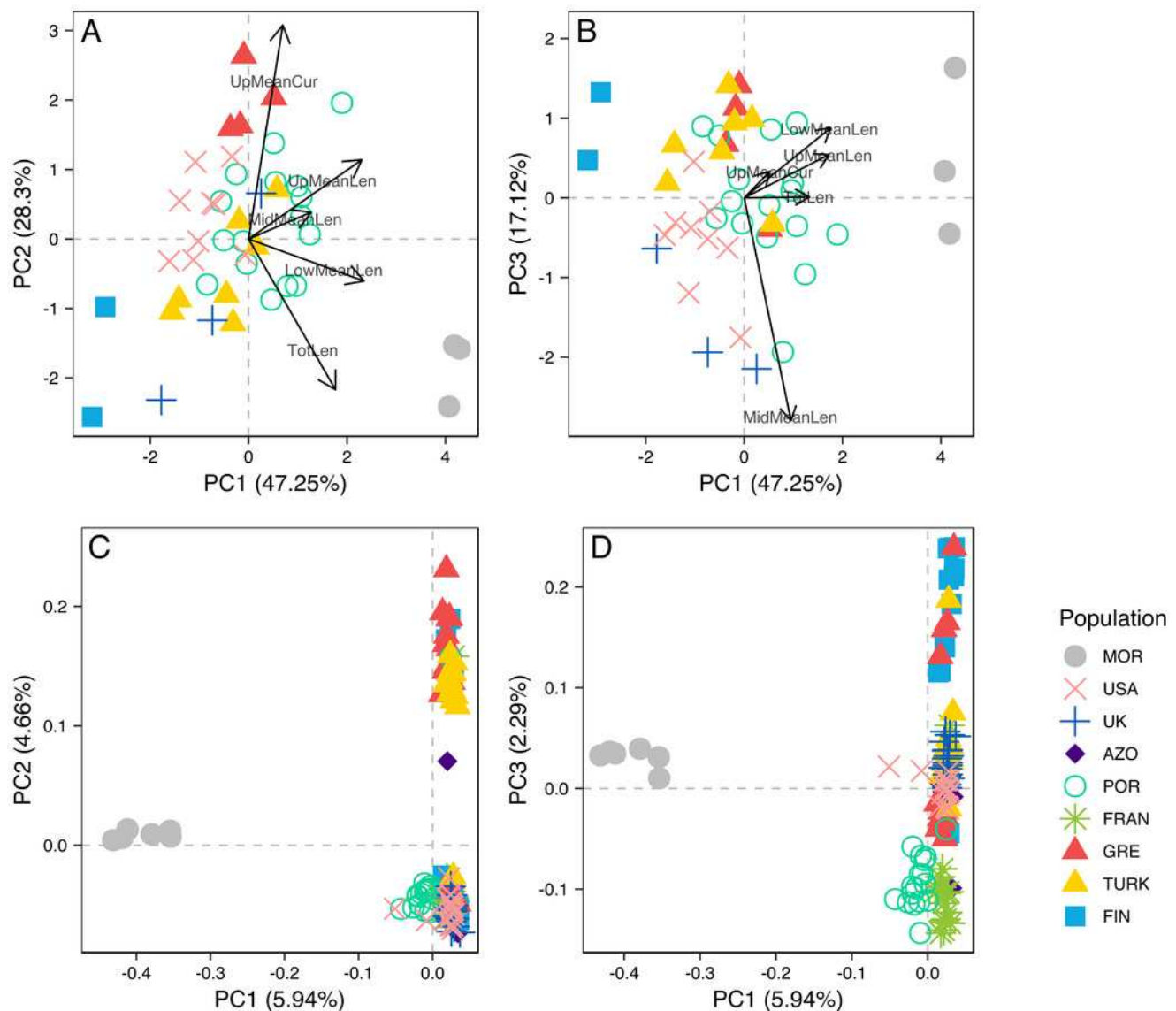



Figure 4

STRUCTURE results for the complete dataset with 9 populations

STRUCTURE results for the dataset including all populations, for K=2 to K=5. The best K according to Evanno et al. (2005) method was K=4. The colours of the major groupings in STRUCTURE were chosen to correspond loosely to the mitochondrial DNA haplogroups (Rodrigues et al., 2014 and this study), for a better visualization.

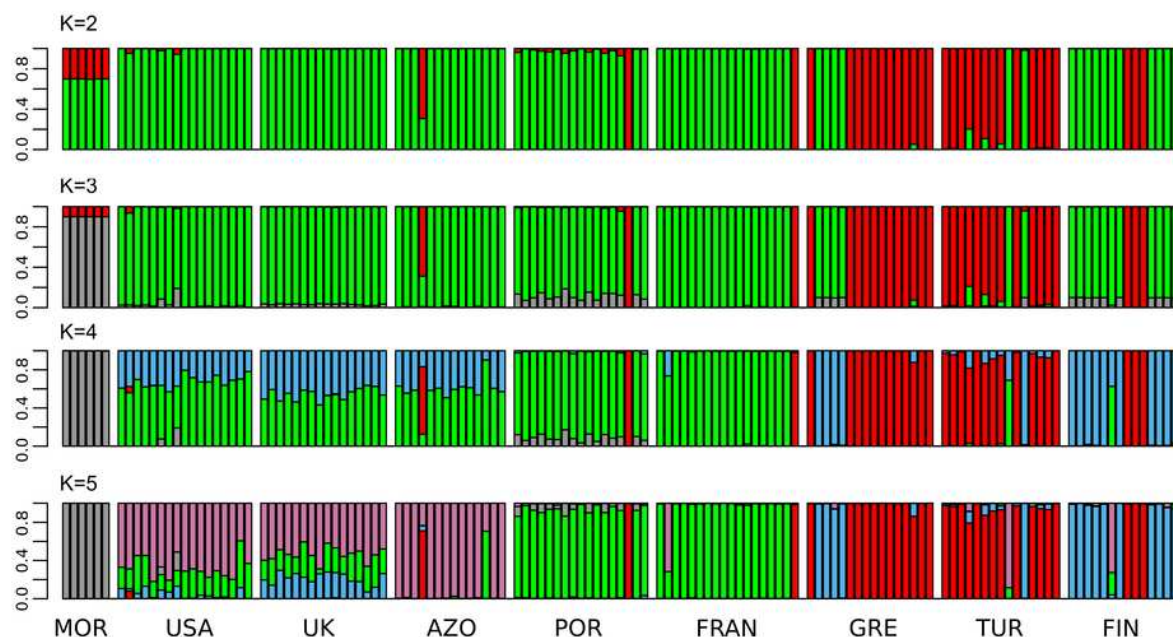



Figure 5

Geographical distance *versus* genetic distance for each pair of populations

Scatterplot of the geographical distance (natural logarithm) *versus* genetic distance ($F_{ST}/(1-F_{ST})$) for each pair of populations. Colours discriminate distances between: Morocco and the other populations (red); European populations (grey); transatlantic populations (USA or Azores) vs. European populations (green).

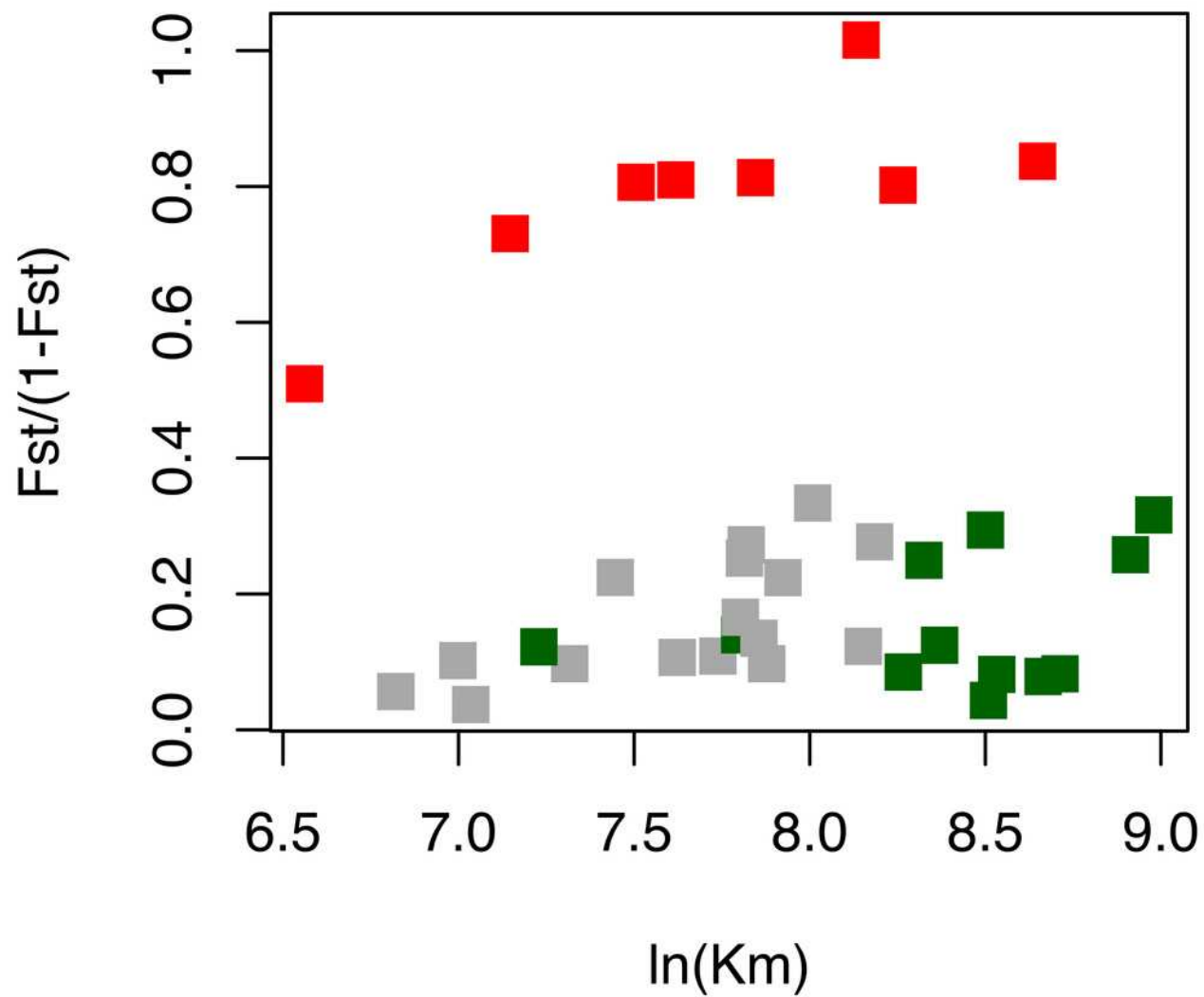
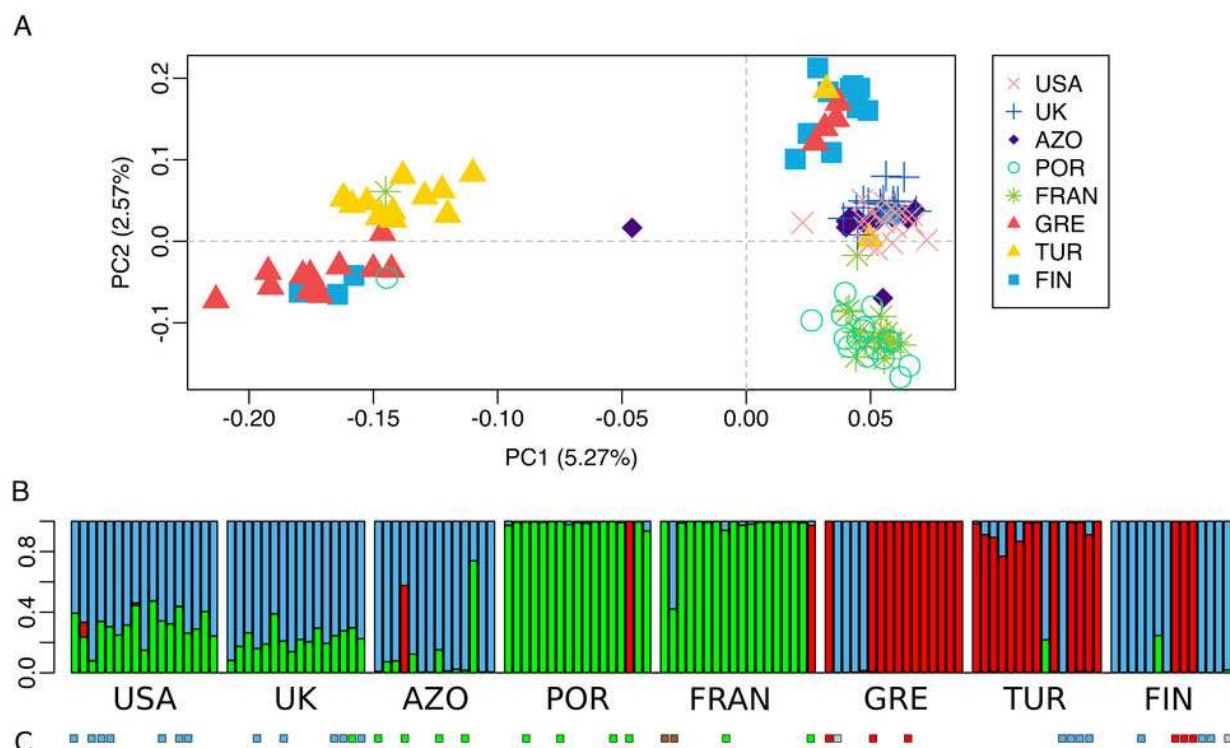



Figure 6

Population genetic structure analysis, excluding the North African population.

(A) Scatterplot of the two first principal components (PC1 and PC2) from the Principal Component Analysis of RAD-seq data for the 8 populations, after excluding the North African population; (B) STRUCTURE results for K=3 (best K according to Evanno et al., 2005). (C) Mitochondrial haplogroups present in each population (Rodrigues et al., 2014 and this study) shown in coloured squares (blue: "North-Eastern", green: "Western"; red: "Eastern-Mediterranean"). The colours of the major groupings in STRUCTURE were chosen to correspond loosely to the mitochondrial DNA haplogroups, for a better visualization

Pairwise F_{ST} matrix and estimates of expected and observed heterozygosity (H_E and H_O, respectively) and F_{IS} for each population.

The triangular matrix shows the F_{ST} values for each pair of populations and the bottom values are the estimates of expected and observed heterozygosity (H_E and H_O, respectively) and F_{IS} for each population.

	MOR	USA	UK	AZO	POR	FRAN	GRE	TUR	FIN
MOR	0								
USA	0.4557	0							
UK	0.4474	0.0415	0						
AZO	0.4464	0.0785	0.098	0					
POR	0.3375	0.0755	0.0887	0.1089	0				
FRAN	0.4222	0.0728	0.0925	0.1226	0.0359	0			
GRE	0.4484	0.205	0.2141	0.1998	0.1831	0.1833	0		
TUR	0.5039	0.2403	0.2504	0.2273	0.2168	0.2018	0.0533	0	
FIN	0.4451	0.0762	0.0964	0.1108	0.1094	0.1188	0.0887	0.1423	0

	MOR	USA	UK	AZO	POR	FRAN	GRE	TUR	FIN
He	0.0373	0.0708	0.0737	0.0641	0.0795	0.0725	0.0808	0.0789	0.0781
Ho	0.0258	0.046	0.0537	0.046	0.0406	0.0475	0.0360	0.0491	0.0396
FIS	0.3313	0.2927	0.2379	0.2599	0.4421	0.2962	0.4795	0.3544	0.4233