Population structure, adaptation and divergence of the meadow spittlebug, *Philaenus spumarius* (Hemiptera, Aphrophoridae), revealed by genomic and morphological data (#53180)

First revision

Guidance from your Editor

Please submit by **3 Apr 2021** for the benefit of the authors .

Structure and Criteria
Please read the 'Structure and Criteria' page for general guidance.

Custom checks
Make sure you include the custom checks shown below, in your review.

Raw data check
Review the raw data.

Image check
Check that figures and images have not been inappropriately manipulated.

Privacy reminder: If uploading an annotated PDF, remove identifiable information to remain anonymous.

Files

Download and review all files from the materials page.

1 Tracked changes manuscript(s)
1 Rebuttal letter(s)
6 Figure file(s)
2 Table file(s)
3 Raw data file(s)
1 Other file(s)

Custom checks

DNA data checks

- Have you checked the authors data deposition statement?
- Can you access the deposited data?
- Has the data been deposited correctly?
- Is the deposition information noted in the manuscript?

Field study

- Have you checked the authors field study permits?
- Are the field study permits appropriate?

Structure and Criteria

Structure your review

The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this PDF and upload it as part of your review

When ready submit online.

Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your guidance page.

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to PeerJ standards, discipline norm, or improved for clarity.
- Figures are relevant, high quality, well labelled & described.
- Raw data supplied (see PeerJ policy).

EXPERIMENTAL DESIGN

- Original primary research within Scope of the journal.
- Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- Impact and novelty not assessed. Negative/inconclusive results accepted. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- All underlying data have been provided; they are robust, statistically sound, & controlled.
- Speculation is welcome, but should be identified as such.
- Conclusions are well stated, linked to original research question & limited to supporting results.

Standout reviewing tips

The best reviewers use these techniques

Tip	*Example*
Support criticisms with evidence from the text or from other sources	*Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.*
Give specific suggestions on how to improve the manuscript	*Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).*
Comment on language and grammar issues	*The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult. I suggest you have a colleague who is proficient in English and familiar with the subject matter review your manuscript, or contact a professional editing service.*
Organize by importance of the issues, and number your points	*1. Your most important issue* *2. The next most important item* *3. …* *4. The least important points*
Please provide constructive criticism, and avoid personal opinions	*I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC*
Comment on strengths (as well as weaknesses) of the manuscript	*I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.*

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

Sofia G 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 G 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), presently the most important 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. This highlights the importance of using integrative approaches in taxonomy. 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-occurrence 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, 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

* 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

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), presently the most important 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. This highlights the importance of using integrative approaches in taxonomy. 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-occurrence 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

60 divergence, gene flow and adaptation in these spittlebugs and led to several hypotheses that
61 require further local investigation.

62

63 ## Introduction

64 Speciation involves the evolution of reproductive isolation and the buildup of genetic
65 differentiation through selection and drift, but gene flow can counteract such divergence by
66 homogenizing allelic variation and also by allowing recombination to oppose or break
67 associations between loci underlying isolating traits (Smadja & Butlin, 2011; Sousa & Hey,
68 2013). However, several mechanisms may favour divergence in the face of gene flow, such as
69 ecologically driven selection or sexual selection (Smadja & Butlin, 2011; Nosil, 2012).
70 According to the genic model of speciation, at the start of the speciation process, and in the
71 presence of gene flow, a few localized regions in the genome subject to divergent selection will
72 differentiate, while the remaining genome continues to be freely exchanged between populations
73 (Wu, 2001). Genome-wide analyses have allowed the detection of these "genomic islands" of
74 differentiation in several systems (e.g., Malinsky et al., 2015; Vijay et al., 2017), although other
75 processes not related to speciation or reproductive isolation may also be responsible for them,
76 such as linked selection, variable recombination rates and/or density of targets of selection (Wolf
77 & Ellegren, 2017). As populations diverge through the action of selection and drift, a genome-
78 wide differentiation will emerge and eventually lead to full reproductive isolation and
79 diversification. Designated species may thus lie somewhere in this "speciation continuum", with
80 different levels of divergence and gene flow (Hendry, Bolnick, Berner, & Peichel, 2009;
81 Peccoud, Ollivier, Plantegenest, & Simon, 2009; Renaut et al., 2012; Riesch et al., 2017).

82

83 Distinguishing taxa and understanding the patterns of gene flow and local adaptation in insect
84 species that transmit diseases are crucial for better management of those diseases (Busvine,
85 1980; Pélissié, Crossley, Cohen, & Schoville, 2018; Bahrndorff et al., 2020). *Philaenus*
86 *spumarius* (Linnaeus, 1758) (Insecta, Hemiptera, Auchenorrhyncha, Aphrophoridae), the
87 meadow spittlebug, is a xylem-feeding vector of *Xylella fastidiosa* Wells et al. 1987, a plant
88 pathogenic bacterium of South American origin that is emergent in Europe (Saponari et al.,
89 2014). Olive quick decline syndrome (OQDS), caused by *X. fastidiosa*, was first detected in
90 Apulia, southern Italy in 2013, where it soon became clear that *P. spumarius* [illegible] the most

important vector (Saponari et al., 2014; Cornara et al., 2017). Since then, *X. fastidiosa* has been detected in several other European countries and is a cause of major concern (EFSA et al., 2019). *X. fastidiosa* is native to the Americas, where it causes important diseases such as Pierce's disease of grapevine, citrus variegated chlorosis, almond leaf scorch and several others in perennial crops and ornamental plants (Baldi & La Porta, 2017). There the main vectors are sharpshooters (another xylem-feeding Auchenorrhyncha group, the Cicadellidae Cicadellinae), while spittlebugs appear to have a small but perhaps not negligible epidemiological importance (Almeida et al., 2019; Cornara et al., 2019; Beal et al., 2021). One of the main vectors of Pierce's disease of grapevines in California is the glassy-winged sharpshooter *Homalodisca coagulata* (Say). It is native to the southern United States and it became established in late 1990´s in California, being a costly invasive species to agriculture. Population genetic structure studies based on DNA fingerprinting and mitochondrial DNA on this species have revealed highly differentiated geographic groups in the natural range and indicated that the likely sources of the California insects were in Texas (Léon, Jones & Morgan, 2004; Smith, 2005), which led, for example, to further work in Texas for a better understanding of the natural population dynamics (Yoon et al., 2014). This demonstrates the potential importance of knowledge of the population genetic structure of *P. spumarius* for understanding the dynamics of the spread of *X. fastidiosa* in Europe. Since the vectors are the only means of natural dissemination of *X. fastidiosa* (Sicard et al., 2018), this information is crucial for the successful management of this pathogen and should be included in models of risk assessment (EFSA Panel on Plant Health, 2015).

P. spumarius is a polyphagous xylem-feeding insect, widespread in the Holarctic, whose nymphs produce a protective foam (spittle masses) from their liquid excretion. Humidity and temperature are particularly limiting in the earlier nymphal stages (Weaver & King, 1954). In general, adults live during one reproductive season in spring/summer, and then at the end of summer/autumn the females oviposit and the eggs overwinter in the vegetation until they hatch in the following spring/summer (Halkka & Halkka, 1990). This species is thought to have a Palaearctic origin, and to have recently colonised North America, the Azorean islands, Hawaii and New Zealand. These introductions were likely mediated by humans (Halkka & Halkka, 1990; Rodrigues et al., 2014), although natural colonisation cannot be excluded for the S. Miguel island in the Azores (Borges et al., 2018; Rodrigues et al., 2014) as the populations in this island are restricted to to high elevation native vegetation of the oriental and geologically oldest part of this island. In parts of North America it has been a crop pest (Weaver & King, 1954), but surveys

of the spittle masses along coastal California have revealed a recent population decline of this species, very accentuated in some places (Karban & Strauss, 2004; Karban & Huntzinger, 2018) and it has also declined the Wonalancet, New Hampshire population sampled for this report (V Thompson, unpublished data).

Previous studies based on mitochondrial and nuclear DNA genes have revealed the major phylogeographic patterns in *Philaenus spumarius* (Maryańska-Nadachowska, Kajtoch & Lachowska, 2011; Rodrigues et al., 2014). Two main mitochondrial lineages have initially diverged during the Pleistocene: the "Western", currently found in the Mediterranean region and also in Central and Northern Europe, and the "North-Eastern", currently found from Eastern Asia to Central and Northern Europe. The "Western" lineage is further differentiated into sublineages: the ''Western'', predominant in the Iberian Peninsula but also in western parts of Central and Northern Europe; and the ''Eastern-Mediterranean'' present in the Balkans and Middle East and others around Black Sea and in the Caucasus (Maryańska-Nadachowska et al., 2011; Rodrigues et al., 2014). These lineages also co-occur in several contact zones (Lis et al., 2014; Rodrigues et al., 2014). The occurrence of different infection rates of the maternally inherited endosymbiont *Wolbachia* in the different lineages of *P. spumarius* has pointed to a possible mechanism to explain the maintenance of genetic differentiation in the Carpathians contact zone (Lis et al., 2015). Previous studies have also revealed close relationships and even shared haplotypes between samples from North America, Azores, New Zealand and those from Great Britain (Rodrigues et al., 2014), indicating a recent human-assisted colonization, as previously suggested for North America and New Zealand (Hamilton, 1979; Yurtsever, 2002).

Seven other species of the genus *Philaenus* occur in the Mediterranean area, having a much more restricted distribution ranges, which partially overlapping that of *P. spumarius* (Drosopoulos, 2003; Maryańska-Nadachowska, Drosopoulos et al., 2010). One such species is *P. tesselatus* Melichar, 1889, which was originally described from Tunisia and was later synonymized with *P. spumarius* (Nast, 1972), being considered a geographic subspecies. Later the synonymy was re-assessed based on morphological evaluation, with the best diagnostic characters being the size and shape of the appendages of the male aedeagus (Drosopoulos & Quartau, 2002). However, geographic variation in the curvature of the aedeagal apical appendages in *P. spumarius* has been

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

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

Materials & Methods

Sampling

Adults and nymphs of *Philaenus spumarius* were collected in 2010 and 2011 from eight populations (Figure 1) across the distribution range of the species: Cerkes, Anatolia, Turkey (TUR); Mount Parnassus, Greece (GRE); Haapanaki-Keuruu, Finland (FIN); Fitou, South of France (FRAN); Gouveia/Fontanelas, Sintra, Portugal (POR); Aberdare, South Wales, United Kingdom (UK); S. Miguel island, Azores (AZO); Wonalancet, New Hampshire, United States of America (USA). *P. tesselatus* was sampled in Morocco (MOR), from three main localities, near Azrou, near Rabat and near Ceuta. In total, 170 specimens were sequenced, including 20 – 22 individuals from each sampling site except Morocco, from where only 7 specimens were included (Supplementary Table S1). Sampling in this last location was initially intended for a phylogeographic characterization (Rodrigues et al., 2014) and only a few individuals were collected from each site.

Nymphs were hand collected from the spittle masses they produce. Adults were collected by sweeping the vegetation with an entomological net. Both nymphs and adults occur and feed on large numbers of host plant species and no particular hosts were selected during sweeping and hand collection. Efforts by others to show associations between common hosts and the genetically determined color polymorphism gave negative results (Halkka et al., 1967) and we know of no evidence suggesting host-specific genetic differentiation.

Insects were preserved in absolute or 96% ethanol until they were subjected to DNA extraction after up to one year.

Morphological characters

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

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

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

DNA extraction and mitochondrial DNA analyses

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

A subset of 48 specimens from the nine areas were sequenced for mitochondrial DNA (Supplementary Table S1) from which we amplified an 800 bp fragment of the 3'-end of the mitochondrial gene cytochrome *c* oxidase subunit I (COI) by polymerase chain reaction (PCR). Primers used were: C1-J-2195 (5'–TTGATTTTTTGGTCATCCAGAAGT–3') and TL2-N-3014 (5'–TCCAATGCACTAATCTGCCATATTA–3') (Simon et al., 1994). PCR was performed in a 12.5 uL reaction volume containing: 1x Colorless GoTaq Flexi Buffer, 2 mM $MgCl_2$, 0.2 mM dNTPs, 0.6 mg/ml of BSA, 0.5 µM of each primer, 0.0375 U GoTaq DNA Polimerase (Promega) and approximately 30 ng of DNA. PCR conditions were: an initial denaturation step

244 at 95ºC for 5 min, followed by 35 cycles of denaturation at 95ºC for 45 s, annealing at 50ºC for
245 35 s and extension at 72ºC for 2 min, with a final extension period at 72ºC for 10 min.
246
247 Chromatograms were verified and edited using SEQUENCHER v. 4.0.5 (Gene Codes
248 Corporation), they were aligned using CLUSTAL W on BIOEDIT v. 7.0.9 (Thompson, Higgins,
249 & Gibson, 1994; Hall, 1999) and subsequently trimmed to the same length. We followed the
250 designation of haplotypes of Rodrigues et al. (2014). A median-joining haplotype network was
251 constructed using PopART version 1.7 (Bandelt, Forster, & Rohl, 1999; Leigh & Bryant, 2015).
252

253 *RAD libraries preparation and sequencing*
254 RAD libraries were prepared using a protocol by Etter et al. (2011), with modifications as in
255 Rodrigues et al. (2016). The restriction enzyme used was *SbfI* (New England Biolabs). Six
256 libraries were prepared, with 28 to 31 individually barcoded samples multiplexed. The libraries
257 were sequenced on three lanes of Illumina HiSeq 2000 in paired-end mode (2 x 100 bp) at
258 Genepool (Ashworth Laboratories) (http://genomics.ed.ac.uk/). The individuals from each
259 population were distributed over the different libraries and lanes to avoid library or lane-specific
260 biases.
261

262 *Assembly and SNP calling*
263 The sequence reads from each run were examined by *process_radtags* from STACKS version
264 1.45 (Catchen et al., 2013), to remove those with uncalled bases and low quality scores (phred
265 score lower than 10), to check that the barcode and restriction site were intact in each read and to
266 demultiplex the samples based on the barcode identification. Reads were trimmed at the 3' end,
267 using TRIMMOMATIC v. 0.38, to keep only 87 bases, since preliminary analyses using the
268 entire read revealed a high number of (possibly false) SNPs at the 3' end after this number of
269 bases (data not shown). This may be due to higher sequencing errors towards the end of the reads
270 characteristic of Illumina sequencing (Dohm et al., 2008). The trimmed reads were *de novo*
271 assembled into "stacks" (identical sets of reads, called loci) for each individual using the
272 STACKS module *ustacks*. The minimum depth of coverage to build a stack (-m) was set to 10,
273 the maximum number of nucleotide differences allowed between stacks to form a locus (-M) was

set to 2. Then the Stacks module *cstacks* was used to build a catalog by merging stacks (loci) from multiple individuals, using the default options. The module *sstacks* was used to match loci from an individual against the catalog. Stacks with very high coverage were removed since they may represent highly repetitive regions and that may include non-orthologous sequences.

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

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

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

VCFtools was used to calculate summary statistics of coverage and percentage of missing data. Genetix v. 4.05.2 was used to obtain expected and observed heterozygosity, as well as F_{IS} in each population. Pairwise differentiation between populations (F_{ST}) were calculated in Arlequin 3.5.1.3, and the significance of F_{ST} was obtained from permutation tests with 10000 repetitions. Mantel tests between $F_{ST}/(1-F_{ST})$ matrices and the natural logarithm of the geographical distance

(Rousset, 1997) were performed with ape package version 5.0 (http://ape-package.ird.fr/) in R version 3.4.0, using 9999 permutations.

Population structure

Principal Components Analysis (PCA) was used as an exploratory tool of the population structure (Novembre & Stephens, 2008). Computations of PCA were performed in R using package SNPRelate version 1.12.0 (Zheng et al., 2012). Population structure was further examined using the model-based clustering algorithm implemented in STRUCTURE v. 2.3.4 (Falush, Stephens, & Pritchard, 2003; Pritchard, Stephens, & Donnelly, 2000). We obtained the coefficients of ancestry using the admixture model and assuming correlated allele frequencies among populations, and K from 1 to 9, with 10 replicate runs of each, applying 50,000 steps of burnin and 1,000,000 MCMC steps after burnin. Structure_threader version 1.2.2 (Pina-Martins et al., 2017) was used to parallelize the runs and to find the K best explaining the data by calculating Delta K on STRUCTURE HARVESTER (Evanno, Regnaut, & Goudet, 2005; Earl & vonHoldt, 2012). Clumpp version 1.1.2 (Jakobsson & Rosenberg, 2007) was then used to obtain the optimal alignment of ancestry proportions, by permuting the 10 replicate runs of STRUCTURE for each value of K.

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

Detection of selection – outlier analyses and environmental associations

In order to detect loci with signs of selection for the *P. spumarius* RAD-seq dataset (without Morocco), two approaches were taken: one that detects outlier loci departing from expectation under neutral demographic models (Foll & Gaggiotti, 2008; Vitalis et al., 2014), and another that

detects loci associated with environmental variation between populations (Coop et al., 2010; Gautier, 2015).

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

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

After finding the candidate SNPs for selection and environmental association, two datasets were created: a “neutral” dataset, for which we excluded the candidate SNPs, and a “candidate”

dataset, that contained only the candidate SNPs. STRUCTURE analyses were also performed on these two datasets.

RAD tags with candidate SNPs were queried against the available *P. spumarius* partial draft genome and transcriptome (Rodrigues et al., 2016), using blastn with an e-value threshold of 1E-30. We obtained a longer sequence (100 bp extended from each end of the RAD tag) from the genome alignment, which was then queried against the NCBI nucleotide database (nr/nt) using BLASTN version 2.9.0 (Altschul et al., 1997), setting a threshold e-value of 1E-5.

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

Results

Morphology of male aedeagus

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

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

Mitochondrial DNA

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

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

427

428 *RAD sequencing*

429 A total of 838,730,936 reads was obtained from the Illumina sequencing. The *process_radtags*
430 step in Stacks retained 647,870,180 reads. This corresponds to an average of 3,811,001 ±
431 3,524,799 (standard deviation) reads per individual. Thirty-seven individuals with lower
432 numbers of reads (<500,000 reads) or large amounts of missing data (>60%) were excluded from
433 the analysis (Supplementary Table S1), leaving a total of 133 individuals, for which the number
434 of reads ranged from 736,248 to 23,798,148 (average 4,507,853 ± 3,668,879 sd). Raw reads after
435 demultiplexing were deposited in SRA database with accession PRJNA606428. The *population*
436 Stacks module, followed by filtering, produced 1691 SNPs, with a mean coverage of 105.5 reads
437 per locus per individual (Supplementary Figure S4) and mean percentage of missing data per
438 individual of 12.3 % (Supplementary Table S1). We applied relatively stringent filtering criteria
439 to avoid having large amounts of missing data per individual resulting from the large genome
440 size of this species. This has produced a relatively small number of SNPs but that have a good
441 representation across individuals and that are expected to be scattered across the genome. Since
442 the draft genome is still incomplete and very scattered we were not able to assess this
443 distribution.
444

445 *Population structure*

446 Principal Component 1 in the PCA clusters Morocco individuals away from the others (Figure
447 3C and D). When testing other assembly and filter parameters we obtained similar patterns in the
448 groupings of samples (Supplementary Figure S5). Also, Structure analysis gave support to a
449 genetic group solely comprising Moroccan samples. Best K according to Evanno et al. 2005 was
450 4; Figure 4). The average F_{ST} of Morocco *versus* other populations was 0.4, much higher than
451 average F_{ST} of other populations' comparisons (0.13) (Table 1). In all population-pairwise F_{ST}
452 calculations involving Morocco, there were a considerable number of SNPs that were fixed or
453 nearly fixed for one allele in Morocco and for the other allele in all the other populations, as seen
454 in the relatively high frequency of high F_{ST} values on the histograms in all comparisons and in
455 the high correlations between F_{ST} values among population pairs (Supplementary Figure S6a).
456 There was neither such a high number of fixed SNPs nor such high correlations between F_{ST}

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

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

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

The dataset without Moroccan samples consisted of 127 individuals and 2083 SNPs. For this dataset, PCA revealed two distinct clusters along PC1, generally separating Greece and Turkey from the remaining populations (Figure 6A). The latter were separated along PC2 in three groups, one including mainly Portugal and France, another including mainly Finland. A third

one, in between these two, including USA, UK and the Azores (Figure 6A). This structure had already been detected in PC3 of the analysis of the dataset that included Morocco (Figure 3D). For this dataset, when excluding the USA and the Azores populations, there was again significant isolation-by-distance for the European populations (Mantel test, z=19.49793, p=0.0229). However, there were a few individuals that were genetically more similar to geographically more distant individuals, which is also seen in the STRUCTURE analysis (Figure 6). In Greece and Finland, there were no admixed individuals between the two main clusters ("Eastern-Mediterranean" in red, and "North-Eastern" in blue) – they were either from one or the other group, with a few exceptions (Figure 6B). An analysis of the Turkish population revealed the presence of possibly admixed individuals from these two groups (with a smaller contribution from the "North-Eastern" cluster). In the USA and UK populations, all individuals showed some level of admixture between the "North-Eastern" (blue) and the "Western" (green) clusters. The Azores allies a small contribution from the "Western" group to a major one belonging to the "North-Eastern" group and one individual was admixed between "North-Eastern" and "Eastern-Mediterranean" (also seen in PCA). The admixture in USA, UK and the Azores is also apparent from their intermediate position between the Portugal+France group and the Finland+Greece+Turkey group in the PCA (Figure 6A). The best K in the STRUCTURE analysis, according to the method by Evanno et al. (2005), was 3.

The majority of specimens for which COI sequence was available, had a correspondence between the mtDNA and the genomic cluster. However, there were some specimens showing a mismatch consisting of a mtDNA haplotype belonging in a different genomic cluster (Figure 6B and C; Supplementary Figure S3). For example, one individual from UK (UK6) bearing a mtDNA haplotype (H24) belonging to "Western" haplogroup (green), turned up "North-Eastern" (blue) in the genome analysis. Two individuals from France, bearing mtDNA H49 haplotype (intermediate between "Western" and "Eastern-Mediterranean"), came up as differentiated at genomic markers, one "Western" (green) and the other intermediate "Western" / "North-Eastern". In Greece and Finland, COI sequenced individuals show both a mtDNA and genomic makeup belonging to either "Eastern-Mediterranean" or "North-Eastern", except two individuals assessed as "North-Eastern" in genomic markers but with a mtDNA haplotype in intermediate position in the network, between "Western" and "North-Eastern" haplogroups. The four

519 individuals from Turkey sequenced for COI belonged in “North-Eastern” haplotypes. While one
520 of them had full ancestry from “North-Eastern” group, the other three had their largest
521 proportion of ancestry from “Eastern-Mediterranean”, based on the genome-wide markers.
522
523 For samples for which both morphometric and RAD-seq data was available (N=32), we
524 computed Principal Components Analysis (Supplementary Figure S8) and calculated the
525 correlation between PC1 and PC2 scores for both analyses. There was a significant correlation
526 between PC1 from morphometry and PC2 from RAD-seq (r_S=0.63, p=1E-04), while all the
527 remaining were low and non significant (r_S=-0.22, p=0.2194 between PC1 from each; r_S=0.1,
528 p=0.5927 between PC2 from each; r_S=0.29, p=0.102 between PC2 from morphometry and PC1
529 from RAD-seq).
530
531

532 *Detection of selection – outlier analysis and environmental associations*
533 Candidate SNPs for positive selection were identified by detection of highly differentiated
534 outliers: eight were detected by Bayescan (Supplementary Figure S9); and 25 by KLD (quantile
535 99% KLD 2.037087) in SelEstim (Supplementary Figure S10; Supplementary Table S3). Six
536 outlier SNPs were common to both analyses. No outlier SNPs for balancing selection were
537 detected in the Bayescan analysis (Supplementary Figure S9).
538
539 The BayPass analysis detected 163 SNPs associated with environmental variables (BF > 15)
540 (Supplementary Table S3 and S4; Supplementary Figure S11), one of them common to the
541 candidate SNPs detected with Bayescan. Variables showing association were: Longitude,
542 Temperature Annual Range, Precipitation of Driest Quarter, Precipitation of Wettest Quarter,
543 Mean Temperature of Warmest Quarter, Mean Diurnal Range (Mean of monthly (max temp –
544 min temp)). Spearman correlations between these 6 variables were generally low (absolute
545 values below 0.6), with only two values above 0.6 (Supplementary Table S4c).
546
547 When excluding these candidate SNPs (188 in total) from the “full” dataset, creating a “neutral”
548 dataset, the main pattern of structuring was maintained, differing only in admixture proportions
549 at higher values of K (4 and 5) (PCA in Supplementary Figure S12 and Structure in

Supplementary Figure S13). When analysing the "candidates" dataset, the PCA showed a separation that corresponded generally to longitude variation along PC1 and to latitude along PC2 (Supplementary Figure S12). The STRUCTURE analysis, although artificial for the dataset of loci under selection, revealed similar structuring when compared to the other datasets, but with less admixture (Supplementary Figure S13). This is an expected outcome considering that these candidate loci have similar allelic variation within each population and different allelic variation between populations. The fact that there are still differentiated individuals within populations in this dataset, consistently assigned to the same groups as in the other datasets, is a reflection of the methods for detecting selection, based on population allelic variation.

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

Discussion

RAD sequencing analysis revealed the genetic distinction of North-African relative to other samples here analysed, which matched the morphological differences at the male genitalia, identifying these as *Philaenus tesselatus*. This genetic differentiation was however not detected at the mitochondrial DNA level, since *P. tesselatus* and *P. spumarius* share mtDNA haplotypes, as described in previous studies (Maryańska-Nadachowska et al., 2011; Rodrigues et al., 2014). These results thus reinforce the importance of taking an integrative approach when studying the taxonomy of a group of species, especially cryptic ones (Edwards & Knowles, 2014; Dejaco et al., 2016; Borges et al., 2017; Tonzo et al., 2019).

The fact that there are mitochondrial DNA haplotypes shared between *P. tesselatus* and *P. spumarius*, while the nuclear genome is differentiated, may indicate selection on mtDNA following introgressive hybridization (Gompert et al., 2008). One possible mechanism for selection on mtDNA described in several insects, is the occurrence of maternally inherited endosymbionts, including *Wolbachia*, associated with certain haplotypes. These endosymbionts are known to manipulate reproductive output, mainly through cytoplasmic incompatibility: no

viable offspring are produced when an infected male fertilizes an uninfected female, or a female infected with a different strain (Werren, Baldo, & Clark, 2008). Since both mitochondria and the symbiont are maternally transmitted, haplotypes associated with the *Wolbachia* infection could thus spread, hitchhiking through the population. Mitochondrial introgression between closely related species caused by *Wolbachia* has been described in several species of Diptera and Lepidoptera (Jiggins, 2003; Narita et al., 2006; Rousset & Soulignac, 1995; Whitworth et al., 2007). In these cases, different species share the same mitochondrial haplotypes, making DNA barcoding ineffective. Such a scenario would be interesting to investigate, as *Wolbachia* infection has already been detected in *P. spumarius* across Europe and North America (Lis et al., 2015; Kapantaidaki et al., 2021; Wheeler et al., 2021, unpublished data).

The admixture from the Moroccan genetic group detected in all the individuals from the Portuguese population (located in the Central-West part of the Iberian Peninsula) suggests some level of recent or ongoing gene-flow between *P. spumarius* and *P. tesselatus*. Despite previous doubts about the taxonomic status of these two taxa, our data point towards them being closely related but independent gene-pools, probably early in the speciation "continuum" (Seehausen et al., 2014). Both taxa co-occur in some locations in southern Iberian Peninsula (personal observation by J. A. Quartau and A. C. Neto, based on identification by male aedeagus morphology) and it will be important to study these sympatric areas. Genital traits are relevant since they may contribute to reproductive isolation, either structural or sensory, if differences in genital morphology between species prevent or reduce the success of copulation and insemination (Masly, 2012). Structural isolation has been shown, for example, in the species pair *Drosophila yakuba* and *D. santomea* (Kamimura & Mitsumoto, 2012), but in many species no convincing evidence for such isolation has been found so far (Masly, 2012). Morphological variation in female genitalia, as well as behavioural and physiological responses during mating may also aid in understanding potential mechanisms of reproductive isolation, particularly in sympatry. We recognise that species identification based on male genitalia characteristics may be insufficient when there is intraspecific variation with some overlap between species. Although our small *P. tesselatus* samples limit our understanding of the range of its variation, genomic data allow higher resolution in detecting genetic differentiation, but this is not enough to infer species status (Tonzo et al., 2019). A more comprehensive study on morphology, mtDNA and

genome-wide variation of a wider sample from the Mediterranean region of both *P. spumarius* and *P. tesselatus* is required.

The morphometric geographical variation detected in *P. spumarius* showed some correlation with genetic variation, although the nature of this association was not fully clear. Clinal latitudinal variation, as well as elevation variation, in the shape of male genitalia had already been described in European populations of *P. spumarius* (Wagner, 1955; 1959). Shorter and less curved upper appendages were found in the north compared to the south, and in higher than in lower altitudes in the same geographical regions (Wagner, 1955; 1959). RAD sequencing data permitted detection of finer population genetic structure within *P. spumarius* than previously known from mtDNA and a limited number of nuclear genes. Although there was a pattern of isolation-by-distance in European populations, there were clear distinctions between groups in the PCA and STRUCTURE analyses not related to geographical distance. The most likely K of three in STRUCTURE corresponded loosely to the three mitochondrial haplogroups already described in Rodrigues et al. (2014) and Maryańska-Nadachowska et al. (2011) but we detected some degree of admixture along contact zones. We found admixed individuals in France, Turkey and Finland, and several other individuals belonging to a different genetic group, with no admixture. This may suggest recent migration or the maintenance of reproductive barriers. In particular, there was almost no admixture between the “Eastern-Mediterranean” and the “North-Eastern” groups. Maryańska-Nadachowska et al. (2011) described a contact zone in the Carpathians between North-Eastern and South-Western haplogroups (this last group corresponds to our “Western” and “Eastern-Mediterranean” together) and detected heteroplasmic mitochondrial DNA, likely caused by paternal leakage from hybridization between members of these two clades. Interestingly, Lis, Maryańska-Nadachowska and Kajtoch (2015) have found different levels of *Wolbachia* infection between the different mitochondrial lineages of *P. spumarius*. The North-Eastern clade showed a higher proportion of infected individuals than the South-Western. In the Carpathian contact zone, infection was more prevalent in both groups, although they harboured different supergroups of *Wolbachia.* The authors suggest that there may be limited gene-flow between genetically distinct populations through a mechanism of cytoplasmic incompatibility. This could explain the low level of admixture detected in our study

between the two genetic groups. A genome-wide survey with a wider sampling of both genetic groups will allow testing these hypotheses.

The intermediate position of UK, USA and Azores individuals in the PCA analysis, as well as the admixture detected in STRUCTURE, suggest they are the result of mixed gene pools. It further corroborates the mtDNA results of Rodrigues et al. (2014) which showed the occurrence of mixed mitochondrial lineages in the UK and USA and that Azores and some USA samples were genetically similar to those from the UK. Across North America, variation in the morphology of male aedeagus in *P. spumarius* was reported by Hamilton (1979) and different mtDNA haplogroups were detected by Rodrigues et al. (2014), leading to the suggestion of multiple colonization events. The analysed population from New Hampshire (USA) showed very low genome-wide differentiation from the UK population (mean F_{ST}=0.042) compared to other pairwise comparisons in this study, and also a large number of COI haplotypes belonging to the "North-Eastern" haplogroup (5 in the 7 samples analysed for mtDNA). This supports a likely origin of the North American *P. spumarius* from Northern Europe, perhaps with multiple colonisation events and with large effective population sizes. In S. Miguel Island (Azores), only two COI haplotypes, differing by one substitution, have been found so far (in 6 samples, Rodrigues et al., 2014 and this study), which are closely related to the UK haplotypes from the "Western" haplogroup. From the genomic results, this population was more differentiated from UK and USA (mean F_{ST}=0.098 between AZO and UK, and mean F_{ST}=0.078 between AZO and USA) than these two were from one another (mean F_{ST}=0.042 between UK and USA), showing the lowest genetic diversity (expected heterozygosity) of all *P. spumarius* populations here analysed. These results suggest a likely origin of the Azores colonization from Northern Europe, and that this colonization is likely to have involved a bottleneck event leading to reduced genetic diversity. The low number of colour morphs found in S. Miguel (Borges et al., 2018) in this highly polymorphic species further supports this hypothesis, although selective processes may also be involved. Expanding the sampling and analyses will allow more precise determination of the origin, mode (whether or not mediated by man) and eventually the timing of these transatlantic colonisations.

Population genomics approaches provide genome-wide information that is expected to reflect a baseline of neutral processes and, at the same time, allow detection of loci with signatures of selection, deviating from this baseline (Hohenlohe, Phillips, & Cresko, 2010). We focused on detecting local adaptation, by finding those loci that are more differentiated (F_{ST}) between populations than expected from the neutral background, and also by finding loci that have allelic variation correlated with environmental variation. When discarding such candidate loci for selection from our dataset, the population structure patterns remained very similar to the neutral dataset, which means that these 9% of loci are not affecting the genome-wide neutral pattern of population structure. Despite the usefulness of RAD sequencing for detection of selection in natural populations (Catchen et al., 2017), this analysis is limited by the number of SNPs analysed and also by the fact that RAD tags are usually distributed non-uniformly across the genome (Lowry et al., 2016). The large genome size of *P. spumarius* (2.58 Gbases; Rodrigues et al., 2016) makes it more difficult to have a good genomic representation with these scattered markers. Also, genetic signatures of selection in individual loci can be weak and not easily detected in cases of soft selective sweeps (adaptation from standing genetic variation), epistatic interactions among loci or genotype-by-environment interactions (Hohenlohe et al., 2010). Whole-genome analyses, by analysing patterns of diversity, differentiation and linkage disequilibrium along the genome, will be essential to better understand the evolutionary forces of selection and recombination shaping genomic variation (Ellegren, 2014). The fact that we did not detect loci under balancing selection in the Bayescan analysis may also be related to low marker density. *P. spumarius* is particularly known for its balanced polymorphism for dorsal colour forms and the assessment of population variation in the colour-associated loci (Rodrigues et al., 2016) remains to be done.

In the environmental association analysis, the associated variables were primarily longitude and those related to the extreme values and range variation in temperature and precipitation. This analysis is tentative, since we have a low number of populations from a wide geographic range. The low number of hits of the candidate loci with the *P. spumarius* transcriptome may indicate their location was mostly in non-coding regions, while the low number of hits with the partial genome denotes its incompleteness (Rodrigues et al., 2016). A more complete draft genome is now available (Biello et al., 2020), and new genomic and transcriptomic resources will be soon

generated for *P. spumarius* which will provide important tools to further explore the molecular basis of adaptation in this species.

Understanding species divergence and the population genetic structures of *P. spumarius* and related species of the genus *Philaenus* is relevant to the management of the eventual progression of the plant pathogenic bacterium *X. fastidiosa*, since their dispersal patterns might aid or constrain disease transmission. Also, the ecological characteristics of different taxa or local populations may be different, for example in host plant preference, ease of acquisition of *X. fastidiosa* or transmission efficiency. Understanding the specific ecology of the vectors has been shown to be crucial in the management of *X. fastidiosa* diseases in America (Redak et al., 2004). Integrating this information is important for epidemiological models of *X. fastidiosa* in Europe and other Mediterranean countries. The risk of *X. fastidiosa* transmission and disease progression is generally expected to be related to long-range human-assisted movements of infected plants and with shorter-range natural dispersal by vectors (EFSA et al., 2019). Genetic studies of *P. spumarius* have shown that it does not constitute a panmictic population and geographical distance is not the only factor restricting gene flow. Other factors have to be taken into account, including habitat fragmentation, barriers to gene exchange such as endosymbionts or behavioural differences, rapid climate changes that may cause major shifts in distribution ranges, as well as unpredictable adaptive responses (Kellermann & Van Heerwaarden, 2019). Even without detectable gene flow, adults of *P. spumarius* may be able to migrate occasionally, or consistently but without reproductive outcome, and spread the bacterium. More ecological studies on the abundance and distribution of this insect vector through the seasons and across years are needed to understand the dispersal patterns across geographical regions and the potential for disease spread. It will be important to understand the dispersal patterns from South Italy, which is a potential source of contamination by *X. fastidiosa* and particularly its subspecies *pauca*, which is the strain associated with OQDS. This is especially true as previous work detected haplotypes from distinct haplogroups in Italy. For example, Rodrigues et al. (2014) detected both “Eastern-Mediterranean” and “Western” haplogroups in North and Central Italy, as well as in Sicily, unveiling the pivotal role of that region in the dispersal patterns of *P. spumarius* among the Mediterranean peninsulas.

.

734

735 ## Conclusions

736 In this study, morphological and genomic analysis allowed a more detailed view of the
737 divergence between *P. spumarius* and *P. tesselatus*, as well as of the population structure and
738 adaptation in *P. spumarius.* We found genome-wide divergence between these two species,
739 despite the lack of mitochondrial DNA differentiation between them. The population genomics
740 approach taken here showed admixture but also co-occurrence of non-admixed individuals in
741 contact zones of diverging mitochondrial lineages of *P. spumarius*. The potential role of
742 *Wolbachia* in shaping these patterns of divergence and introgression should be further explored.
743
744 The findings on species divergence and population structure described here point to the need for
745 elucidating the dispersal and ecological requirements of the different taxa and local populations
746 of these vectors for a better management of *X. fastidiosa* progression.
747
748

749

750

751 ## Acknowledgements

752 We thank Carla Pereira, Carla Rego, Rui Nunes and Luis Ferreira for help in field-work; Karim
753 Gharbi and Mark Blaxter for help with the RAD sequencing; Vera Nunes and Juan Farinas for
754 help in the laboratory; the Faculty of Sciences of the University of Lisbon's Microscopy Facility,
755 a node of the Portuguese Platform for Bioimaging, reference PPBI-POCI-01-0145-FEDER-
756 022122, where images from insect specimens were acquired.

757

758

759

760

761

762 ## References

763 Almeida, R. P. P., Fuente, L. D. La, Koebnik, R., Lopes, J. R. S., Parnell, S., & Scherm, H. (2019).
764 Addressing the New Global Threat of *Xylella fastidiosa*. *Phytopathology*, *109*, 172–174.
765 https://doi.org/10.1094/PHYTO-12-18-0488-FI
766
767 Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J.
768 (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.
769 *Nucleic Acids Research, 25,* 3389–3402. doi: 10.1093/nar/25.17.3389.
770
771 Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing
772 the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*,
773 *17*(2), 81–92. doi: 10.1038/nrg.2015.28.
774
775 Baldi, P., & La Porta, N. (2017). *Xylella fastidiosa*: Host Range and Advance in Molecular
776 Identification Techniques. *Frontiers in Plant Science*, *8*, 944.
777 https://doi.org/10.3389/fpls.2017.00944
778
779 Bahrndorff, S., Ruiz-González, A., Jonge, N. De, Nielsen, J. L., Skovgård, H., & Pertoldi, C.
780 (2020). Integrated genome-wide investigations of the housefly, a global vector of diseases reveal
781 unique dispersal patterns and bacterial communities across farms. *BMC Genomics*, *21*, 66. doi.
782 10.1186/s12864-020-6445-z.
783
784 Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E.
785 U., Cresko, W. A., & Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using
786 sequenced RAD markers. *PloS One*, *3*(10), e3376. doi: 10.1371/journal.pone.0003376.
787
788 Bandelt, H. J., Forster, P., & Rohl, A. (1999). Median-joining networks for inferring intraspecific
789 phylogenies. *Molecular Biology and Evolution*, *16*(1), 37–48. doi:
790 10.1093/oxfordjournals.molbev.a026036.
791
792 Beal, D. J., Cooper, M., Daugherty, M. P., Purcell, A. H., & Almeida, R. P. P. (2021). Seasonal
793 abundance and infectivity of *Philaenus spumarius* (Hemiptera: Aphrophoridae), a vector of

Xylella fastidiosa in California vineyards. *Environmental Entomology* 10 pp. nvaa178, https://doi.org/10.1093/ee/nvaa178

Biello, R., Mathers, T. C., Mugford, S. T., Liu, Q., Rodrigues, A. S. B., Neto, A. C., Rebelo, M. T., Paulo, O. S., Seabra, S. G., & Hogenhout, S. A. (2020). Draft genome assembly version 1 of the meadow spittlebug *Philaenus spumarius* (Linnaeus, 1758) (Hemiptera, Aphrophoridae) (Version version 1) [Data set]. *Zenodo*. doi: 10.5281/zenodo.3368385.

Borges, P. A. V., Amorim, I. R., Terzopoulou, S., Rigal, F., Emerson, B., & Serrano, A. R. M. (2017). Cryptic diversity in Azorean beetle genus *Tarphius* Erichson, 1845 (Coleoptera: Zopheridae): An integrative taxonomic approach with description of four new species. *Zootaxa,* 4236, 401–449. doi: 10.11646/zootaxa.4236.3.1.

Borges, P. A. V, Rodrigues, A. S. B., Silva, S. E., Seabra, S. G., Paulo, O. S., & Quartau, J. A. (2018). New data on polymorphism of the meadow spittlebug *Philaenus spumarius* (L.) (Hemiptera: Aphrophoridae) from the island of São Miguel (Azores), *Zootaxa*, *4369*(1), 144–150. doi: 10.11646/zootaxa.4369.1.9.

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

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

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

825 Coop, G., Witonsky, D., Di Rienzo, A., & Pritchard, J. K. (2010). Using environmental
826 correlations to identify loci underlying local adaptation. *Genetics*, *185*(4), 1411–1423. doi:
827 10.1534/genetics.110.114819.
828
829 Cornara, D., Saponari, M., Zeilinger, A. R., Stradis, A. De, Boscia, D., Loconsole, G., Bosco, D.,
830 Martelli, G. P., Almeida, R. P. P., & Porcelli, F. (2017). Spittlebugs as vectors of *Xylella*
831 *fastidiosa* in olive orchards in Italy. *Journal of Pest Science*, *90*(2), 521–530. doi:
832 10.1007/s10340-016-0793-0.
833
834 Cornara, D., Morente, M., Markheiser, A., Bodino, N., Tsai, C.-W., Fereres, A., Redak, R. A.,
835 Perring, T. M., & Lopes, J. R. S. (2019). An overview on the worldwide vectors of *Xylella*
836 *fastidiosa*. *Entomologia Generalis*, *39*(3–4), 157–181.
837 https://doi.org/10.1127/entomologia/2019/0811
838
839 Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R.
840 E., Lunter, G., Marth, G, T., Sherry, S. T., McVean, G., Durbin, R., & Group, 1000 Genomes
841 Project Analysis. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–
842 2158. doi: 10.1093/bioinformatics/btr330.
843
844 Dejaco, T., Gassner, M., Arthofer, W., Schlick-Steiner, B. C., & Steiner, F. M. (2016).
845 Taxonomist ' s nightmare … Evolutionist ' s delight: an integrative approach resolves species
846 limits in jumping bristletails despite widespread hybridization and parthenogenesis. *Systematic*
847 *Biology*, *65*(6), 947–974. doi: 10.1093/sysbio/syw003.
848
849 Dohm, J. C., Lottaz, C., Borodina, T., & Himmelbauer, H. (2008). Substantial biases in ultra-
850 short read data sets from high-throughput DNA sequencing. *Nucleic Acids Research, 36,* e105.
851 doi: 10.1093/nar/gkn425.
852
853 Drosopoulos, S. (2003). New data on the nature and origin of colour polymorphism in the
854 spittlebug genus *Philaenus* (Hemiptera: Aphorophoridae). *Annales de la Société Entomologique*
855 *de France*, *39*(1), 31–42. doi: 10.1080/00379271.2003.10697360.

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

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

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

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

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

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

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

885 Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals
886 using the software STRUCTURE: a simulation study. *Molecular Ecology*, *14*, 2611–2620. doi:
887 10.1111/j.1365-294X.2005.02553.x.
888
889 Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using
890 multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, *164*, 1567–
891 1587.
892
893 Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for
894 both dominant and codominant markers: a bayesian perspective. *Genetics*, *180*, 977–993. doi:
895 10.1534/genetics.108.092221.
896
897 Gautier, M. (2015). Genome-wide scan for adaptive divergence and association with population-
898 specific covariates. *Genetics*, *201*(4), 1555–1579. doi: 10.1534/genetics.115.181453.
899
900 Gompert, Z., Forister, M. L., Fordyce, J. A., & Nice, C. C. (2008). Widespread mito-nuclear
901 discordance with evidence for introgressive hybridization and selective sweeps in Lycaeides.
902 *Molecular Ecology*, *17*, 5231–5244. doi: 10.1111/j.1365-294X.2008.03988.x.
903
904 Halkka, O., & Halkka, L. (1990). Population genetics of the polymorphic meadow spittlebug,
905 *Philaenus spumarius* (L.). *Evolutionary Biology (USA), 24,* 149–191.
906

907 Halkka, O., Raatikainen, M., Vasarainen, A., & Heinonen, L. (1967). Ecology and ecological
908 genetics of *Philaenus spumarìus* (L.) (Homoptera). *Annales Zoologici Fennici*, *4*, 1-18.

909
910 Hall, T.A. (1999). BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis
911 Program for Windows 95/98/NT. *Nucleic Acids Symposium Series, 41*, 95–98.
912
913 Hamilton, K. G. A. (1979). Synopsis of North American Philaenini (Rhynchota: Homoptera:
914 Cercopidae) with a new genus and four new species. *The Canadian Entomologist*, *111*, 127–141.

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

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

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

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

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

Kapantaidaki, D. E., Antonatos, S., Evangelou, V., Papachristos, D. P., & Milonas, P. (2021). Genetic and endosymbiotic diversity of Greek populations of *Philaenus spumarius*, *Philaenus signatus* and *Neophilaenus campestris*, vectors of *Xylella fastidiosa*. *Scientific Reports*, *11*(1), 3752. doi: 10.1038/s41598-021-83109-z.

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

Karban, R., & Strauss, S. (2004). Physiological tolerance, climate change, and a northward range shift in the spittlebug, *Philaenus spumarius*. *Ecological Entomology*, *29*, 251–254. doi: 10.1111/j.1365-2311.2004.00576.x.

946
947 Kellermann, V., & Van Heerwaarden, B. (2019). Terrestrial insects and climate change: adaptive
948 responses in key traits. *Physiological Entomology, 44,* 99–115. doi: 10.1111/phen.12282.
949
950 Leigh, J. W., & Bryant, D. (2015). PopART: Full-feature software for haplotype network
951 construction. *Methods in Ecology and Evolution, 6*(9), 1110–1116. doi: 10.1111/2041-
952 210X.12410.
953
954 León, J., Jones, W. & Morgan, D. (2009). Population genetic structure of *Homalodisca*
955 *coagulata* (Homoptera: Cicadellidae), the vector of the bacterium *Xylella fastidiosa* causing
956 Pierce's disease in grapevines. *Annals of the Entomological Society of America, 97*(4), 809-818.
957
958 Linck, E., & Battey, C. J. (2019). Minor allele frequency thresholds strongly affect population
959 structure inference with genomic data sets, *Molecular Ecology Resources, 19*(3), 639–647. doi:
960 10.1111/1755-0998.12995.
961
962 Lis, A., Maryańska-Nadachowska, A., & Kajtoch, Ł. (2015). Relations of *Wolbachia* infection
963 with phylogeography of *Philaenus spumarius* (Hemiptera: Aphrophoridae) populations within
964 and beyond the Carpathian contact zone. *Microbial Ecology*, *70*(2), 509–521. doi:
965 10.1007/s00248-015-0570-2.
966
967 Lis, A., Maryańska-Nadachowska, A., Lachowska-Cierlik, D., & Kajtoch, L. (2014). The
968 secondary contact zone of phylogenetic lineages of the *Philaenus spumarius* (Hemiptera:
969 Aphrophoridae): an example of incomplete allopatric speciation. *Journal of Insect Science*, *14*,
970 227. doi: 10.1093/jisesa/ieu089.
971
972 Lischer, H. E. L., & Excoffier, L. (2012) PGDSpider: An automated data conversion tool for
973 connecting population genetics and genomics programs. *Bioinformatics, 28*, 298–299. doi:
974 10.1093/bioinformatics/btr642.
975

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer, A. (2016). Breaking RAD: An evaluation of the utility of restriction site associated DNA sequencing for genome scans of adaptation. *Molecular Ecology Resources*, *17*, 142–152. doi: 10.1111/1755-0998.12596

Malinsky, M., Challis, R. J., Tyers, A. M., Schiffels, S., Terai, Y., Ngatunga, B. P., Miska, E. A., Durbin, R., Genner, M. J., & Turner, G. F. (2015). Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science (New York, N.Y.)*, *350*(6267), 1493–1498. https://doi.org/10.1126/science.aac9927

Maryańska-Nadachowska, A., Drosopoulos, S., Lachowska, D., Kajtoch, L., & Kuznetsova, V. G. (2010). Molecular phylogeny of the Mediterranean species of *Philaenus* (Hemiptera: Auchenorrhyncha: Aphrophoridae) using mitochondrial and nuclear DNA sequences. *Systematic Entomology*, *35*, 318–328. doi: 10.1111/j.1365-3113.2009.00510.x.

Maryańska-Nadachowska, A., Kajtoch, Ł., & Lachowska, D. (2011). Genetic diversity of *Philaenus spumarius* and *P. tesselatus* (Hemiptera, Aphrophoridae): implications for evolution and taxonomy. *Systematic Entomology*, *37*(1), 55–64. doi: 10.1111/j.1365-3113.2011.00592.x.

Masly, J. P. (2012). 170 Years of "Lock-and-Key": genital morphology and reproductive isolation. *International Journal of Evolutionary Biology*, *2012*, 247352. doi: 10.1155/2012/247352.

Nadeau, N. J., Whibley, A., Jones, R. T., Davey, J. W., Dasmahapatra, K. K., Baxter, S. W., Quail, M. A., Joron, M., ffrench-Constant, R. H., Blaxter, M. L., Mallet, J., & Jiggins, C. D. (2012). Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*, 343–353. doi: 10.1098/rstb.2011.0198.

1006 Narita, S., Nomura, M., Kato, Y., & Fukatsu, T. (2006). Genetic structure of sibling butterfly
1007 species affected by *Wolbachia* infection sweep : evolutionary and biogeographical implications.
1008 *Molecular Ecology*, *15*, 1095–1108. doi: 10.1111/j.1365-294X.2006.02857.x.
1009
1010 Nast, J. (1972). *Palaearctic Auchenorrhyncha (Homoptera): An Annotated Check List*. Institute
1011 of Zoology, Polish Academy of Sciences, Warsaw: Polish Scientific Publisher, 550 pp.
1012
1013 Nosil, P. (2012). *Ecological Speciation*. Oxford Series in Ecology and Evolution. 300 pp.
1014
1015 Novembre, J., & Stephens, M. (2008). Interpreting principal component analyses of spatial
1016 population genetic variation. *Nature Genetics*, *40*(5), 646–649. doi: 10.1038/ng.139.Interpreting.
1017
1018 Peccoud, J., Ollivier, A., Plantegenest, M., & Simon, J.-C. (2009). A continuum of genetic
1019 divergence from sympatric host races to species in the pea aphid complex. *Proceedings of the*
1020 *National Academy of Sciences of the United States of America*, *106*(18), 7495–7500. doi:
1021 10.1073/pnas.0811117106.
1022
1023 Pélissié, B., Crossley, M. S., Cohen, Z. P., & Schoville, S. D. (2018). Rapid evolution in insect
1024 pests: the importance of space and time in population genomics studies *Current Opinion in*
1025 *Insect Science*, *26*, 8–16. doi: 10.1016/j.cois.2017.12.008.
1026
1027 Pina-Martins, F., Silva, D. N., Fino, J., & Paulo, O. S. (2017). Structure_threader: An improved
1028 method for automation and parallelization of programs structure, fastStructure and MavericK on
1029 multicore CPU systems. *Molecular Ecology Resources*, *17*(6), e268–e274. Doi: 10.1111/1755-
1030 0998.12702.
1031
1032 Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using
1033 multilocus genotype data. *Genetics*, *155*, 945–959.
1034
1035 R Core Team (2017). R: A language and environment for statistical computing. R Foundation for
1036 Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

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

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

Riesch, R., Muschick, M., Villoutreix, R., Comeault, A. A., Farkas, T. E., Lucek, K., Hellen, E., Soria-Carrasco, V., Dennis, S. R., de Carvalho, C. F., Safran, R. J., Sandoval, C. P., Feder, J., Gries, R., Crespi, B.J., Gries, G., Gompert, Z., & Nosil, P. (2017). Transitions between phases of genomic differentiation during stick-insect speciation. *Nature Ecology & Evolution*, *1*, 0082. doi: 10.1038/s41559-017-0082.

Rodrigues, A. S. B., Silva, S. E., Marabuto, E., Silva, D. N., Wilson, M. R., Thompson, V., Yurtsever, S., Halkka, A., Borges, P. A. V., Quartau, J. A., Paulo, O. S., & Seabra, S. G. (2014). New mitochondrial and nuclear evidences support recent demographic expansion and an atypical phylogeographic pattern in the spittlebug *Philaenus spumarius* (Hemiptera, Aphrophoridae). *PLoS ONE*, *9*(6), e98375. doi: 10.1371/journal.pone.0098375.

Rodrigues, A. S. B., Silva, S. E., Pina-Martins, F., Loureiro, J., Castro, M., Gharbi, K., Johnson, K. P., Dietricht, C. H., Borges, P. A. V., Quartau, J. A., Jiggins, C. D., Paulo, O. S., & Seabra, S. G. (2016). Assessing genotype-phenotype associations in three dorsal colour morphs in the meadow spittlebug Philaenus spumarius (L.) (Hemiptera: Aphrophoridae) using genomic and transcriptomic resources. *BMC Genetics*, *17*(1), 144. doi: 10.1186/s12863-016-0455-5.

Rousset, F. (1997). Genetic differentiation and estimation of gene flow from F-Statistics under isolation by distance. *Genetics*, *145*, 1219–1228.

Rousset, F., & Solignac, M. (1995). Evolution of single and double *Wolbachia* symbioses during speciation in the *Drosophila simulans* complex. *Proceedings of the National Academy of Sciences of the United States of America*, *92*, 6389–6393.

Saponari, M., Loconsole, G., Cornara, D., Yokomi, R. K., Stradis, A. D. E., Boscia, D., Bosco, D., Martelli, G. P., Krugner, R., & Porcelli, F. (2014). Infectivity and transmission of *Xylella fastidiosa* by *Philaenus spumarius* (Hemiptera: Aphrophoridae) in Apulia, Italy. *Journal of Economic Entomology*, *107*(4), 1–4. doi: 10.1603/ec14142.

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., … Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews Genetics*, *15*(3), 176–192. doi: 10.1038/nrg3644.

Sicard, A., Zeilinger, A. R., Vanhove, M., Schartel, T. E., Beal, D. J., Daugherty, M. P., & Almeida, R. P. P. (2018). *Xylella fastidiosa*: insights into an emerging plant pathogen. *Annual Review of Phytopathology*, *56*, 181–202. doi: 10.1146/annurev-phyto-080417-045849.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., & Flook, P. (1994). Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*, *87*, 651–701.

Smadja, C. M., & Butlin, R. K. (2011). A framework for comparing processes of speciation in the presence of gene flow. *Molecular Ecology*, *20*, 5123–5140. doi: 10.1111/j.1365-294X.2011.05350.x.

Smith P. T. (2005). Mitochondrial DNA variation among populations of the glassy-winged sharpshooter, *Homalodisca coagulata*. *Journal of insect science (Online)*, *5*, 41. https://doi.org/10.1093/jis/5.1.41

1099 Sousa, V., & Hey, J. (2013). Understanding the origin of species with genome-scale data:
1100 modelling gene flow. *Nature Reviews. Genetics*, *14*(6), 404–14. doi: 10.1038/nrg3446.
1101
1102 Thompson, J. D., Higgins, D. G., Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity
1103 of progressive multiple sequence alignment through sequence weighting, position-specific gap
1104 penalties and weight matrix choice. *Nucleic Acids Research*, *22,* 4673–4680. doi:
1105 10.1093/nar/22.22.4673.
1106
1107 Tonzo, V., Papadopoulou, A., & Ortego, J. (2019). Genomic data reveal deep genetic structure
1108 but no support for current taxonomic designation in a grasshopper species complex. *Molecular*
1109 *Ecology*, *28*(17), 3869–3886. doi: 10.1111/mec.15189.
1110
1111 Turner, T. L., Hahn, M. W., & Nuzhdin, S. V. (2005). Genomic islands of speciation in
1112 *Anopheles gambiae*. *PLoS Biology*, *3*(9), e285. doi: 10.1371/journal.pbio.0030285.
1113
1114 Vijay, N., Bossu, C. M., Poelstra, J. W., Weissensteiner, M. H., Suh, A., Kryukov, A. P., &
1115 Wolf, J. B. W. (2016). Evolution of heterogeneous genome differentiation across multiple
1116 contact zones in a crow species complex. *Nature Communications*, *7*, 1–10.
1117 https://doi.org/10.1038/ncomms13195
1118
1119 Vitalis, R., Gautier, M., Dawson, K. J., & Beaumont, M. A. (2014). Detecting and measuring
1120 selection from gene frequency data. *Genetics*, *196*(3), 799–817. doi:
1121 10.1534/genetics.113.152991.
1122
1123 Wagner, W. (1955). Die Bewertung morphologischer Merkmale in den unteren taxonomischen
1124 Kategorien, aufgezeigt an Beispielen aus der Taxionomie der Zikaden. Mitt. Hamburg. Zool.
1125 Mus. Inst. 53, 75–108.
1126
1127 Wagner, W. (1959). Homoptera. In *Zoologische Studien in Westgriechenland. IX* (pp. 583–605).
1128 Wien.
1129

1130 Weaver, C., & King, D., (1954). *Meadow Spittlebug*. Research Bulletin, Ohio Agricultural
1131 Experiment Station, 741, pp. 99
1132
1133 Werren, J. H., Baldo, L., & Clark, M. E. (2008). *Wolbachia*: master manipulators of invertebrate
1134 biology. *Nature Reviews. Microbiology*, *6*, 741–751. doi: 10.1038/nrmicro1969.
1135
1136 Whitworth, T. L., Dawson, R. D., Magalon, H., & Baudry, E. (2007). DNA barcoding cannot
1137 reliably identify species of the blowfly genus *Protocalliphora* (Diptera: Calliphoridae).
1138 *Proceedings of the Royal Society B: Biological Sciences*, *274*, 1731–1739. doi:
1139 10.1098/rspb.2007.0062.
1140
1141 Wickham, H. (2016). *ggplot2: elegant graphics for data analysis*. Springer: Verlag, New York.
1142
1143 Wolf, J. B. W., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in
1144 light of speciation. *Nature Reviews Genetics*, *18*, 87–100. doi: 10.1038/nrg.2016.133.
1145
1146 Wu, C.-I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology*,
1147 *14*, 851–865.
1148
1149 Yoon, J.-.I., Hrynkiv, V., Morano, L., Nguyen, A. T., Wilder, S., Mitchell, F. (2014).
1150 Mathematical modelling of glassy-winged sharpshooter population. *Mathematical Biosciences*
1151 *and Engineering,* 11, 667-677.
1152
1153 Yurtsever, S. (2002). Hybrid crosses of the meadow spittlebug *Philaenus spumarius* (L.)
1154 (Homoptera: Cercopidae) between New Zealand and Welsh populations. *New Zealand Journal*
1155 *of Zoology*, *29*, 245–251.
1156
1157 Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-
1158 performance computing toolset for relatedness and principal component analysis of SNP data.
1159 *Bioinformatics*, *28*(24), 3326–3328. doi: 10.1093/bioinformatics/bts606.
1160

1161

1162

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.

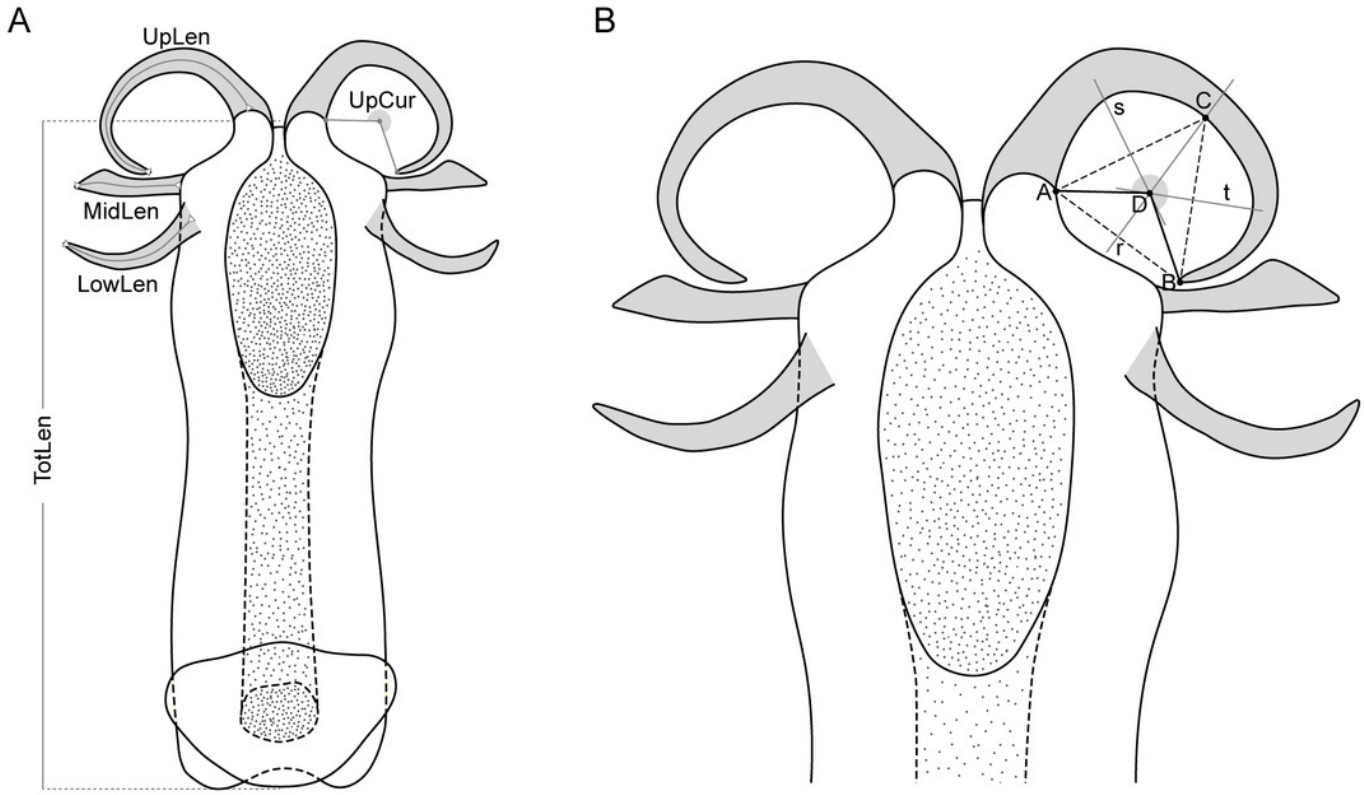

Figure 3

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).

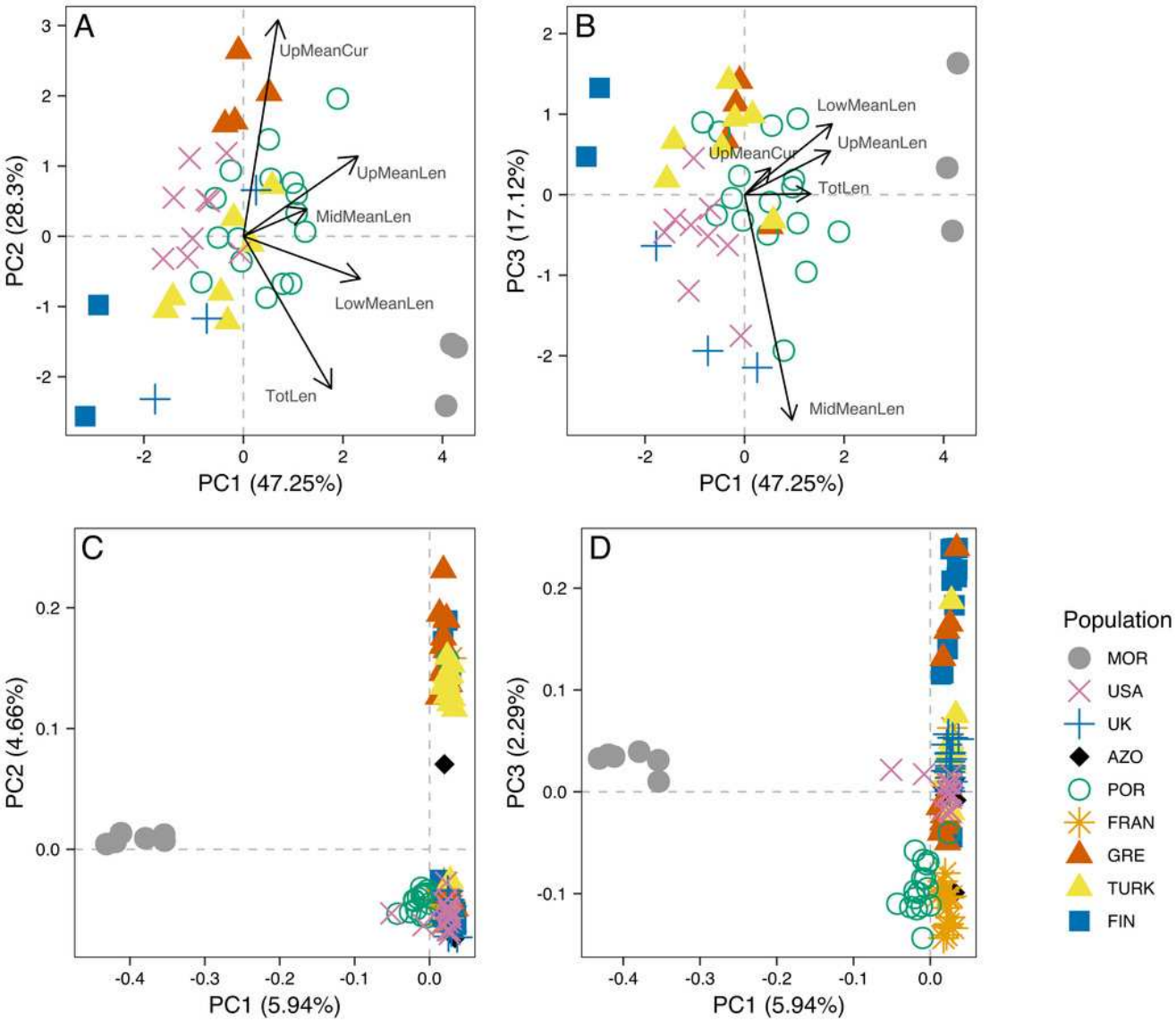

Figure 4

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.

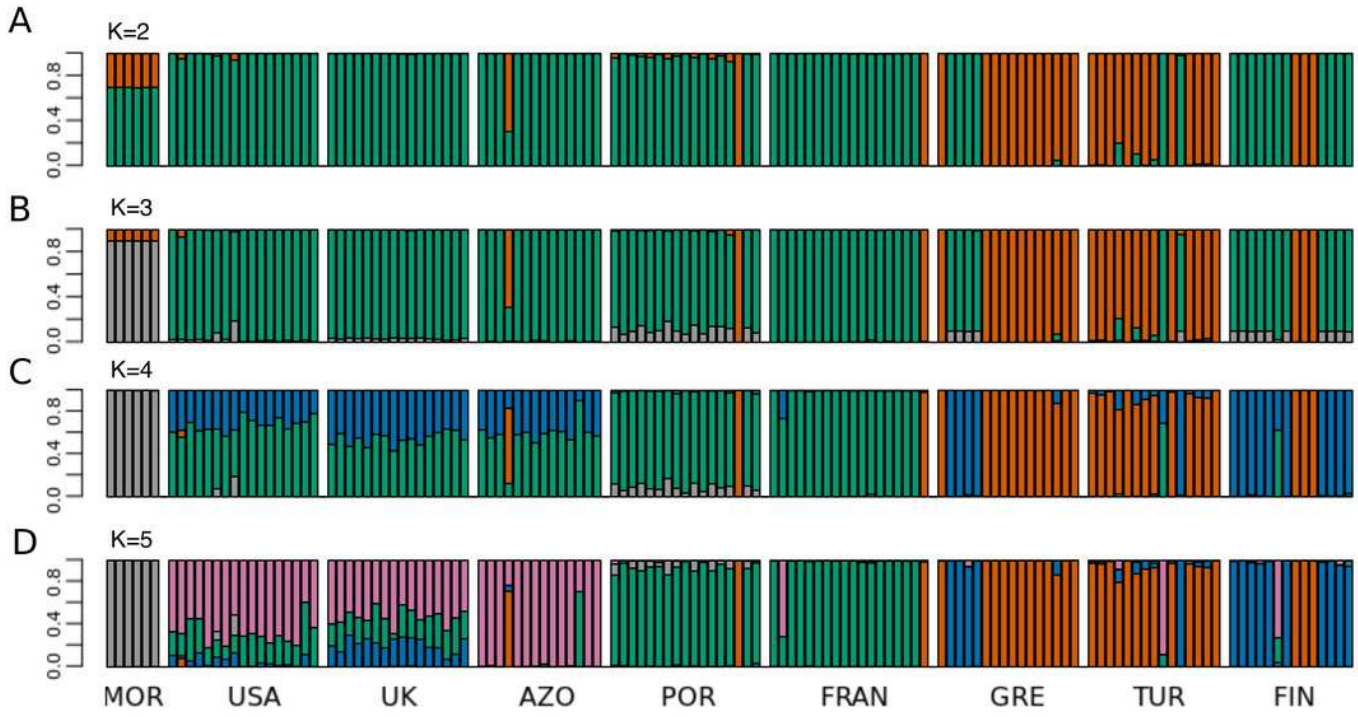

Figure 5

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 (circles); European populations (diamond); transatlantic populations (USA or Azores) vs. European populations (crosses).

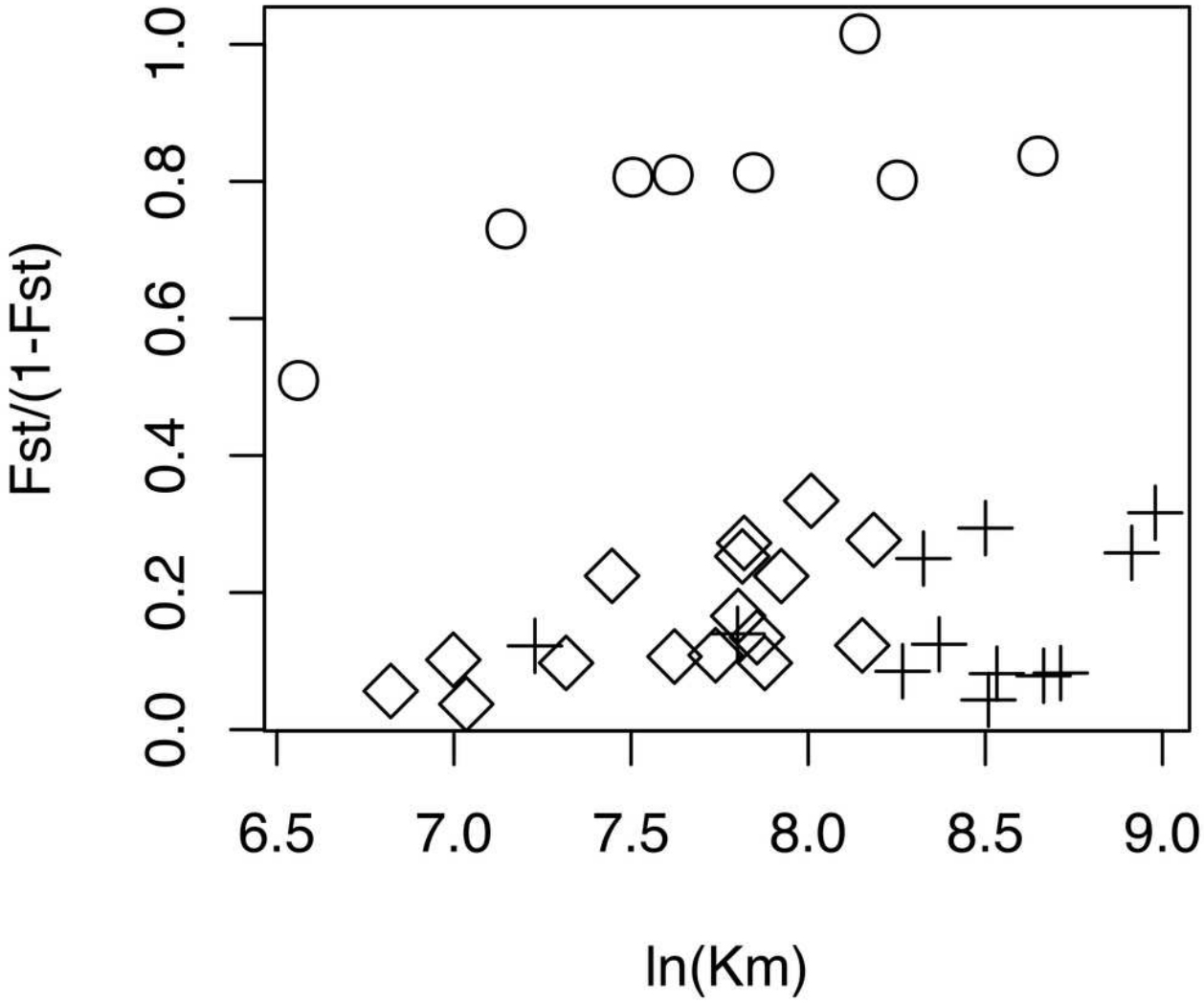
Fst/(1-Fst)
0.0
0.2
0.4
0.6
0.8
1.0
6.5
7.0
7.5
8.0
8.5
9.0
ln(Km)

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

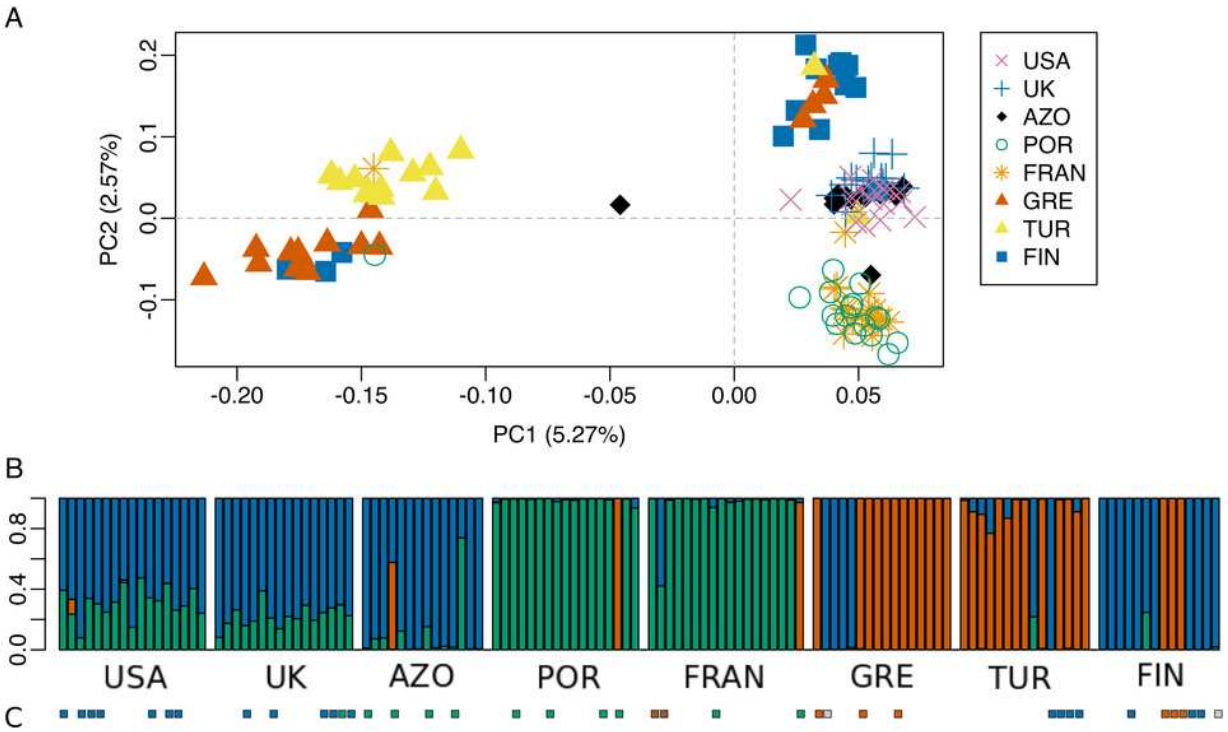

Table 1 (on next page)

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	0.0980	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	0.2050	0.2141	0.1998	0.1831	0.1833	0		
TUR	0.5039	0.2403	0.2503	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	0.0460	0.0537	0.0460	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

1