Genetic effects on life-history traits in the Glanville fritillary butterfly

Anne Duplouy [Corresp., 1] , **Swee C Wong** [1] , **Jukka Corander** [2, 3] , **Rainer Lehtonen** [4] , **Ilkka Hanski** [5]

[1] Department of Biosciences, Metapopulation Research Centre, University of Helsinki, Helsinki, Finland

[2] Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland

[3] Department of Biostatistics, Institute of Basic Medical Sciences, Faculty of Medecine, University of Oslo, Oslo, Norway

[4] Institute of Biomedicine & Genome-scale Biology Research Program, Biomedicum, Faculty of Medicine, University of Helsinki, Helsinki, Finland

[5] Department of Biosciences, University of Helsinki, Helsinki, Finland

Corresponding Author: Anne Duplouy
Email address: anne.duplouy@helsinki.fi

Background. Adaptation to local habitat conditions may lead to the natural divergence of populations in life-history traits such as body size, time of reproduction, mate signaling or dispersal capacity. Given enough time and strong enough selection pressures, populations may experience local genetic differentiation. The genetic basis of many life-history traits, and their evolution according to different environmental conditions remain however poorly understood.

Methods. We conducted an association study on the Glanville fritillary butterfly, using material from five populations along a latitudinal gradient within the Baltic Sea region, which show different degrees of habitat fragmentation. We investigated variation in ten principal components, cofounding in total 21 life-history traits, according to two environmental types, and 33 genetic SNP markers from 15 candidate genes.

Results. We found that nine SNPs from five genes showed strong trend for trait associations (*p*-values under 0.001 before correction). These associations, yet non-significant after multiple test corrections, with a total number of 1086 tests, were consistent across the study populations. Additionally, these nine genes also showed an allele frequency difference between the populations from the northern fragmented versus the southern continuous landscape.

Discussion. Our study provides further support for previously described trait associations within the Glanville fritillary butterfly species across different spatial scales. Although our results alone are inconclusive, they are concordant with previous studies that identified these associations to be related to climatic changes or habitat fragmentation within the Åland population.

1   Genetic effects on life-history traits in the Glanville fritillary butterfly

3   Anne Duplouy[1]*, Swee Chong Wong[1]**, Jukka Corander[2,3], Rainer Lehtonen[4] and Ilkka

4   Hanski[1]†

6   [1] Department of Biosciences, Metapopulation Research Centre, University of Helsinki, Helsinki,

7   Finland

8   [2] Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland

9   [3] Department of Biostatistics, Institute of Basic Medical Sciences, Faculty of Medicine,

10   University of Oslo, Oslo, Norway

11   [4] Institute of Biomedicine & Genome-scale Biology Research Program, Faculty of Medicine,

12   University of Helsinki, Helsinki, Finland

13   † Deceased

15   **Corresponding authors:** * anne.duplouy@helsinki.fi and ** sweechongwong@gmail.com

16   (fax: +358 9191 57694)

18   ***Running title*:** Genetic effects on butterfly life-history traits

20   **Abstract**

21   **Background.** Adaptation to local habitat conditions may lead to the natural divergence of

22   populations in life-history traits such as body size, time of reproduction, mate signaling or

23   dispersal capacity. Given enough time and strong enough selection pressures, populations may

24   experience local genetic differentiation. The genetic basis of many life-history traits, and their

25   evolution according to different environmental conditions remain however poorly understood.

26   **Methods.** We conducted an association study on the Glanville fritillary butterfly, using material

27   from five populations along a latitudinal gradient within the Baltic Sea region, which show

28   different degrees of habitat fragmentation. We investigated variation in ten principal

29   components, cofounding in total 21 life-history traits, according to two environmental types, and

30   33 genetic SNP markers from 15 candidate genes.

31   **Results.** We found that nine SNPs from five genes showed strong trend for trait associations (p-

32   values under 0.001 before correction). These associations, yet non-significant after multiple test

33   corrections, with a total number of 1086 tests, were consistent across the study populations.

34   Additionally, these nine genes also showed an allele frequency difference between the

35   populations from the northern fragmented versus the southern continuous landscape.

36   **Discussion.** Our study provides further support for previously described trait associations within

37   the Glanville fritillary butterfly species across different spatial scales. Although our results alone

38   are inconclusive, they are concordant with previous studies that identified these associations to

39   be related to climatic changes or habitat fragmentation within the Åland population.

41   **Key words**: *Melitaea cinxia*; life-history; association study; *Pgi*; fragmented population

43 **Introduction**

44 Human land use has led to natural habitat loss and fragmentation such that numerous species and

45 populations are now at risk of extinction (Hanski 2005; Hughes et al. 1997). The impacts of the

46 remaining patch matrix structure or metapopulations, of the size and the number of remaining

47 suitable habitat patches, of the level of connectivity or the isolation degree between the patches

48 on the persistence and evolutionary dynamics of species and populations have been intensely

49 investigated (Hanski & Mononen 2011; Hanski & Ovaskainen 2000). Life-history traits such as

50 body size, reproductive output or dispersal capacity are known to often vary accordingly to the

51 degree of fragmentation experienced by the different populations (Hanski 2011; Saastamoinen

52 2007b; Wheat et al. 2011). Although the type and the strength of natural selection in fragmented

53 habitat might vary locally, given enough time and strong enough selection pressures, persisting

54 populations might adapt to fragmentation and experience genetic differentiation. To date, the

55 evolution of the genetic basis of many life-history traits according to habitat fragmentation

56 remain however poorly understood. This also remains true in the case of the Glanville fritillary

57 butterfly, *Melitaea cinxia* (L.), despite that the ecology of the species has been intensively

58 investigated over the last 20 years in the Åland islands, and the species is now considered as a

59 model organism in the dynamics and evolution of natural (meta)population.

60 Allelic variation in metabolic genes is often associated with life-history traits in natural

61 populations (Marden 2013). A much-studied example is the gene *Phosphoglucose isomerase*

62 (*Pgi*) in butterflies but also in other insects (Dahlhoff & Rank 2000) and plants (Mitchell-Olds &

63 Pedersen 1998). In the Orange Sulfur butterfly (*Colias eurytheme,* Pieridae), *Pgi* allozyme alleles

64 are associated with survival in the field (Watt 1977) and reproductive success (Watt et al. 1985),

65 while in the Glanville fritillary butterfly (*Melitaea cinxia,* Nymphalidae), a SNP in *Pgi* is

66 associated with flight metabolic rate (Niitepold 2010; Niitepõld et al. 2009) and dispersal rate in

67   the field (Hanski & Mononen 2011; Niitepõld et al. 2009), population growth rate (Hanski &

68   Saccheri 2006), female fecundity (Saastamoinen 2007b), body temperature at flight

69   (Saastamoinen & Hanski 2008), lifespan (Orsini et al. 2009; Saastamoinen et al. 2009) and larval

70   development (Saastamoinen et al. 2013). Considering insect studies on other candidate genes,

71   Saastamoinen et al. (Saastamoinen et al. 2013) found genetic associations between the incidence

72   of an extra larval instar and allelic variation in *Pgi*, *Serpin-1* and *Vitellin-degrading protease*

73   *precursor* genes. Importantly, many of these associations involve a significant interaction

74   between genotype and ambient (Niitepold 2010) or acclimation temperature (Wong et al. 2016),

75   consistent with the hypothesis of temperature-dependent enzyme kinetics.

76   Apart from association studies with candidate genes, genetic effects on life-history

77   traits in the Glanville fritillary are demonstrated by significant heritabilities of many traits (de

78   Jong et al. 2014; Klemme & Hanski 2009; Kvist et al. 2015; Mattila & Hanski 2014). For

79   instance, recent studies have shown high heritabilities for both post-diapause larval development

80   time (h^2=0.37) and lifetime offspring production (h^2=0.62) (de Jong et al. 2014; Kvist et al.

81   2013). These studies do not discriminate between single gene and multigene effects, though de

82   Jong et al. (2014) found that genetic polymorphism in the cytochrome P450 gene *CYP337*

83   accounted for 14% of the estimated heritability in egg production.

84   A generic problem with many candidate gene studies on wild populations is relatively

85   small sample sizes and lack of replication, which cast doubt on the conclusions even in the case

86   of significant results. Exceptions that demonstrate convincing single-locus associations include

87   the study of the *Sdh* gene polymorphism (*Sdh* is a subunit of a *Succinate dehydrogenase*,

88   SDH, gene), which is associated with a reduction in SDH enzyme activity in flight muscles and

89   an increase of metabolic rate during flight in the Glanville fritillary (Marden et al. 2013; Wheat

90   et al. 2011). Variation in the same locus is associated with longevity in hyperoxic conditions in

91   *Drosophila* (Walker et al. 2006).

92   Here, we aimed to test whether habitat fragmentation might drive genetic

93   differentiation between five natural populations of the Glanville fritillary butterfly, using

94   numerous SNP marker-phenotypic trait associations. We show evidence of associations at

95   various loci, through a candidate gene study involving a total of 49 Single Nucleotide

96   Polymorphisms (SNPs) from 26 candidate genes (metabolic genes). All SNPs were picked

97   mostly based on previous similar, but more limited studies in terms of total number of traits,

98   populations or markers included. We investigated 21 different life-history traits, selected from

99   traits previously recorded by a study by Duplouy et al. (2013) investigating phenotypic

100   differentiation of butterfly populations under landscape fragmentation, and that we found to have

101   high importance in three principal components analyses. Importantly, the present study is

102   replicated by including butterfly samples from five different regional populations in the Baltic

103   area, one population in Finland, three in Sweden and one in Estonia. Additionally, we also

104   provide data from a pilot study, providing a temporal semi-replication to the main study (only on

105   larval and pupal development traits). Several studies have shown that some differences in life-

106   history traits (Duplouy et al. 2013), allelic frequencies (Fountain et al. 2016) and gene expression

107   profiles (Somervuo et al. 2014) occur between the populations inhabiting fragmented versus

108   continuous landscapes. As both northern regional populations inhabit a highly fragmented

109   landscape, including the well-studied metapopulation in the Åland Islands (Hanski 2011), while

110   the three remaining southern regional populations inhabit a more continuous habitat, we

111   incorporated an environment factor confounding both the degree of habitat fragmentation and the

112   latitude (or climatic region) of the populations, to our analyses to test if and how such factor may

113 affect the Glanville fritillary butterfly at the genetic level. Although other major population

114 characteristics, such as the different degree of relatedness between the populations, might play a

115 role, our results present various potential genetic associations that despite being non-significant

116 after accounting for all tests included in the study (N=1086) are consistent across the

117 populations, and the two environments.

118

119 **Material and methods**

120 ***Study material***

121 We used the material reported in the phenotypic life-history study of Duplouy et al. (Duplouy et

122 al. 2013). We included the results from a pilot study that took place in 2007, as they provide

123 independent replicate to the findings from the main 2010 study. Sample sizes for each year in

124 given in the Supplementary Table S1.

125   Pre-diapause larvae of the Glanville fritillary were sampled from five regional

126 populations in the Baltic Sea region (Fig. 1) in September 2006 and September 2009. Both the

127 Åland Islands (ÅL) metapopulation in Finland (60°07'N, 19°54'E) and the Swedish east-coast

128 Uppland (UP, 59°90'N, 17°50'E) population occur in highly fragmented landscapes, where the

129 meadow habitat occurs in small fragments (average area: 0.19 ha in ÅL, Ojanen et al. 2013),

130 which cover only about one percent of the total landscape area (Hanski 2011). In contrast, in

131 Öland (ÖL, 56°39'N, 16°38'E) and Gotland (GO, 57°47'N, 18°49'E) in Sweden, and in

132 Saaremaa (SA, 58°05'N, 22°15'E) in Estonia, there are extensive continuous calcareous

133 meadows (alvars), often exceeding several hundred hectares in size. The larval host plants

134 *Plantago lanceolata* and *Veronica spicata* are the same in all five populations (Duplouy et al.

135 2013; Nieminen et al. 2004). Females lay large clutches of ca 150 eggs and the larvae live in

136   groups of full sibs (Nieminen et al. 2004). Despite generally colder monthly temperature minima

137   recorded across the years for both northern fragmented populations (Supplementary Fig. S1), the

138   butterfly phenologies are generally considered similar between the five populations of this study

139   (Duplouy et al. 2013). We anyway confounded both the degree of fragmentation and the latitude

140   of the population into one environment factor: northern fragmented environment versus southern

141   continuous environment. For more detailed description of the species' ecology and eco-

142   evolutionary dynamics see Nieminen et al. (2004) and Hanski (2011), respectively.

143   For the pilot study ran in the summer 2007, our samples originated from different larval

144   family groups in each population ($N_{\text{ÅL}}$=9, N_{UP}=13, N_{GO}=2 and N_{SA}=12), with several individuals

145   from the same family included. In contrast, for the experiment ran in the summer 2010, our

146   samples originated from 50 larval family groups from each population, widely distributed across

147   the respective study areas, and for which the degree of relatedness was previously estimated to

148   be low (estimate of genetic similarity by Duplouy et al. 2013). Most butterflies belong to the

149   eastern mitochondrial clade, originating from the Asian glacial refuge area, with the minor

150   exception of a few butterflies from ÖL belonging to the south-eastern clade (Duplouy et al. 2013;

151   Wahlberg & Saccheri 2007). According to the IUCN Red List, the butterfly is not classified as a

152   threatened species in the five study regions, hence no permits were required for sampling.

153   For a detailed description of the rearing conditions and measuring of the phenotypic

154   traits see Duplouy et al. (2013). In brief, the caterpillars diapaused in the constant temperature of

155   3°C in incubators, the post-diapause larvae were reared indoors under 12:12 L/D light and

156   25/15°C day/night temperature and were fed with the leaves of *P. lanceolata,* cultivated in the

157   laboratory. Adult butterflies were individually marked and released into an outdoor population

158   cage to collect data on behavioural and reproductive traits as described below.

159

160   ***Phenotypic traits***

161   *Larval and pupal development.* For each individual in the 2007 pilot study, we only recorded the

162   sex, the pupal weight (mg) and the post-diapause larval period (from 5th instar molt to pupation,

163   in days). For each individual collected in the fall 2009, we recorded the sex and the weights of

164   the 5th, 6th, and 7th instar larvae and pupae, the time period between successive molts, and the

165   pupal period.

166

167   *Outdoor cage experiment.* We did not collect any adult traits from the 2007 pilot study. Twenty-

168   four hours after eclosion, butterflies emerging from the larvae collected in the fall 2009, were

169   marked with a unique number written on the hind wing and released into a large outdoor

170   population cage (32 x 26 x 3 m) at the Lammi biological station in southern Finland (Duplouy et

171   al. 2013; Hanski et al. 2006; Saastamoinen 2008). Data on each mating and oviposition were

172   collected by monitoring the butterflies continuously between 8 am and 6 pm for ten sunny days

173   ranging between the 4th and the 20th of June 2010. A sunny day is defined as a day with no or

174   little overcast, thus allowing the warming-up of the butterflies' body temperature and the

175   initiation of activity in the cage (flight and oviposition are initiated at ambient temperature

176   T>13°C, Saastamoinen 2007a; Saastamoinen 2007b; Saastamoinen & Hanski 2008). The

177   temperatures experienced in June 2010 ranged from 11°C to 23.5°C (monthly average=16.6°C),

178   which remains well within the range of the average temperatures recorded in the five populations

179   (Supplementary Fig. S1). At 9, 11, 13, 15 and 17 hours, we recorded the position in the cage (for

180   a map of the cage see: Duplouy et al. 2013), and whether the encountered butterfly was flying or

181   not (probability of the butterfly flying in the cage). The average distance between two

182   observations of the same individual was calculated as the cumulative distance between all

183   consecutive observations divided by the total number of observations for the butterfly. For both

184   males and females we calculated the age at first mating and mating success (total numbers of

185   observed copulations). The survival of each individual was defined as the time between eclosion

186   and its last observation, dead or alive.

187   The cage had 200 potted host plants, 100 plants of *P. lanceola* and *V. spicata* each.

188   Each egg clutch laid by a female was placed individually on a petri dish labelled with the female

189   identity, the host plant species and the date of oviposition, and incubated in the laboratory. The

190   host plants were monitored for egg-laying females intensively, to record each egg clutch

191   (Duplouy et al. 2013). Eggs were counted three days after oviposition and the caterpillars were

192   counted three days after eclosion. For each female, we calculated the age at 1st oviposition, total

193   number of clutches, eggs and caterpillars, the host plant preference (the proportion of clutches

194   laid on *P. lanceolate*), and the egg hatch rate of each clutch (percentage of eggs that hatched).

195   As many of the traits recorded were highly intercorrelated (Table S2, Pearson's

196   correlation), and to avoid the use of a large number of tests and the subsequent increased risk of

197   type I error, we reduced our analyses to three independent principal component analyses (PCAs),

198   and ten principal components (PCs). Chosen PCs all had *eigen*-values over 1, and proportion of

199   variance values over 0.1. The first PCA (PCA$_1$) included all the larval and pupal traits from both

200   sexes using only those individuals for which information was available for all developmental

201   traits (as described in Duplouy et al. 2013). PCA$_1$ provided four PCs (PC$_{1-1}$ to PC$_{1-4}$). The second

202   PCA (PCA$_M$) included only male adult traits and produced three PCs (PC$_{M1}$ to PC$_{M3}$), while the

203   third PCA (PCA$_F$) included only female adult traits, and produced three PCs (PC$_{F1}$ to PC$_{F3}$) (see

204   Supplementary Table S3, S4 and S5 for details on each PCAs, and description of the phenotypic

205   traits included in each PCA). The probability of a butterfly being observed flying, the hatch rate

206   of all clutches and of the 1st clutch only, and the host plant preference were treated as

207   proportions, and were ArcSin-transformed prior being integrated into the PCAs. The ten

208   analysed PCs accounted for the majority of the total variation in the traits included in the

209   respective PCAs (74% in developmental-traits, and over 58% in male and female adult-traits).

210   Normality of the ten PCs was assessed using Shapiro-Wilk tests (α=0.01).

211

212   ***Genotyping***

213   A small fragment of the hind wing (3.14 mm^2) was sampled from each butterfly using sterile

214   biopsy punch (Tamro, Vantaa). The biopsy does not impair the flight capacity of the butterfly

215   (Hanski et al. 2006). Wing tissues, plus all body tissues from individuals recovered at the end of

216   the experiment, were stored individually in 99.6% ethanol. The DNA was extracted at the

217   Institute of Biotechnology, University of Helsinki, using Nucleo Spin® 96 Tissue kit (Macherey-

218   Nagel GmbH & Co. KG, Germany) following manufacturer's protocol. DNA was amplified

219   using the Illustra GenomiPhi V2 DNA Amplification Kit according to manufacturer's protocol

220   (GE Healthcare, USA). All DNA extracts were genotyped for two SNPs from the *Pgi* gene, or all

221   49 SNPs from 26 candidate genes, for the pilot and main experiment samples, respectively

222   (Supplementary Table S6).

223   The candidate genes were selected based on a study by Ahola *et al*. (Ahola et al. 2015),

224   which compared transcriptomic variation between individuals from Åland Islands and China, and

225   revealed genes with significantly reduced variation in the Åland Islands, possibly due to a

226   selective sweep or local adaptation. The resulting gene set was tested by gene set enrichment

227   analysis. Among the overrepresented GO groups, genes belonging to serine proteases, to serine

228 protease inhibitors and to cuticular proteins GO groups were chosen for genotyping (Ahola et al.

229 2015). Following selection, 31 SNP markers from nine genes fulfilled the quality criteria and

230 were retained for further analysis. To this list, we added the following markers: (i) three SNPs

231 from the *Phosphoglucose isomerase* (*Pgi*) gene (Orsini et al. 2009), (ii) eight SNPs with minor

232 allele frequency (MAF) > 0.2 in an annotated EST library of *Pgi* pathway associated genes

233 (upstream and downstream)(Ahola et al. 2015), and (iii) ten SNPs with interesting expression

234 patterns in a chip expression dataset (Kvist et al. 2013). Supplementary Table S6 provides the list

235 of the all selected markers. All SNPs included in this study were bi-allelic SNPs. SNP

236 genotyping was performed at the Institute for Molecular Medicine Finland (FIMM, Helsinki,

237 Finland) using Sequenom iPLEX Gold chemistry. The results were validated for seven

238 independent samples by direct genomic sequencing with AB 3730 (Applied Biosystems Europ,

239 Bleiswijk, The Netherlands). We used MassARRAY Assay Design 3.1 to design the primers for

240 Sequenom, while the validation primers and extension probes were designed using Primer3

241 (Rozen & Skaletsky 2000).

242   The quality of SNPs, including deviation from the Hardy-Weinberg equilibrium

243 (HWE), was assessed using an in-house quality control pipeline (Wong 2011). False discovery

244 rate calculation was used to correct for multiple testing (α=0.05) (Hochberg & Benjamini 1990).

245 For the pilot study, one of the two *Pgi* SNPs genotyped did not reach our quality criteria, while

246 for the samples collected in 2009, out of the 49 SNPs genotyped, 16 did not reach the quality

247 criteria; they were hence excluded from subsequent analyses. The criteria included unclear

248 heterozygote/homozygote cluster patterns and no heterozygote cluster. We also remove SNPs

249 with significant deviations from the HWE or a minimum allele frequency (MAF) less than 0.05,

250 or both in two or more of the populations. Similarly, if only one of the four populations did not

251   comply with the HWE or the MAF at a certain SNP, we removed that population from the

252   analyses (Supplementary Table S6).

253

254   ***Data analyses***

255   *Phenotype-genotype associations.* Subsequent analyses were carried out using R 3.1.3 software

256   (R Core Team 2013). For the 2007 pilot study, we used linear mixed models to analyse

257   separately the effect of the *Pgi* SNP (*Pgi*:c.331A>C) on the pupal weight and post-diapause

258   larval period. Sex, environment type and genotype were used as explanatory variables.

259   Population and family were added as a random factors nested within environment in the models.

260   For the samples collected in 2009, we used linear mixed models to separately test the

261   effect of 33 SNPs on the four principal components (PC_{1-1} to PC_{14}) describing post-diapause

262   larval and pupal development. Sex, environment type, genotype and environment by genotype

263   interaction were used as explanatory variables to those models, while population was included as

264   a random factor nested within the environment type. Similar models, excluding the sex factor,

265   were used to test genetic associations to the three principal components from both PCA_F and

266   PCA_M. Initial models with interaction term were subsequently simplified by the removal of non-

267   significant terms to give a final minimal adequate model.

268   We accounted for additive, dominant, recessive, and over-recessive/dominant effects of

269   the alleles, and conducted false discovery rate calculation (Hochberg & Benjamini 1990) to

270   correct for multiple testing for all tests. This means that each raw *p*-value was corrected for a

271   total of 1086 tests conducted (including 4 inheritance models x 10 PCs x 33 SNPs that reached

272   our quality criteria, Table S6). A FDR adjusted *p*-value of 0.05 means 5% of the uncorrected

273   significant discovery will result in false positives.

274   In the Glanville fritillary, the linkage disequilibrium (LD) between genes (r^2) is

275   currently estimated to reach a level of 0.45 at about 300bp distance (Ahola et al. 2014b; Rastas et

276   al. 2013), but may extend over thousands of base pairs across the genome of the butterfly

277   (r^2=0.3), based on data from the Åland island only). For simplicity of the analysis, we decided to

278   investigate a 300bp window from both ends of each gene with significant association with

279   phenotypic trait(s), thus to identify neighbouring genes and functional variants (Ahola et al.

280   2014a) potentially linked to the marker SNPs. Correlation between SNPs (LD) was not taken

281   into account in the genetic association analyses. We also performed power analyses using genetic

282   power calculations (Purcell et al. 2003), including the allele frequency of the tested SNP (0.3),

283   the allele frequency of the causal SNP (0.3), the LD between target and causal SNP (0.1 to 0.5),

284   our sample size (N=171) and the estimated effect size for either small or large effect (total

285   QTL=0.1 or 0.5, respectively), and for both additive and dominant models.

286

287   *Allele frequency.* A directed permutation test was used to assess the strength of evidence for

288   environment type having influenced the allele frequencies at a particular SNP. To account for

289   multiple testing, we applied a conservative Bonferroni correction at 1% significance level on the

290   nominal *p*s derived using 1,000,000 random permutations of the sample labels. Given the data at

291   a locus, let p_{11}, p_{12} be the maximum likelihood estimates of the minor allele frequency (MAF) in

292   the northern fragmented populations (ÅL, UP). Let p_{21}, p_{22} be the corresponding estimates in the

293   southern continuous populations (ÖL/GO, SA). For each locus, the test statistic was calculated as

294   the vector of MAF differences: p_{11}-p_{21}, p_{11}-p_{22}, p_{12}-p_{21}, p_{12}-p_{22}. Thus, signs and absolute values

295   of the elements of the test statistic reflect collectively whether the MAFs are consistently smaller

296   or greater in the northern fragmented environment, and how large the differences are. For every

297   locus in which the differences were all consistently positive or negative, the random

298   permutations were used to calculate the probability that at least as large a difference in the same

299   direction was observed by chance, which resulted in the nominal empirical *p*-values for the

300   locus. As the allele frequency differences between environments may be confounded by the

301   distance separating the populations (isolation by distance), we measured pairwise genetic

302   distances between populations (FSTs, α=0.05) using the allelic profiles of the 33 SNPs that

303   fulfilled our quality criteria, and the software GenoDive 2.0b27 (Meirmans & Van Tienderen

304   2004).

305   Additionally, the genotypes of 36 SNPs (33 SNPs that passed our quality criteria and

306   three extra SNPs) were encoded into numerical form. We then implemented the "heatmap.2"

307   function from the gplots R package (Warnes et al. 2009) with the numerical data from either all

308   SNPS (Supplementary Fig. S2A), or only from the 15 SNPs identified as showing significant

309   allele frequency differences between environment types using direct permutation tests

310   (Supplementary Fig. S2B). In the heatmap.2, the relative count of alleles for each SNP and each

311   sample is calculated to allow the hierarchical clustering of the samples based on their allele

312   count. Samples similarity is represented as a dendogram on the left side of each heatmap.

313

314   ***Data accessibility***

315   All data are available in the public archive *Dryad* under the provisional data identifier: doi

316   XXXX. Temperature data for the 1992-2001 period from Uppland coastal region, Öland and

317   Gotland were extracted from the Metadata file ISO 19115:2003 ECDS (File identifier:

318   7085c74e-58a4-49f0-b8c0-b5a0eb3c74f9) publicly available from the Swedish Meteorological

319   and Hydrological Institute (SMHI) website (http://www.smhi.se/en). The publicly available

320   temperature data from Åland and Saaremaa were provided by the Finnish Meteorological

321   Institute (http://en.ilmatieteenlaitos.fi/), and extracted from the RIMFROST database

322   (www.rimfrost.no), respectively.

323

Results

325   After correction, no significant results were conserved (α=0.1, all *p*-values>0.12). Due to the

326   large number of tests conducted and corrected for in this study, and to the relatively small sample

327   size, our analyses might however have generated false negative results. Thus, we present below

328   the results from the most significant associations before correction (Table 1, and Supplementary

329   Table S7). In order to achieve 80% power, our sample size for testing a large effect SNP should

330   be of 84 or 59, for dominant or additive model, respectively, when LD is of 0.5. However, as our

331   SNPs are more likely to be small effect SNPs, to achieve 80% power, our sample size should be

332   of 439 and 310 for dominant and additive models, respectively, with LD of 0.5. In all case, the

333   sample size would increase to N=11670 for dominance model, and N=7845 for additive model,

334   when the LD is small (LD=0.1).

335

336   ***Associations involving larval and pupal traits***

337   The traits of larval and pupal development were summarized by the first four principal

338   components of PCA_1 (Supplementary Table S3)(Duplouy et al. 2013). Of the four principal

339   components, PC_{1-1} is positively correlated with pupal weight, but negatively correlated with

340   pupal period. The other PC's show strongest correlations with the weights and durations of

341   specific larval instars. The sex of the larvae had a strong significant effect on PC_{1-1} and PC_{1-3} (*p*-

342   value $< 5.13^{e-6}$), with females showing lowest PC_{1-1} and PC_{1-3} values. Sex had no effect on

343   neither PC_{1-2} nor PC_{1-4} (Supplementary Table S7, Duplouy et al. 2013).

344   One association was found between *Pgi*:c.331A>C and PC_{1-3} (*p*-value = 3.84e-4,

345   dominant effect). Individuals with one or two copies of the C allele had higher values of PC_{1-3}

346   than the AA homozygotes (Fig. 2). PC_{1-3} is primarily negatively correlated with the weights of

347   the 5th instar post-diapause larvae (Supplementary Table S3), thus suggesting that AA

348   homozygotes tend to be heavier after diapause than the other genotypes (CC 4.24 mg, CA 4.04

349   mg and AA 5.27 mg, Fig. S3A). This is true for all four populations, even if the frequency of the

350   C allele is significantly higher in the southern continuous than in northern fragmented

351   environments (*p*-value<10^-6, Supplementary Table S8). Similarly, the pupae with one or two

352   copies of the C allele were heavier than the AA homozygotes in the 2007 pilot material (CC 192

353   mg, CA 193 mg and AA 176 mg, *p*-value=0.0032, dominant effect, Supplementary Fig. S3B).

354   Both SNPs c480_est:926G>A and c480_est:1051G>A (125 bp's apart) in the

355   *Endocuticle structural glycoprotein* gene (*SgAbd-8*) showed significant interactions with

356   environment type in affecting PC_{1-2} (*p*-value= 0.0023 and 0.002, respectively, dominant effects,

357   Fig. 3). There were no significant differences between the genotypes in the southern continuous

358   environments (*p*-value= 0.28), but the individuals from the northern fragmented environments

359   carrying at least one G allele had lower values of PC_{1-2} than the AA homozygotes (*p*-value=

360   0.019, Fig. 3). PC_{1-2} is mostly negatively correlated with the weight of the 7th larval instar

361   (Supplementary Table S3).

362   Finally, there was a significant effect of two SNPs in the *Hemolymph proteinase-5* gene

363   (89 bp's apart) on PC_{1-2} (*p*-value=1.15e-3, and *p*-value=1.21e-3, over-dominant effects,

364   Supplementary Fig. S4) and a SNP in the *Heat shock protein* gene on PC_{1-1} (*p*-value=2.19e-4,

365   over-dominant effect, Supplementary Fig. S5). In each case, the homozygotes show higher PCs

366   values than heterozygote individuals.

367

368   ***Associations involving adult traits***

369   The male adult-traits were summarized by the first three principal components of PCA$_M$

370   (Supplementary Table S4) to reflect on male quality. PC$_{M1}$ is negatively correlated with both

371   survival and average distance flown per day, indicating that males of short survival do fly less.

372   PC$_{M2}$ is negatively correlated with pupal weight but positively correlated with age at 1st mating,

373   indicating that smaller males mate at older age. Finally, PC$_{M3}$ positively correlated flight activity

374   (probability to fly) with mating success (total number of copulations and age at 1st mating), thus

375   male flying more mate more, despite starting to mate at older age. There were no direct effects of

376   environment on any of the PCs.

377   We found an association between PC$_{M1}$ and the SNP c3917_est:386A>C in the *Serine*

378   *proteinase-like protein* gene (*p*-value=1.64e-4, recessive effect, Fig. S6). Butterflies with at least

379   one copy of the A allele showed lower PC$_{M1}$ values (AA=-3.07, AC=0.13 and CC=15.09).

380   We found an association between SNPs in the *SgAbd-8* gene and PC$_{M2}$ (SNP

381   c480_est:1003G>C, *p*-value=2.15e-4, dominant effect of G allele, Fig. 4A), and PC$_{M3}$ (SNP

382   *c480_est:926G>A* and SNP *c480_est:1051G>A*, 125 bp's apart, *p*-values=8.56e-4, dominant

383   effects of G allele). Additionally, the direct effects of the two SNPs in the *SgAbd-8* gene were

384   coupled with the effect of environment by genotype interactions (*p*-values=1.87e-4, over-

385   dominant effects, Fig. 4B). There were no significant differences between the genotypes in the

386   northern fragmented environments, but the heterozygote individuals from the southern

387   continuous environments had higher PC$_{M3}$ values than the homozygote individuals (Fig. 4B).

388   There was also a significant association between PC_{M3} and the SNP hsp_1:206T>G in

389   the *Heat shock 70kDa protein* coding gene (*p*-value=1.33e-4, recessive effect, Fig. S7A). In

390   general, individuals that carry at least one T allele show lower PC_{M3} values (GG=0.45, GT=-0.14

391   and TT=-0.07). Finally, there were significant associations between PC_{M3} and the SNP

392   c50_est:824A>G in the *Hemolymph proteinase-5* gene. Homozygote CC males show highest

393   PC_{M3} values (CC=0.18, AC= -0.02 and AA= -0.16, *p*-value=2.35e-4, additive effect, Fig. S7B).

394   The female adult and oviposition related-traits were summarized by the first three

395   principal components (PC_{F1} to PC_{F3}) of a female PCA_F (Supplementary Table S5). PC $_{F1}$ is

396   negatively correlated with female offspring production (both total eggs and total caterpillars

397   production), suggesting that females laying small clutches produce less caterpillars. PC $_{F2}$ shows

398   positive correlation with 1st clutch size and hatch rate (ArcSin corrected), suggesting that larger

399   clutches show higher hatch rate success. Finally, PC $_{F3}$ is strongly positively correlated to female

400   age at 1st oviposition, and to a lower degree to the size of the 1st clutch laid, suggesting that the

401   first clutch of older females are larger. There were no effects of the environment nor the

402   genotype on neither of the three PCs from PCA_F.

403

404   ***Allele frequencies in northern fragmented versus southern continuous environments***

405   There were 15 SNPs in eight genes with significant allele frequency differences between the

406   northern fragmented and southern continuous environments, using directed permutation

407   calculations. The eight genes are *Phosphoglucose isomerase* (*Pgi*, for both years), *Serine*

408   *proteinase-like protein-*1, *Hemolymph proteinase-5*, *Vitellin-degrading protease precursor,*

409   *Endocuticle structural glygoprotein* (*SgAbd-8*)*, Cytochrome P450* (*CYP337*)*, Heat shock 70kDa*

410   *protein* (*Hsp70*) and *Glucose-6-phosphate-1-dehydrogenase* (*G6PD*). Five of these genes also

411   showed a significant association with larval development traits and/or male adult traits (as

412   further described above), namely *Pgi, Hemolymph proteinase-5*, *Vitellin-degrading protease*

413   *precursor*, *HSP70*, and *SgAbd-8* (Supplementary Table S8). Additionally, pairwise FSTs (Table

414   2) revealed that the four populations used in the main experiment were not significantly different

415   or similar to each other. We also did not identify any other genes or apparent functional variants

416   within a 300bp LD (r^2=0.45) window from spanning any of these eight genes, or any of the other

417   genotyped SNPs. As there is no data available on the occurrence of epistasis in the Glanville

418   fritillary butterfly, we also could not rule out the effect of hiding or masking gene(s) on the

419   different associations tested in this study.

420   The hierarchical clustering of samples used to generate the heatmaps supports the

421   evidence of environment-specific variation in allele count for several SNPs with significant allele

422   frequency differences between the northern fragmented and southern continuous environments,

423   when using the directed permutation calculations described above. The environment-specific

424   differentiations illustrated in the heatmaps (Supplementary Fig. S2) are especially obvious for

425   the SNPs in the *Serine proteinase-like protein*-1 gene, the *Hemolymph proteinase-5* gene, the

426   *CYP377* gene*, the SgAbd-8* gene, the *Pgi* gene*,* and the *Glucose-6-phosphate-1-dehydrogenase*

427   (*G6PD*) gene.

428

429   **Discussion**

430   We highlight several associations between SNPs in candidate genes and some of the principal

431   components from PCAs on various life-history traits in the Glanville fritillary butterfly (Table 1),

432   however these associations did not hold up following correction for multiple testing. One reason

433   for the lack of significant results may remain the small sample size for each population, and the

434   large amount of tests done, leading to false negative results. We therefore believe many of the

435   positive results before correction, are robust and most likely driven by the fact that the selection

436   of candidate genes was largely based on previous studies (discussed below), in which these

437   genes had already shown significant or nearly significant effects. Furthermore, we used

438   phenotypic traits, some of which have been highly heritable in previous studies (de Jong et al.

439   2014; Kvist et al. 2013) or have shown phenotypic variation among environments suggestive of

440   local adaption (Duplouy et al. 2013). Although functional validations are needed to confirm our

441   findings, the present results provide a starting point for genome-wide association studies on the

442   Glanville fritillary. Such studies have remained impractical so far, but they are becoming feasible

443   in the future, with the help of the now available genome (Ahola et al. 2014a), linkage map

444   (Rastas et al. 2013) and other tools for the Glanville fritillary, as well as the decreasing cost of

445   whole-genome sequencing.

446   The material for the present study originates from five distinct populations, over two

447   different collection years. Consistent results across this material can be considered to be more

448   robust than would be comparable results for a single population in the Glanville fritillary

449   butterfly. However, because the phylogenetic relatedness varies between the five populations

450   (see Somervuo et al. 2014, and Table 2 for estimates of population divergence; and Wahlberg &

451   Saccheri 2007, for description of the different mitochondrial clades), it is not possible to

452   conclude with certitude that the genes with known or presumed function in life-history ecology

453   show significant differences in allele frequencies due to variable natural selection in the two

454   environments (northern fragmented versus southern continuous). Other non-investigated

455   population characteristics, such as the historical relatedness or geographical closeness of the

456   populations from similar environments, might be of similar or higher importance in this context.

457   Although the pairwise genetic distances between populations, calculated based on the genotypes

458   at 33 SNPs of 176 butterflies are not significantly grouping the populations in the two

459   environments, future association studies in the Glanville fritillary butterfly should further address

460   this issue by including additional neutral markers. Anyway, the fact that out of the 49 SNPs from

461   26 genes investigated, 15 SNPs from eight genes seem to show differences in allele frequencies

462   between the two environmental extremes suggests their potential role in local habitat adaptation.

463   In each case there is support from previous studies on the Glanville fritillary, or from

464   other insect studies, and to some extend from the present study, that these genes may influence

465   life-history traits or other phenotypic traits. For example, while working on four of the five

466   populations included in the present study (Åland, Uppland, Öland and Saaremaa), Somervuo et

467   al. (2014) identified almost 2,000 genes with dissimilar expression levels between the northern

468   fragmented and southern continuous environments. This list of genes in Somervuo et al. (2014)

469   includes the eight genes for which we found diverging allele frequencies (at one or several

470   SNPs) between environmental types. The key published results supporting our findings regarding

471   the association of these eight genes to variation in life-history traits or other important

472   phenotypic traits include the following:

473   1. The glycolytic *Phosphoglucose isomerase* (*Pgi*) gene, for which genetic

474   polymorphism has previously been linked to variation in a large number of traits (Dahlhoff &

475   Rank 2000; Hanski & Saccheri 2006; Niitepõld et al. 2009; Orsini et al. 2009; Saastamoinen et

476   al. 2009) as further discussed below, and for which our results support for associations with

477   larval development.

478   2. The *Cytochrom p450* (*CYP337*) gene, which has been previously associated with

479   life-time egg and larval production, and host plant preference in the Glanville fritillary (de Jong

480   et al. 2014).

481   3. The *Vitellin-degrading protease precursor* gene, for which polymorphism has been

482   associated with variation in larval development (Ahola et al. 2015, this study) and the incidence

483   of an extra larval instar in the Glanville fritillary (Saastamoinen et al. 2013).

484   4. The *Heat shock 70Kda protein* (*Hsp70*) gene, for which a previous study shown that

485   gene expression is up-regulated during thermal stress (Luo et al. 2014; Mattila 2015), as HSPs

486   may provide cells with protection against damage from extreme temperature conditions

487   (Sørensen et al. 2003). Furthermore, in the Glanville fritillary, polymorphism in HSP genes have

488   been previously associated with variation in fecundity traits, including egg hatch rate, age at 1st

489   oviposition and 1st clucth size and in mobility (de Jong et al. 2014) and in male flight metabolic

490   rate (Mattila 2015). In the present study, polymorphism in the SNPs *hsp_1:106G>A* and

491   *hsp_1:206T>G* are associated to variation in PC_{1-1} and PC_{M3} in males, respectively, thus

492   suggesting association with larval development, or male flight activity and mating success.

493   5 and 6. The *Hemolymph proteinase-5* and the *Serine proteinase-like* genes are two

494   genes linked to immune and stress responses in Lepidoptera (An et al. 2009; Han et al. 2013) and

495   other insects (Ashida & Brey 1995), and to larval development traits in the Glanville fritillary

496   (Ahola et al. 2015, this study). Again, the correlations we observe in larvae and in males between

497   PC_{1-2} (larval development) or PC_{M3} (male mating success and flight activity) and polymorphism

498   in the *Hemolymph protease-5* gene might suggest variation in the quality, through for example

499   variation in immune functions, between individuals in the Glanville fritillary.

500   7. The *Endocuticle structural glycoprotein* (*SgAbd-8*) gene, described as a structural

501   gene associated with larval development in insects (e.g. in bumblebees, Colgan et al. 2011),

502   showed genetic associations with male PCs, as well as a gene by environment type interaction on

503   larval development in the present study.

504   8. The *Glucose-6-phosphatase dehydrogenase* (*G6PD*) gene, in which polymorphism

505   has been associated with mating success in *Colias* butterflies (Carter 1988).

506   The results of the genetic association for the SNP (Pgi):c.331A>C are a case in point.

507   We found significant associations with larval development. The AA homozygotes have heavier

508   larvae following diapause, but smaller pupae than the other genotypes. These results were

509   consistent across the four populations, even though there are great differences in the frequency of

510   the C allele in the populations: the C allele is much more common in the two northern

511   populations from continuous environment (around 50%) than in the southern populations from

512   fragmented environment (around 20%).

513   Several previous studies conducted in the Åland Islands have demonstrated that the

514   (Pgi):c.331A>C AC heterozygotes have superior flight metabolic rate and dispersal rate in the

515   field compared with the AA homozygotes (Niitepold et al. 2011; 2009). Both empirical and

516   modeling studies have suggested that natural selection in highly fragmented landscapes favors

517   heterozygosity because of associated superior dispersal and colonization abilities (Fountain et al.

518   2016; Hanski & Mononen 2011), thus allowing the individuals to differently cope with spatial

519   heterogeneity, and the species to best occupy the available habitat. Similarly, other phenotypic

520   traits (e.g. larval and fecundity) may be affected by the prevailing environmental conditions

521   (Saastamoinen 2007b; 2013). Working on larval development, Saastamoinen et al. (2013) found

522   that individuals with the C allele develop more often via an extra larval instar before pupation.

523   As the extra instar is common in larvae that are small after diapause, this finding is consistent

524   with our results showing that the AA larvae are heavy after diapause, as well as with results from

525   Kallioniemi and Hanski (2011) showing that AA homozygotes produce heavier pupae than the

526   other genotypes. Furthermore, the homozygote AA individuals have higher survival in low

527   (stressful) temperature in the laboratory (Kallioniemi & Hanski 2011), while the small CC

528   individuals generally showed reduced survival (Orsini et al. 2009). These results are consistent

529   with studies on other butterfly species. In the copper butterflies, *Lycaena tityrus*, Karl et al.

530   (2008) found that *Pgi* genotypes are associated to several traits of larval and pupal development

531   and cold stress resistance, while in *Colias eurytheme Pgi* heterozygote individuals have higher

532   survival than the homozygotes in natural populations (Watt 1977).

533   Although, the current study did not highlight any genetic and environmental association

534   with female adult traits (including survival, fecundity and flight activity traits), previous studies

535   have shown that *Pgi*-AC heterozygous females (SNP (Pgi):c.331A>C) lay larger clutches than

536   AA homozygous females especially at low ambient temperatures (Saastamoinen 2007a;

537   Saastamoinen & Hanski 2008). In *Colias*, *Pgi* heterozygous individuals in allozyme studies had

538   similarly high male mating success and high female fecundity (Carter 1988; Watt 1992; Watt et

539   al. 2003). The CC homozygous individuals are uncommon in the Åland Islands and hence

540   previously little studied. We suggest that this lack of association with female traits may be due to

541   the large number of traits included in the PCA, rather than the true lack of association with

542   genetic polymorphism in the two environment types.

543   In summary, several studies on the Glanville fritillary examining associations between

544   the SNP (Pgi):c.331A>C and life-history traits have found consistent results for several traits,

545   including larval development. The individuals with the AA genotype do not appear to be as good

546   dispersers as the heterozygotes (Niitepõld et al. 2009), but might re-allocate their resources

547   towards other fitness traits, such as their own body size, making them competitive individuals

548   within their habitat patches (e.g. in many butterflies, females select for larger males) and the

549   environmental conditions (e.g. fluctuating temperatures). There may also be trade-offs between

550   dispersal behavior and immune responses, as suggested by the significant associations observed

551   between another gene, the *Heat shock 70Kda protein* gene, and adult male activity (PC_{M3}).

552   Active flight causes oxidative stress, which may become an important factor for dispersive

553   butterflies. Somervuo et al. (2014) showed that butterflies from fragmented landscapes are more

554   tolerant of hypoxia, which is suggested to be another likely genomic adaptation to dispersal.

555

556   **Conclusion**

557   The current literature shows multiple evidence pointing to the possibility that

558   population-level variations in the environmental conditions, in the Baltic Sea region, may

559   influence life-history evolution and thereby allele frequencies and gene expression in the

560   Glanville fritillary butterfly. In the context of habitat fragmentation, traits such as dispersal and

561   re-colonization, which are key processes facilitating the persistence of metapopulations in

562   fragmented landscapes, are among the traits that have previously been suggested to be affected

563   (Hanski 2011). Equally, traits influencing performance in large, stable populations (in continuous

564   landscapes) may be important. Studies on allelic variation in *Pgi* and *sdhd* genes have similarly

565   demonstrated significant differences between new versus old local populations in the highly

566   fragmented landscape in the Åland metapopulation (Hanski & Saccheri 2006; Hanski 2011;

567   Wheat et al. 2011; Zheng et al. 2009), supporting the notion that population turnover, extinctions

568   and re-colonizations, impose selection in fragmented landscapes. Similarly, other gene

569   expression studies demonstrated significant allelic differences in various genes between

570   individuals experiencing different ambient temperature conditions (Kvist et al. 2013;

571   Saastamoinen & Hanski 2008; Saastamoinen et al. 2013). The differences between newly-

572   established versus old local populations in fragmented landscapes, and between individuals

573   evolving in cold versus warmer environments reflect some of the selection processes that might

574   operate at the population level. Although, we could predict that the difference between

575   environments should reveal the outcome of such processes during longer periods of time, further

576   functional analyses are needed to fully tease apart the relative importance of the confounding and

577   stochastic evolutionary processes related to the degree of habitat fragmentation and the

578   geographical positions of the five populations used in our study.

579

580   **References:**

581   Ahola V, Koskinen P, Wong SC, Kvist J, Paulin L, Auvinen P, Saastamoinen M, Frilander MJ, Lehtonen R,
582   and Hanski I. 2015. Temperature- and sex-related effects of serine protease alleles on larval
583   development in the Glanville fritillary butterfly. *J Evol Biol* 28:2224-2235. 10.1111/jeb.12745
584   Ahola V, Lehtonen R, Somervuo P, Salmela L, Koskinen P, Rastas P, Valimaki N, Paulin L, Kvist J, Wahlberg
585   N, Tanskanen J, Hornett EA, Ferguson LC, Luo S, Cao Z, de Jong MA, Duplouy A, Smolander OP,
586   Vogel H, McCoy RC, Qian K, Chong WS, Zhang Q, Ahmad F, Haukka JK, Joshi A, Salojarvi J, Wheat
587   CW, Grosse-Wilde E, Hughes D, Katainen R, Pitkanen E, Ylinen J, Waterhouse RM, Turunen M,
588   Vaharautio A, Ojanen SP, Schulman AH, Taipale M, Lawson D, Ukkonen E, Makinen V, Goldsmith
589   MR, Holm L, Auvinen P, Frilander MJ, and Hanski I. 2014a. The Glanville fritillary genome retains
590   an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera. *Nat Commun*
591   5:4737. 10.1038/ncomms5737
592   Ahola V, Lehtonen R, Somervuo P, Salmela L, Koskinen P, Rastas P, Valimaki N, Paulin L, Kvist J, Wahlberg
593   N, Tanskanen J, Hornett EA, Ferguson LC, Luo SQ, Cao ZJ, de Jong MA, Duplouy A, Smolander OP,
594   Vogel H, McCoy RC, Qian K, Chong WS, Zhang Q, Ahmad F, Haukka JK, Joshi A, Salojarvi J, Wheat
595   CW, Grosse-Wilde E, Hughes D, Katainen R, Pitkanen E, Ylinen J, Waterhouse RM, Turunen M,
596   Vaharautio A, Ojanen SP, Schulman AH, Taipale M, Lawson D, Ukkonen E, Makinen V, Goldsmith
597   MR, Holm L, Auvinen P, Frilander MJ, and Hanski I. 2014b. The Glanville fritillary genome retains
598   an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera. *Nature*
599   *Communications* 5.
600   An C, Ishibashi J, Ragan EJ, Jiang H, and Kanost MR. 2009. Functions of *Manduca sexta* hemolymph
601   proteinases HP6 and HP8 in two innate immune pathways. *J Biol Chem* 284:19716-19726.
602   10.1074/jbc.M109.007112

603 Ashida M, and Brey PT. 1995. Role of the integument in insect defense: pro-phenol oxidase cascade in
604   the cuticular matrix. *Proc Natl Acad Sci U S A* 92:10698-10702.
605 Carter PA. 1988. Adaptation at specific loci. V. Metabolically adjacent enzyme loci may have very distinct
606   experiences of selective pressures. *Genetics* 119:913-924.
607 Colgan TJ, Carolan JC, Bridgett SJ, Sumner S, Blaxter ML, and Brown MJ. 2011. Polyphenism in social
608   insects: insights from a transcriptome-wide analysis of gene expression in the life stages of the
609   key pollinator, *Bombus terrestris*. *BMC Genomics* 12:623. 10.1186/1471-2164-12-623
610 Dahlhoff EP, and Rank NE. 2000. Functional and physiological consequences of genetic variation at
611   phosphoglucose isomerase: heat shock protein expression is related to enzyme genotype in a
612   montane beetle. *Proc Natl Acad Sci U S A* 97:10056-10061. 10.1073/pnas.160277697
613 de Jong MA, Wong SC, Lehtonen R, and Hanski I. 2014. Cytochrome P450 gene CYP337 and heritability of
614   fitness traits in the Glanville fritillary butterfly. *Mol Ecol* 23:1994-2005. 10.1111/mec.12697
615 Duplouy A, Ikonen S, and Hanski I. 2013. Life history of the Glanville fritillary butterfly in fragmented
616   versus continuous landscapes. *Ecol Evol* 3:5141-5156. 10.1002/ece3.885
617 Fountain T, Nieminen M, Siren J, Wong SC, and Hanski I. 2016. Predictable allele frequency changes due
618   to habitat fragmentation in the Glanville fritillary butterfly. *Proc Natl Acad Sci U S A* 113:2678–
619   2683. 10.1073/pnas.1600951113
620 Han P, Jin F, Dong X, Fan J, Qiu B, and Ren S. 2013. Transcript and protein profiling analysis of the
621   Destruxin A-induced response in larvae of *Plutella xylostella*. *PLoS One* 8:e60771.
622   10.1371/journal.pone.0060771
623 Hanski I. 2005. *The Shrinking World*. Oldendorf/Luhe, Germany: International Ecology Institute.
624 Hanski I, and Mononen T. 2011. Eco-evolutionary dynamics of dispersal in spatially heterogeneous
625   environments. *Ecol Lett* 14:1025-1034. 10.1111/j.1461-0248.2011.01671.x
626 Hanski I, and Ovaskainen O. 2000. The metapopulation capacity of a fragmented landscape. *Nature*
627   404:755-758. 10.1038/35008063
628 Hanski I, Saastamoinen M, and Ovaskainen O. 2006. Dispersal-related life-history trade-offs in a butterfly
629   metapopulation. *J Anim Ecol* 75:91-100.
630 Hanski I, and Saccheri I. 2006. Molecular-level variation affects population growth in a butterfly
631   metapopulation. *PLoS Biol* 4:e129. 10.1371/journal.pbio.0040129
632 Hanski IA. 2011. Eco-evolutionary spatial dynamics in the Glanville fritillary butterfly. *Proc Natl Acad Sci*
633   *U S A* 108:14397-14404. 10.1073/pnas.1110020108
634 Hochberg Y, and Benjamini Y. 1990. More powerful procedures for multiple significance testing. *Stat*
635   *Med* 9:811-818.
636 Hughes JB, Daily GC, and Ehrlich PR. 1997. Population diversity: Its extent and extinction. *Science*
637   278:689-692.
638 Kallioniemi E, and Hanski I. 2011. Interactive effects of Pgi genotype and temperature on larval growth
639   and survival in the Glanville fritillary butterfly. *Functional Ecology* 25:1032-1039.
640 Karl I, Schmitt T, and Fischer K. 2008. Phosphoglucose isomerase genotype affects life-history traits and
641   cold stress resistance in a Copper butterfly. *Functional Ecology* 22:887-894.
642 Klemme I, and Hanski I. 2009. Heritability of and strong single gene (*Pgi*) effects on life-history traits in
643   the Glanville fritillary butterfly. *J Evol Biol* 22:1944-1953.
644 Kvist J, Mattila AL, Somervuo P, Ahola V, Koskinen P, Paulin L, Salmela L, Fountain T, Rastas P,
645   Ruokolainen A, Taipale M, Holm L, Auvinen P, Lehtonen R, Frilander MJ, and Hanski I. 2015.
646   Flight-induced changes in gene expression in the Glanville fritillary butterfly. *Mol Ecol* 24:4886-
647   4900. 10.1111/mec.13359
648 Kvist J, Wheat CW, Kallioniemi E, Saastamoinen M, Hanski I, and Frilander MJ. 2013. Temperature
649   treatments during larval development reveal extensive heritable and plastic variation in gene
650   expression and life history traits. *Mol Ecol* 22:602-619. 10.1111/j.1365-294X.2012.05521.x

651 Luo S, Chong Wong S, Xu C, Hanski I, Wang R, and Lehtonen R. 2014. Phenotypic plasticity in thermal
652   tolerance in the Glanville fritillary butterfly. *J Therm Biol* 42:33-39.
653   10.1016/j.jtherbio.2014.02.018
654 Marden JH. 2013. Nature's inordinate fondness for metabolic enzymes: why metabolic enzyme loci are
655   so frequently targets of selection. *Mol Ecol* 22:5743-5764. 10.1111/mec.12534
656 Marden JH, Fescemyer HW, Schilder RJ, Doerfler WR, Vera JC, and Wheat CW. 2013. Genetic variation in
657   HIF signaling underlies quantitative variation in physiological and life-history traits within
658   lowland butterfly populations. *Evolution* 67:1105-1115. 10.1111/evo.12004
659 Mattila AL, and Hanski I. 2014. Heritability of flight and resting metabolic rates in the Glanville fritillary
660   butterfly. *J Evol Biol* 27:1733-1743. 10.1111/jeb.12426
661 Mattila ALK. 2015. Thermal biology of flight in a butterfly: genotype, flight metabolism, and
662   environmental conditions. *Ecol Evol* 5:5539-5551.
663 Meirmans PG, and Van Tienderen PH. 2004. GENOTYPE and GENODIVE: two programs for the analysis of
664   genetic diversity of asexual organisms. *Molecular Ecology Notes* 4:792-794.
665 Mitchell-Olds T, and Pedersen D. 1998. The molecular basis of quantitative genetic variation in central
666   and secondary metabolism in *Arabidopsis*. *Genetics* 149:739-747.
667 Nieminen M, Siljander M, and Hanski I. 2004. Structure and dynamics of *Melitaea cinxia*
668   metapopulations. In: Ehrlich PR, and Hanski I, eds. *On the wings of checkerspots: a model system*
669   *for population biology*. Oxford: Oxford University Press, 63-91.
670 Niitepold K. 2010. Genotype by temperature interactions in the metabolic rate of the Glanville fritillary
671   butterfly. *J Exp Biol* 213:1042-1048. 10.1242/jeb.034132
672 Niitepold K, Mattila AL, Harrison PJ, and Hanski I. 2011. Flight metabolic rate has contrasting effects on
673   dispersal in the two sexes of the Glanville fritillary butterfly. *Oecologia* 165:847-854.
674   10.1007/s00442-010-1886-8
675 Niitepõld K, Smith AD, Osborne JL, Reynolds DR, Carreck NL, Martin AP, Marden JH, Ovaskainen O, and
676   Hanski I. 2009. Flight metabolic rate and Pgi genotype influence butterfly dispersal rate in the
677   field. *Ecology* 90:2223-2232. 10.1890/08-1498.1
678 Ojanen SP, Nieminen M, Meyke E, Poyry J, and Hanski I. 2013. Long-term metapopulation study of the
679   Glanville fritillary butterfly (Melitaea cinxia): survey methods, data management, and long-term
680   population trends. *Ecol Evol* 3:3713-3737. 10.1002/ece3.733
681 Orsini L, Wheat CW, Haag CR, Kvist J, Frilander MJ, and Hanski I. 2009. Fitness differences associated
682   with *Pgi* SNP genotypes in the Glanville fritillary butterfly (*Melitaea cinxia*). *J Evol Biol* 22:367-
683   375. 10.1111/j.1420-9101.2008.01653.x
684 Purcell S, Cherny SS, Sham PC. 2003. Genetic Power Calculator: design of linkage and association genetic
685   mapping studies of complex traits. *Bioinformatics* 19:149-150. 10.1093/bioinformatics/19.1.149
686 R Core Team. 2013. *R: A language and environment for statistical computing*. R Foundation for Statistical
687   Computing.
688 Rastas P, Paulin L, Hanski I, Lehtonen R, and Auvinen P. 2013. Lep-MAP: fast and accurate linkage map
689   construction for large SNP datasets. *Bioinformatics* 29:3128-3134.
690   10.1093/bioinformatics/btt563
691 Rozen S, and Skaletsky H. 2000. Primer3 on the WWW for general users and for biologist programmers.
692   In: Misener S, and Krawetz SA, eds. *Bioinformatics Methods and Protocols*. Totowa, NJ: Humana
693   Press Inc, 365-386.
694 Saastamoinen M. 2007a. Life‐history, genotypic, and environmental correlates of clutch size in the
695   Glanville fritillary butterfly. *Ecological Entomology* 32:235-242.
696 Saastamoinen M. 2007b. Mobility and lifetime fecundity in new versus old populations of the Glanville
697   fritillary butterfly. *Oecologia* 153:569-578. 10.1007/s00442-007-0772-5

698 Saastamoinen M. 2008. Heritability of dispersal rate and other life history traits in the Glanville fritillary
699   butterfly. *Heredity (Edinb)* 100:39-46. 10.1038/sj.hdy.6801056
700 Saastamoinen M, and Hanski I. 2008. Genotypic and environmental effects on flight activity and
701   oviposition in the Glanville fritillary butterfly. *Am Nat* 171:701-712. 10.1086/587531
702 Saastamoinen M, Ikonen S, and Hanski I. 2009. Significant effects of *Pgi* genotype and body reserves on
703   lifespan in the Glanville fritillary butterfly. *Proc Biol Sci* 276:1313-1322. 10.1098/rspb.2008.1464
704 Saastamoinen M, Ikonen S, Wong SC, Lehtonen R, and Hanski I. 2013. Plastic larval development in a
705   butterfly has complex environmental and genetic causes and consequences for population
706   dynamics. *J Anim Ecol* 82:529-539. 10.1111/1365-2656.12034
707 Somervuo P, Kvist J, Ikonen S, Auvinen P, Paulin L, Koskinen P, Holm L, Taipale M, Duplouy A,
708   Ruokolainen A, Saarnio S, Siren J, Kohonen J, Corander J, Frilander MJ, Ahola V, and Hanski I.
709   2014. Transcriptome analysis reveals signature of adaptation to landscape fragmentation. *PLoS*
710   *One* 9:e101467. 10.1371/journal.pone.0101467
711 Sørensen JG, Kristensen TN, and Loeschcke V. 2003. The evolutionary and ecological role of heat shock
712   proteins. *Ecol Lett* 6:1025-1037.
713 Wahlberg N, and Saccheri I. 2007. The effects of Pleistocene glaciations on the phylogeography of
714   *Melitaea cinxia* (Lepidoptera: *Nymphalidae*) *European Journal of Entomology* 104:675-684.
715 Walker DW, Hajek P, Muffat J, Knoepfle D, Cornelison S, Attardi G, and Benzer S. 2006. Hypersensitivity
716   to oxygen and shortened lifespan in a *Drosophila* mitochondrial complex II mutant. *Proc Natl*
717   *Acad Sci U S A* 103:16382-16387. 10.1073/pnas.0607918103
718 Warnes GR, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, Lumley T, Maechler M, Magnusson
719   A, Moeller S, Schwartz M, and Venables B. 2009. gplots: various R programming tools for
720   plotting data. *R package version* 2.
721 Watt WB. 1977. Adaptation at specific loci. I. Natural selectin on phosphoglucose isomerase of Colias
722   butterflies: biochemical and population aspects. *Genetics* 87:177-194.
723 Watt WB. 1992. Eggs, enzymes, and evolution: Natural genetic variants change insect fecundity. *Proc*
724   *Natl Acad Sci U S A* 89:10608-10612.
725 Watt WB, Carter PA, and Blower SM. 1985. Adaptation at specific loci. IV. Differential mating success
726   among glycolytic allozyme genotypes of Colias butterflies. *Genetics* 109:157-175.
727 Watt WB, Wheat CW, Meyer EH, and Martin J-F. 2003. Adaptation at specific loci. VII. Natural selection,
728   dispersal and the diversity of molecular-functional variation patterns among butterfly species
729   complexes (Colias: Lepidoptera, *Pieridae*). *Mol Ecol* 12:1265-1275.
730 Wheat CW, Fescemyer HW, Kvist J, Tas E, Vera JC, Frilander MJ, Hanski I, and Marden JH. 2011.
731   Functional genomics of life history variation in a butterfly metapopulation. *Mol Ecol* 20:1813-
732   1828. 10.1111/j.1365-294X.2011.05062.x
733 Wong SC. 2011. Large-scale genotyping pipe-line for non-model organisms Master. University of
734   Helsinki.
735 Wong SC, Oksanen A, Mattila AL, Lehtonen R, Niitepold K, and Hanski I. 2016. Effects of ambient and
736   preceding temperatures and metabolic genes on flight metabolism in the Glanville fritillary
737   butterfly. *Journal of Insect Physiology* 85:23-31. 10.1016/j.jinsphys.2015.11.015
738 Zheng C, Ovaskainen O, and Hanski I. 2009. Modelling single nucleotide effects in phosphoglucose
739   isomerase on dispersal in the Glanville fritillary butterfly: coupling of ecological and evolutionary
740   dynamics. *Philos Trans R Soc Lond B Biol Sci* 364:1519-1532. 10.1098/rstb.2009.0005
741

742 **Acknowledgements**

743   We thank V. Ahola, G. Blanchet, SW. Geange, P. Somervuo, A. Tack, T. Schulz and E.

744   Numminen for help with the analyses, M. Saastamoinen for precious discussions on the topic of

745   the study, and H. Körnich for providing us with the proper SMHI weather data. Suvi Ikonen and

746   all the assistants at the Lammi Biological Station are thanked for help with data collection. The

747   authors want to dedicate this manuscript to the memory of Prof. I. Hanski for his supportive

748   leadership and numerous major contributions to Ecology.

Description of the SNP-phenotypic trait associations with lowest *p*-values before correction for multiple testing, with model of inheritance.

Both *p*-values from before and after FDR are given. The asterisk (*) identifies the significant result from the 2007 pilot study, while all other results are from the 2010 experiment.

Gene	SNP	Trait	Figure	Inheritance	p-value	FDR adjusted p-value
Pgi	(Pgi):c.331A>C	PC_{1-3}	Fig.2	Dominant	3.84e-4	0.12
Pgi	(Pgi):c.331A>C *	Pupal weight	Fig. S3B	Dominant	0.0032	0.12
SgAbd-8	c480_est:926G>A	PC_{1-2}	Fig.3	Recessive	2.27e-3	0.12
SgAbd-8	c480_est:1003G>C	PC_{M2}	Fig.4A	Dominant	2.15e-3	0.12
SgAbd-8	c480_est:926G>A	PC_{M3}	Fig.4B	Dominant	8.56e-4	0.12
SgAbd-8	c480_est:1051G>A	PC_{M3}	Fig.4B	Dominant	8.56e-4	0.12
Hemolymph proteinase-5	c50_est:735A>G	PC_{1-2}	Fig. S4	Over-dominant	1.15e-3	0.12
Hemolymph proteinase-5	c50_est:824A>G	PC_{1-2}	Fig. S4	Over-dominant	1.21e-3	0.12
Hemolymph proteinase-5	c50_est:824A>G	PC_{M3}	Fig. S7B	Additive	2.35e-3	0.12
Heat shock protein	hsp_1:106G>A	PC_{1-1}	Fig. S5	Over-dominant	2.18e-3	0.12
Heat shock protein	hsp_1:206T>G	PC_{M3}	Fig. S7A	Recessive	1.33e-3	0.12
Serine proteinase-like protein	c3917_est:386A>C	PC_{M1}	Fig. S6	Recessive	1.64e-3	0.12

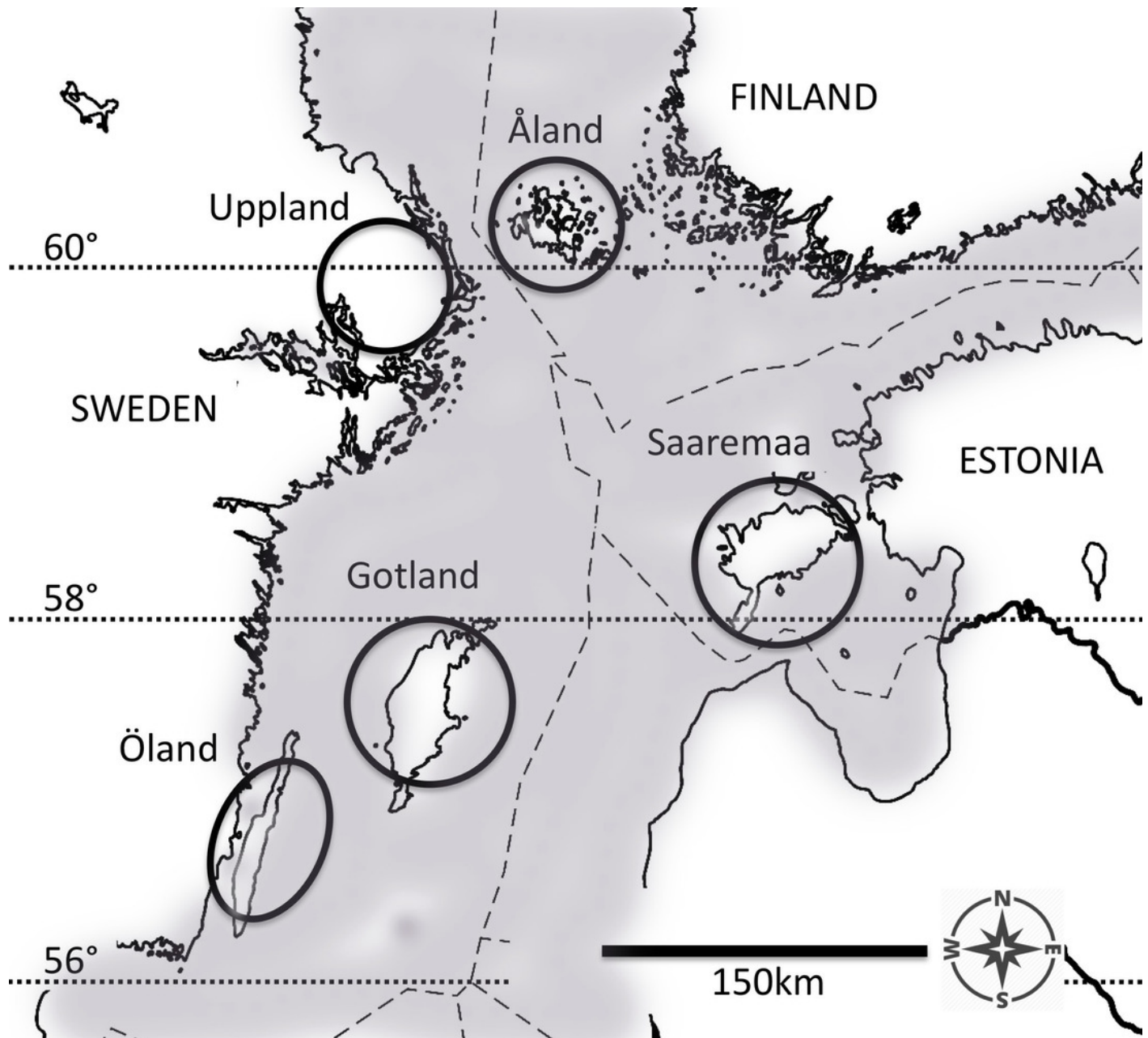
Pairwise FSTs (in bold & grey) and pairwise standardised FSTs (in white) between the four populations collected in 2009, as calculated using GENODIVE (Meirmans and van Tiederen, 2004).

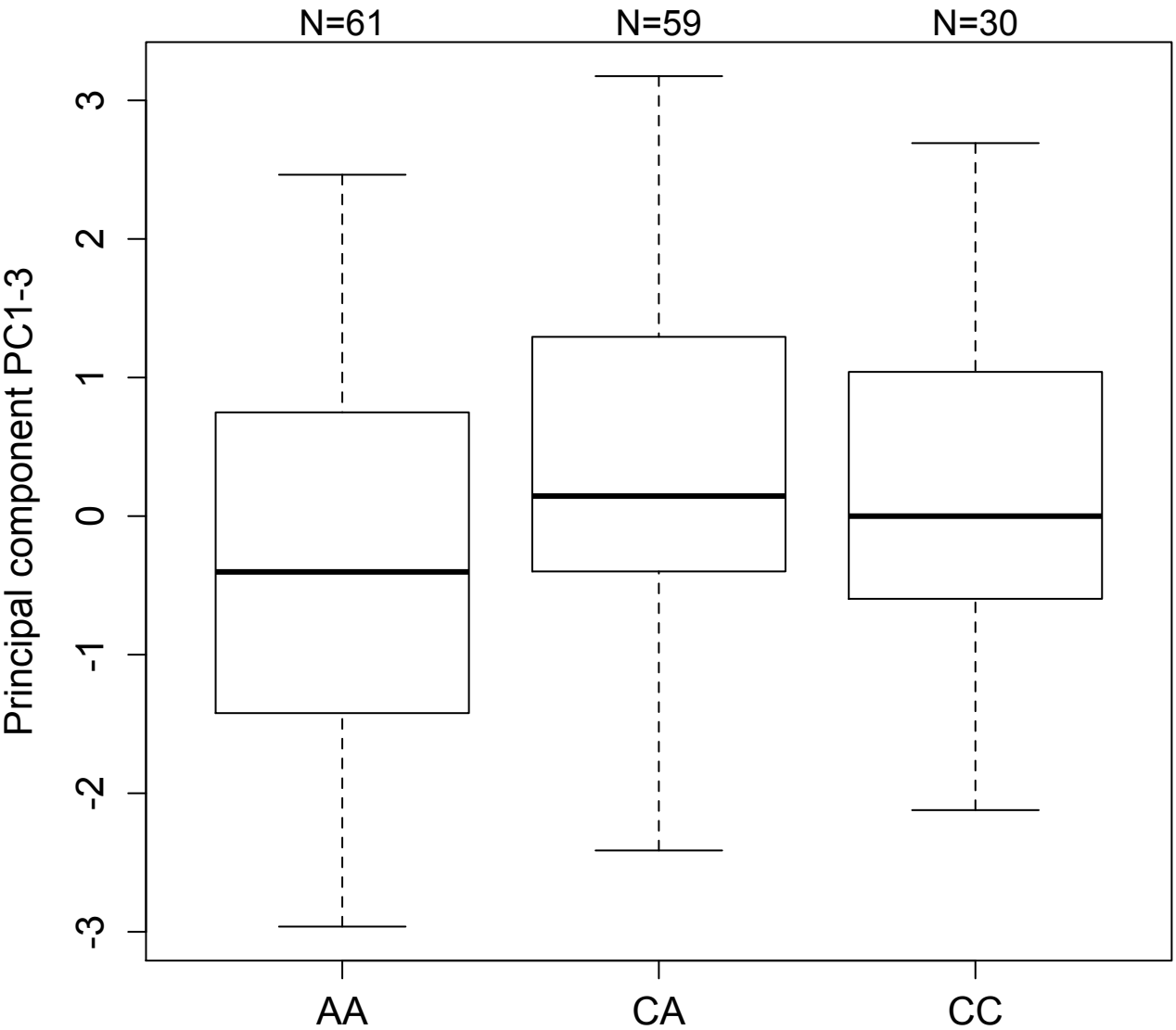
(ÅL) stands for the Finnish population in the Åland islands, (ÖL) and (UP) for the Swedish populations in the Öland island and the Uppland coastal region, respectively, and (SA) for the Estonian population in the island of Saaremaa.

Populations	ÅL	ÖL	SA	UP
Sample size (N=)	35	39	58	44
ÅL	-	**0.084**	**0.269**	**0.072**
ÖL	0.130	-	**0.170**	**0.164**
SA	0.383	0.223	-	**0.328**
UP	0.115	0.251	0.459	-

Figure 1

The five populations of the Glanville fritillary butterfly in the Baltic Sea Region.

The Finnish population on the Åland islands, and the Swedish population on the coastal region of Uppland inhabit northern fragmented environments. The Swedish populations on the Öland and Gotland islands, and the Estonian population on the Saaremaa island evolve in southern continuous environments. The map was created by A. Duplouy using a modified version of the image "Location map of the Baltic Sea" (https://commons.wikimedia.org/wiki/File:Baltic_Sea_location_map.svg, under the license CC BY-SA 3.0, https://creativecommons.org/licenses/by-sa/3.0/) by NordNordWest/Wikipedia. The original image was cropped using Adobe Photoshop CS6 (Version: 13.06x64, http://www.adobe.com/products/photoshop.html).

Values of development PC1-3 for the different genotypes of the SNP Pgi:c.331A>C in the gene *Pgi*.

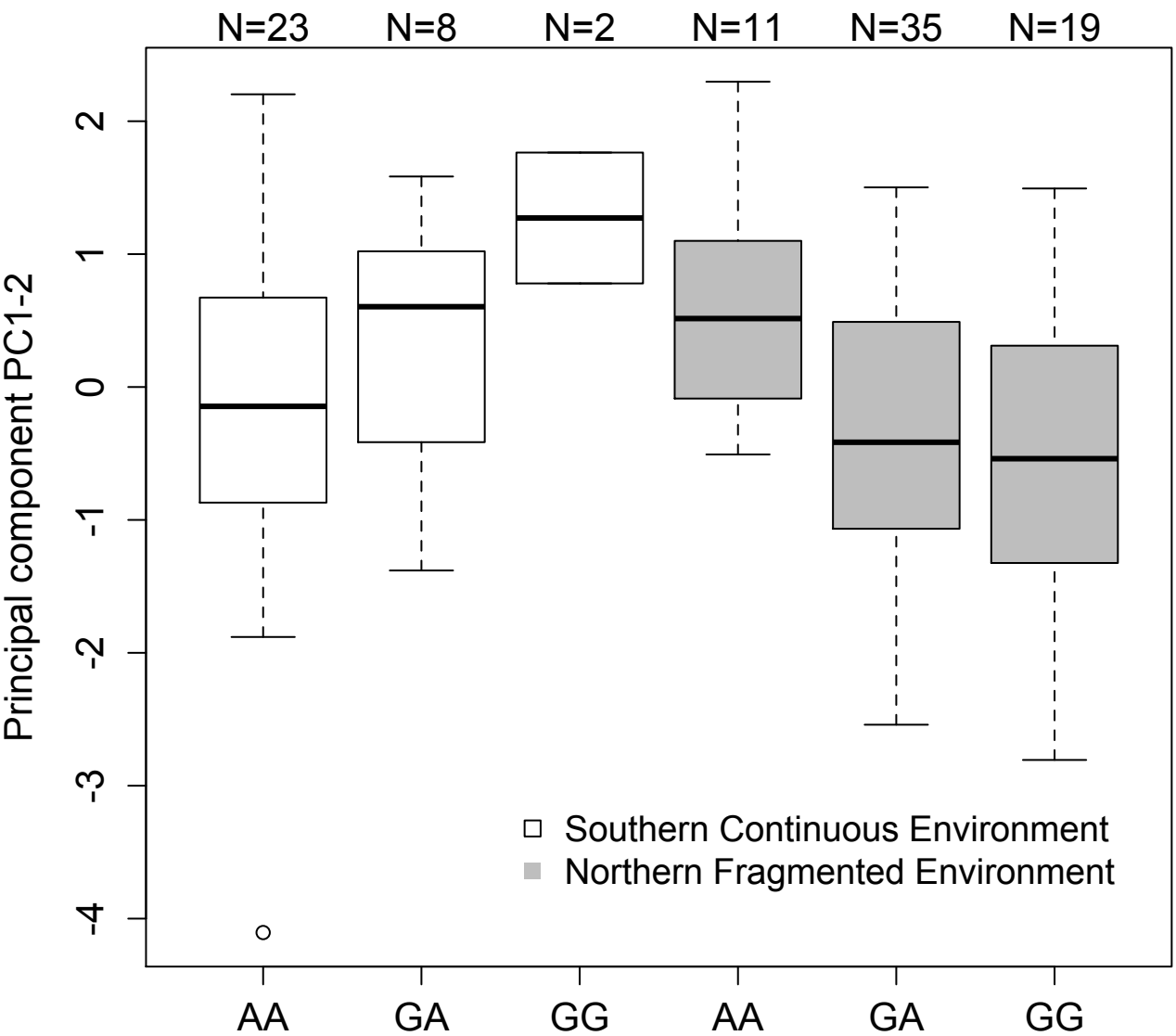
Sample size is given by the number above the bar. Heavy horizontal lines represent median values, boxes give interquartile ranges, whiskers and the dot give the minimum and maximum, and outlier values, respectively.

Differences in the development PC1-2 values for different butterfly genotypes in two SNPs in the *SgAbd-8* gene, evolving in southern continuous (white bars) or northern fragmented environment (gray).

Sample size is given by the number above the bar. Heavy horizontal lines represent median values, boxes give interquartile ranges, whiskers and dots give the minimum and maximum, and outlier values, respectively.

Figure axis and plot content:

N=23   N=8   N=2   N=11   N=35   N=19

□ Southern Continuous Environment
■ Northern Fragmented Environment

AA   GA   GG   AA   GA   GG

SgAbd-8 (c480_est:926G>A and 1051G>A)

Principal component PC1-2

Variations in PCs-values of males of the different genotypes in three SNPs in the *SgAbd-8* gene.

(A) PCM2 values per genotypes in the SNP c480_est:1003G>C. (B) Effect of environment on PCM3 values per genotypes in the SNPs c480_est:926G>A and c480_est:1051G>A. Individuals from southern continuous environment are shown in white, while individuals from northern continuous environment are shown in gray. Sample size is given by the number above the bar. Heavy horizontal lines represent median values, boxes give interquartile ranges, whiskers and dots give the minimum and maximum, and outlier values, respectively.