Original Article

The contribution of Anatolia to European phylogeography: the centre of origin for the meadow grasshopper *Chorthippus parallelus*

E. Mahir Korkmaz[1]*, David H. Lunt[2], Battal Çıplak[3] Naci Değerli[1] and Hasan H. Başibüyük[4]

[1]Department of Molecular Biology and Genetics, Faculty of Science, Cumhuriyet University, Sivas, 58140, Turkey.

[2]Department of Biological, Biomiomedical and Environmental Sciences, University of Hull, Hull, HU6 7RX, UK.

[3]Department of Biology, Faculty of Science, Akdeniz University, Antalya, 07058, Turkey.

[4]Department of Biology, Faculty of Science, Cumhuriyet University Sivas, 58140, Turkey.

*Correspondence: E. Mahir Korkmaz

Department of Molecular Biology and Genetics, Faculty of Science, Cumhuriyet University, Sivas, 58140, Turkey.

E-mail: ekorkmaz@cumhuriyet.edu.tr

Short running header Phylogeography of the meadow grasshopper *Chorthippus parallelus*

Abstract

Aim *Chorthippus parallelus* is one of the classic model systems for studying genetic structure and phylogeography of the Western Palaearctic. Here, we investigate the regional genetic differentiation of *C. parallelus* and evaluate historical and evolutionary processes responsible for such genetic structuring; test the nature of the Marmara water body as barriers to dispersal, and explore the contribution of Anatolian populations to the biodiversity of the West Palaearctic including its likely expansion routes.

Location West Palaearctic.

Methods We have incorporated sequence data from detailed sampling of the phylogeographically important Anatolian region using both previous and newly obtained sequence data of the nuclear *cpnl-1* and the mitochondrial *COI-tRNAleu-COII*. A total of 1049 sequences of *cpnl-1* from 33 regions were analysed to investigate the genetic diversity, genetic structuring and phylogeography of *C. parallelus* across its distribution range. The mtDNA region was additionally utilised to test the barrier position of the Marmara water body.

Results The analyses revealed that not all southern refugial populations of *C. parallelus* have contributed equally to the postglacial recolonization of Europe. Four genetic clusters across the species range were recovered: Cluster A (east part of the Anatolian Diagonal), Cluster B (west part of the Anatolian Diagonal), Cluster C (Spain, Italy, Southern Balkans, west part of Anatolia and Russia), and Cluster D (current distribution range of the species). The Marmara water body has been a weak barrier to dispersal by *C. parallelus*, allowing gene flow from Anatolia to the Balkans.

Main conclusions The current patterns of genetic structuring of *C. parallelus* were best explained by multiple expansion and contraction events. Anatolia has been well connected to the Balkans, contributing genetically to the establishment of central and northern European populations, and its expansion prior to the Holocene will have extended as far north as Scandinavia. The Anatolian refugium is suggested to be the centre of origin for west Palaeoarctic *C. parallelus* diversity rather than a Balkan refugium.

Keywords

Anatolia, *Chorthippus parallelus*, genetic diversity, genetic structuring, glacial refugia, Marmara Sea, phylogeography, range expansion waves, West Palaearctic.

Introduction

Phylogeographic studies of Western Palaearctic lineages have allowed several important generalisations concerning the history of temperate species (Hewitt, 1996, 2000; Taberlet *et al.,* 1998). One such is that the Pleistocene glacial cycles have had significant effects on the present distribution of many species as the biodiversity of northern areas are largely removed during glacial periods and re-established by founder populations from southern refugia during the warmer interglacials (Hewitt, 1996, 1999, 2000; Taberlet *et al.,* 1998; Schmitt, 2007). Iberia, Italy, and the Balkans have been identified as the likely glacial refugia of the Western Palaearctic (Hewitt, 1996; Atkinson *et al.,* 2007; Médail & Diadema, 2009; Çıplak *et al.,* 2010). Three common routes of recolonization from southern refugia to northern areas have been recognized (Hewitt, 1999); however additional routes may also have played a role (Taberlet *et al.,* 1998; Rokas *et al.,* 2003; Schmitt, 2007).

Although not rigorously tested, Anatolia has also been suggested as an important glacial refugium (Hewitt, 1996; Rokas *et al.,* 2003; Çıplak, 2004, 2008; Korkmaz *et al.,* 2010; Ansell *et al.,* 2011). The geographic location of Anatolia together with its complex paleogeographic history, heterogeneous topography and climate contributes to the high levels of endemism and global importance of its biodiversity (Şekercioğlu *et al.,* 2010). Albeit data is still limited, studies focusing on several lineages reveal that Anatolia harbours a high level of genetic diversity (Rokas *et al.,* 2003; Gündüz *et al.,* 2007; Dubey *et al.,* 2007; Fritz *et al.,* 2009; Stamatis *et al.,* 2009; Akın *et al.,* 2010; Çıplak *et al.,* 2010; Mutun, 2011). Anatolia, however, has been underrepresented in regional phylogeographic studies with specifically Anatolian lineages, or representatives of Anatolian populations for wide-ranging forms, rarely, inadequately, or only recently being included. With the effects of the climatic shifts of Pliocene/Pleistocene, vertical range changes may be expected depending on the heterogeneous topography. In particular during the warm interglacials the populations

72 preferring cold conditions would have been isolated and possibly diverged on top of high
73 mountains or ‘sky islands’ and this feature may suggest a complex refugial system for
74 Anatolia (Çıplak, 2004, 2008).
75 The meadow grasshopper *Chorthippus parallelus* (Zettersted, 1821) has been a key example
76 of phylogeographic structuring and indeed gives its name to one of Hewitt's three refugial
77 expansion models (Hewitt, 1999). Phylogeographic studies of *C. parallelus* have utilized both
78 mitochondrial DNA and nuclear loci and suggest three major genomic units in Iberia,
79 Calabria (Southern Italy), and the populations in remaining range of the species (Cooper *et*
80 *al.,* 1995; Lunt *et al.,* 1998). Besides the acceptance that the present European range has
81 originated by postglacial expansion from the Balkans, Hewitt (1996) also suggested that
82 Anatolian populations may have survived the glacial cycles and contributed to the
83 repopulation of Europe. Some Anatolian specimens were included in the study by Cooper *et*
84 *al.* (1995), but this sampling was insufficient to properly represent either genetic diversity or
85 geographic fragmentation of the species within Anatolia (Korkmaz *et al.,* 2010). The
86 implication of Hewitt (1996) that Anatolia could be the ultimate origin of the species has
87 remained therefore largely untested. An investigation of the relationship among Anatolian
88 populations may contribute to our understanding of the origins of *C. parallelus* diversity
89 during the expansion periods. The other issue to be considered is the direction of range
90 change during either cold or warm periods. It is generally accepted that during interglacial
91 periods the northern edge of the population was the leading or expanding edge and the
92 southern edge was the eroding or rear edge, while the reverse directions were the case in cold
93 periods (Çıplak, 2004; Hampe & Petit, 2005). However some assumptions of the rear
94 edge/leading edge hypothesis appear to be in conflict with the characteristics of a complex
95 refugial system (Médail & Diadema, 2009; Korkmaz *et al.,* 2010). Therefore, determining the

characteristics of leading edge populations and gene flow between different regions within the refugial system requires particular attention.

We note that since the preparation of this paper, *C. parallelus* has been transferred to a newly established genus *Pseudochorthippus* (Defaut, 2012) to maintain consistency we use the traditional name *Chorthippus parallelus* until the whole Gomphocerini lineage can be reconsidered on the basis of a larger dataset.

Here we revisited the *C. parallelus* phylogeographic system, using both previous and newly obtained sequence data of the nuclear *cpnl-1* region, and with particular attention to expand the sampling and analysis of its Anatolian populations. Our first objective is to examine the structure of diversity and differentiation between this region and the others and to investigate the historical and evolutionary processes responsible for such genetic structuring. Second, we aim to examine the potential role of Anatolia as a centre of origin for diversity within *C. parallelus* by including a larger and more geographically diverse collection of Anatolian samples. Thirdly, we investigate patterns of gene flow and likely expansion routes connecting Anatolia with the rest of the current distribution. In particular, we investigate whether the Bosphorus and Dardanelles, and Marmara Sea (hereafter is called the Marmara water body) acts as a barrier using a large dataset of both nuclear DNA (*cpnl-1*) and newly generated mtDNA (*COI-tRNALeu-COII*) sequences. Finally, we aim to explore the contribution of Anatolian forms to biodiversity of West Palaearctic more generally based on the evolutionary history of *C. parallelus.*

Materials and Methods

Sampling

The species used in this study is not endangered or protected, nor are the sampling locations; therefore no specific permits were required. Specimens were collected from June to August

between 2005 and 2009, using several keys to check identification (Bei-Bienko & Mistshenko, 1951). The number of individuals per locality varied due to availability of specimens and those collected were preserved in EtOH for DNA extraction (see Fig. 1 & Appendix S1 in Supporting Information for sampling localities of both previous and new datasets).

Two different datasets were generated from both nuclear and mitochondrial genes. The nuclear data included *cpnl-1* sequences of 248 individuals primarily from W. Europe and Russia from published studies (Cooper & Hewitt, 1993; Cooper *et al.,* 1995); and that of 558 individuals representing 33 populations from Anatolia, three from Thracian Turkey and two from Bulgaria generated in this work (Appendix S1). The mitochondrial data set generated in this study was used in testing the barrier statues of the Marmara water body and investigating expansion patterns and times of Anatolian and Thracian populations. It comprised of *COI-tRNALeu-COII* sequences from 201 individuals representing eight populations from northwest Anatolia and five populations from Thracia (Appendix S1). The mtDNA region utilised in this study is not comparable with the region that was used in previous studies (Lunt *et al.,* 1998).

DNA extraction, PCR and sequencing

Total genomic DNA was extracted from the hind leg of specimens using the Chelex-100 method described by Walsh *et al.* (1991). The amplification and single strand conformation polymorphism (SSCP) analysis of *cpnl-1* (410 bp) were carried out by the method outlined in Korkmaz *et al.* (2010). The mitochondrial *COI-tRNALeu-COII* gene fragment (1433 bp) was amplified using newly designed primers: COIMF (5'-TTGACCCAGCTGGAGGTGGAGAC-3') and COIIMR (5'-TGATTCCAATAGCAGGAACTGCTC-3'). Amplification was carried out in 50-μL volumes containing 0.5 U of *Taq* polymerase, 5 μL of 10 × reaction buffer (100 mM Tris-HCl, pH 8.8, 500 mM KCl, and 0.8% Nonidet P-40), 10 pmol of each of the primers, 0.2 mM of each of the four dNTPs, 1.5 mM $MgCl_2$, and 1 μL of DNA template (50-100 ng).

PCR cycle conditions were: 94 ^{O}C, 5 min; 35 × (94 ^{O}C, 1 min; 60.5 ^{O}C, 1 min; 72 ^{O}C, 30 s); 72 ^{O}C, 5 min. Sequencing reactions were carried out in both directions using the same primers as in PCR reactions. The forward and reverse nucleotide sequences were assembled, edited and aligned by eye using the CODONCODE ALIGNER 3.5.6 (CodonCode Corporation).

Data analysis

Genetic diversity

The sampling locations of *C. parallelus* were divided into 33 regions, based on the geographic proximity of populations and the location of geographic barriers, in Europe (including three Thracian regions), Russia and Anatolia (Fig. 1 & Appendix S1). Anatolia was also divided into four regions (including a total 33 of populations) considering the positions of the Anatolian Diagonal as well as Taurus and Pontids Mts ranges. The distribution of genetic diversity across the study area was evaluated in order to detect signals for refugial position. Nucleotide diversity (π), haplotype diversity (h), number of segregating/indel sites (S), number of unique haplotypes and average number of nucleotide differences (k) were calculated using DNASP 5 (Librado & Rozas, 2009) and ARLEQUIN 3.5 (Excoffier *et al.,* 2005). We also compared and characterised the distribution of haplotypes containing indel positions. The components of genetic diversity (C_S) and differentiation (C_D) were also calculated using the computer software CONTRIB 1.02 (Petit *et al.,* 1998), to estimate the contribution of each region to genetic diversity and genetic divergence.

Genetic structure

The levels of genetic subdivision and gene flow among regions were quantified and tested applying the sequence statistic, K_{ST} (Hudson *et al.,* 1992). K_{ST} was computed as K_{ST} =1–(K_S/K_T); where K_S = weighted average number of differences among sequences in a region and K_T = average number of differences between two sequences in the dataset. The pairwise K_{ST} values against Euclidian geographic distances of regions (Hutchison & Templeton, 1999)

was tested using the Mantel option with 10.000 random permutations of genetic distance matrix in the GENALEX 6.3 (Peakall & Smouse, 2006).

In order to detect global genetic structure and to define the most divergent region(s) within the species range without prior region boundary assumptions, a spatial analysis of molecular variance (SAMOVA) based on the proportion of total genetic variance (Φ-statistics) was carried out by SAMOVA 1.0 (Dupanloup *et al.,* 2002). This program requires the number of *K* and we ran SAMOVA on our datasets on different *K*, ranging from 2 to 7, with 250 simulated annealing processes. Significance levels of Φ-statistics were estimated by 1000 random permutations. To establish spatial genetic structure in *C. parallelus*, we also performed a discriminant analysis of principal components (DAPC) (Jombart *et al.,* 2010). See Appendix S2 for a more detailed explanation of the DAPC analysis using the R package ADEGENET 1.3-1 (Jombart, 2008). Twenty nine indel positions were also observed with many consisting of two or more sites. We therefore defined a total of 13 indel events (see Appendix S1). The distribution of the indel events were superimposed on the clusters generated by the DAPC analysis.

Phylogeography and demographic history

Three datasets were generated: (i) sequence data of all haplotypes; (ii) sequence data of the unique haplotypes; and (iii) presence and absence of haplotypes which were coded as (1) and (0) respectively and analysed under different assumptions by application of maximum likelihood, maximum parsimony, Bayesian inference and minimum evolution (NJ) approaches. We also carried out network analyses to investigate the genealogical relationship between regions. We constructed a haplotype network using NETWORK 4.5.1.6 (available at http://www.fluxux-technology.com). Network was calculated by the median-joining (MJ) method ($\varepsilon = 0$) (Bandelt *et al.,* 1999) and a subsequent maximum parsimony calculation was applied. To reduce complexity and evaluate historic demographic expansion events in the

196 network, we also performed an additional analysis using star contraction algorithm
197 implemented in the same program.
198 We then investigated the demographic history of the clusters defined by the DPCA. We firstly
199 examined Tajima's *D* (Tajima, 1989), Fu' s *Fs* (Fu, 1997) and the raggedness index (based
200 on the mismatch distribution) (Harpending, 1994), using DNASP, with 10.000 coalescent
201 simulations, to detect population growth and infer population demographic events. Second,
202 the mismatch distributions and the likely expansion times of the clusters were estimated from
203 demographic expansion factor static Tau ($\tau = 2\mu kt$) with 95% confidence intervals by using
204 ARLEQUIN 3.5, where k is the length of sequences and μ is the mutation rate per nucleotide.
205 Because there is no estimate on mutation rate of the non-coding *cpnl-1*, we utilised the same
206 mutation rate used in mtDNA which is the proposed average rate of neutral single copy
207 nuclear DNA in insect (Papadopoulou *et al.,* 2010).
208 **The barrier position of the Marmara water body**
209 We used three complementary approaches to test the barrier position of the Marmara water
210 body. Thracian and North-West Anatolian groups were created by inclusion of five and eight
211 populations respectively (Appendix S1). The *COI-tRNALeu-COII* (1433 bp) (GenBank
212 accession numbers: KC107629 - KC107654, see Appendix S1 for haplotypes) and *cpnl-1*
213 sequences were utilised in the analyses. First, the relationships among both mtDNA and *cpnl-*
214 *1* haplotypes belonging to Thracian and North-West Anatolian groups were separately
215 constructed using NETWORK 4.5.1.6. Second, we estimated the likely expansion times of
216 these two groups. Third, Markov Chain Monte Carlo (MCMC) simulations of the Isolation
217 with Migration model using the IMA2 software (Hey, 2010) were carried out to assess the
218 potential barrier status of the water body, splitting time, and rate of gene flow to each
219 direction. See Appendix S3 for a more detailed explanation of the IMA2 analysis.
220

221 **Results**

222 **Spatial distribution of genetic diversity**

223 The 1049 *cpnl-1* sequences represented 33 regions across the species range in Europe, Russia
224 and Anatolia (Fig. 1 & Appendix S1). After alignment and trimming, the remaining length of
225 sequences was 314 bp. Overall, 116 variable positions, composed of 87 single bp
226 substitutions, 29 indels, and 40 parsimony informative sites were observed defining a total of
227 188 *cpnl-1* haplotypes (GenBank accession numbers were given in Appendix S1). *h* was
228 0.848 $\pm$ 0.011 and generally higher than 0.50 in all regions ranging from 0.53 to 0.94 except
229 in three regions (SSpn, Blg2 and Blg3) (Table 1). π was 0.0090 $\pm$ 0.0003 and *k* was 2.512. π,
230 in contrast, was relatively low for many regions, ranging between 0.000 (SSpn) and 0.039
231 (ERs) with Blg2 and Blg3 having a score of 0.002 (Table 1). The most common haplotype
232 was Hap_2 (found in 18.7% of sequences). The 136 haplotypes (72.3% of the overall
233 haplotypes, Table 1) were definable with a singleton mutation difference and were unique to a
234 single region (Appendix S1). Each region of Anatolia, Thracia, Bulgaria and Greece had
235 unique haplotypes in varying frequencies (Appendix S1). The frequencies of the shared
236 haplotypes were presented in Appendix S1.

237 The variability levels of genetic divergence among regions were based mostly on the number
238 of haplotypes detected per region. The contribution of those regions to total diversity is due to
239 possession of both unique and divergent haplotypes (Fig. 2). Haplotypes found in all
240 Anatolian, Thracian, Russian, and most northern European regions provided relatively low
241 positive contribution to total diversity while negative contribution of the diversity components
242 was mostly due to Spanish and Balkans haplotypes (Fig. 2).

243 **Genetic structure**

244 Average genetic differentiation (K_{ST}) of all regions was 0.28 and ranged from -0.35 to 0.97
245 indicating a variable genetic differentiation among regions (Table 2). Pairwise K_{ST} values

between regions showed that the high levels of differentiation were mainly due to remarkably divergent regions of Spain, Italy and Russia. On the other hand, low levels of differentiation (< 0.1) were commonly observed among Anatolian, Thracian, Balkans and northern European regions as well as some regions having significant *P* values (Table 2). The Mantel test indicated no significant correlation between K_{ST} and geographic distances of the species range ($R^2 = 0.005$, $P > 0.05$).

We applied successive SAMOVA analyses to the data matrix of regions to see possible structuring in a geographic basis (Fig. 3). The results showed that F_{CT} value was the highest when all regions were subdivided into two groups; two Spanish regions (SSpn-PSpn) and all other remaining regions ($F_{CT} = 0.491$, Fig. 3). When five groups were included in SAMOVA, F_{CT} and F_{ST} values overlapped and PSpn, SSpn, CSpn, SIt were separated from all remaining (Fig. 3). A similar pattern was observed in the pairwise K_{ST} values [from 0.21 (SSpn-CSpn) to 0.97 (SSpn-Blg2) for SSpn, from 0.23 (PSpn-CSpn) to 0.77 (PSpn-PFrc) for PSpn, from 0.12 (CSpn-NWIt) to 0.69 (CSpn-PFrc) for CSpn and from 0.07 (SIt-NSpn) to 0.72 (SIt-PFrc) for SIt, but not among them (Table 2)].

The result of DAPC analyses indicated a more complex history of the species. The retained PCA components explained 79.9% of the total variance observed. The DAPC analysis partitioned the all individuals of *C. parallelus* into four clusters (Fig. 4 & see Appendix S2 for a scatterplot graph). Moreover, the posterior probability of assignment for each individual to the correct genetic cluster was 99%, indicating a high robustness of the analysis (Appendix S2). The Cluster A and B divided Anatolia into two parts, of which one corresponded to individuals from east part of the Anatolian Diagonal (58 individuals, 5 haplotypes; Fig. 4a); the other corresponded to west part of Anatolian Diagonal (103 individuals, 32 haplotypes; Fig. 4b). On the other hand, the remaining two clusters displayed relatively a broad distribution. Cluster C (152 individuals; 25 haplotypes) was comprised a mixture of

individuals mostly from southern range of the species (Spain, Italy, Balkans, Russia and western Anatolia) (Fig. 4c); while Cluster D (736 individuals, 127 haplotypes) corresponded to the current distribution of the species with individuals mostly from Anatolian, Thracian, Balkans, and northern European regions (Fig. 4d).

Phylogeography and demographic history

None of our diverse phylogenetic analyses of the three sets of *cpnl-1* haplotype alignments resulted in a well-resolved phylogenetic tree, which is not unusual for population level diversity. The length of indels exhibited a great variation from one to 10 bp in the fragment (Appendix S1). Number of indels per sequence showed a clinal tendency from east to west across the species range. The highest numbers of total indel positions per sequence was observed in the SEAnt and NEAnt (with total of 21 indel positions) (Appendix S1). The presence of two indel events at site 257 (GAGA), and at site 260 (A) was observed in all individuals of Cluster A, while an event including three nucleotides at sites 8 (ACT) was detected in all individuals of Cluster B (Fig. 4ab). The most common indel event at site 217 (AACTT) was found in all individuals belong to Cluster C indicating mostly a southern refugial distribution (Fig. 4c). Cluster C also contained two indel events at site 217 (AACGT) in Russia and at site 230 in NSpn. On the other hand, the remaining seven indel events were observed only in certain individuals of Cluster D (Fig. 4d). Moreover, five indel events (at sites 6 (A; from NEAnt), 127 (TAT; from NWAnt and SWAnt), 130 (from NEAnt and SEAnt), 219 (CTT; from NWAnt and SWAnt) and 261 (CAGAGA; from SEAnt) were only from Anatolia, and two indel events were from Thracia [at site 209 (from Thr2 and Thr3)] and Balkans [at sites 178 (from PeGrc) and 209 (A; from Blg2)].

We also investigated network approaches to the data and present a median-joining haplotype network in Fig. 5. The figure shows the high diversity of *cpnl-1* haplotypes recovered, and reveals that the network is not dominated either by high-frequency alleles nor structured by

large divergences between haplotype groups. The main geographic ranges, illustrated in the figure by separate colours, do not fall into single sections of the network indicating inheritance of haplotypes that are not necessarily each others' closest relatives. Anatolian haplotypes (red) are found throughout the network, closely related to almost all other haplotypes, and most of the low-frequency haplotypes have only been recovered from Anatolia. Results of neutrality tests and output from the mismatch distribution analysis including expansion time estimates based on *cpnl-1* sequences for all clusters are provided in of Appendix S3. With a generation time of one year, the estimated times using τ statistic of expansion for Cluster A-D respectively were 0.041, 0.259, 0.234 and 0.067 mya (Appendix S3).

The barrier position of the Marmara water body

Prior to the analyses, we investigated several features of the mtDNA in order to eliminate the possibility that the data may have included nuclear copies of mitochondrial DNA (numts). The presence of single peaks in each chromatograph, typical A+T bias observed for insect mtDNA, and the absence of indel and premature stop codons provide good evidence that the aligned sequences correspond to a functional mitochondrial region.

All of the findings of network analyses and demographic statistic tests indicated an undifferentiated population pattern across both sides of the Marmara water body indicating a recent expansion event or gene flow (See Appendix S3 for a more detailed comparison of the findings of network analyses and demographic statistic tests). The three IMA simulations produced similar results and all multidimensional peak locations fell within the 95% confidence intervals of the marginal posterior density distributions (Table 3). The estimates of divergence time for the Thracia-NWAnt were $\tau_{\text{Thracia-NWAnt}} = 0.039$ mya (0.024-0.178) (Table 3 & Fig. 6). In addition to the divergence time, we estimated the effective population size and the rate of gene flow between groups to test the possible effects of the current barrier

to population parameters. The effective population size of the NWAnt [θ_2 = 281650 (146737-519112)] was greater than that of the Thracian group [θ_1 = 158025 (28700-315237)], whereas the ancestral population was significantly smaller [θ_A = 58025 (100-325200); Table 3]. The rate of gene flow from the NWAnt to the Thracia was approximately 20 times greater (m_2 = 2.245) than that in the opposite direction (m_1 = 0.115) (Table 3 & Fig. 6).

Discussion

We observe a remarkable genetic diversity and genetic structuring of *C. parallelus* in Anatolia. This supports previous studies suggesting Anatolia either as a centre of diversity or a glacial refugium (Hewitt, 1999; Çıplak, 2008; Çıplak *et al.,* 2010; Ansell *et al.,* 2011; Sekercioğlu *et al.,* 2011). The evidences presented here are consistent with Anatolia as the most diverse of all European regions in terms of haplotype and nucleotide diversity (Table 1) as well as indel events (Fig. 4 & Appendix S1); although increased sampling of the whole distribution range is needed for more quantative comparisons. Multiple expansion and contraction events during the Pleistocene glacial cycles appear to be responsible for the present structuring and distribution of the species, with the formation of the differentiated genetic clusters (Fig. 4). Our data also reveals that Anatolian populations have had persistent contact and interaction with European populations, specifically with exchange between Balkans populations via Thracia and western Anatolia. These exchanges were also supported by confirmation that the Marmara water body was not a significant barrier to *C. parallelus* until the Holocene allowing a high level of gene flow from Anatolia to the Balkans (Table 3 & Fig.6). The implications of our results are discussed in detail below.

Spatial distribution of genetic diversity

345 According to the rear/leading edge concept, within/between population diversity in glacial
346 ancestral resource regions (or rear edge) are expected to be higher than that of expanding edge
347 (or leading edge) (Hampe & Petit, 2005; Diekman & Serrão, 2012). The findings of the
348 present study are not consistent with the main hypothesis that small populations isolated in the
349 rear edge are genetically depleted. The meadow grasshopper regions studied here were
350 characterized by high genetic diversity at the regional levels especially in refugia (Table 1)
351 and the results also indicated that most of the total genetic diversity was present within
352 regions (Table 1 & Appendix S1). Although haplotype diversity was high, low nucleotide
353 diversity values indicated only small differences between haplotypes (Table 1 & Fig. 5). The
354 locations of the regions across the species range might have triggered the variability of unique
355 haplotypes at regional levels. Indeed, the percent of unique haplotypes was highest in the
356 southern locations—Anatolia (78.29% of all detected haplotypes), Thracia (62.5%) and Spain
357 (54.55%) (Table 1). We also recorded a relatively high level of nucleotide diversity and
358 average number of nucleotide differences in southern regions (Table 1). The findings
359 suggested that while rear edge regions of *C. parallelus* are reservoirs of unique genetic
360 variation, this is not the case for the corresponding leading edge regions.
361 The spatial distribution of genetic diversity can also provide valuable insights into how rear
362 edge regions have contributed to postglacial recolonizations (Cooper *et al.,* 1995; Hewitt,
363 1996, 2000; Provan & Maggs, 2012). Italian and Iberian refugial populations of *C. parallelus*
364 have not contributed to the postglacial recolonization of Europe, probably because of the
365 Pyrenees and Alps acting as significant barriers to dispersal (Cooper & Hewitt, 1993; Lunt *et*
366 *al.,* 1998). The occurrence of low diversity values in some Spanish (Sspn, PSpn) and Balkan
367 (Blg2 and Blg3) regions (Table 1 & Fig. 2) might be explained by being separated from
368 nearby populations by unsuitable habitat; therefore high genetic drift and low gene flow are
369 likely drivers of their higher genetic differentiation (Diekmann & Serrão, 2012). Anatolian

370 regions show a more complex scenario however, exhibiting the highest level of genetic
371 diversity (Table 1), and the presence of a high number of indel positions (Fig. 4 & Appendix
372 1). The occurrence of a high nucleotide and haplotype diversity, as well as high number of
373 indels, in Anatolia represents a stable rear edge that has not become genetically depauperate
374 (Cooper *et al.,* 1995; Hewitt, 1996; Dubey *et al.,* 2007; Korkmaz *et al.,* 2010). Similar
375 conclusions were also reported for several other organisms from Anatolia (Moghaddam *et al.,*
376 2000; Rokas *et al.,* 2003; Dubey *et al.,* 2007).

377 **Geographical structure of genetic differentiation**

378 It is likely that the genetic patterns observed have been greatly influenced by climatic changes
379 associated with the Pleistocene glacial periods as these are known to influence drift and gene
380 flow, which will combine particularly strongly in the postglacial recolonization phase
381 (Hewitt, 1999). For wide-ranging species, climate fluctuations may have caused range
382 contractions in some parts of the range and expansions in others (Çıplak, 2004, 2008), which
383 together may have formed the distinct genetic lineages that we know today (Hewitt, 2004;
384 Ehrich *et al.,* 2007). In addition to the possible effect of climatic changes, topographic
385 structure and mountain chains of a certain region may have led to genetic differentiation.
386 These are consistent with our findings as the analyses related with genetic structuring indicate
387 that populations of *C. parallelus* are not genetically homogenous across distribution ranges
388 (Table 2 & Figs. 3-4). The formations of the cluster A and B suggest that the Anatolian
389 Diagonal acts as an effective barrier to gene flow between the west and east regions of
390 Anatolia (Fig. 4ab). This is also supported by the presence of indel positions specific to both
391 clusters (Fig. 4ab & Appendix S1). The barrier position of Anatolian Diagonal was also
392 observed in many taxa in Anatolia (Gündüz *et al.,* 2007 and references therein). On the other
393 hand, the mainly southern distribution of the species together with some Russian individuals
394 appears to represent a relatively old genetic unit (Cluster C; Fig. 4c & Appendix S3). This

structuring possibly has been formed by the direct effect of glacial periods rather than the geographical relationships among regions. This is further supported by presence of the most common indel at site 217 in the cluster C and again could be attributed to its old age, possibly resulting from a dispersal wave out of Anatolia (Fig. 4c). Moreover, this conclusion is consistent with the primary distinction of some southern refugial regions from all other in SAMOVA (Fig. 3) and also the component contributions (Fig. 2). At the same time, the geographic position of the Cluster C may point out to a historical connection among Spain, Italy, Balkans and Anatolia (Fig. 4c). Finally, the Cluster D consists of the current distribution area of the species and the absence of any indel events except in some Anatolian individuals suggests a relatively new dispersal wave out of Anatolia/Balkans (Fig. 4d).

Phylogeography, demographic history and the barrier position of the Marmara water body

Thanks to the studies of Hewitt and his colleagues, we already have considerable knowledge of the phylogeography of *C. parallelus* (Butlin *et al.,* 1991; Cooper & Hewitt, 1993; Cooper *et al.,* 1995; Hewitt, 1996, 1999, 2000; Lunt *et al.,* 1998). In these studies it is assumed that the present population in Europe, excluding Spain and southern Italy, is founded by a population dispersing from a Balkans refugium during the last interglacial period. The Spanish and Italian populations are thought to be separate units, originating from their own refugial stocks, arriving to these two peninsulas in an earlier interglacial.

Our data is partly consistent with these findings but add more details (see Tables 1-2 & Appendix S1). Our results support that the Iberia is structured into multiple groups as suggested earlier (Figs. 3-4; Gomez & Lunt, 2007). A reported relationship between some Spanish and Italian forms by Bella *et al.* (2007) is also accord with our findings. One of our striking results is the presence of some Russian individuals within the cluster C (Fig. 4c). This

419 may be attributed to a contraction and isolation event after an old dispersal wave since there is
420 no indication of a recent expansion (Appendix S3).
421 Although a detailed discussion of the interrelationships of the populations within Anatolia is
422 out of the scope of this current study our data shows the presence of Anatolian individuals in
423 all clusters and indicates multiple expansion events from Anatolia (Fig. 4). The hypothesis of
424 gene flow between western Anatolia and the Balkans is tested by the IMA analysis, with the
425 results indicating that substantial gene flow between the Balkans and Anatolia has indeed
426 continued since the establishment of these populations. Additional mitochondrial DNA data
427 also appear to support this conclusion (Korkmaz *et al.,* 2011), and suggest that the
428 populations of SEAnt are a source of genotypes involved in this expansion. The high rate of
429 gene flow from Anatolia to Thracia indicated that Anatolia has consistently acted as an
430 important source of diversity for European regions until the Holocene (Table 3 & Fig. 6). The
431 marine barrier presented to *C. parallelus* by the Bosphorus, Dardanelles and Marmara Sea has
432 of course been influenced greatly by climate cycles. A significant decrease in sea level during
433 the last glacial maximum (Ergin *et al.,* 2007) would have led to a large part of the Marmara
434 Sea becoming a land mass across which grasshoppers may have dispersed. These results are
435 in agreement with the above suggestions that the Anatolian population contributed to the
436 establishment of central and northern European populations and in this way the genetic
437 signatures of Anatolia may have extended before the Holocene as far north as Scandinavia
438 (Fig. 4d). This conclusion is fully accord with indels distributions (Fig. 4).
439 Comprehensive data for Anatolian regions has now brought the classic model to a new stage.
440 Hewitt (1996) has previously suggested that the ultimate ancestral region is probably in
441 Anatolia, and indeed our results suggest very strongly that Anatolia is the original source of
442 European populations. We add a number of novel aspects to the classic model of *C. parallelus*
443 phylogeography including the connectedness of the Balkans and Anatolia despite the apparent

444 marine barrier. The presence of a high number of indels in Anatolia with a clinal decline from
445 Anatolia to Balkan and Europe is in support of these assumptions (Fig. 4 & Appendix 1).
446 Thus, rather than a simple, small, bottlenecked Balkans refugium we have shown that a larger,
447 more permeable Anatolian centre of origin is a much more realistic model for west
448 Palaeoarctic diversity of *C. parallelus*.
449
450 **Acknowledgements**
451 This study is dedicated to the memory of Professor Godfrey Hewitt (1940-2013) who first
452 advocated *C. parallelus* as a study system, spoke often of the likely importance of Anatolia to
453 European biota, and through mentoring, encouragement, and research greatly advanced the
454 study of Eurasian phylogeography as a discipline. The financial support of this study was
455 provided by Cumhuriyet University via a research project numbered as CUBAP, F-245. A
456 part of this project was presented as a PhD thesis of the first author (EMK). We are also
457 grateful to The Council of Higher Education (Turkey) for providing a bursary of six months
458 to EMK to carry out a part of his PhD research project in the Bioinformatics laboratory of Dr.
459 Dave Lunt (University of Hull, UK). We are also grateful to Prof. Dr. Günter Köhler
460 (Friedrich Schiller University, Germany) and Dr. Dragan Chobanov (Bulgarian Academy of
461 Science, Bulgaria) for supplying some of the specimens studied. Thank to Sarp Kaya, Mahir
462 Budak and Mahir Yıldırım for their help and companionships during field studies.
463
464 **References**
465 Akın, Ç., Bilgin, C.C., Beerli, P., Westaway, R., Ohst, T., Litvinchuk, S.N., Uzzell, T., Bilgin,
466 M., Hotz, H., Guex, G.D. & Plötner, J. (2010) Phylogeographic patterns of genetic diversity
467 in eastern Mediterranean water frogs were determined by geological processes and climate
468 change in the Late Cenozoic. *Journal of Biogeography*, **37**, 2111-2124.

Ansell, S.W., Stenøien, H.K., Grundmann, M., Russell, S.J., Koch, M.A., Schneider, H. & Vogel, J.C. (2011) The importance of Anatolian mountains as the cradle of global diversity in *Arabis alpina*, a key arctic‐alpine species. *Annals of Botany*, **108**, 241-252.

Atkinson, R.J., Rokas, A. & Stone, G.N. (2007) Longitudinal patterns in species richness and genetic diversity in European oaks and oak gallwasps. *Phylogeography of southern European refugia* (ed. by S. Weiss, N. Ferrand), pp. 127-151. Springer, New York.

Bandelt, H.J., Forster, P. & Röhl, A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37-48.

Bei-Bienko, G.J. & Mistshenko, L.L. (1951) *The grasshoppers of the fauna the USSR and adjacent countries*. Zoologicheskii institut: Akademiia Nauk SSSR.

Bella, J.L., Serrano, L., Orellana, J. & Mason, P.L. (2007) The origin of the *Chorthippus parallelus* hybrid zone: chromosomal evidence of multiple refugia for Iberian populations. *Journal of Evolutionary Biology*, **20**, 568-76.

Butlin, R.K., Ritchie, M.G. & Hewitt, G.M. (1991). Comparisons among morphological characters and between localities in the *Chorthippus parallelus* hybrid zone (Orthoptera: Acrididae). *Philosophical Transactions of the Royal Society B: Biological Sciences*, **334**, 297-3087.

Cooper, S.J.B. & Hewitt, G.M. (1993) Nuclear DNA sequence divergence between parapatric subspecies of the grasshopper *Chorthippus parallelus*. *Insect Molecular Biology*, **2**, 185-194.

Cooper, S.J.B., Ibrahim, K.M. & Hewitt, G.M. (1995) Postglacial expansion and genome subdivision in the European grasshopper *Chorthippus parallelus*. *Molecular Ecology*, **4**, 49-60.

Çıplak, B. (2004) Systematics, phylogeny and biogeography of *Anterastes* (Orthoptera, Tettigoniidae, Tettigoniinae): evolution within a refugium. *Zoologica Scripta*, **33**, 19-44.

Çıplak, B. (2008) The analogy between glacial cycles and global warming for the glacial

494 relicts in a refugium: a biogeographic perspective for conservation of Anatolian Orthoptera.
495 *Insect Ecology and Conservation* (ed. by S. Fattorini), pp. 135-163. Research Signpost.
496 Çıplak, B. Kaya, S. & Gündüz, İ. (2010) Phylogeography of *Anterastes serbicus* species
497 group (Orthoptera, Tettigoniidae): phylogroups correlate with mountain belts, but not with the
498 morphospecies. *Journal of Orthoptera Research*, **19**, 29-40.
499 Defaut, B. (2012) Implications taxonomiques et nomenclaturales de publications récentes en
500 phylogénie moléculaire: 1. Les Gomphocerinae de France (Orthoptera, Acrididae). *Matériaux*
501 *Orthoptériques et Entomocénotiques*, **17**, 15-20.
502 Diekmann, O.E. & Serrão, E.A. (2012) Range-edge genetic diversity: locally poor extant
503 southern patches maintain a regionally diverse hotspot in the seagrass *Zostera marina*.
504 *Molecular Ecology*, **21**, 1647-1657.
505 Dubey, S., Cosson, J.F., Vohralík, V., Krystufek, B., Diker, E. & Vogel, P. (2007) Molecular
506 evidence of Pleistocene bidirectional faunal exchange between Europe and the Near East: the
507 case of the bicoloured shrew (*Crocidura leucodon*, Soricidae). *Journal of Evolutionary*
508 *Biology,* **20**, 1799-1808.
509 Dupanloup, I., Schneider, S. & Excoffier, L. (2002) A simulated annealing approach to define
510 the genetic structure of populations. *Molecular Ecology*, **11**, 2571-2581.
511 Ehrich, D., Gaudeul, M., Assefa, A, Koch, M.A., Mummenhoff, K., Nemomissa, S.,
512 Intrabiodiv Consortium & Brochmann, C. (2007) Genetic consequences of Pleistocene range
513 shifts: contrast between the Arctic, the Alps and the East African mountains. *Molecular*
514 *Ecology*, **16**, 2542-2559.
515 Ergin, M., Uluadam, E., Sarıkavak, K., Keskin, Ş., Gökaşan, E. & Tur, H. (2007). Late
516 Quaternary sedimentation and tectonics in the submarine Şarköy Canyon, western Marmara
517 Sea (Turkey). *The geodynamics of the Aegean and Anatolia* (ed. by T. Taymaz, Y. Yılmaz
518 and Y. Dilek) pp. 231-257, Geological Society, Special Publications, London.

519 Excoffier, L., Laval, G. & Schneider, S. (2005) Arlequin ver. 3.0: An integrated software
520 package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47-50.
521 Fritz, U., Ayaz, D., Hundsdörfer, A.K., Kotenko, T., Guicking, D., Wink, M., Tok, C.V.,
522 Cicek, K. & Buschbom, J. (2009) Mitochondrial diversity of European pond turtles (*Emys*
523 *orbicularis*) in Anatolia and Ponto-Caspian region: multiple old refuges, hotspots of extant
524 diversification and critically endangered endemics. *Organisms, Diversity & Evolution,* **9**, 100-
525 114.
526 Fu, Y.X. (1997) Statistical neutrality of mutations against population growth, hitchhiking and
527 background selection. *Genetics*, **147**, 915-925.
528 Gomez, A & Lunt, D.H. (2007) Refugia within refugia: patterns of phylogeographic
529 concordance in the Iberian Peninsula. *Phylogeography of southern European* refugia (ed. by
530 S. Weiss, N. Ferrand). pp. 155-188, Springer, Dordrecht, The Netherlands.
531 Gündüz, I., Jaarola, M., Tez, C., Yeniyurt, C., Polly P.D. & Searle, J.B. (2007) Multigenic
532 and morphometric differentiation of ground squirrels (*Spermophilus*, Sciuridae, Rodentia) in
533 Turkey. *Molecular Phylogenetics and Evolution*, **43**, 916-935.
534 Hampe, A. & Petit, R.J. (2005) Conserving biodiversity under climate change: the rear edge
535 matters. *Ecology Letters*, **8**, 461-467.
536 Harpending, H.C. (1994) Signature of ancient population-growth in a lowresolution
537 mitochondrial-DNA mismatch distribution. *Human Biology,* **66**, 591-600.
538 Hewitt, G.M. (1996) Some genetic consequence of ice ages, and their role in diverging and
539 speciation. *Biological Journal of Linnean Society*, **58**, 247-276.
540 Hewitt, G.M. (1999) Post-glacial re-colonization of European biota. *Biological Journal of*
541 *Linnean Society*, **68**, 87-112.
542 Hewitt, G.M. (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907-913.
543 Hewitt, G.M. (2004) Genetic consequences of climatic oscillations in the Quaternary.

544 *Philosophical Transactions of the Royal Society B: Biological Sciences*, **359**, 183-195.
545 Hey, J. (2010) Isolation with migration models for more than two populations. *Molecular*
546 *Biology and Evolution*, **27**, 905-920.
547 Hudson, R.R., Boos, D.D. & Kaplan, N.L. (1992) A statistical test for detecting geographic
548 subdivision. *Molecular Biology and Evolution,* **9**, 138-151.
549 Hutchison, D.W. & Templeton, A.R. (1999) Correlation of pairwise genetic and geographic
550 distance measures: inferring the relative influences of gene flow and drift on the distribution
551 of genetic variability. *Evolution*, **53**, 1898-1914.
552 Jombart, T. (2008) ADEGENET: a R package for the multivariate analysis of genetic
553 markers. *Bioinformatics,* **24**, 1403-1405.
554 Jombart, T., Devillard, S. & Balloux, F. (2010) Discriminant analysis of principal
555 components: a new method for the analysis of genetically structured populations. *BMC*
556 *Genetics,* **11**, 94.
557 Korkmaz, E.M., Sari, M. & Basibuyuk, H.H. (2010) Genetic structure of *Chorthippus*
558 *parallelus* (Orthoptera: Acrididae: Gomphocerinae) populations in Anatolia: a stable rear
559 edge population. *Annals of the Entomological Society of America*, **103**, 625-634.
560 Korkmaz, E.M. (2011) Phylogeography and landscape genetics of Anatolian populations of
561 the meadow grasshopper *Chorthippus parallelus* (Zetterstedt, 1821) (Acrididae: Orthoptera).
562 PhD Thesis, Cumhuriyet University, Sivas.
563 Librado, P. & Rozas, J. (2009) DnaSP v5: a software for comprehensive analysis of DNA
564 polymorphism data. *Bioinformatics*, **25**, 1451-1452.
565 Lunt, D.H., Ibrahim, K.M. & Hewitt, G.M. (1998) mtDNA phylogeography and postglacial
566 patterns of subdivision in the meadow grasshopper *Chorthippus parallelus*. *Heredity*, **80**, 633-
567 641.
568 Médail, F. & Diadema, K. (2009) Glacial refugia influence plant diversity patterns in the

569 Mediterranean Basin. *Journal of Biogeography*, **36**, 1333-1345.
570 Moghaddam, M., Ehdaie, B. & Waines, J.G. (2000). Genetic diversity in populations of wild
571 diploid wheat *Triticum urartu* Tum. ex. Gandil. revealed by isozyme markers. *Genetic*
572 *Resources and Crop Evolution*, **47**, 323-334.
573 Mutun, S. (2011) Intraspecific genetic diversity of the oak gall wasp *Andricus lucidus*
574 (Hymenoptera: Cynipidae) populations in Anatolia. *Turkish Journal of Zoology*, **35**, 559-570.
575 Papadopoulou, A., Anastasiou, I. & Vogler, A.P. (2010) Revisiting the insect mitochondrial
576 molecular clock: the mid-aegean trench calibration. *Molecular Biology and Evolution,* **27**,
577 1659-1672.
578 Peakall, R. & Smouse, P.E. (2006) GENALEX 6: genetic analysis in Excel. Population
579 genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288-295.
580 Petit, R.J., Mousadik, A.E. & Poas, O. (1998) Identifying populations for conservation on the
581 basis of genetic markers. *Conservation Biology*, **12**, 844-855.
582 Provan, J. & Maggs, C.A. (2012) Unique genetic variation at a species' rear edge is under
583 threat from global climate change. *Proceedings of the Royal Society B: Biological Sciences*,
584 **279**, 39-47.
585 Rokas, A., Atkinson, R.J., Webster, L.W.I., Csóka, G. & Stone, G.N. (2003) Out of Anatolia:
586 longitudinal gradients in genetic diversity support an eastern origin for a circum-
587 Mediterranean oak gallwasp *Andricus quercustozae*. *Molecular Ecology,* **12**, 2153-2174.
588 Schmitt, T. (2007) Molecular biogeography of Europe: Pleistocene cycles and postglacial
589 trends. *Frontiers in Zoology*, **4**, 11.
590 Stamatis, C., Suchentrunk, F., Moutou, K.A., Giacometti, M., Haerer, G., Djan, M., Vapa, L.,
591 Vukovic, M. & Tvrtkovic, N. (2009) Phyloegography of the brown hare (*Lepus europaeus*) in
592 Europe: a legacy of south-eastern Mediterranean refugia? *Journal of Biogeography*, **36**, 515-
593 528.

Şekercioğlu, Ç.H., Anderson, S., Akçay, E., Bilgin, R., Can, Ö.E., Semiz, G., Tavsanoglu, Ç., Yokes, M.B., Soyumert, A., Ipekdal, K., Saglam, I.K., Yücel, M. & Dalfes H.N. (2011) Turkey's globally important biodiversity in crisis. *Biological Conservation*, **144**, 2752-2769.

Taberlet, P., Fumagalli, L., Wust-Saucy, A.G. & Cosson, J.F. (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453-464.

Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585-595.

Walsh, P.S., Metzger, D.A. & Higuchi, R. (1991) Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques*, **10**, 506-1.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Sampling data information, accession numbers and indel variants of *cpnl-1* haplotypes and shared *cpnl-1* haplotypes between regions of *Chorthippus parallelus*.

Appendix S2 DAPC on R package to identify and describe all possible genetic structuring across the distribution area of *C. parallelus*.

Appendix S3 The results of demographic history of the clusters, IMA and the network and demographic analyses of Thracian and North West Anatolian groups to investigate the barrier position of the Straits;

Biosketch

Dr. E. Mahir Korkmaz an Assistant Professor at the Cumhuriyet University is interested in systematics and molecular evolution of Western Palaearctic species. In particularly, he is focused on Anatolian historical biogeography and mitochondrial genome evolution in insects.

Research interests of the co-authors include molecular phylogeny, phylogeography, bioinformatics and biogeography. For further information please see: www.cumsag.com; http://biyo.fen.akdeniz.edu.tr/tr.i54.prof-dr-battal-ciplak; http://davelunt.net/

Author contributions: A core team (E.M.K., B.Ç., H.H.B.) conceived the principal ideas and the main structure of the manuscript and led the writing; E.M.K., B.Ç. and N.D. collected the specimens and data; D.L. helped with the analyses and writing.