



HAL
open science

Museomics identifies genetic erosion in two butterfly species across the 20th century in Finland

Jérémy Gauthier, Mila Pajkovic, Samuel Neuenschwander, Lauri Kaila, Sarah Schmid, Ludovic Orlando, Nadir Alvarez

► **To cite this version:**

Jérémy Gauthier, Mila Pajkovic, Samuel Neuenschwander, Lauri Kaila, Sarah Schmid, et al..
Museomics identifies genetic erosion in two butterfly species across the 20th century in Finland.
Molecular Ecology Resources, 2020, 20 (5), pp.1191-1205. 10.1111/1755-0998.13167. hal-03030325

HAL Id: hal-03030325

<https://hal.science/hal-03030325>

Submitted on 30 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

26

27 **Abstract**

28 Erosion of biodiversity generated by anthropogenic activities has been studied for decades and in
29 many areas at the species level, using taxa monitoring. In contrast, genetic erosion within species
30 has rarely been tracked, and is often studied by inferring past population dynamics from
31 contemporaneous estimators. An alternative to such inferences is the direct examination of past
32 genes, by analyzing museum collection specimens. While providing direct access to genetic
33 variation over time, historical DNA is usually not optimally preserved, and it is necessary to apply
34 genotyping methods based on hybridization-capture to unravel past genetic variation. In this
35 manuscript, we apply such a method, i.e., HyRAD, to large time series of two butterfly species in
36 Finland, and present a new bioinformatic pipeline, namely PopHyRAD, that standardizes and
37 optimizes the analysis of HyRAD data at the within-species level. We show in the localities for
38 which the data retrieved has sufficient power to accurately examine genetic dynamics through time,
39 that genetic erosion has increased across the last hundred years, as revealed by signatures of allele
40 extinctions and heterozygosity decreases, despite local variations. In one of the two butterflies, i.e.,
41 *Erebia embla*, isolation by distance also increased through time, revealing the effect of greater
42 habitat fragmentation over time.

43

44 **Keywords**

45 Population Dynamics, HyRAD, Past Gene Frequencies, Lepidoptera, Museomics

46

47

48 **Funding**

49 This research was funded by the Swiss National Science Foundation grants PP00P3_144870 and
50 PP00P3_172899 awarded to Nadir Alvarez.

51 **Acknowledgements**

52 We are grateful to Christophe Dufresnes, Tomasz Suchan, Camille Pitteloud and Kimmo Saarinen
53 for their invaluable help in the field, to Nils Arrigo for bioinformatic support, to Stéphanie Manel
54 for advice in distance-based redundancy analysis, and to Jérôme Goudet for support during early
55 development of the PopHyRAD pipeline and for commenting on this manuscript. We thank the
56 Lausanne Genomic Technologies Facility (LGTF) for the sequencing service as well as two
57 anonymous reviewers for their useful comments that led us to improve our manuscript.

58 **Data accessibility statement**

59 Sequence reads are archived at Zenodo : <http://doi.org/10.5281/zenodo.3668644> for *Erebia embla*
60 and <http://doi.org/10.5281/zenodo.3668660> for *Lycaena helle*. Scripts for the whole analytic
61 process have been uploaded to Github
62 (https://github.com/JeremyLGauthier/Scripts_Gauthier_et.al_2019_MER). The PopHyRAD
63 pipeline is constantly under development and improvement. Current version can be found at
64 <https://github.com/JeremyLGauthier/PHyRAD>).

65 **Author contributions**

66 NA and MP designed the study. MP and LK performed sampling. MP performed labwork. JG
67 analyzed the molecular data, with contributions from SN, MP, NA, LO and SS. All authors took
68 part in discussions concerning the analyses and result interpretations. JG and NA wrote the paper,
69 with contributions from all authors.

70

71 **Introduction**

72 An increasing number of studies reveals that a dramatic collapse of biodiversity has been on-going
73 since the early 20th century, and that decay rates have accelerated over the past 50 years (e.g.
74 (Johnson et al., 2017); (de Oliveira Roque et al., 2018). There is no doubt that this global on-going
75 crisis of biodiversity loss is due to anthropogenic factors that have transformed the species' habitat,
76 and caused a decrease in population sizes, with potential long-term consequences on local or global
77 extinction. Many studies have highlighted the consequences of this extinction in terms of erosion
78 of species diversity for decades (e.g. Ceballos, Ehrlich, & Dirzo, 2017; Ehrlich, 1995), which
79 provided the basis for international reports assessing the state of biodiversity and the ecosystem
80 services it provides to society (e.g. the Intergovernmental Platform on Biodiversity and Ecosystem
81 Services, IPBES; <https://bit.ly/IPBESReport>). It is, however, only recently that observations and
82 empirical approaches have shown that biomass, and by extension, population sizes have also
83 declined significantly over the last 50 years, especially for insects, resulting in multi-trophic
84 cascades affecting the biomass of many insectivorous species (Karp et al., 2013). In particular,
85 Hallmann and colleagues (2017) have measured insect biomass during 27 years in 63 German
86 localities, and found a more-than-three-quarters decline in this short time span, which represents
87 an average 3% drop per year. Similarly, Lister & Garcia (2018) reported a decrease of more than
88 90% of the terrestrial arthropod biomass and 80% of that of the canopy over the last 36 years in
89 Puerto Rico, which has in turn strongly reduced the abundance of predator populations of lizards,
90 frogs and birds. Anthropogenic factors including land-use change (habitat destruction), climate
91 change, pesticides and fertilizers use in agriculture, chemical contamination (Whiteside & Marvin
92 Herndon, 2018), light pollution, invasion by exogenous species (Sánchez-Bayo & Wyckhuys,
93 2019; van Strien, van Swaay, van Strien-van Liempt, Poot, & WallisDeVries, 2019), and wireless

94 communication systems (Thielens et al., 2018), are strong candidates for the recognized
95 insectageddon.

96 These various anthropic factors will impact the populations at different intensities ranging from
97 reduction of their size to local extinction. The reduction of population sizes will have several
98 consequences. A reduction of effective population size (N_e) will cause a loss of the genetic
99 diversity, i.e. the number of alleles in the populations, thus decreasing both neutral variation and
100 adaptive potential. In addition, shrinking populations will experience increased allele extinction
101 due to inbreeding that causes allele fixation by genetic drift. In addition, inbred populations will
102 tend to accumulate deleterious alleles and such a mechanism, referred to as inbreeding depression,
103 will further decrease the average fitness of a population (Keller & Waller, 2002; Kristensen,
104 Pedersen, Vermeulen, & Loeschcke, 2010). Furthermore, the extinction dynamics will weaken the
105 connectivity network between remaining populations and thus reduce overall gene flow. This will
106 increase their divergence and prevent their capacity to exchange potential adaptive alleles. The
107 combination of these mechanisms will further decrease the population persistence likelihood over
108 time (Bouzat, 2010).

109 Empirical data providing time series of abundance data and genetic diversity remain, however,
110 scarce, which limits our ability to precisely infer the recent demogenetic trajectory of most
111 populations. A variety of indirect proxies can help overcome the general lack of long-term
112 abundance data across taxa to estimate the recent demographic trajectories. For example, the
113 genetic information present in museum specimens collected across the last decades or even
114 centuries provides a unique opportunity to obtain temporal snapshots of past population genetic
115 diversity and quantify the extent and dynamics of the current biodiversity crisis (Jensen et al., 2018;
116 Meineke, Davies, Daru, & Davis, 2018; Ryan et al., 2018). It is estimated that the number of

117 museum specimens collected around the world exceeds 1 billion individuals, and covers
118 approximately 1.2 million species (Pyke & Ehrlich, 2010). The molecular diversity present in these
119 specimens can help assess population sizes through time but signatures can also help track
120 adaptation to changing environments (Hoffmann, Sgrò, & Kristensen, 2017). These approaches
121 can thus both help define conservation priorities and estimate the future resilience to ongoing
122 environmental change.

123

124 Analyzing DNA contained in Museum collections (hereafter referred to as historical DNA), is,
125 however, not devoid of difficulties owing to post-mortem decay reactions fragmenting the DNA
126 backbone and modifying the chemical nature of individual nucleotidic bases (Dabney, Meyer, &
127 Pääbo, 2013). The extensive fragmentation of historical DNA molecules precludes widely-used
128 conventional genetic analyses, such as high-density array genotyping (Decker et al., 2009) and
129 RAD-Seq (Linck, Hanna, Sellas, & Dumbacher, 2017). Over the past decade, a number of
130 alternative molecular methods have been developed to gather historical DNA data at population
131 and genome-wide scales (reviewed in Burrell, Disotell, & Bergey, 2015; Horn, 2012; Orlando,
132 Gilbert, & Willerslev, 2015). These most often rely on the construction of next-generation
133 sequencing DNA libraries, and the capture of those DNA library templates annealing to short
134 synthetic nucleic acid baits spread across pre-defined loci of interest. This rationale was applied in
135 2016 to target GBS or RADseq loci, either through bench-top synthesized probes (ie, HyRAD;
136 Suchan et al., 2016), or commercially-produced synthetic GBS or RADseq oligonucleotides (Ali
137 et al., 2016; Boucher, Casazza, Szövényi, & Conti, 2016; Hoffberg et al., 2016; Sánchez Barreiro
138 et al., 2017). These methods generally improve the quality of the genotypic information retrieved
139 by increasing the overall average depth-of-coverage and reducing the fraction of loci for which no

140 data could be obtained (Suchan et al., 2016). The HyRAD method, which leverages standard
141 ddRAD protocol (Mastretta-Yanes et al., 2015) to prepare a set of DNA probes from fresh samples,
142 has proven especially versatile in its applications to ancient (Schmid et al., 2017) and historical
143 DNA (Schmid et al., 2018; Suchan et al., 2016), and for obtaining genome-wide data at the
144 population scale in a cost-effective manner.

145

146 The museum collections of Finland, in particular the Finnish Museum of Natural History in
147 Helsinki (Luonnontieteellinen keskusmuseo Luomus), contain approximately 22 million
148 specimens collected during the last centuries (Tegelberg, Haapala, Mononen, Pajari, & Saarenmaa,
149 2012). This provides a fantastic opportunity to obtain time series molecular data across a range of
150 species, including butterfly taxa that have declined over the past century. This is notably the case
151 of *Erebia embla*—a northern European and eastern Palearctic species found in bogs—which has
152 experienced a strong decrease in Southern Finland, following the extensive drainage of its habitat
153 (Rassi, Hyvärinen, Juslén, & Mannerkoski, 2010). The same holds true for *Lycaena helle*, an arctic-
154 alpine Palearctic species inhabiting fresh, damp meadows, which shows a patchy distribution
155 throughout Europe and has declined throughout most of its range, especially in Finland, where it
156 remains present in only two sites, Kuivaniemi Simo and Kuusamo (Heino, Poykko, & Itames,
157 1998).

158 In this study, we aim to examine how the genetic structures and diversities of *E. embla* and *L. helle*
159 have been impacted by recent environmental changes and human activities, relying on HyRAD
160 data generated from 118 and 165 museum specimens, respectively, collected across the last century
161 as well as from extant populations. To achieve this, we developed a pipeline, namely `PopHyRAD`,
162 that (1) aligns each sequence read against the probe catalog (including reference genomes when

163 available, or probe sequences provided by users), (2) identifies and controls for deamination
164 patterns (typical of historical DNA), (3) eliminates putative paralogs, PCR duplicates, low quality
165 genotypes and indels, and; (4) keeps only bi-allelic loci for downstream analyses. We then
166 investigated the spatial and temporal dynamics of genetic diversity indices as well as isolation by
167 distance in the two above mentioned butterflies, both locally and regionally.

168

169 **Material and Methods**

170 *Sampling*

171 Historical samples of *L. helle* and *E. embla* were sampled among the Lepidoptera collection of the
172 Finnish Museum of Natural History in 2014 and 2015. Fresh samples were collected following
173 field work in Summer 2015, by capturing flying adults with a net, sampling and storing a single
174 leg in EtOH 95%, before releasing alive individuals. The sampling details are given in Table 1, and
175 illustrated in Figure 2.

176

177 *DNA extraction and HyRAD protocol*

178 The HyRAD protocol was carried out according to Suchan et al. (2016). Briefly, for historical and
179 fresh samples, DNA was extracted from a leg using the QIAamp DNA Micro kit (Qiagen,
180 Hombrechtikon, Switzerland). The probe precursors were prepared using a double-digestion RAD
181 protocol (ddRAD) applied to six fresh samples from each focal species (Mastretta-Yanes et al.,
182 2015; Peterson, Weber, Kay, Fisher, & Hoekstra, 2012). Total genomic DNA was digested with
183 the restriction enzymes SbfI and MseI, DNA adapters were ligated and the resulting library was
184 purified, size-selected with a range of 200bp to 250bp and amplified by PCR for 30 cycles. An
185 aliquot of the final library was sequenced on one lane of Illumina MiSeq 150 bp paired-end at the

186 Lausanne Genomic Technology Facility (LGTF) in order to obtain a sequence catalog of the loci
187 represented in the ddRAD probes, and the rest of the library was converted into probes by removing
188 adapter sequences.

189 Individual Illumina DNA libraries were prepared from the fresh and museum specimens based on
190 a published protocol for degraded DNA samples (Tin, Economo, & Mikheyev, 2014).
191 Hybridization-capture and enrichment was performed as described in Schmid et al. (2017), using
192 a dual-indexing tagging, i.e., with different combinations of barcodes and indexes, allowing
193 extensive multiplexing of samples on a single sequencing lane : in *L. helle*, 9 and 25 indexes and
194 barcodes were used, respectively; while 10 and 25 were used in *E. embla*, respectively. For each
195 butterfly species, the final (capture-hybridized) enriched libraries were sequenced on two lanes of
196 HiSeq 100 bp paired-end at the LGTF.

197

198 *popHyRAD bioinformatic pipeline*

199 The sequence pairs generated from the ddRAD probe libraries, were first cleaned and overlapping
200 reads were collapsed using `AdapterRemovalv2` (Schubert, Lindgreen, & Orlando, 2016). Then
201 a first dataset reduction was performed following the first step of the `dDocent` pipeline (Puritz,
202 Hollenbeck, & Gold, 2014), keeping only alleles covered by at least four reads and, thus, removing
203 the vast majority of sequencing errors. We merged intra-individual loci using `cd-hit-est` from
204 the `CD-HIT` tool (Fu, Niu, Zhu, Wu, & Li, 2012) using a minimum identity threshold of 90%. This
205 step was repeated across samples to produce a combined catalog of all loci identified for a given
206 species. Loci shared by at least half of the probe samples (i.e. three out of six) were kept to remove
207 uninformative specific loci. The absence of contamination in the catalog was evaluated using
208 `Centrifuge` (Kim, Song, Breitwieser, & Salzberg, 2016).

209 Reads from each historical sample were cleaned using `Trimmomatic` (Bolger, Lohse, & Usadel,
210 2014) and individually mapped on the loci catalog generated above using `bwa aln` (Li & Durbin,
211 2009). PCR duplicates were removed using `MarkDuplicates` from the `Picard` toolkit
212 (<http://broadinstitute.github.io/picard>). Nucleotide mis-incorporation patterns were investigated
213 using `MapDamage2.0` (Jónsson, Ginolhac, Schubert, Johnson, & Orlando, 2013), and base
214 quality scores were rescaled according to their probability to represent a post-mortem DNA
215 deamination event in order to reduce the impact of DNA decay on downstream analyses. Finally,
216 individual genotypes were called using `freeBayes` (Garrison & Marth, 2012), and considered
217 for further analyses when (1) showing qualities higher than 100, (2) shared by at least 80% of the
218 samples, and (3) biallelic (Figure 1). This complete pipeline has been implemented under the name
219 `PopHyRAD`, with “Pop” standing for `Population` genetics.

220 Although `bwa aln` is the recommended tool for ancient DNA analysis (Schubert et al., 2012) ,
221 we evaluated the performance of two other mapping tools, namely `bwa mem` (Li, 2013) and
222 `bowtie2` (Langmead & Salzberg, 2012). Similarly, we evaluated the possible impact of the
223 genotyping method on downstream analyses using `varscan` (Koboldt et al., 2009).

224

225 *Population genetics analyses*

226 Allelic frequencies, Observed gene diversity (H_s) and Inbreeding level (F_{IS}) were estimated using
227 the R package `hierfstat` (Goudet, 2005). The frequency of fixed alleles, which is a proxy of
228 the frequency of allele extinction, was estimated using a custom script. For these estimations only
229 localities with more than 8 samples were kept. These estimations were performed after the samples
230 were binned within temporal groups (or time slices). More precisely, we merged samples from
231 each geographic location within three time slices: before 1950, between 1950 and 2000 and present

232 samples (i.e. 2015; see details in Table 1). Additionally, we merged all samples from all locations
233 within the same temporal bins to identify patterns at the whole scale of Finland.

234 Population genetic structure was investigated using a subset of SNPs, where only one SNP per
235 locus was considered so as to minimize linkage disequilibrium (i.e., to minimize redundant
236 information), as generally recommended (Falush, Stephens, & Pritchard, 2003). We used Bayesian
237 admixture analysis implemented in `Structure` (Pritchard, Stephens, & Donnelly, 2000) to
238 estimate admixture proportions, that is, the proportion of each individual's genome inherited from
239 each of K hypothetical source populations. We ran analyses with K from 1 to 5 with 10 independent
240 Markov chains each, using 1,000,000 steps and including 50,000 burn-in steps. We checked
241 visually the results obtained in each run to assess whether the Markov chains were convergent. The
242 most likely number of clusters was identified using Evanno's method (Evanno, Regnaut, & Goudet,
243 2005) implemented in `Structure Harvester` (Earl & vonHoldt, 2012). Then each sample
244 was associated to the corresponding cluster and the proportion of the samples present in each cluster
245 was drawn on the Finland map using pie charts.

246 Isolation-by-distance was tested by examining the correlation between genetic differentiation,
247 $F_{ST}/1-F_{ST}$, using F_{ST} estimations between population pairs performed with `hierfstat` (Goudet,
248 2005), and the log-transformed distance, as suggested by Rousset (1997). Due to the ongoing
249 debate in using Mantel tests for IBD examination (Diniz-Filho et al., 2013), we used a simple
250 Spearman's correlation. To corroborate the role of geographic distance as well as time in the
251 differentiation between localities by using another approach than a simple Spearman's correlation,
252 we also performed distance-based redundancy analysis (dbRDA) integrating geographic variables
253 and year of collection, in an individual-based approach, following recommendations from Laura
254 Benestan (unpublished data; <https://github.com/laurabenestan/db-RDA-and-db-MEM>). First, we

255 created spatial variables using Moran Eigenvector's Maps (MEMs) implemented in `adespatial`
256 R package (Dray, Blanchet, Borcard, Guenard, & Jombart, 2016). Second, a Principal Coordinates
257 Analysis (PCoA) was performed on the Euclidean genetic distances based on genotypes of each
258 sample. Finally a global dbRDA was applied by integrating the year of collection as an additional
259 variable to spatial components, and an ANOVA with 1000 permutations was performed to assess
260 significance of each variable within the model using `vegan` R package (Oksanen, Kindt,
261 Legendre, & O'Hara, 2016).

262

263 **Results**

264 *HyRAD efficiency*

265 The HyRAD wetlab and `popHyRAD` bioinformatic pipeline developed for this study were
266 particularly efficient in retrieving DNA sequences of historical samples collected up to more than
267 100 years ago and stored in museum conditions. The analysis of ddRAD data used for designing
268 the probes allowed to obtain sufficient data at 3,826 loci for *E. embla* and 3,443 for *L. helle*, which
269 allowed us to undertake population genomic analyses. In addition, the identification of deamination
270 patterns at the extremities of the reads from the collection samples and not in the fresh samples
271 confirmed the endogenous nature of DNA present in these historical samples (Supplementary
272 Figure 1).

273

274 *Mapping and SNPs calling comparison*

275 We first aimed at testing whether different read alignment procedures could impact on the amount
276 of HyRAD data retrieved. `Bwa aln` was able to align only a limited number of reads against the
277 catalog of HyRAD probes, representing a mean of 11.2% (sd 6.7) across all samples This

278 proportion was improved to 37.1% (sd 8.1) when using `bwa mem` (Supplementary Table 2).
279 `Bowtie2` showed intermediate efficiency with a mean value of 18.3% (sd 7.9) across all samples.
280 These differences, however, decreased after quality filtering, with `bwa aln` showing 6.9% (sd
281 2.0) of mapped reads, `bwa mem` 12.1% (sd 4.8) and `bowtie2` 11.0% (sd 2.8).
282 We also observed broad differences between methodologies aimed at calling genotypes. We
283 counted the total number of positions covered by at least 80% of the samples and showing biallelic
284 SNP polymorphisms. This number was maximal when combining `bwa mem` alignments and
285 `varscan` genotype calling, representing a total of 10,116 SNPs from 1,289 loci in *E. embla* and
286 11,534 SNPs from 1,241 loci for *L. helle*. The number was minimal when `bwa aln` alignments
287 and `freebayes` genotype calls were combined, which led to the identification of 2,742 SNPs
288 from 869 loci for *E. embla* and 2,549 SNPs from 1,015 loci for *L. helle*.
289 For each of these six combinations (three mappers and two SNP callers), we applied the complete
290 pipeline and subsequently estimated genetic diversity statistics. We observed that these statistics
291 were extremely variable according to the SNP calling tool used, especially the inbreeding
292 coefficient (F_{IS}) (Supplementary Figure 2). With `varscan`, a large proportion of the SNPs
293 demonstrate an inbreeding coefficient below zero with a non-normal distribution, while the normal
294 distribution obtained from `freebayes` is more consistent with biological expectations.
295 Therefore, we decided to utilize the conservative dataset obtained using `freebayes` on the
296 mapping generated by `bwa aln`, despite being associated with an overall smaller number of SNPs;
297 and implemented `freebayes` and `bwa aln` as the standard caller*mapper combination in the
298 PopHyRAD pipeline.

299

300 *Patterns of population declines*

301 To investigate genetic patterns associated with population size reduction, we focused our analyses
302 on three descriptive statistics, (i) the observed gene diversity (H_s), (ii) the frequency of fixed
303 alleles, and (iii) the inbreeding coefficient (F_{IS}). Results are depicted in Figure 3.

304 For *E. embla*, sufficient temporal data could only be retrieved for three localities, i.e. Kuusamo,
305 Muonio and Pelkosenniemi. These indicated reduction of the observed gene diversity (H_s) and the
306 frequency of fixed alleles through time. This overall reduction is, paradoxically, associated with a
307 sharp increase of the inbreeding coefficient at only one locality, i.e. Kuusamo. For *L. helle*, only
308 one locality, i.e. Kuusamo Liikasenvaara, was sampled at different time points. It showed a
309 dynamic diversity trend where the observed gene diversity (H_s) and the frequency of fixed alleles
310 were relatively stable between the 1920s and the 1990s, but declined after the 1990s. In contrast,
311 the inbreeding coefficient (F_{IS}) was found to decrease between the 1920s and the 1990s and to
312 sharply increase since.

313 Grouping all samples within three main time slices (i.e. before 1950, between 1950 and 2000 and
314 in 2015) increased the sample size of each locality and reinforced the observations made above
315 (Figure 4). Indeed, we observed a decrease of the observed allelic richness (H_s) over time in both
316 species, concomitant with an increase in the frequency of fixed alleles.

317 To investigate whether the patterns observed at the local level were retrieved at the global level
318 (i.e. throughout Finland), we merged all the samples within the three time slices considered above,
319 regardless of their geographic origin. This provided an opportunity to retrace the temporal
320 trajectory of the Finland-wide genetic diversity present in both focal species. We found that the
321 observed gene diversity decreased through time and that the frequency of fixed alleles increased,
322 especially in the most recent time period (Figure 5).

323

324 *Genetic structure and isolation by distance*

325 To investigate the mechanisms that may explain the observed decrease in genetic diversity and the
326 increase in the inbreeding coefficient (F_{IS}), we studied the spatial genetic structure of populations,
327 which only revealed a faint structuring (Figure 6). In contrast, a more marked pattern of spatial
328 structuring was retrieved in the Isolation by distance (IBD) analysis, revealing a varying correlation
329 between genetic and geographic distances over time (Figure 7). Before 1950, no correlation
330 between genetic and geographic distances was found in any of the two species investigated.
331 Between the 1950s and the 2000s, a significant correlation between genetic and geographic
332 distances was found for *E. embla*. The level of the correlation and the associated slope increased
333 further when considering the modern time period (year 2015), suggesting an increasing spatial
334 structuring from 1950 onwards, likely in relation with increased habitat fragmentation. No
335 significant correlation could be retrieved in *L. helle* in any of the two historical time periods
336 considered (this correlative analysis could not be carried out for modern times, as only two extant
337 populations are known) (Figure 7). Despite an overall low level of genetic structuring through
338 space and time (overall R^2 of 0.95% and 1.52%, based on 1 temporal and 8 spatial axes, for *E.*
339 *embla* and *L. helle*, respectively), the dbRDA approach indicated a significant impact of time ($P =$
340 0.002) to explain populations differentiation, as well as a fainter effect of 3/8 spatial variables for
341 *E. embla* and 2/8 spatial variables for *L. helle*, although with respective contributions of spatial
342 variables to the overall R^2 remaining $< 0.5\%$ (see Supplementary Table 3).

343

344 **Discussion**

345 *A direct estimation of genetic variation across the past*

346 The study of past population dynamics has received a lot of scholar attention in recent decades
347 (e.g., Bi et al., 2019; Nadachowska-Brzyska, Li, Smeds, Zhang, & Ellegren, 2015; Tallavaara,
348 Luoto, Korhonen, Järvinen, & Seppä, 2015). Those studies classically identify the most likely
349 demographic model underlying the allelic frequency spectrum measured in modern specimens
350 (Csilléry, Blum, Gaggiotti, & François, 2010; Espíndola et al., 2012; François & Durand, 2010).
351 These demographic inference approaches are, however, often limited as different models can
352 produce similar allelic frequency spectrum and summary statistics, and cannot be discriminated
353 (Lapierre, Lambert, & Achaz, 2017). In contrast, genomic data from historical specimens catch
354 evolution red-handed, and can help overcome such limitations by providing direct snapshots of the
355 past genetic diversity present in a population.

356 In this study, we collected a large sample set of two butterfly species spread across Finland, and
357 spanning the last ~120 years. This sampling provided us with an unique opportunity to quantify the
358 variation of the genetic diversity in both species at a time when their distribution drastically
359 declined (Rassi et al., 2010). We have benefited from the HyRAD genome-complexity-reduction
360 method to obtain genetic data from these valuable samples. HyRAD has been increasingly used in
361 different labs, not only to identify genetic variation in historical material (Crates et al. 2019;
362 Keighley et al. 2019; Linck et al., 2017; Linck et al. 2019; Schmid et al., 2017), but also in ancient
363 DNA (Schmid et al., 2017). Indeed, these methods based on hybridization capture allow to retrieve
364 very small quantities of degraded DNA, which often remain unquantifiable before capture
365 (Supplementary Table 1). The amount of DNA in historical samples and the ability to extract,
366 capture and sequence it depends on the history of the sample, the conditions of collection, sample
367 preparation including drying, pinning etc., and storage. Unfortunately, for most of our historical
368 samples we do not have access to such an information. However, in this study we were able to

369 perform the entire process from historical specimen subsampling to SNP calling for ca. 75% of the
370 samples analyzed from both species, thus suggesting that it is compliant with most preparation
371 histories.

372 For this study we developed a specific pipeline, PopHyRAD, to exploit to the maximum the genetic
373 information contained in samples. For now, the PopHyRAD computational pipeline released here
374 facilitates HyRAD sequence analysis at the within-species level by automating the steps underlying
375 read cleaning, trimming and merging, as well as read mapping, and probes clustering. This pipeline
376 is versatile, and can be used to analyze any type of hybridization-capture data, using either probes
377 from ddRAD or another RAD-seq protocol, or any tool able to reduce genomic complexity such as
378 selective extraction of organellar genomes or amplification of specific genes. The catalog definition
379 and the rest of the analytical workflow were so far essentially empirically explored (or *a posteriori*
380 chosen), considering the outputs, and thus the tools that provided the best geographic or
381 phylogenetic structure (Schmid et al., 2018). Here, we take the opportunity to test more accurately
382 the performance of aligners and SNP-callers on HyRAD data, using different tool combinations,
383 and using a realistic criterion from the point of view of population genetics, namely the FIS. These
384 results revealed large differences on the SNPs identified and on the estimation of the genetic
385 diversity and the inbreeding coefficient and suggested the `bwa aln` read aligner and the
386 `freebayes` SNP caller as the most conservative combination. The non-biologically relevant FIS
387 values obtained with other combinations are likely to be due, at least partially, to increased false
388 positive alignment rates (e.g. mis-identified paralogs) as well as to the oversplitting of loci, i.e a
389 locus separated in two loci in the catalog. This type of difference has already been highlighted in
390 analyses of standard RAD-Seq data (Shafer et al., 2017) and calls for caution in downstream
391 analyses. An analysis based on data simulation is outside the scope of this study but would likely

392 clarify the specificity and sensitivity of the different aligners and SNP-callers, and help each user
393 to refine the most appropriate parameters for their analyses and model species.

394

395 *Genetic diversity decline in butterfly populations across Finland*

396 The HyRAD data gathered in this study supported an overall erosion of genetic diversity at the
397 country-wide level of Finland in both species (see Figure 5). Despite interpretation of genetic
398 diversity variation should be tempered due to our relatively reduced sampling size per locality and
399 temporal bin, one should keep in mind that in the context of museomics, our sampling remains
400 substantial. This pattern of genetic diversity reduction parallels those found by similar studies on
401 butterflies in Northern Europe (Fountain et al., 2016; Ugelvig, Nielsen, Boomsma, & Nash, 2011)
402 but also more broadly in other taxa (Dufresnes et al., 2018; Schmid et al., 2018). Our data also
403 uncovered strong regional differences, with at least one locality (i.e. Kuusamo) showing a local
404 increase in diversity at a given time point, potentially following migration linked to the persistence
405 of their habitat in these specific localities (Habel, Meyer, & Schmitt, 2014) playing a role of refuge
406 for individuals from other populations carrying genetic diversity (Craioveanu, Sitar, & Rákossy,
407 2014). However, estimations on recent samples (i.e., collected in 2015) still show a decline in
408 genetic diversity in this particular locality. The overall erosion of the genetic diversity, both locally
409 and country-wide, is expected given that most Finnish populations of these two butterflies have
410 gone extinct through the 20th century, as a result of a drastic reduction in habitat availability, with
411 the remaining populations not allowing to maintain genetic diversity to levels that once existed in
412 an area of ~340,000 km², a century, or even a few decades ago.

413 The second striking result of this study is the increase in IBD over time, at least in one of the two
414 species. Indeed, in *E. embla*, when considering time slices that divide the timeframe of collected

415 specimens in three periods, only the two last ones (i.e, 1950 - 2000, and 2015) are associated with
416 a significant IBD, with an increasing slope as we reach contemporaneous times. The effect of time
417 was also retrieved in the dbRDA approach, although due to the fact this analysis is individual-
418 based, it was less representative of genetic variation per deme through space and time, thus
419 explaining the low R^2 retrieved in the overall model.

420 Our main result of an overall increase in the slope of the isolation by distance pattern is likely the
421 consequence of an increase in fragmentation, revealing a reduced number of migrants among
422 demes, and thus an increase in the differentiation of populations, essentially due—given the short
423 timespan involved—to drift. This signature might be also found in *L. helle*, even if our sampling
424 does not allow the estimation of IBD for the most recent period (i.e., only two populations are still
425 extant today). This transition from a virtually countrywide panmictic system to a more marked
426 structuring in space is indicative of the fact that despite acknowledged dispersal capabilities of
427 these butterflies in Finland (Habel, Finger, Schmitt, & Nève, 2011), generally related to a
428 colonization-edge syndrome characteristic of populations found at the northern edge of a species'
429 distribution (Duploux, Wong, Corander, Lehtonen, & Hanski, 2017), the fragmentation of habitats
430 has led to a decrease in these exchanges, and thus to local differentiation.

431 Through their impact on biodiversity, human activities are accelerating the extinction of
432 populations and the differentiation of those that persist. This could be catalyzing lineage
433 divergence, except that habitat destruction is an ongoing process, potentially hampered by
434 geopolitical, but potentially ubiquitous, decisions. Our study of two species of butterflies in Finland
435 indicates that not all species might respond identically to this fragmentation, and that comparative
436 studies, involving a larger number of species represented by fresh but also historical specimens,
437 are needed to understand how life history traits influence the species' population response to

438 anthropogenic habitat disturbance and destruction. With the application of both the wetlab HyRAD
439 protocol to historical and fresh specimens, and the PopHyRAD bioinformatic pipeline as described
440 in this study, access to both past and extant genetic diversity should allow a better understanding
441 and anticipation of the neutral response of populations to drastic habitat loss.

442

443 **References**

- 444 Ali, O. A., O'Rourke, S. M., Amish, S. J., Meek, M. H., Luikart, G., Jeffres, C., & Miller, M. R.
445 (2016). RAD Capture (Rapture): Flexible and Efficient Sequence-Based Genotyping.
446 *Genetics*, 202, 389–400.
- 447 Bi, K., Linderoth, T., Singhal, S., Vanderpool, D., Patton, J. L., Nielsen, R., Moritz, C., Good, J.
448 M. (2019). Temporal genomic contrasts reveal rapid evolutionary responses in an alpine
449 mammal during recent climate change. *PLoS Genetics*, 15, e1008119.
- 450 Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina
451 sequence data. *Bioinformatics*, 30, 2114–2120.
- 452 Boucher, F. C., Casazza, G., Szövényi, P., & Conti, E. (2016). Sequence capture using RAD
453 probes clarifies phylogenetic relationships and species boundaries in *Primula* sect. *Auricula*.
454 *Molecular Phylogenetics and Evolution*, 104, 60–72.
- 455 Bouzat, J. L. (2010). Conservation genetics of population bottlenecks: the role of chance,
456 selection, and history. *Conservation Genetics*, 11, 463–478.
- 457 Burrell, A. S., Disotell, T. R., & Bergey, C. M. (2015). The use of museum specimens with high-
458 throughput DNA sequencers. *Journal of Human Evolution*, 79, 35–44.
- 459 Ceballos, G., Ehrlich, P. R., & Dirzo, R. (2017). Biological annihilation via the ongoing sixth
460 mass extinction signaled by vertebrate population losses and declines. *Proceedings of the*
461 *National Academy of Sciences of the United States of America*, 114, E6089–E6096.
- 462 Craioveanu, C., Sitar, C., & Rákósy, L. (2014). Mobility, behaviour and phenology of the Violet
463 Copper *Lycaena helle* in North-Western Romania. In *Jewels in the mist. A synopsis on the*
464 *endangered Violet Copper butterfly Lycaena helle*. Pensoft, Sofia-Moscow, p 91-105.
- 465 Crates, R., Olah, G., Adamski, M., Aitken, N., Banks, S., Ingwersen, D., Ranjard, L., Rayner, L.,

466 Stojanovic, D. , Suchan, T. , von Takach Dukai, B., & Heinsohn, R. (2019) Genomic impact
467 of severe population decline in a nomadic songbird. *PLoS ONE*, 14, e0223953.

468 Csilléry, K., Blum, M. G. B., Gaggiotti, O. E., & François, O. (2010). Approximate Bayesian
469 Computation (ABC) in practice. *Trends in Ecology & Evolution*, 25, 410–418.

470 Dabney, J., Meyer, M., & Pääbo, S. (2013). Ancient DNA damage. *Cold Spring Harbor*
471 *Perspectives in Biology*, 5.

472 Decker, J. E., Pires, J. C., Conant, G. C., McKay, S. D., Heaton, M. P., Chen, K., Cooper, A.,
473 Vilkki, J., Seabury, C. M., Caetano, A. R., Johnson, G. S., Brenneman, R. A., Hanotte, O.,
474 Eggert, L. S., Wiener, P., Kim, J.-J., Kim, K.-S., Sonstegard, T. S., VanTassell, C. P.,
475 Neibergs, H. L., McEwan, J. C., Brauning, R., Coutinho, L. L., Babar, M. E., Wilson, G. A.,
476 McClure, M. C., Rolf, M. M., Kim, J., Schnabel, R. D., & Taylor, J. F. (2009). Resolving
477 the evolution of extant and extinct ruminants with high-throughput phylogenomics.
478 *Proceedings of the National Academy of Sciences of the United States of America*, 106,
479 18644–18649.

480 de Oliveira Roque, F., Menezes, J. F. S., Northfield, T., Ochoa-Quintero, J. M., Campbell, M. J.,
481 & Laurance, W. F. (2018). Warning signals of biodiversity collapse across gradients of
482 tropical forest loss. *Scientific Reports*, 8, 1622.

483 Dray, S., Blanchet, G., Borcard, D., Guenard, G., & Jombart, T. (2016). adespatial: Multivariate
484 multiscale spatial analysis. *R Package Version 0.0.3*.

485 Dufresnes, C., Mazepa, G., Rodrigues, N., Brelsford, A., Litvinchuk, S. N., Sermier, R.,
486 Lavanchy, G., Betto-Colliard, C., Blaser, O., Borzée, A., Cavoto, E., Fabre, G., Ghali, K.,
487 Grossen, C., Horn, A., Leuenberger, J., Phillips, B. C., Saunders, P. A., Savary, R.,
488 Maddalena, T., Stöck, M., Dubey, S., Canestrelli, D., & Jeffries, D. L. (2018). Genomic

489 evidence for cryptic speciation in tree frogs from the Apennine peninsula, with description
490 of *Hyla perrini* sp. nov. *Frontiers in Ecology and Evolution*, 6, 144.

491 Duploux, A., Wong, S. C., Corander, J., Lehtonen, R., & Hanski, I. (2017). Genetic effects on
492 life-history traits in the Glanville fritillary butterfly. *PeerJ*, 5, e3371.

493 Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for
494 visualizing STRUCTURE output and implementing the Evanno method. *Conservation*
495 *Genetics Resources*, 4, 359–361.

496 Ehrlich, P. R. (1995). The scale of human enterprise and biodiversity loss. *Extinction Rates* (H.
497 Lawton and R. M. May, eds.). Oxford Univ, pp. 214–226.

498 Espíndola, A., Pellissier, L., Maiorano, L., Hordijk, W., Guisan, A., & Alvarez, N. (2012).
499 Predicting present and future intra-specific genetic structure through niche hindcasting
500 across 24 millennia. *Ecology Letters*, 15, 649–657.

501 Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals
502 using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14, 2611–2620.

503 Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using
504 multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, 164, 1567–
505 1587.

506 Fountain, T., Nieminen, M., Sirén, J., Wong, S. C., Lehtonen, R., & Hanski, I. (2016).
507 Predictable allele frequency changes due to habitat fragmentation in the Glanville fritillary
508 butterfly. *Proceedings of the National Academy of Sciences of the United States of America*,
509 113, 2678–2683.

510 François, O., & Durand, E. (2010). Spatially explicit Bayesian clustering models in population
511 genetics. *Molecular Ecology Resources*, 10, 773–784.

512 Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: accelerated for clustering the next-
513 generation sequencing data. *Bioinformatics*, *28*, 3150–3152.

514 Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing.
515 *Arxiv preprint*. Retrieved from <http://arxiv.org/abs/1207.3907>

516 Goudet, J. (2005). Hierfstat, a package for R to compute and test hierarchical F-statistics.
517 *Molecular Ecology Notes*, *5*, 184–186.

518 Habel, J. C., Finger, A., Schmitt, T., & Neve, G. (2011). Changing over time: 15 years of
519 isolation influence the population genetic structure of the endangered butterfly *Lycaena*
520 *helle*. *Journal of Zoological Systematics and Evolutionary Research*, *36*, 125–135.

521 Habel, J. C., Finger, A., Schmitt, T., & Nève, G. (2011). Survival of the endangered butterfly
522 *Lycaena helle* in a fragmented environment: Genetic analyses over 15 years. *Journal of*
523 *Zoological Systematics and Evolutionary Research*, *49*, 25–31.

524 Hallmann, C. A., Sorg, M., Jongejans, E., Siepel, H., Hofland, N., Schwan, H., Stenmans, W.,
525 Müller, A., Sumser, H., Hörden, T., Goulson, D., & de Kroon, H. (2017). More than 75
526 percent decline over 27 years in total flying insect biomass in protected areas. *PloS One*, *12*,
527 e0185809.

528 Heino, J., Poykko, H., & Itämes, J. (1998). Occurrence, biology and conservation possibilities of
529 *Lycaena helle* in the area of Koillismaa, eastern Finland. *Baptria, Helsinki*, *23*:163-168.

530 Hoffberg, S. L., Kieran, T. J., Catchen, J. M., Devault, A., Faircloth, B. C., Mauricio, R., &
531 Glenn, T. C. (2016). RADcap: sequence capture of dual-digest RADseq libraries with
532 identifiable duplicates and reduced missing data. *Molecular Ecology Resources*, *16*, 1264–
533 1278.

534 Hoffmann, A. A., Sgrò, C. M., & Kristensen, T. N. (2017). Revisiting Adaptive Potential,

535 Population Size, and Conservation. *Trends in Ecology & Evolution*, 32, 506–517.

536 Horn, S. (2012). Target enrichment via DNA hybridization capture. *Methods in Molecular*
537 *Biology*, 840, 177–188.

538 Jensen, E. L., Edwards, D. L., Garrick, R. C., Miller, J. M., Gibbs, J. P., Cayot, L. J., Tapia, W.,
539 Caccone, A., & Russello, M. A. (2018). Population genomics through time provides insights
540 into the consequences of decline and rapid demographic recovery through head-starting in a
541 Galapagos giant tortoise. *Evolutionary Applications*, 11, 1811–1821.

542 Johnson, C. N., Balmford, A., Brook, B. W., Buettel, J. C., Galetti, M., Guangchun, L., &
543 Wilmshurst, J. M. (2017). Biodiversity losses and conservation responses in the
544 Anthropocene. *Science*, 356, 270–275.

545 Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., & Orlando, L. (2013).
546 mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters.
547 *Bioinformatics*, 29, 1682–1684.

548 Karp, D. S., Mendenhall, C. D., Sandí, R. F., Chaumont, N., Ehrlich, P. R., Hadly, E. A., &
549 Daily, G. C. (2013). Forest bolsters bird abundance, pest control and coffee yield. *Ecology*
550 *Letters*, 16, 1339–1347.

551 Keighley, M. V., Heinsohn, R., Langmore, N. E., Murphy, S. A., & Peñalba, J. V. (2019).
552 Genomic population structure aligns with vocal dialects in Palm Cockatoos (*Probosciger*
553 *aterrimus*); evidence for refugial late-Quaternary distribution? *Emu-Austral Ornithology*,
554 119, 24-37.

555 Keller, L. F., & Waller, D. M. (2002). Inbreeding effects in wild populations. *Trends in Ecology*
556 *& Evolution*, 17, 230–241.

557 Kim, D., Song, L., Breitwieser, F. P., & Salzberg, S. L. (2016). Centrifuge: rapid and sensitive

558 classification of metagenomic sequences. *Genome Research*, 26, 1721–1729.

559 Koboldt, D. C., Chen, K., Wylie, T., Larson, D. E., McLellan, M. D., Mardis, E. R., Weinstock,
560 G. M., Wilson, R. K., & Ding, L. (2009). VarScan: variant detection in massively parallel
561 sequencing of individual and pooled samples. *Bioinformatics*, 25, 2283–2285.

562 Kristensen, T. N., Pedersen, K. S., Vermeulen, C. J., & Loeschcke, V. (2010). Research on
563 inbreeding in the “omic” era. *Trends in Ecology & Evolution*, 25, 44–52.

564 Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature*
565 *Methods*, 9, 357–359.

566 Lapiere, M., Lambert, A., & Achaz, G. (2017). Accuracy of Demographic Inferences from the
567 Site Frequency Spectrum: The Case of the Yoruba Population. *Genetics*, 206, 439–449.

568 Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
569 *Arxiv Preprint*. Retrieved from <http://arxiv.org/abs/1303.3997>

570 Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler
571 transform. *Bioinformatics*, 25, 1754–1760.

572 Linck, E. B., Hanna, Z. R., Sellas, A., & Dumbacher, J. P. (2017). Evaluating hybridization
573 capture with RAD probes as a tool for museum genomics with historical bird specimens.
574 *Ecology and Evolution*, 7, 4755–4767.

575 Linck, E. B., Freeman, B. G., & Dumbacher, J. P. (2019). Speciation with gene flow across an
576 elevational gradient in New Guinea kingfishers. *bioRxiv*, 589044.

577 Lister, B. C., & Garcia, A. (2018). Climate-driven declines in arthropod abundance restructure a
578 rainforest food web. *Proceedings of the National Academy of Sciences of the United States*
579 *of America*, 115, E10397–E10406.

580 Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T. H., Piñero, D., & Emerson, B. C.

581 (2015). Restriction site-associated DNA sequencing, genotyping error estimation and de
582 novo assembly optimization for population genetic inference. *Molecular Ecology Resources*,
583 *15*, 28–41.

584 Meineke, E. K., Davies, T. J., Daru, B. H., & Davis, C. C. (2018). Biological collections for
585 understanding biodiversity in the Anthropocene. *Philosophical Transactions of the Royal*
586 *Society of London. Series B, Biological Sciences*, *374*, 20170386.

587 Nadachowska-Brzyska, K., Li, C., Smeds, L., Zhang, G., & Ellegren, H. (2015). Temporal
588 Dynamics of Avian Populations during Pleistocene Revealed by Whole-Genome Sequences.
589 *Current Biology*, *25*, 1375–1380.

590 Oksanen, J., Kindt, R., Legendre, P., & O’Hara, B. (2016). The vegan package. R package
591 version 2.4-1.

592 Orlando, L., Gilbert, M. T. P., & Willerslev, E. (2015). Reconstructing ancient genomes and
593 epigenomes. *Nature Reviews. Genetics*, *16*, 395–408.

594 Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest
595 RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and
596 non-model species. *PloS One*, *7*, e37135.

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

599 Puritz, J. B., Hollenbeck, C. M., & Gold, J. R. (2014). dDocent: a RADseq, variant-calling
600 pipeline designed for population genomics of non-model organisms. *PeerJ*, *2*, e431.

601 Pyke, G. H., & Ehrlich, P. R. (2010). Biological collections and ecological/environmental
602 research: a review, some observations and a look to the future. *Biological Reviews of the*
603 *Cambridge Philosophical Society*, *85*, 247–266.

604 Rassi, P., Hyvärinen, E., Juslén, A., & Mannerkoski, I. (2010). The 2010 Red List of Finnish
605 Species. Ministry of the Environment. *Finnish Environment Institute, Helsinki*.

606 Ryan, S. F., Deines, J. M., Scriber, J. M., Pfrender, M. E., Jones, S. E., Emrich, S. J., &
607 Hellmann, J. J. (2018). Climate-mediated hybrid zone movement revealed with genomics,
608 museum collection, and simulation modeling. *Proceedings of the National Academy of*
609 *Sciences of the United States of America*, *115*, E2284–E2291.

610 Sánchez Barreiro, F., Vieira, F. G., Martin, M. D., Haile, J., Gilbert, M. T. P., & Wales, N.
611 (2017). Characterizing restriction enzyme-associated loci in historic ragweed (*Ambrosia*
612 *artemisiifolia*) voucher specimens using custom-designed RNA probes. *Molecular Ecology*
613 *Resources* *17*, 209–220.

614 Sánchez-Bayo, F., & Wyckhuys, K. A. G. (2019). Worldwide decline of the entomofauna: A
615 review of its drivers. *Biological Conservation*, *232*, 8–27.

616 Schmid, S., Genevest, R., Gobet, E., Suchan, T., Sperisen, C., Tinner, W., & Alvarez, N. (2017).
617 HyRAD-X, a versatile method combining exome capture and RAD sequencing to extract
618 genomic information from ancient DNA. *Methods in Ecology and Evolution / British*
619 *Ecological Society*, *8*, 1374–1388.

620 Schmid, S., Neuenschwander, S., Pitteloud, C., Heckel, G., Pajkovic, M., Arlettaz, R., & Alvarez,
621 N. (2018). Spatial and temporal genetic dynamics of the grasshopper *Oedaleus decorus*
622 revealed by museum genomics. *Ecology and Evolution*, *8*, 1480–1495.

623 Schubert, M., Ginolhac, A., Lindgreen, S., Thompson, J. F., Al-Rasheid, K. A. S., Willerslev, E.,
624 Krogh, A., & Orlando, L. (2012). Improving ancient DNA read mapping against modern
625 reference genomes. *BMC Genomics*, *13*, 178.

626 Schubert, M., Lindgreen, S., & Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming,

627 identification, and read merging. *BMC Research Notes*, 9, 88.

628 Shafer, A. B. A., Peart, C. R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C. W., & Wolf, J. B.
629 W. (2017). Bioinformatic processing of RAD-seq data dramatically impacts downstream
630 population genetic inference. *Methods in Ecology and Evolution / British Ecological*
631 *Society*, 8, 907–917.

632 Suchan, T., Pitteloud, C., Gerasimova, N. S., Kostikova, A., Schmid, S., Arrigo, N., Pajkovic, M.,
633 Ronikier, M., & Alvarez, N. (2016). Hybridization Capture Using RAD Probes (hyRAD), a
634 New Tool for Performing Genomic Analyses on Collection Specimens. *PloS One*, 11,
635 e0151651.

636 Tallavaara, M., Luoto, M., Korhonen, N., Järvinen, H., & Seppä, H. (2015). Human population
637 dynamics in Europe over the Last Glacial Maximum. *Proceedings of the National Academy*
638 *of Sciences of the United States of America*, 112, 8232–8237.

639 Tegelberg, R., Haapala, J., Mononen, T., Pajari, M., & Saarenmaa, H. (2012). The development
640 of a digitising service centre for natural history collections. *ZooKeys*, 209, 75–86.

641 Thielens, A., Bell, D., Mortimore, D. B., Greco, M. K., Martens, L., & Joseph, W. (2018).
642 Exposure of Insects to Radio-Frequency Electromagnetic Fields from 2 to 120 GHz.
643 *Scientific Reports*, 8, 3924.

644 Tin, M. M.-Y., Economo, E. P., & Mikheyev, A. S. (2014). Sequencing degraded DNA from
645 non-destructively sampled museum specimens for RAD-tagging and low-coverage shotgun
646 phylogenetics. *PloS One*, 9, e96793.

647 Ugelvig, L. V., Nielsen, P. S., Boomsma, J. J., & Nash, D. R. (2011). Reconstructing eight
648 decades of genetic variation in an isolated Danish population of the large blue butterfly
649 *Maculinea arion*. *BMC Evolutionary Biology*, 11, 201.

650 van Strien, A. J., van Swaay, C. A. M., van Strien-van Liempt, W. T. F. H., Poot, M. J. M., &
651 WallisDeVries, M. F. (2019). Over a century of data reveal more than 80% decline in
652 butterflies in the Netherlands. *Biological Conservation*, 234, 116–122.

653 Whiteside, M., & Marvin Herndon, J. (2018). Previously Unacknowledged Potential Factors in
654 Catastrophic Bee and Insect Die-off Arising from Coal Fly Ash Geoengineering. *Asian*
655 *Journal of Biology*, 6, 1–13.

656 **Figure captions:** 2,742 SNPs from 869 loci for *E. embla* and 2,549

657 **Figure 1.** Schematic PopHyRAD pipeline including the creation of the catalog from probes and the
658 treatment of samples.

659

660 **Figure 2.** a,c Pictures of the two focal species, *Lycaena helle* and *Erebia embla*. b,d Maps of
661 Finland with the sampling localities. The different colors indicate whether the samples considered
662 were historical, with a distinction before and after 1950, or modern (year 2015).

663

664 **Figure 3.** Observed gene diversity (H_s), frequency of fixed alleles and inbreeding coefficient (F_{IS})
665 estimated for each locality and years. For each point the number indicates the number of samples.
666 Vertical lines represent the standard error across all SNPs (2,742 for *Erebia embla* and 2,549 for
667 *Lycaena helle*). Localities sampled at different time points are linked by a line of corresponding
668 color.

669

670 **Figure 4.** Observed gene diversity (H_s) and frequency of fixed alleles over time, i.e., before 1950,
671 between 1950 and 2000, and 2015, and for each locality. Vertical lines represent the standard error
672 across all SNPs (2,742 for *Erebia embla* and 2,549 for *Lycaena helle*). Localities sampled at
673 different time slices are linked by a line of identical color.

674

675 **Figure 5.** Observed gene diversity (H_s) and frequency of fixed alleles have been estimated by time
676 slices, before 1950, between 1950 and 2000 and 2015, all over Finland by merging samples.
677 Vertical lines represent the standard error across all SNPs (2,742 for *Erebia embla* and 2,549 for
678 *Lycaena helle*).

679

680 **Figure 6.** Population structure of all sampled localities according to the three studied periods. The
681 pie charts indicate the proportion of samples associated with each lineage as identified by
682 STRUCTURE with $K=2$.

683

684 **Figure 7.** Isolation by distance was investigated for each time slice by computing the correlation
685 between the genetic differentiation between populations ($F_{ST}/1-F_{ST}$) and the log-transformed
686 geographic distance ($\log(\text{distance})$). Correlations were tested using Spearman's rank correlation,
687 and coefficient of determination, intercept and slopes using a linear model are shown.

Table 1. Sample information comprising locality name, approximate GPS positions, year of collection and number of samples. The three time slices used for the analyses have been separated.

<i>Erebia embla</i>						<i>Lycaena helle</i>					
	Locality	Latitude	Longitude	Year	N		Locality	Latitude	Longitude	Year	N
before 1950	Haminalahti	62.8534641	27.5326783	1909	12		Kuusamo Liikasenvaara	65.9645637	29.1883283	1928	10
	Pirkkala	61.4654497	23.6456252	1909	1		Paanajarvi	66.4555006	28.9798017	1934	1
	Pirkkala	61.4654497	23.6456252	1925	2		Paanajarvi	66.4555006	28.9798017	1935	6
	Pirkkala	61.4654497	23.6456252	1930	5		Ivalo	68.6588185	27.5348114	1937	13
	Pernio	60.2050782	23.1235771	1932	1		Mikkeli	61.6877956	27.2726569	1938	3
	Portom	62.7100207	21.6163442	1937	13		Harmoinen	61.4852477	25.1409736	1940	8
	Muonio	67.9593397	23.6774037	1938	13		Kannus	63.9007773	23.9170363	1940	5
	Nurmes	63.5422079	29.1410100	1941	6		Nurmes	63.5422079	29.1410100	1941	13
	Pernio	60.2050782	23.1235771	1944	2		Ruovesi	61.9856303	24.0703481	1941	13
	Pirkkala	61.4654497	23.6456252	1945	2		Mikkeli	61.6877956	27.2726569	1942	9
	Jakobstad Pietarsaari	63.6666709	22.7000229	1947	1		Haapavesi	64.1378737	25.3658176	1946	5
	Pelkosenniemi	67.1095969	27.5118116	1947	9		Pelkosenniemi	67.1095969	27.5118116	1947	12
	Jakobstad Pietarsaari	63.6666709	22.7000229	1949	4		Paltamo	64.4068668	27.8335512	1949	10
between 1950 and 2000	Jakobstad Pietarsaari	63.6666709	22.7000229	1951	2		Kuusamo Liikasenvaara	65.9645637	29.1883283	1955	1
	Jakobstad Pietarsaari	63.6666709	22.7000229	1953	6		Tohmajarvi	62.2259448	30.3335512	1957	1
	Kuusamo	65.9645637	29.1883283	1955	8		Tohmajarvi	62.2259448	30.3335512	1958	5
	Tyrvanto	61.1546112	24.3283168	1959	1		Kuopio	62.8241424	27.5945615	1959	4
	Karttula	62.8952013	26.9723784	1963	4		Kuusamo Liikasenvaara	65.9645637	29.1883283	1959	1
	Ikaalinen	61.7701493	23.0633777	1965	5		Kuopio	62.8241424	27.5945615	1960	4
	Ikaalinen	61.7701493	23.0633777	1969	5		Tohmajarvi	62.2259448	30.3335512	1960	4
	Tuulos	61.1181656	24.8337064	1970	2		Kuusamo Liikasenvaara	65.9645637	29.1883283	1962	2
	Tuulos	61.1181656	24.8337064	1973	2		Kuopio	62.8241424	27.5945615	1964	1
	Tuulos	61.1181656	24.8337064	1975	1		Kuusamo Liikasenvaara	65.9645637	29.1883283	1975	3
	Kuusamo	65.9645637	29.1883283	1977	4		Kuivaniemi Simo	65.6040217	25.2038392	1980	6
	Mikkeli	61.6877956	27.2726569	1979	2		Kuusamo Liikasenvaara	65.9645637	29.1883283	1985	12
	Kuusamo	65.9645637	29.1883283	1980	1		Kuusamo Liikasenvaara	65.9645637	29.1883283	1991	13
	Kuusamo	65.9645637	29.1883283	1981	3						
	Mikkeli	61.6877956	27.2726569	1983	1						
2015	Ivalo	68.6588185	27.5348114	2015	5		Kuivaniemi Simo	65.6040217	25.2038392	2015	9
	Kuusamo	65.9645637	29.1883283	2015	9		Kuusamo Liikasenvaara	65.9645637	29.1883283	2015	23
	Muonio	67.9593397	23.6774037	2015	9						
	Oulu	65.0118734	25.4716809	2015	12						
	Pelkosenniemi	67.1095969	27.5118116	2015	8						
Rovaniemi	66.4976214	25.7192101	2015	11							

Supplementary material

Supplementary Table 1. For each sample, detailed information including year of sampling, and thus the age of the sample, the DNA concentration measured after the extraction, the DNA concentration after capture, the number of sequenced reads, the percentage of mapped reads and finally the number of called SNPs.

Supplementary Table 2. Number of reads and mapping percentages before and after cleaning for all samples and for the three mapping tools tested.

Supplementary Table 3. Results of the dbRDA analysis for *Erebia embla* and *Lycaena helle*. R^2 values are given for the overall model as well as for each of the 8 spatial and 1 temporal variable. Respective P values are also indicated.

Supplementary Figure 1. Deamination patterns in historical (top) and fresh (bottom) samples.

Supplementary Figure 2. Comparison of FIS distribution between the SNPs obtained from the three mapping tools and the two SNP-calling tools.

Erebia embla

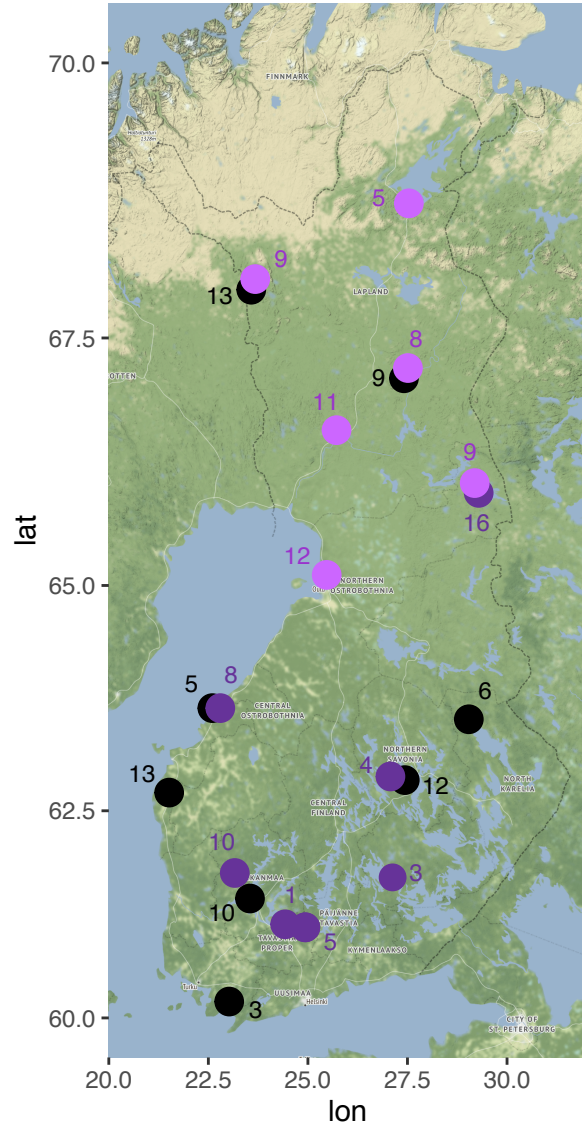
a.



Legend

- 2015
- Between 1950 and 2000
- Before 1950

b.



Lycaena helle

c.



Legend

- 2015
- Between 1950 and 2000
- Before 1950

d.

Figure 1

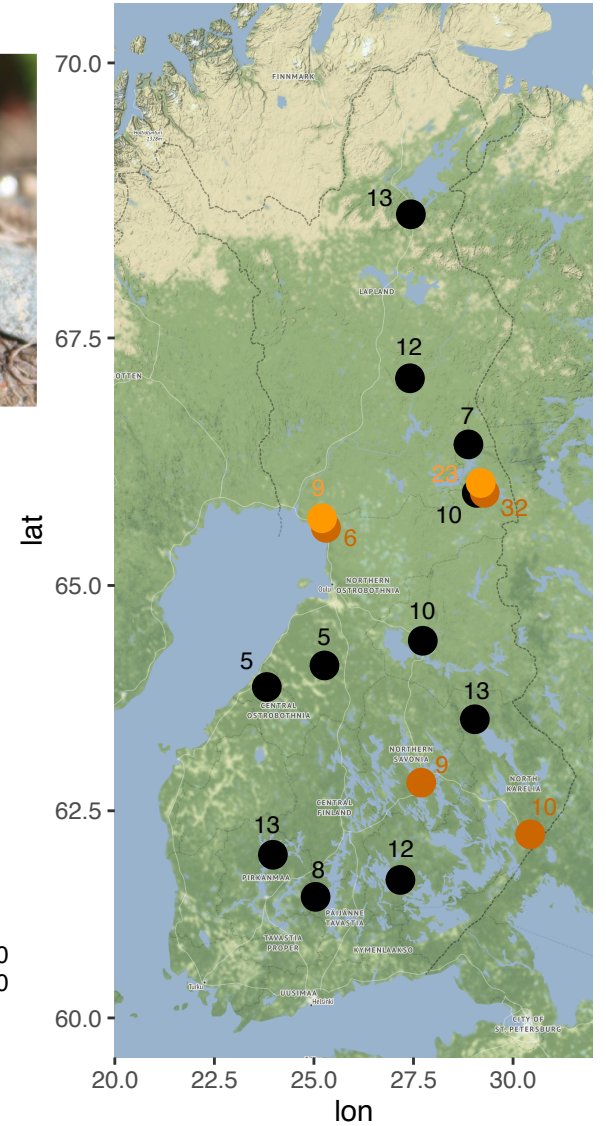
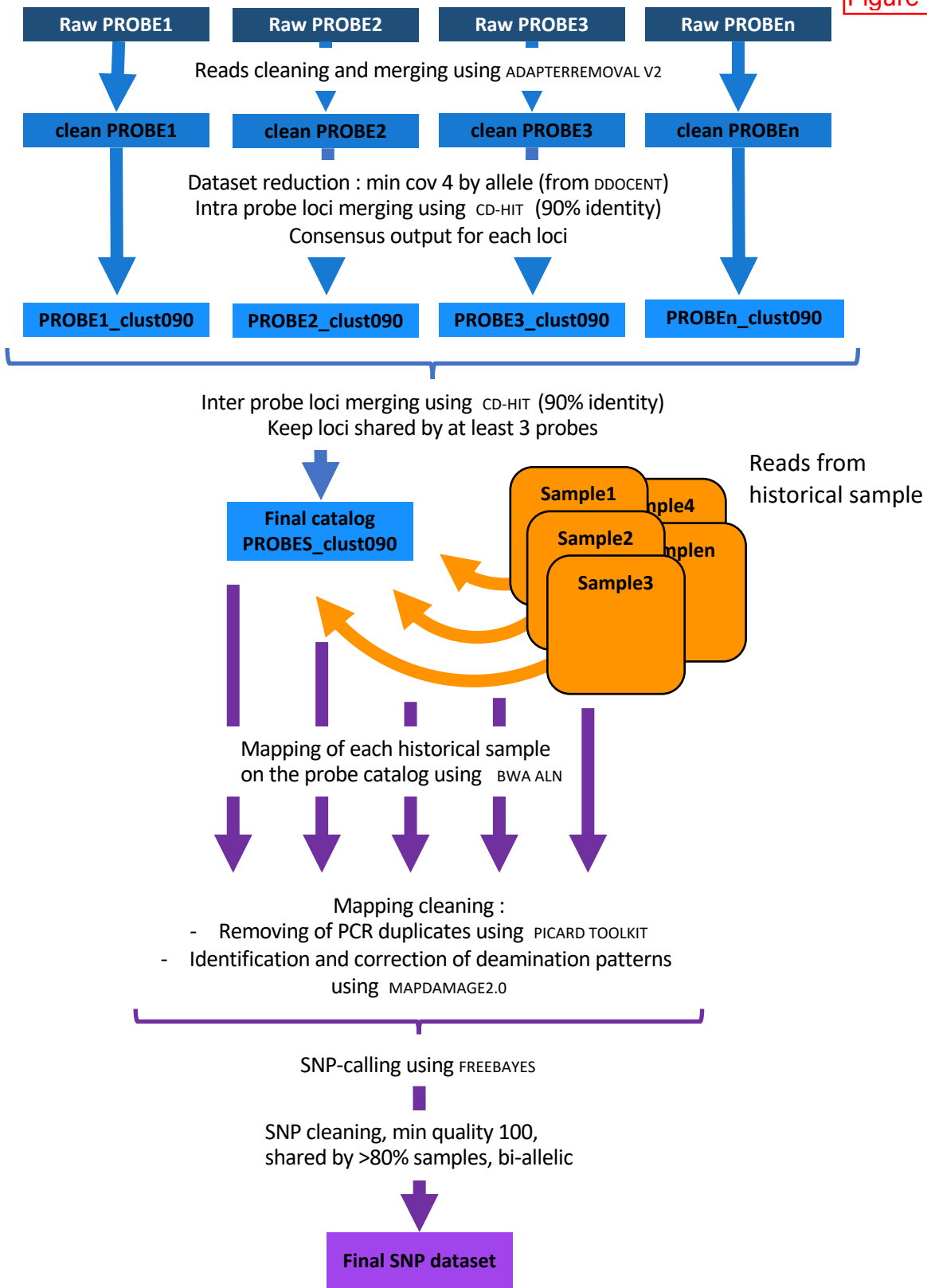
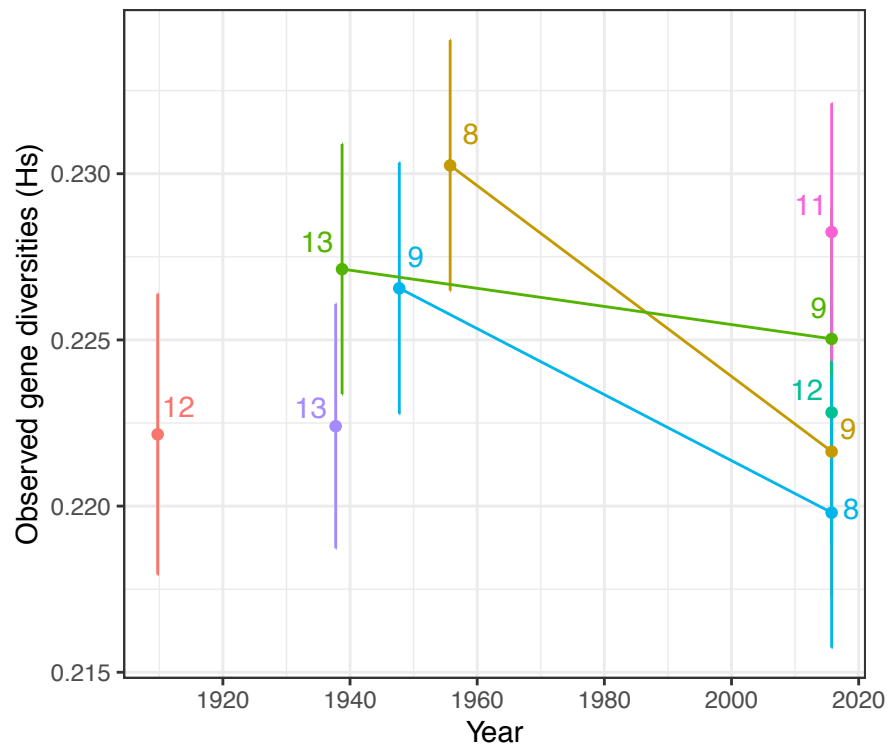


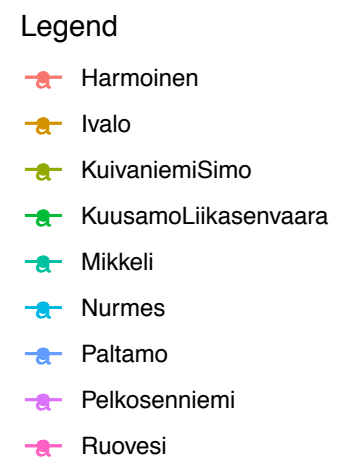
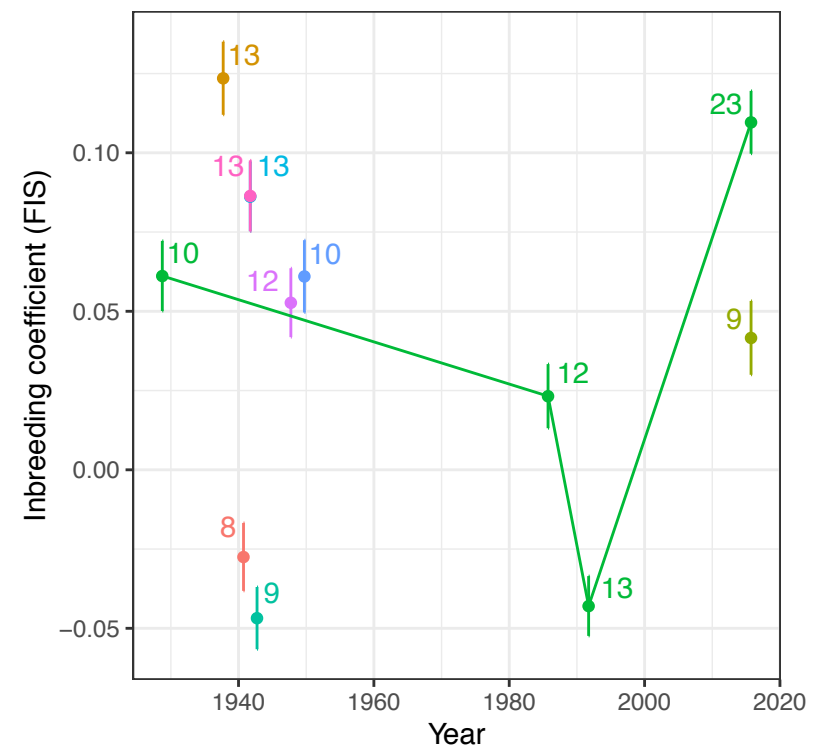
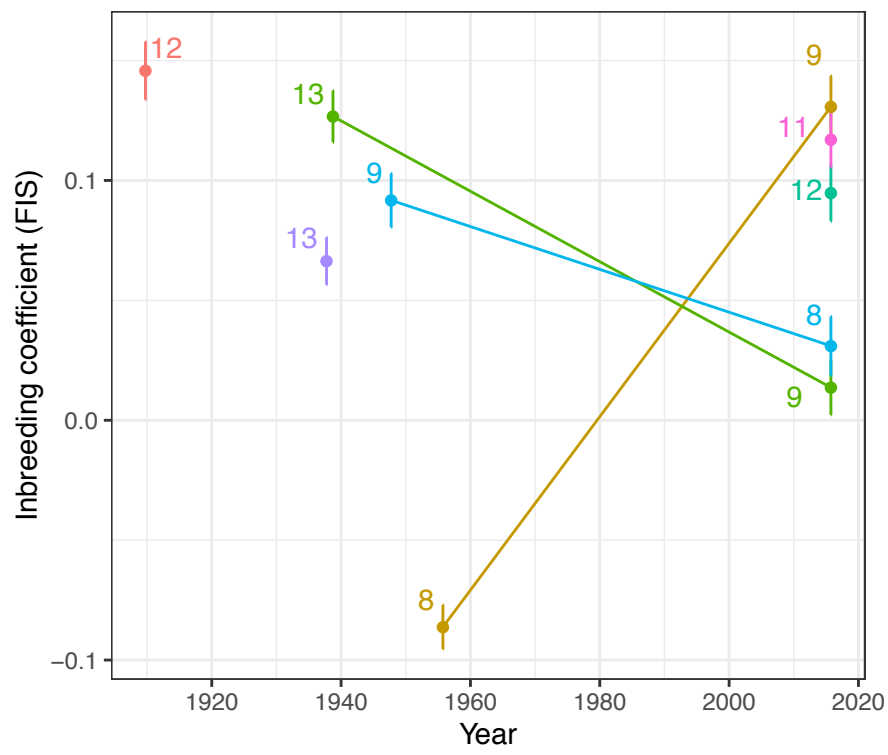
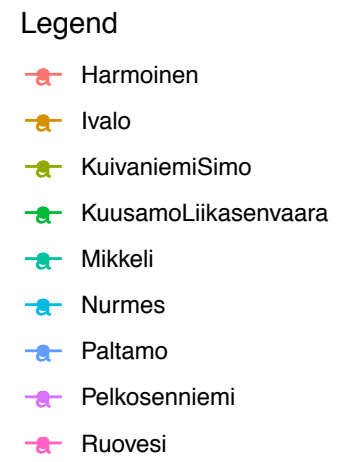
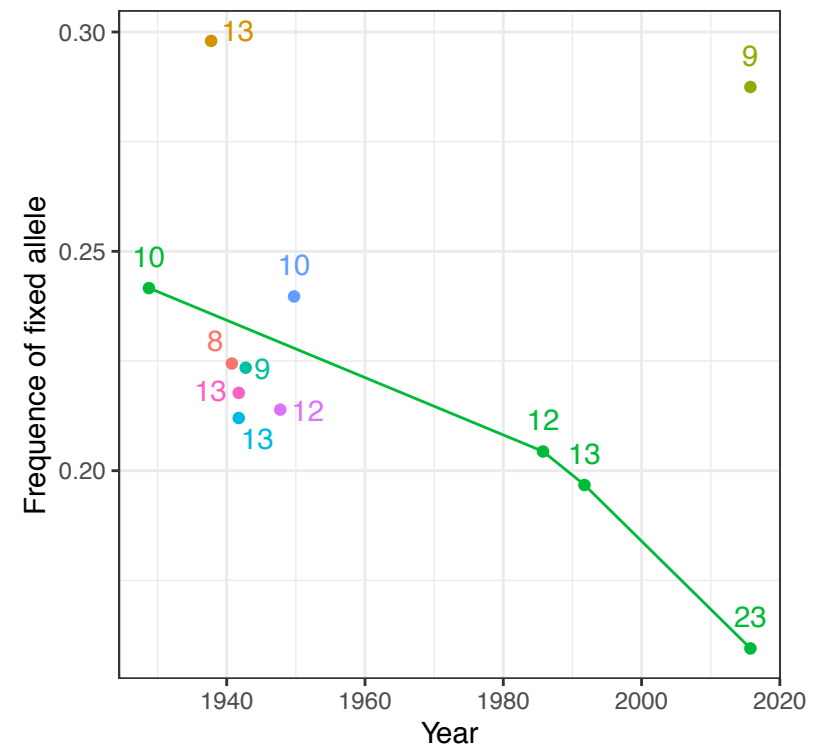
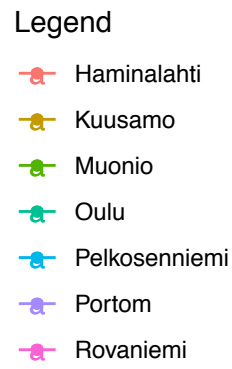
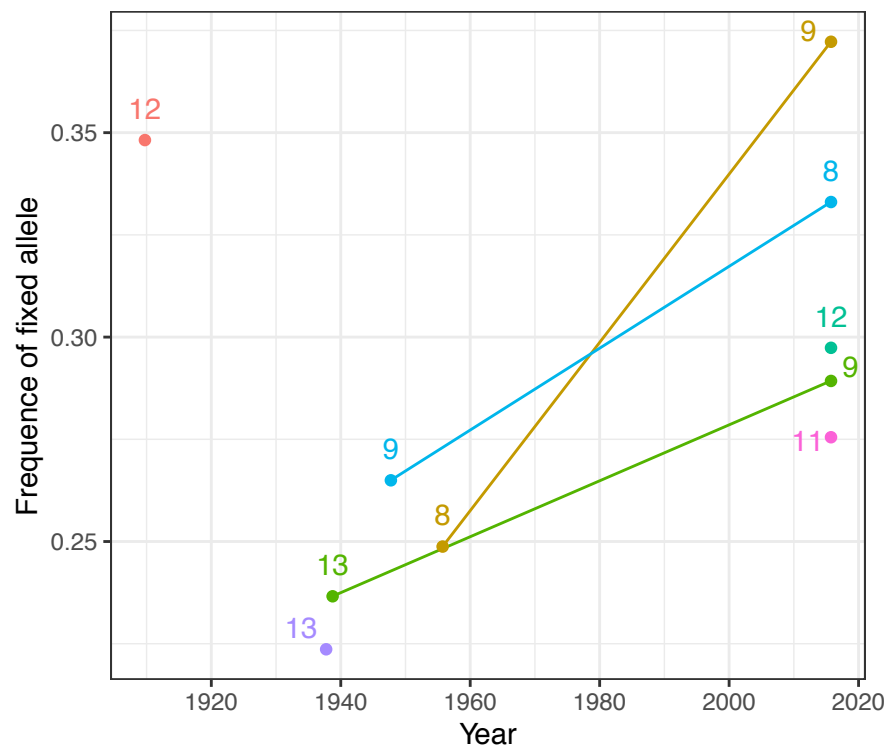
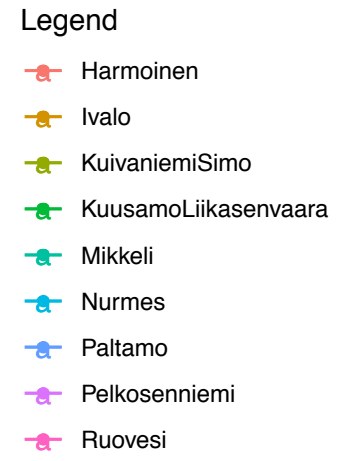
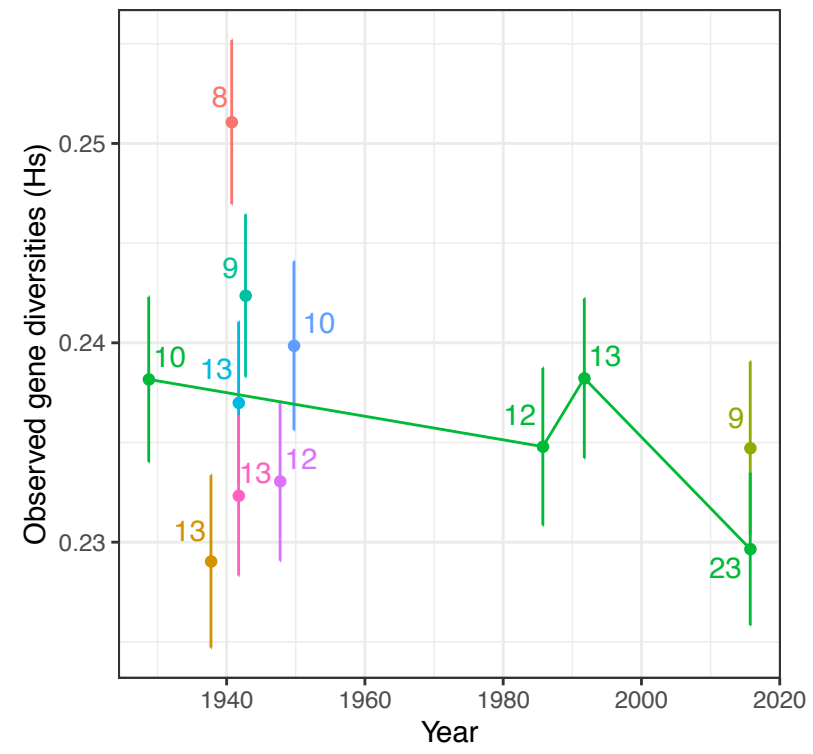
Figure 2



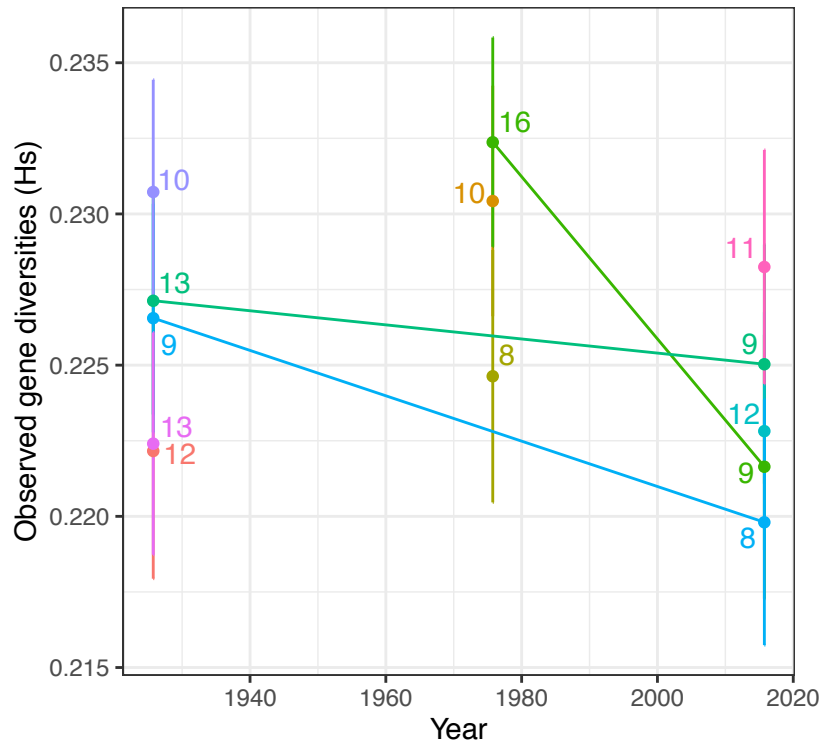
Erebia embla



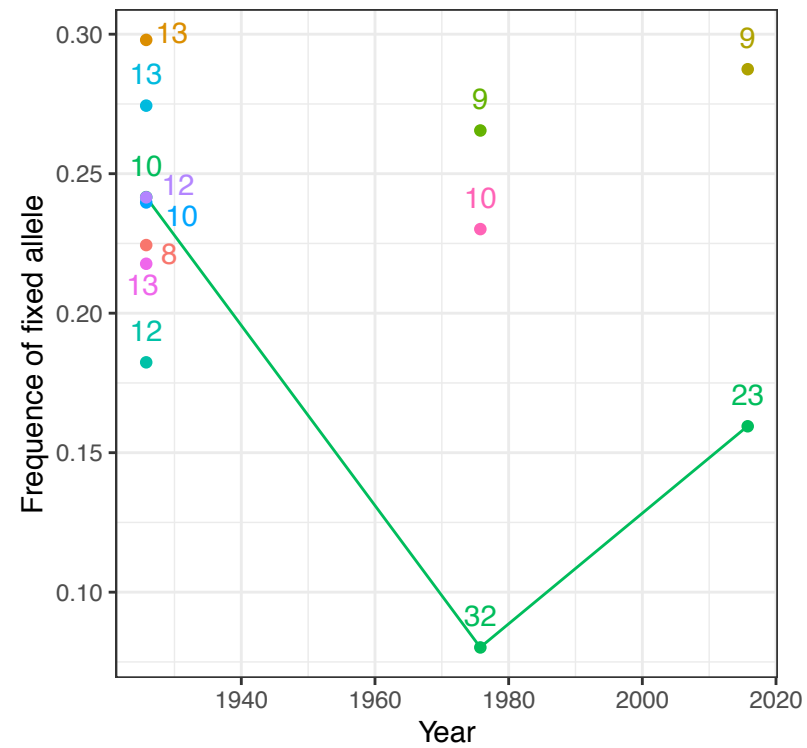
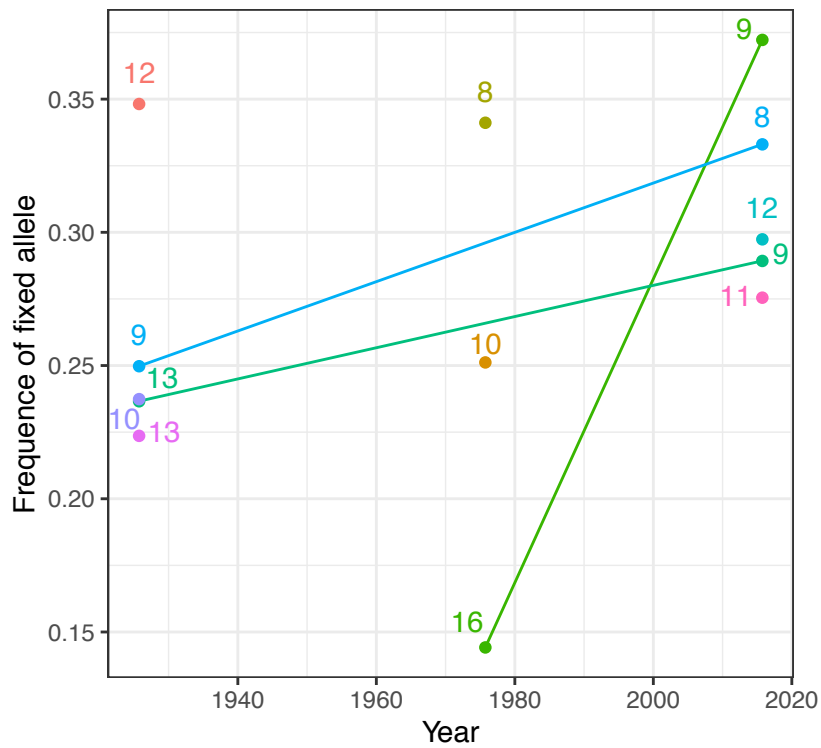
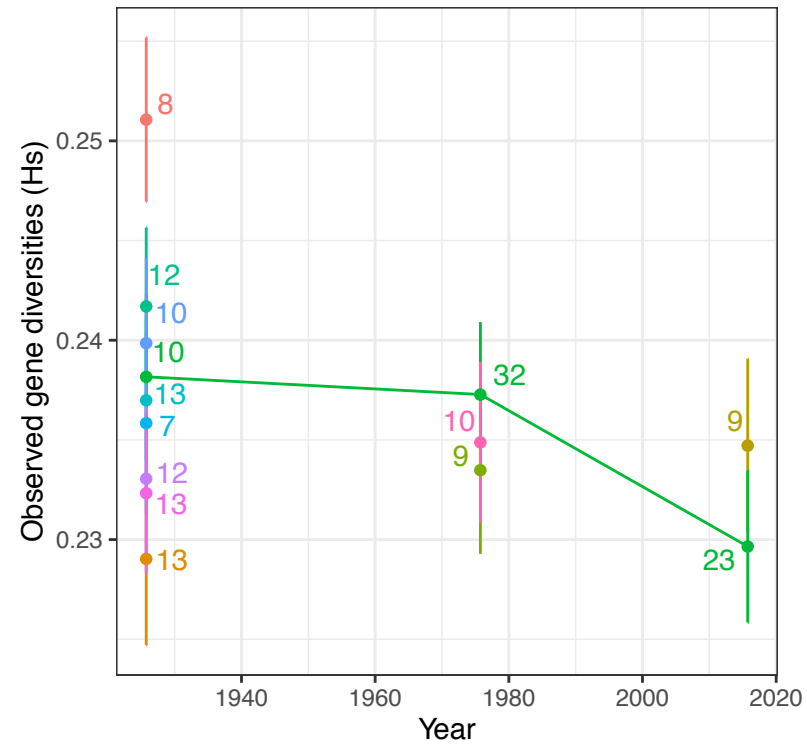
Lycaena helle



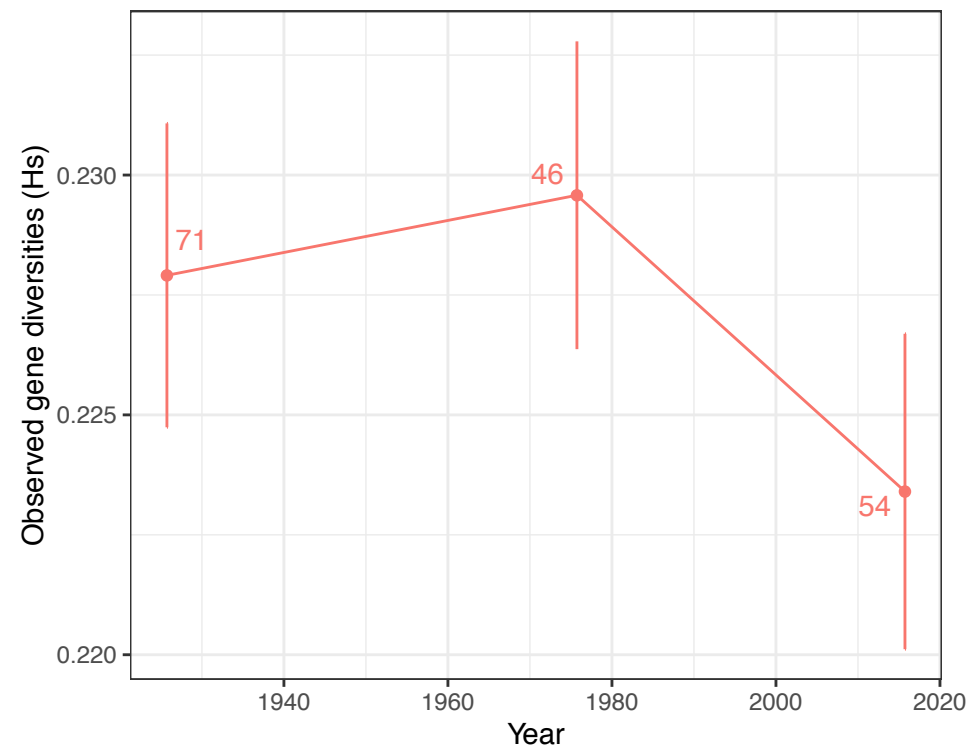
Erebia embla



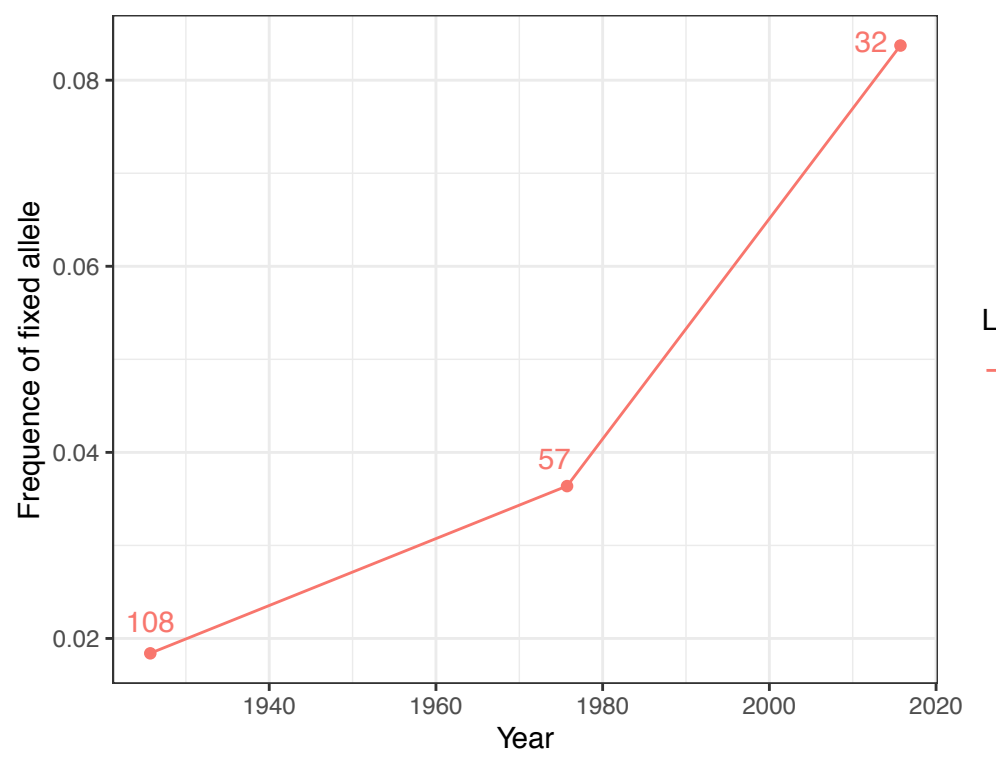
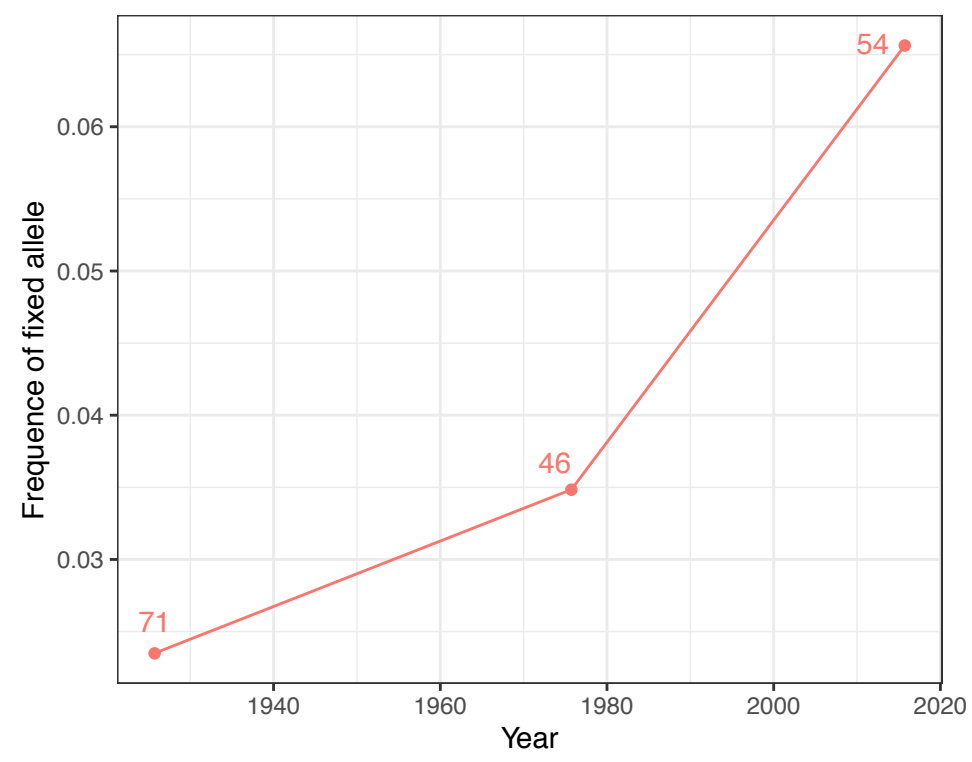
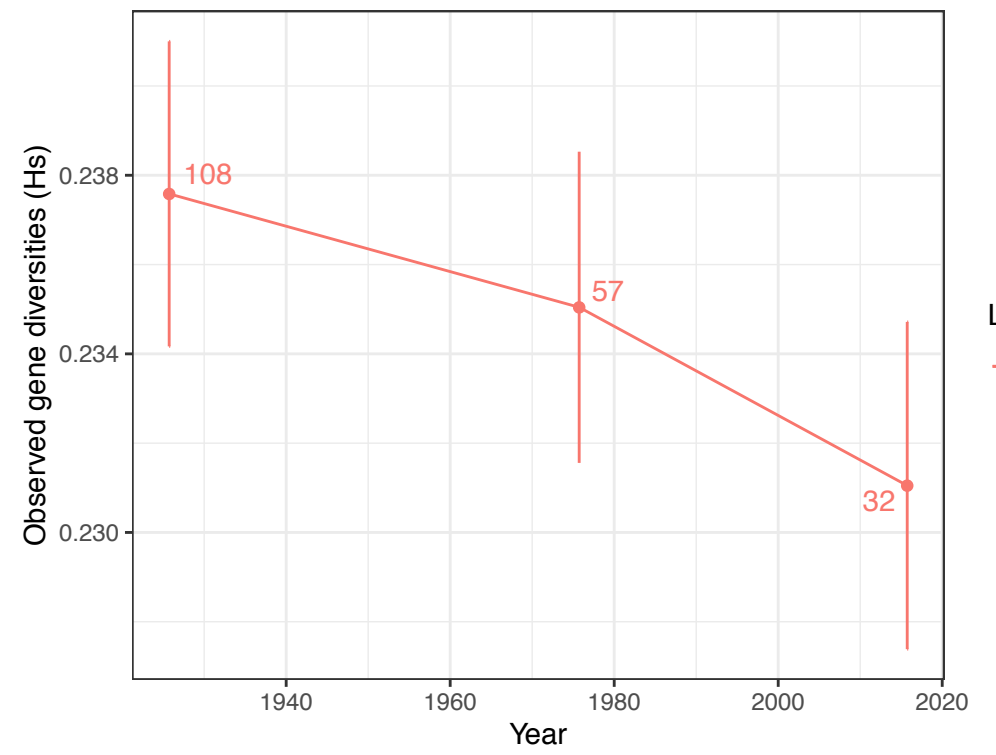
Lycaena helle



Erebia embla



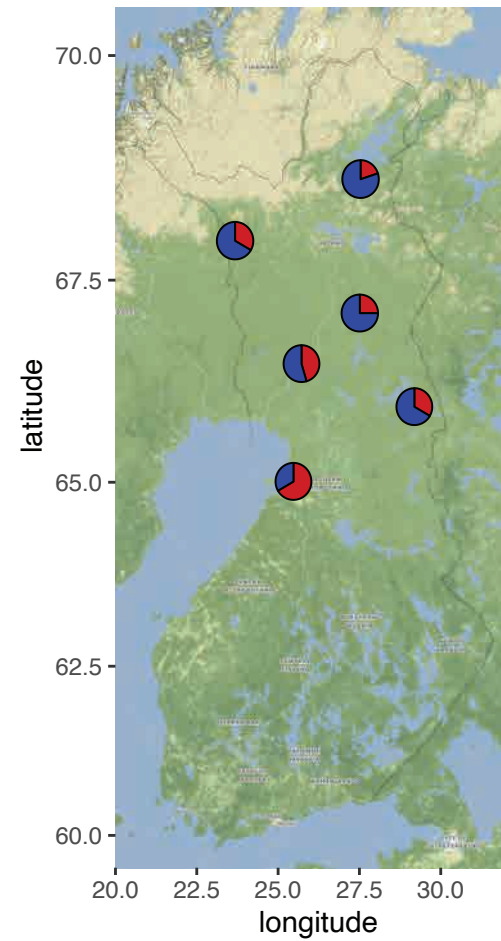
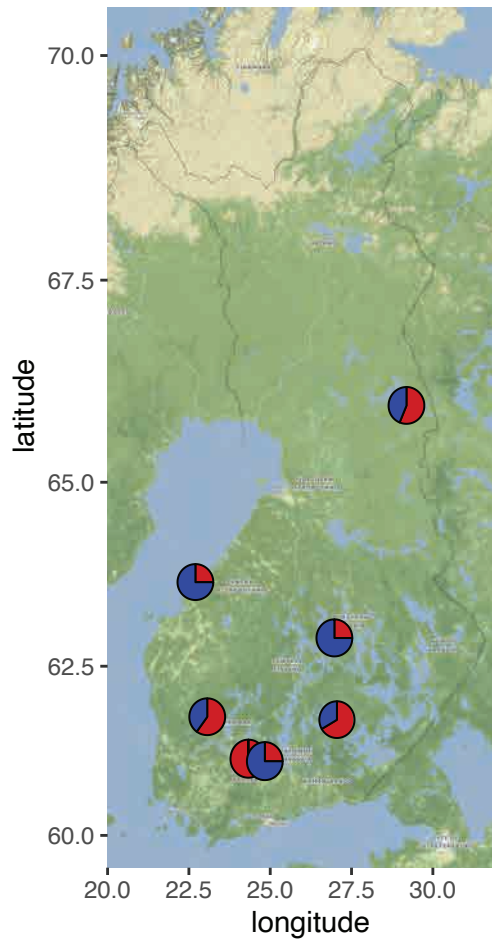
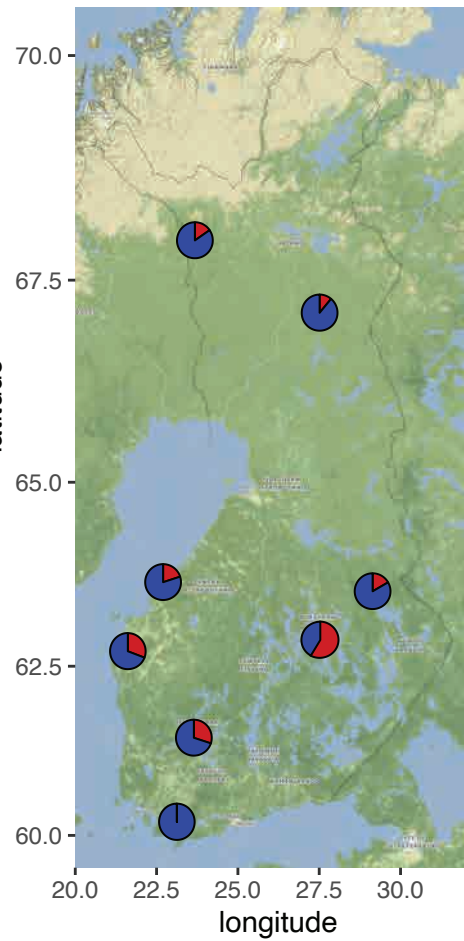
Lycaena helle



Before 1950

Between 1950 and 2000

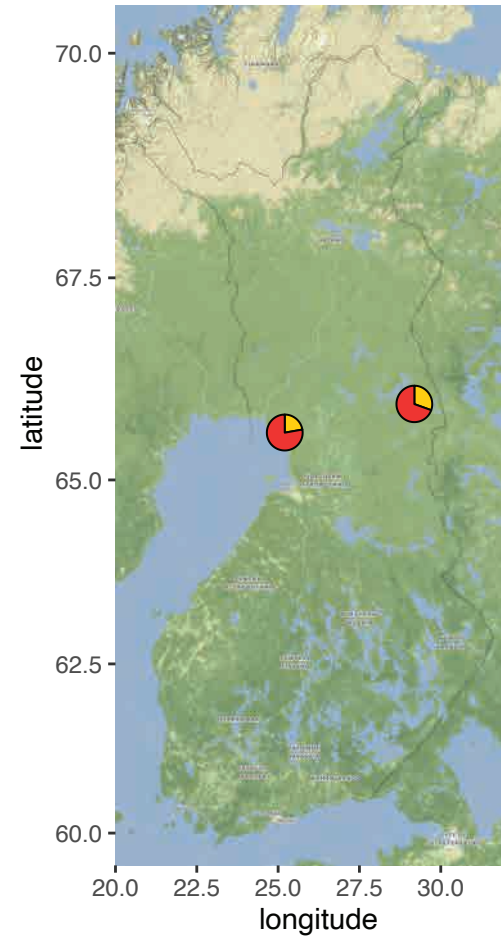
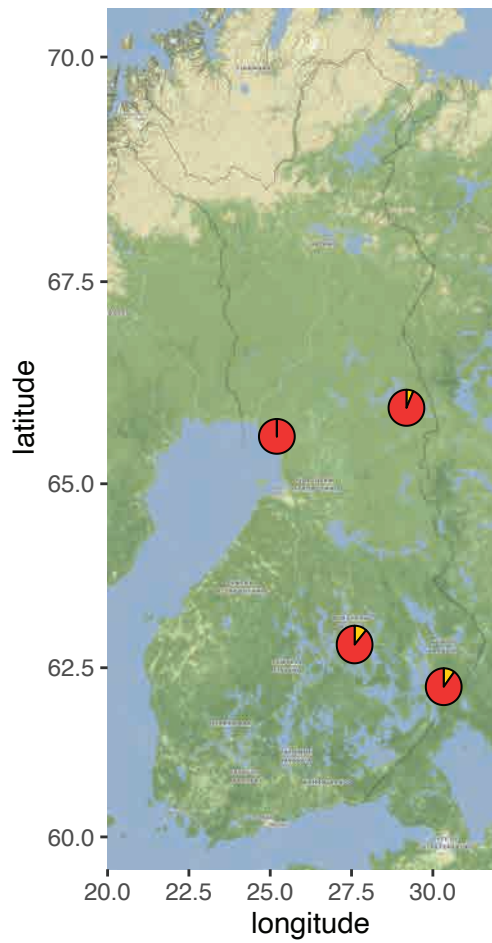
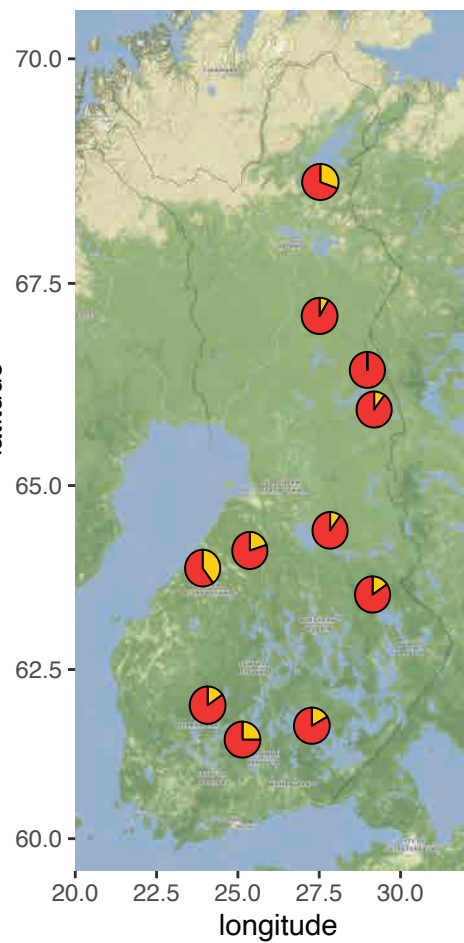
2015



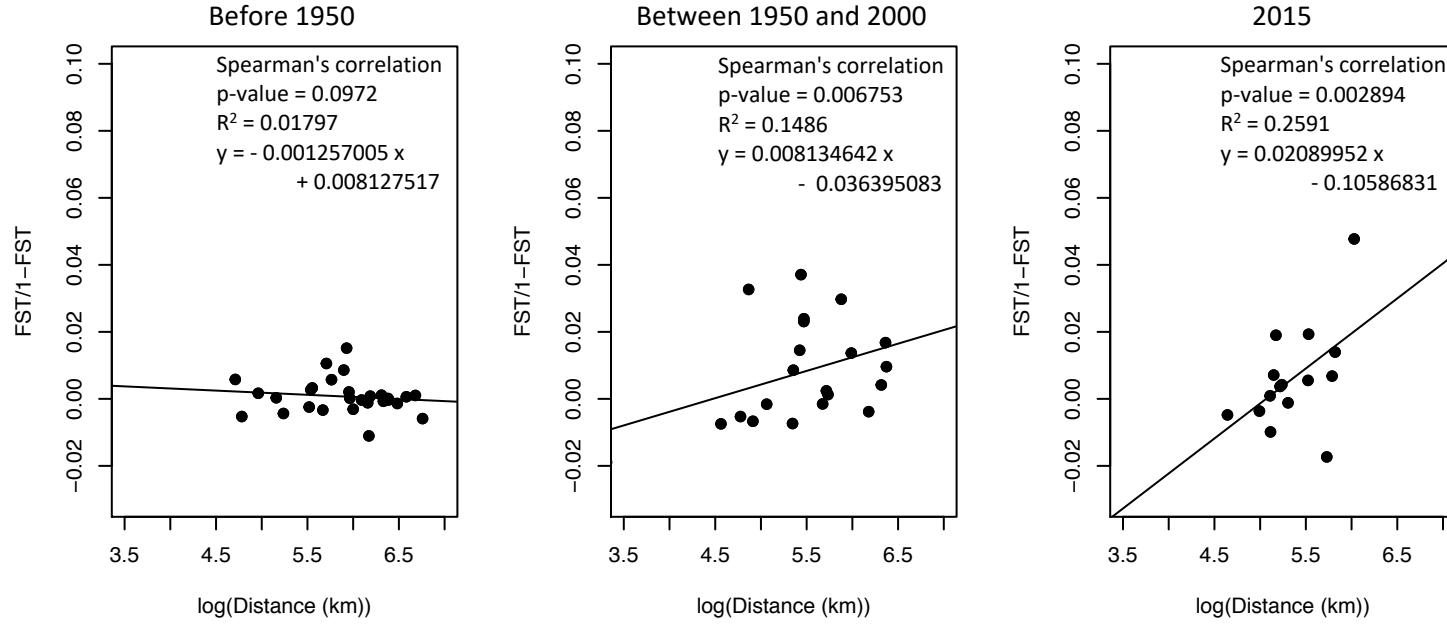
Before 1950

Between 1950 and 2000

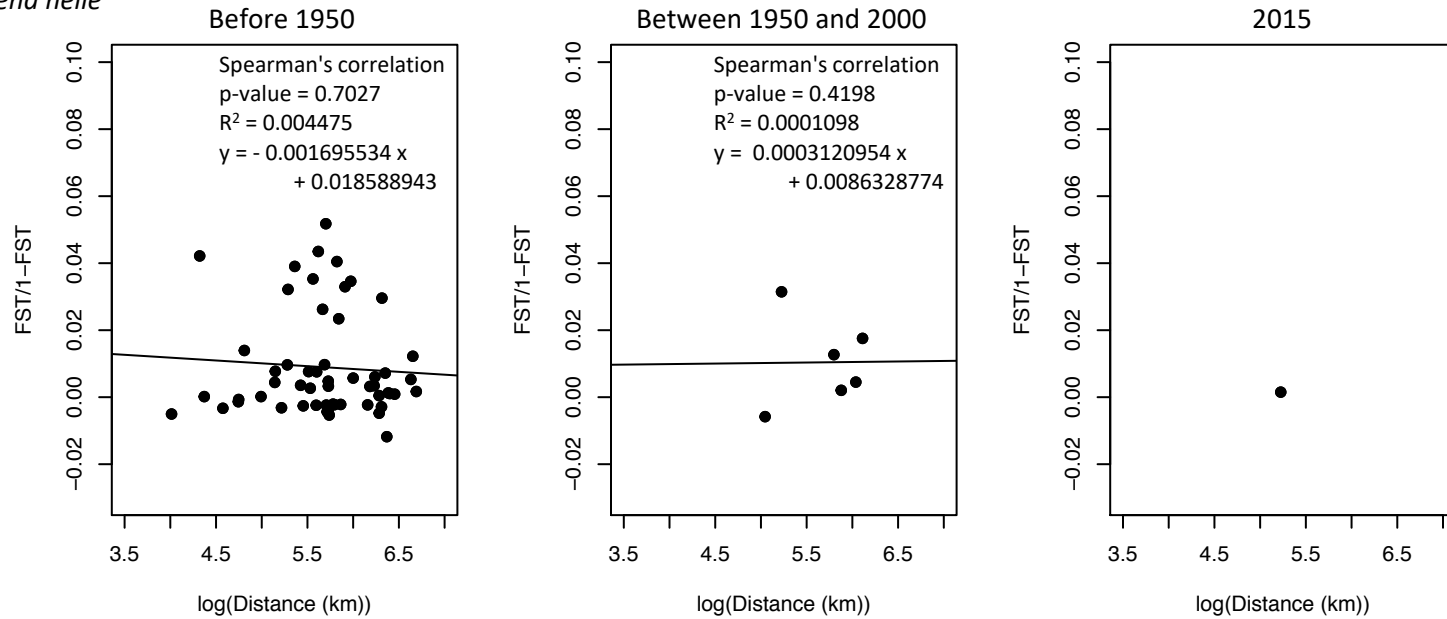
2015



Erebia embla



Lycaena helle



Museomics identifies genetic erosion in two butterfly species across the 20th century in Finland

Jérémy Gauthier^{1§}, Mila Pajkovic^{2§}, Samuel Neuenschwander³, Lauri Kaila⁴, Sarah Schmid^{2,5}, Ludovic Orlando^{6,7},
Nadir Alvarez^{1,2*}

¹ Geneva Natural History Museum, 1208 Geneva, Switzerland

² Department of Ecology and Evolution, University of Lausanne, 1015, Lausanne, Switzerland

³ Vital-IT, Swiss Institute of Bioinformatics, University of Lausanne, 1015, Lausanne, Switzerland

⁴ Finnish Museum of Natural History, Zoology Unit, P.O.Box 17, FI-00014, University of Helsinki, Finland

⁵ Department of Computational Biology, University of Lausanne, 1015, Lausanne, Switzerland

⁶ Laboratoire AMIS CNRS UMR 5288, Faculté de Médecine de Purpan-37 allées Jules Guesde, Bâtiment A, 31000

Toulouse, France

⁷ Globe Institut, Lundbeck Foundation GeoGenetics Centre, University of Copenhagen, Øster Voldgade 5-7, 1350

Copenhagen, Denmark

§ These authors contributed equally to this work and are considered as joint first authors.

SUPPORTING INFORMATION

Supplementary Table 1

Sample	Year	Concentration (ng/ul) in Soul	Concentration of final individual libraries (ng/ul)	Fold change	# clean reads (R1+R2)	% mapped reads after cleaning	# filtered SNP
E1_1	1941	2,42	9,06	3,74	6274098	6,59	2708
E1_10	1955		20,40		4912742	7,81	2624
E1_11	1955				4867924	7,91	2682
E1_12	1955				5600550	6,98	2713
E1_13	1955		16,90		4995844	6,97	2603
E1_14	1955		13,30		3845782	8,17	2694
E1_15	1955		20,00		6019518	7,56	2652
E1_16	1955		15,50		4524970	7,87	2647
E1_17	1955		25,80		6240728	6,69	2660
E1_18	1981	2,30	15,90	6,91	4052028	8,32	2521
E1_19	1981		22,20		6363462	6,98	2691
E1_2	1941		7,00		5958694	6,88	2673
E1_20	1981		19,20		5235898	7,67	2586
E1_23	1980		10,20		2358164	8,94	2443
E1_24	1941	5,36	11,50	2,15	2898804	9,04	2554
E1_3	1941		22,60		4985722	7,09	2528
E1_4	1941		5,85		4460490	7,37	2515
E1_7	1941		20,40		4031012	6,79	2520
E2_1	1977		14,80		2843360	7,29	2633
E2_10	1947		12,70		2273482	8,15	2451
E2_11	1947		9,84		1843272	8,08	2459
E2_12	1947		9,10		1982636	7,02	2556
E2_13	1947		9,48		1868244	7,54	2320
E2_14	1938	1,36	6,42	4,72	1085620	8,27	2369
E2_15	1938		15,00		3138372	7,29	2494
E2_16	1938		9,54		1908792	9,32	2500
E2_17	1938		10,30		2198776	8,08	2382
E2_18	1938		9,22		1622480	7,82	2218
E2_19	1938		6,80		1348634	8,12	2417
E2_2	1977		11,10		2243018	7,06	2501
E2_20	1938		7,54		1326580	7,81	2029
E2_23	1977	4,46	10,30	2,31	1917306	7,57	2418
E2_24	1947	1,59	3,84	2,42	633876	8,93	2080
E2_3	1977		8,84		1522732	8,12	2188
E2_4	1947	4,06	14,40	3,55	2808914	7,40	2537
E2_5	1947		12,70		2652342	7,51	2621
E2_6	1947		9,98		2140374	7,53	2382
E2_7	1947		9,98		2292774	7,89	2615
E3_1	1938		6,08		3312352	9,17	2611
E3_10	1983				4043060	7,89	2518
E3_11	1953	1,34			1095756	9,08	2164
E3_12	1953		8,34		3019946	7,29	2655
E3_13	1953				1855404	8,44	2160
E3_14	1953				1962472	8,65	2512
E3_15	1953				2764178	8,37	2444
E3_16	1953				1331806	10,01	2204
E3_17	1951				153588	8,62	1007
E3_18	1949				2465348	8,26	2458
E3_19	1949				2282456	8,36	2541
E3_2	1938				3053450	7,25	2472
E3_20	1949				2638836	8,37	2484
E3_23	1949				1807054	8,44	2365
E3_24	1951				2826202	7,51	2543
E3_25	1979	0,80	9,42	11,83	3235218	6,36	2286
E3_3	1938				3602164	8,12	2479
E3_4	1938				3656306	7,69	2444
E3_5	1938				1442824	9,49	2252
E3_6	1938				2165366	9,08	2182
E3_7	1979				2940532	7,60	2549
E4_1	1947		6,94		7153656	6,02	2708
E4_10	1909		0,62		52204	9,69	523
E4_11	1909		2,42		5022302	7,23	2582
E4_12	1909		0,96		318170	10,28	1735
E4_13	1909		1,14		63704	10,87	559
E4_14	1909		2,10		779010	9,44	2105
E4_15	1909		0,74		175078	10,60	881
E4_16	1963	0,58	2,10	3,60	1359650	8,33	2120
E4_17	1963		20,20		8718416	4,68	2683
E4_18	1963		15,50		1582634	7,31	2047
E4_19	1969		13,80		5021072	5,50	2567
E4_2	1909	1,04	2,36	2,27	1108648	8,45	1961
E4_20	1969		13,80		1464438	6,03	1758
E4_23	1969		13,50		4326408	5,12	2397
E4_24	1969		20,00		5721668	5,16	2536
E4_25	1963	3,14	8,46	2,69	4305210	5,62	2282
E4_3	1909		3,02		4643934	7,44	2389
E4_4	1909		4,28		986004	9,13	1647
E4_5	1909		3,28		800208	9,01	1912
E4_6	1909		6,42		2503388	7,80	2122
E4_7	1909		5,98		3614076	7,29	2536
E5_1	1969				4652616	7,09	2671
E5_10	1944				5503658	7,37	2639
E5_11	1944		10,10		5214270	7,39	2686
E5_12	1937				2514184	8,06	2559
E5_13	1937				3439540	8,28	2464
E5_14	1937				4521138	7,81	2676
E5_15	1937				4308886	7,51	2548
E5_16	1937				3941046	8,39	2596
E5_17	1937				2877568	8,76	2441
E5_18	1937				956952	10,24	1923
E5_19	1937				1087554	11,13	2096
E5_2	1965				2447904	7,96	2375
E5_23	1937				2547098	8,74	2526
E5_24	1932	3,36	14,20	4,23	3604268	8,61	2446
E5_3	1965				3513424	7,58	2433
E5_4	1965				4060266	7,31	2525
E5_5	1965				4409522	7,53	2654
E5_6	1965		7,32		2892478	7,82	2437
E6_1	1937				1430436	7,49	2395
E6_10	1925				3751650	7,09	2524
E6_11	1930				4070974	6,65	2641
E6_12	1925				4793404	6,45	2690
E6_13	1930				3015066	6,97	2329
E6_14	1945				3488684	6,93	2647
E6_15	1945				58630	9,01	619
E6_16	1909	4,24			3746084	7,05	2609
E6_17	1973		15,70		1692992	9,08	2090
E6_19	1975				3694326	7,35	2625
E6_2	1937				4945946	6,97	2651
E6_20	1973				4021688	7,79	2504
E6_23	1970	3,18			1528378	9,17	2268
E6_24	1959	17,60			3325858	7,60	2509
E6_3	1937		27,80		3964948	7,23	2514
E6_4	1937		17,00		2484022	7,31	2066
E6_5	1930				3766494	7,26	2647
E6_6	1930				4115920	7,28	2531
E6_7	1930		25,00		3611308	7,48	2626
EF1_1	2015				239362	12,12	1985
EF1_10	2015				1270630	8,11	2027
EF1_11	2015				1737076	7,24	2329
EF1_12	2015				1862364	7,79	2477
EF1_13	2015				1652880	6,46	1879
EF1_14	2015				3027234	6,32	2535
EF1_15	2015		4,20		2298388	6,70	2168
EF1_16	2015				397526	9,67	1913
EF1_17	2015				2128270	6,99	2326
EF1_18	2015				1550866	7,07	2083
EF1_19	2015				59824	15,90	1455
EF1_2	2015		2,64		1088524	8,66	2250
EF1_20	2015				38606	10,51	770
EF1_3	2015				5913768	6,52	2596
EF1_4	2015				1166920	8,54	1977
EF1_5	2015				2465244	7,65	2612
EF1_6	2015				63106	10,40	804
EF1_7	2015				1195596	8,17	2355
EF2_1	2015				2674570	6,19	2566
EF2_10	2015				1521018	7,19	2185
EF2_11	2015				226252	7,81	1561
EF2_12	2015				5594713	5,32	2669
EF2_13	2015				4211932	5,29	2467
EF2_14	2015				2414746	6,00	2456
EF2_15	2015				3739156	5,50	2351

Supplementary Table 1 (continued)

Sample	Year	Concentration (ng/ul) in 50ul	Concentration of final individual libraries (ng/ul)	Fold change	# clean reads (R1+R2)	% mapped reads after cleaning	# filtered SNP
EF2_16	2015				82812	12,13	1281
EF2_17	2015				3159958	5,77	2337
EF2_18	2015				3595612	4,51	2138
EF2_19	2015				2241176	5,78	2468
EF2_2	2015				898840	6,60	2069
EF2_20	2015				2489030	6,14	2362
EF2_3	2015				3757564	6,72	2490
EF2_4	2015				4092346	6,20	2519
EF2_5	2015				2576260	6,31	2542
EF2_6	2015				1408612	6,34	2018
EF2_7	2015		7,98		680116	7,35	2083
EF3_1	2015		10,40		3229342	5,96	2649
EF3_10	2015				2877954	6,08	2388
EF3_11	2015				1152450	5,89	2019
EF3_12	2015				7075988	5,28	2717
EF3_13	2015				2281572	5,92	2222
EF3_14	2015				3936410	5,85	2649
EF3_15	2015				3547252	5,41	2388
EF3_16	2015				295012	8,50	1758
EF3_17	2015				2800858	5,15	2168
EF3_18	2015				693276	6,51	1718
EF3_19	2015				4260710	5,41	2614
EF3_2	2015		9,32		3216122	6,18	2613
EF3_20	2015				2373064	5,74	2351
EF3_3	2015		13,40		5284192	5,74	2616
EF3_4	2015		9,56		2405280	5,88	2217
EF3_5	2015		15,70		5045496	5,85	2672
EF3_6	2015		10,10		3056964	5,80	2446
EF3_7	2015		7,00		2195742	6,78	2520
H1_1	1946				1841358	7,32	2140
H1_10	1940				3527668	6,44	2090
H1_11	1940				6430464	5,98	2421
H1_12	1940	1,69			4951236	5,11	2388
H1_13	1940				7472816	5,05	2214
H1_14	1940				10879786	5,60	2527
H1_15	1940				11252344	5,17	2420
H1_16	1940				10018622	6,12	2487
H1_17	1937	0,11			8408572	5,29	2292
H1_18	1937	0,16			3984438	5,31	1900
H1_19	1937	0,16			7256516	5,67	2333
H1_2	1946				58542	15,43	1164
H1_20	1937	0,18			3279158	5,67	1924
H1_23	1937	0,67			16222420	10,62	2527
H1_24	1937				5822132	6,63	2075
H1_3	1946				516538	9,09	1080
H1_5	1946				1213738	8,40	1810
H1_6	1946				3433392	6,79	2128
H1_7	1940				1613156	7,36	2132
H2_1	1937				2999432	2,59	1963
H2_10	1940				173216	4,87	729
H2_11	1940				1156346	3,39	1538
H2_12	1940				3202992	2,69	2066
H2_13	1940	2,32			2772200	2,37	1433
H2_14	1940				3830620	2,70	2158
H2_15	1980				1127098	2,83	961
H2_16	1980				4434852	2,31	1780
H2_17	1980				4960868	2,11	1654
H2_18	1980				5202584	1,99	1657
H2_19	1980				6177164	2,05	2150
H2_2	1937				31646	14,72	891
H2_20	1980				4900596	2,17	1785
H2_23	1960				9277828	9,49	2469
H2_24	1960				2430908	2,91	1250
H2_3	1937	0,12			96960	5,23	403
H2_4	1937				1126308	3,31	1083
H2_5	1937				61370	8,82	858
H2_6	1937				1297016	3,09	1130
H2_7	1937				1680592	3,31	1777
H3_1	1959	0,72			4215206	5,97	2389
H3_10	1991				5078962	6,23	2458
H3_11	1985				3728212	6,85	2393
H3_12	1991				4577536	6,25	2496
H3_13	1991				2165706	7,63	2015
H3_14	1985				4519544	7,06	2489
H3_15	1985				4640896	6,76	2362
H3_16	1985				47564	14,02	887
H3_17	1985				4289172	6,13	2344
H3_18	1975				4151678	7,10	2321
H3_19	1991				2214108	6,02	2161
H3_2	1959				3365096	6,20	2355
H3_20	1985				3834762	6,58	2397
H3_23	1975	0,56			11479794	11,87	2506
H3_24	1985				1595036	6,86	1799
H3_25	1959				1226276	3,49	1699
H3_3	1964				3618722	6,81	2180
H3_4	1959				5376922	6,13	2420
H3_5	1960				2067214	6,98	2074
H3_6	1959				2009604	5,55	1834
H3_7	1960				4064222	6,54	2459
H4_1	1928	0,57			3548298	5,84	2346
H4_10	1955				5552482	6,40	2390
H4_11	1991				5733808	5,92	2483
H4_12	1991				7167150	5,93	2520
H4_13	1991				4776204	5,64	2354
H4_14	1991				5244092	5,61	2504
H4_15	1991				5328548	5,99	2379
H4_16	1991				4745940	5,93	2472
H4_17	1991				4627788	6,04	2373
H4_18	1991				4769868	6,09	2416
H4_19	1991				5945332	5,92	2503
H4_2	1985				3105138	5,19	2274
H4_20	1928				2468690	5,72	1985
H4_23	1928	0,18			11653212	8,35	2525
H4_24	1928				1338108	5,57	1470
H4_25	1928				4395294	5,03	2299
H4_3	1985				4622760	6,36	2321
H4_4	1985	0,67			5517244	6,68	2402
H4_5	1985				4391746	6,34	2406
H4_6	1975				4046232	6,89	2257
H4_7	1985				6356752	6,44	2502
H5_1	1928				3298886	5,71	2352
H5_10	1942				5081914	7,78	2411
H5_11	1938				4703488	7,37	2442
H5_12	1942				3043574	5,46	2407
H5_13	1942				6538038	7,39	2459
H5_14	1942				6684662	6,63	2523
H5_15	1942				5420298	7,92	2413
H5_16	1942				5952828	7,74	2483
H5_17	1942	0,47			4179616	6,47	2320
H5_18	1942				1411660	5,55	1650
H5_19	1938				2480920	7,20	2310
H5_2	1928				1653414	5,57	2014
H5_20	1938				1116284	6,59	1773
H5_23	1942				18271766	9,60	2544
H5_24	1941				3591724	7,02	2259
H5_25	1941				7769956	5,16	2493
H5_3	1928				3291736	5,74	2076
H5_4	1928				4695248	5,93	2362
H5_5	1928				5711234	5,98	2444
H5_6	1962				7341442	7,01	2500
H5_7	1962				4597664	7,03	2504
H6_1	1941				2422322	6,45	2161
H6_10	1941				3678322	5,96	2129
H6_11	1941				4835074	6,31	2340
H6_12	1941				2923382	5,57	2269
H6_13	1941				2946738	5,57	1779
H6_14	1934				5013760	7,10	2509
H6_15	1935				2432694	5,75	1711
H6_16	1935				2305636	6,47	2197
H6_17	1935				4257362	7,87	2333
H6_18	1935				3861956	7,52	2280
H6_19	1935				2455440	6,03	2269
H6_2	1941				4958876	5,73	2397
H6_20	1935				3754426	6,32	2290
H6_23	1949				15888874	10,79	2539

Supplementary Table 1 (continued)

Sample	Year	Concentration (ng/μl) in Soul	Concentration of final individual libraries (ng/μl)	Fold change	# clean reads (R1+R2)	% mapped reads after cleaning	# filtered SNP
H6_24	1949				2432588	6,97	1945
H6_25	1949				2458392	5,69	2291
H6_3	1941	0,57			5215798	6,52	2298
H6_4	1941				3302266	6,08	2085
H6_5	1941				4806846	5,52	2375
H6_6	1941				3440968	5,76	2165
H6_7	1941				2268384	5,91	2153
H7_1	1949				2317186	6,75	2281
H7_10	1947				3712668	6,22	2214
H7_11	1947				3212386	6,44	2304
H7_12	1947				4285912	6,41	2473
H7_13	1947				3319194	6,63	2187
H7_14	1947	0,98			3522578	6,37	2432
H7_15	1947				2914808	6,57	2017
H7_16	1947				5124778	5,84	2422
H7_17	1947				3146916	6,75	2109
H7_18	1947				3667748	6,94	2187
H7_19	1947				3516984	6,57	2404
H7_2	1949				2566484	6,12	2245
H7_20	1947				2620428	6,36	2162
H7_23	1947				15278012	11,05	2529
H7_24	1941				690894	6,99	1213
H7_25	1941				541886	8,31	1581
H7_3	1949				4162812	5,83	2259
H7_4	1949				4348734	6,75	2296
H7_5	1949				2620972	7,37	2148
H7_6	1949				2578934	6,70	2058
H7_7	1949				2863004	6,98	2345
H8_1	1941				713192	6,01	1730
H8_10	1941				3824422	6,66	2266
H8_11	1941				3818800	6,70	2339
H8_12	1941				3643828	6,25	2427
H8_13	1941				2970156	6,63	1973
H8_14	1960				4685686	6,89	2499
H8_15	1958				6608828	6,09	2436
H8_16	1958				1559348	8,68	2110
H8_17	1958				5784810	6,57	2413
H8_18	1958				6324854	6,75	2423
H8_19	1960				2900834	6,94	2326
H8_2	1941				3886454	5,80	2345
H8_20	1958				3762006	6,85	2284
H8_23	1957				16152434	10,99	2532
H8_24	1960				1491276	6,04	1450
H8_25	1960				1957834	6,33	2074
H8_3	1941				4273384	6,90	2280
H8_4	1941				4453366	6,75	2325
H8_5	1941				3598844	7,01	2288
H8_6	1941				3726878	6,09	2152
H8_7	1941				3620186	6,45	2437
HF1_1	2015	0,77			468886	8,21	1770
HF1_10	2015				5898566	3,98	2185
HF1_11	2015				68276	18,75	1167
HF1_12	2015				8475212	4,16	2502
HF1_13	2015				5826816	4,47	2302
HF1_14	2015				3273376	5,74	2389
HF1_15	2015	7,82			6774896	4,27	2291
HF1_16	2015				3104514	6,99	2307
HF1_17	2015				9943940	4,07	2459
HF1_18	2015				48134	9,13	703
HF1_2	2015				4103496	4,81	2317
HF1_3	2015				3295688	3,40	1866
HF1_4	2015				14383772	4,26	2488
HF1_5	2015				1089206	5,95	1928
HF1_6	2015				9825776	4,94	2447
HF1_7	2015				3998344	6,37	2433
HF2_1	2015				364834	8,40	1708
HF2_10	2015				1407532	5,79	1698
HF2_11	2015				638522	6,83	1717
HF2_12	2015				2496028	5,18	2291
HF2_13	2015				1195648	6,13	1655
HF2_14	2015	8,06			3146516	5,54	2405
HF2_15	2015				1714208	5,22	1805
HF2_16	2015				3154778	6,35	2339
HF2_17	2015				2372966	5,50	2069
HF2_18	2015				2735676	5,88	2176
HF2_2	2015				1376192	5,90	2026
HF2_3	2015				5349778	5,81	2402
HF2_4	2015				2454848	6,26	2151
HF2_5	2015				2491360	5,67	2279
HF2_6	2015				1754070	5,69	1969
HF2_7	2015				538050	7,67	1836

Supplementary Table 2

Sample	# reads	bwaaln %mapping	bwaaln %mapping_after cleaning	bwamem %mapping	bwamem %mapping_after cleaning	bowtie2 %mapping	bowtie2 %mapping_after cleaning
E1_1	6274098	10.11	6.59	37.37	9.60	17.54	11.01
E1_10	4912742	11.59	7.81	37.60	10.58	19.12	12.63
E1_11	4867924	11.41	7.91	36.63	10.64	18.57	12.61
E1_12	5600550	10.45	6.98	37.40	10.27	17.46	11.04
E1_13	4995844	10.29	6.97	37.68	10.73	17.73	11.60
E1_14	3845782	11.65	8.17	37.90	12.05	19.19	13.23
E1_15	6019518	11.11	7.56	38.45	10.06	18.61	12.12
E1_16	4524970	11.24	7.87	38.00	11.31	18.64	12.57
E1_17	6240728	10.29	6.69	37.54	9.70	17.49	10.82
E1_18	4052028	11.73	8.32	37.48	11.50	19.15	13.38
E1_19	6363462	10.57	6.98	37.64	9.67	18.17	11.53
E1_2	5958694	10.65	6.88	38.84	9.81	18.36	11.42
E1_20	5235898	11.26	7.67	37.86	10.40	18.83	12.52
E1_23	2358164	11.85	8.94	36.18	13.91	19.22	14.58
E1_24	2898804	12.30	9.04	34.76	12.25	19.20	14.07
E1_3	4985722	11.51	7.09	41.14	10.60	19.93	11.94
E1_4	4460490	10.64	7.37	37.70	11.14	18.10	12.29
E1_7	4031012	10.91	6.79	36.81	10.72	18.37	11.39
E2_1	2843360	9.53	7.29	33.92	12.98	16.14	12.52
E2_10	2273482	10.56	8.15	35.87	14.72	17.07	13.04
E2_11	1843272	10.19	8.08	33.37	15.24	16.23	12.94
E2_12	1982636	9.37	7.02	34.42	15.09	15.44	11.39
E2_13	1868244	9.56	7.54	33.82	15.32	15.78	12.42
E2_14	1085620	10.56	8.27	34.50	18.15	17.06	13.70
E2_15	3138372	9.77	7.29	36.82	13.08	16.80	12.38
E2_16	1908792	11.95	9.32	35.81	15.33	18.84	15.02
E2_17	2198776	10.33	8.08	35.66	14.53	17.49	13.95
E2_18	1622480	9.81	7.82	35.88	16.39	16.74	13.46
E2_19	1348634	10.19	8.12	36.73	17.71	17.44	14.07
E2_2	2243018	8.98	7.06	33.21	14.21	15.32	12.10
E2_20	1326580	9.92	7.81	36.18	17.48	17.00	13.62
E2_23	1917306	9.68	7.57	33.16	14.89	15.95	12.54
E2_24	633876	11.02	8.93	34.95	21.04	17.80	14.80
E2_3	1522732	10.12	8.12	36.77	17.32	17.27	14.00
E2_4	2808914	9.65	7.40	36.25	13.85	16.30	12.15
E2_5	2652342	9.75	7.51	35.86	14.07	16.49	12.46
E2_6	2140374	9.53	7.53	36.02	15.33	16.30	12.79
E2_7	2292774	10.18	7.89	36.76	15.29	16.63	12.38
E3_1	3312352	13.38	9.17	42.14	13.01	21.62	14.39
E3_10	4043060	11.22	7.89	37.52	11.85	18.54	12.82
E3_11	1095756	11.99	9.08	40.88	19.67	20.29	15.66
E3_12	3019946	10.26	7.29	36.72	13.07	17.19	11.89
E3_13	1855404	11.33	8.44	38.63	15.83	19.10	14.39
E3_14	1962472	11.63	8.65	38.51	15.79	19.24	14.44
E3_15	2764178	11.48	8.37	38.78	13.80	19.40	14.08
E3_16	1331806	12.88	10.01	40.62	18.71	21.26	16.81
E3_17	153588	10.02	8.62	30.32	23.71	16.07	13.90
E3_18	2465348	11.03	8.26	37.09	14.16	18.42	13.79
E3_19	2282456	11.22	8.36	35.85	14.34	18.05	13.35
E3_2	3053450	10.32	7.25	31.88	11.36	16.30	11.49
E3_20	2638836	11.20	8.37	35.93	13.59	18.14	13.51
E3_23	1807054	11.03	8.44	36.69	15.57	18.50	14.21
E3_24	2826202	10.64	7.51	38.11	13.49	18.35	12.91
E3_25	3235218	12.22	6.36	38.07	11.32	19.68	10.27
E3_3	3602164	11.69	8.12	40.38	12.45	19.61	13.29
E3_4	3656306	10.98	7.69	34.49	11.22	17.73	12.40
E3_5	1442824	12.63	9.49	39.62	17.31	20.40	15.49
E3_6	2165366	12.11	9.08	39.39	15.08	20.13	15.11
E3_7	2940532	10.96	7.60	38.47	13.46	18.67	12.87
F4_1	7153656	9.92	6.02	36.72	8.63	17.03	9.68
E4_10	52204	10.90	9.69	32.18	26.38	16.99	14.59
E4_11	5022302	11.98	7.23	36.38	9.33	19.02	11.15
E4_12	318170	15.44	10.28	36.64	20.92	22.25	14.15
E4_13	63704	12.60	10.87	29.43	23.50	17.68	15.18
E4_14	779010	12.88	9.44	33.71	17.68	19.07	13.97
E4_15	175078	12.93	10.60	35.11	24.70	19.73	16.30
E4_16	1359650	11.39	8.33	34.20	15.80	17.70	12.97
E4_17	8718416	9.84	4.68	37.02	7.66	16.69	7.37
E4_18	1582634	10.70	7.31	35.03	15.04	17.10	11.57
E4_19	5021072	10.35	5.50	33.55	8.82	16.65	8.61
E4_2	1108648	11.27	8.45	32.95	15.94	17.63	13.46
E4_20	1464438	9.76	6.03	34.25	13.99	16.33	10.09
E4_23	4326408	10.15	5.12	33.04	8.81	16.31	7.96
E4_24	5721668	10.21	5.16	34.08	8.15	16.82	8.23
E4_25	4305210	10.53	5.62	33.39	9.32	16.64	8.67
E4_3	4643934	11.55	7.44	39.40	10.36	19.55	12.24
E4_4	986004	11.99	9.13	38.35	18.55	19.57	15.09
E4_5	800208	12.29	9.01	37.62	18.95	19.67	14.52
E4_6	2503388	11.34	7.60	38.98	13.18	19.26	12.80
E4_7	3614076	11.18	7.29	39.05	11.59	19.16	12.25
E5_1	4652616	10.39	7.09	35.55	10.56	17.41	11.63
E5_10	5503658	11.15	7.37	38.38	10.20	18.82	12.04
E5_11	5214270	10.98	7.39	37.39	10.34	18.45	12.10
E5_12	2514184	11.52	8.06	39.13	14.02	19.41	13.52
E5_13	3439540	11.98	8.28	39.83	12.43	20.24	13.88
E5_14	4522138	11.59	7.81	38.18	10.97	19.38	12.82
E5_15	4308886	11.57	7.51	40.21	11.35	19.70	12.47
E5_16	3941046	12.03	8.39	38.57	11.67	19.98	13.79
E5_17	2877568	11.76	8.76	36.55	8.76	19.19	14.35
E5_18	956952	12.64	10.24	37.47	13.03	20.08	16.61
E5_19	1087554	13.58	11.13	35.54	17.91	20.55	17.23
E5_2	2447904	10.66	7.96	31.66	12.57	16.78	12.79
E5_23	2547098	12.06	8.74	38.64	13.86	20.30	14.77
E5_24	3604268	12.01	8.61	35.06	11.39	18.96	13.51
E5_3	3513424	10.40	7.58	34.33	11.73	17.08	12.43
E5_4	4060266	10.43	7.31	34.71	11.04	17.34	12.08
E5_5	4409522	11.03	7.53	36.23	10.84	18.24	12.26
E5_6	2892478	10.66	7.82	36.28	13.27	17.55	12.64
E6_1	1430436	10.24	7.49	34.88	15.79	16.95	12.73
E6_10	3751650	9.67	7.09	34.91	11.44	16.62	12.18
E6_11	4070974	9.38	6.65	35.22	10.90	16.31	11.40
E6_12	4793404	9.23	6.45	34.69	10.26	16.01	11.03
E6_13	3015066	9.39	6.97	34.45	11.99	16.37	12.24
E6_14	3488684	9.45	6.93	34.22	11.70	16.12	11.79
E6_15	58630	9.33	9.01	34.12	31.26	15.98	15.52
E6_16	3746084	9.59	7.05	33.67	11.03	16.39	12.15
E6_17	1692992	11.10	9.08	31.69	14.50	17.32	14.68
E6_18	0	na	na	na	na	na	na
E6_19	3694326	9.71	7.35	30.50	10.67	15.79	12.08
E6_2	4945946	9.83	6.97	35.47	10.17	17.01	12.04
E6_20	4021688	10.46	7.79	33.64	11.05	17.26	13.00
E6_23	1528378	11.46	9.17	28.76	13.88	16.85	14.07
E6_24	3325858	10.26	7.60	33.80	11.71	16.68	12.15
E6_3	3964948	9.92	7.23	37.63	11.75	17.45	12.58
E6_4	2484022	10.13	7.31	35.88	12.87	17.35	12.80
E6_5	3766494	9.99	7.26	36.92	11.72	17.39	12.62
E6_6	4115920	10.01	7.28	35.95	11.12	17.18	12.41
E6_7	3611308	10.25	7.48	36.24	11.83	17.72	13.12
EF1_1	239362	19.20	12.12	46.85	25.75	27.44	16.96
EF1_10	1270630	11.72	8.11	44.24	18.18	18.25	11.17
EF1_11	1733706	10.95	7.24	42.51	16.03	17.39	10.18
EF1_12	1862364	11.70	7.79	41.13	15.94	17.92	10.81
EF1_13	1652880	10.48	6.46	42.64	15.50	16.94	9.34
EF1_14	3027234	11.31	6.32	43.64	12.32	18.07	8.98

Supplementary Table 2 (continued)

Sample	# reads	bwaaln.%mapping	bwaaln.%mapping_after_cleaning	bwamem.%mapping	bwamem.%mapping_after_cleaning	bowtie2.%mapping	bowtie2.%mapping_after_cleaning
EF1_15	2298388	9,98	6,70	40,73	14,47	16,22	9,74
EF1_16	397526	12,43	9,67	42,99	26,50	19,10	14,17
EF1_17	2128270	10,24	6,99	41,08	15,20	16,20	9,62
EF1_18	1550866	10,86	7,07	43,67	16,58	18,21	11,14
EF1_19	59824	24,90	15,90	49,93	33,03	32,10	20,02
EF1_2	1088524	11,75	8,66	42,23	19,30	18,80	12,92
EF1_20	38606	13,78	10,51	39,65	29,18	20,84	14,98
EF1_3	5913768	10,37	6,52	39,86	9,89	15,91	8,27
EF1_4	1166920	11,66	8,54	41,59	18,32	18,35	12,48
EF1_5	2465244	11,57	7,65	42,86	14,52	17,69	9,94
EF1_6	63106	11,77	10,40	40,10	33,21	18,77	16,21
EF1_7	1195596	11,47	8,17	42,20	19,02	17,32	10,74
EF2_1	2674570	9,22	6,19	33,40	12,35	14,82	9,43
EF2_10	1521018	9,68	7,19	36,17	16,18	16,05	11,52
EF2_11	226252	10,06	7,81	34,47	24,02	15,81	12,20
EF2_12	5594718	9,39	5,32	35,51	8,97	15,25	7,82
EF2_13	4211932	8,70	5,29	35,20	10,30	14,35	7,87
EF2_14	2414746	9,56	6,00	35,33	12,44	16,00	9,92
EF2_15	3739156	9,10	5,50	34,90	10,48	14,84	8,28
EF2_16	82812	14,14	12,13	37,39	30,04	20,62	17,84
EF2_17	3159958	9,65	5,77	36,45	11,44	15,61	8,68
EF2_18	3595612	8,74	4,51	35,75	9,77	14,83	7,28
EF2_19	2241176	10,15	6,78	36,37	13,41	16,70	10,80
EF2_2	898840	8,73	6,60	32,81	17,54	14,34	10,54
EF2_20	2489030	9,68	6,14	36,07	12,43	16,28	10,02
EF2_3	3757564	9,29	6,72	33,63	11,50	15,14	10,40
EF2_4	4092346	9,08	6,20	34,58	11,08	15,04	9,75
EF2_5	2676260	9,61	6,31	36,02	12,68	15,62	9,60
EF2_6	1408612	9,14	6,34	35,85	15,40	15,47	10,47
EF2_7	680116	9,77	7,35	35,04	19,49	16,05	12,11
EF3_1	3229342	8,81	5,96	33,77	11,95	14,27	8,96
EF3_10	2877954	9,66	6,08	36,93	12,37	15,97	9,39
EF3_11	1152450	8,71	5,89	33,47	15,45	14,45	9,62
EF3_12	7075988	9,16	5,28	35,63	8,73	14,69	7,27
EF3_13	2281572	8,76	5,92	34,52	13,19	14,72	9,51
EF3_14	3936410	9,38	5,85	35,21	10,75	15,66	9,27
EF3_15	3547252	8,58	5,41	33,40	10,80	14,25	8,58
EF3_16	295012	10,40	8,50	36,69	25,10	16,28	12,96
EF3_17	2800858	8,28	5,15	33,43	11,44	14,06	8,40
EF3_18	693276	8,53	6,51	33,25	17,98	14,68	11,13
EF3_19	4260710	8,81	5,41	34,40	10,31	14,35	7,99
EF3_2	3216122	9,01	6,18	34,47	12,16	14,58	9,18
EF3_20	2373064	8,79	5,74	34,21	12,65	14,74	9,20
EF3_3	5284192	9,43	5,74	36,83	10,04	15,48	8,43
EF3_4	2405280	8,77	5,88	34,34	13,18	14,48	9,17
EF3_5	5045496	9,62	5,85	36,65	10,17	15,66	8,61
EF3_6	3056964	9,00	5,80	35,72	12,24	15,10	9,10
EF3_7	2195742	9,81	6,78	36,35	14,09	16,07	10,58
H1_1	1841358	10,94	7,32	34,60	11,55	17,88	11,85
H1_10	3527668	10,08	6,44	33,53	8,74	16,84	10,57
H1_11	6430464	10,57	5,98	36,75	7,19	18,30	9,78
H1_12	4951236	8,54	5,11	28,60	6,89	14,32	8,23
H1_13	7472816	9,46	5,05	34,17	6,16	16,48	8,25
H1_14	10879786	10,05	5,60	36,15	6,08	17,70	9,00
H1_15	11252344	9,94	5,17	34,69	5,39	16,87	8,05
H1_16	10018622	10,88	6,12	39,13	6,50	19,47	9,90
H1_17	8406572	9,20	5,29	30,72	5,70	15,10	8,18
H1_18	3984438	8,19	5,31	26,99	7,11	13,40	8,45
H1_19	7256516	10,92	5,67	37,37	6,42	19,25	9,18
H1_2	58542	24,19	15,43	48,33	26,61	33,17	19,15
H1_20	3279158	9,82	5,67	35,25	8,52	17,43	9,77
H1_23	16222420	50,30	10,62	73,80	2,82	63,26	3,66
H1_24	5822132	12,55	6,63	39,69	7,16	21,16	10,42
H1_3	516538	11,19	9,09	31,88	15,83	17,62	14,53
H1_5	1213738	11,23	8,40	33,74	13,17	18,21	13,57
H1_6	3433392	10,25	6,79	34,86	9,32	17,61	11,38
H1_7	1613156	10,39	7,36	31,88	11,75	16,82	11,83
H2_1	2999432	3,87	2,59	20,45	7,25	7,53	5,02
H2_10	173216	5,55	4,87	22,94	17,00	9,99	8,86
H2_11	1156346	4,71	3,39	22,55	10,58	9,11	6,57
H2_12	3202992	4,10	2,69	19,65	6,87	7,46	4,82
H2_13	2772200	3,31	2,37	18,54	7,15	6,53	4,67
H2_14	3830620	4,03	2,70	19,98	6,61	7,64	5,05
H2_15	1127098	3,61	2,83	18,31	9,35	6,95	5,57
H2_16	4434852	3,28	2,31	17,66	5,75	6,33	4,41
H2_17	4960868	3,05	2,11	17,66	5,44	6,12	4,19
H2_18	5202584	2,88	1,99	17,17	5,28	5,84	4,00
H2_19	6177164	3,06	2,05	17,68	5,09	6,02	3,90
H2_2	31646	20,23	14,72	39,04	24,75	26,99	17,76
H2_20	4900596	3,13	2,17	17,75	5,54	6,24	4,30
H2_23	9277828	41,15	9,49	61,78	3,03	52,26	2,99
H2_24	2430908	3,93	2,91	18,40	7,26	7,32	5,50
H2_3	96960	5,79	5,23	23,34	18,67	10,41	9,45
H2_4	1126308	4,55	3,31	23,68	10,77	9,21	6,86
H2_5	61370	11,18	8,82	30,60	22,84	17,28	13,23
H2_6	1297016	4,21	3,09	22,10	9,96	8,54	6,45
H2_7	1680592	4,68	3,31	21,48	9,02	8,70	6,15
H3_1	4215206	9,10	5,97	33,15	9,06	15,90	10,26
H3_10	5078962	9,51	6,23	37,71	9,60	16,88	10,55
H3_11	3728212	10,35	6,85	37,12	10,24	18,01	11,72
H3_12	4577536	10,00	6,25	37,01	9,26	17,67	10,71
H3_13	2163706	11,30	6,25	37,12	11,86	18,83	12,84
H3_14	4516954	11,22	7,06	38,99	9,60	19,56	11,87
H3_15	4640896	10,22	6,76	38,32	9,68	18,48	11,88
H3_16	47564	18,19	14,02	38,95	25,67	25,56	18,18
H3_17	4289172	9,29	6,13	35,56	9,44	16,61	10,77
H3_18	4151678	10,72	7,10	37,69	9,76	18,73	12,13
H3_19	2214108	9,31	6,02	32,97	10,48	15,62	9,94
H3_2	3363096	9,31	6,20	33,65	9,87	16,09	10,71
H3_20	3834762	9,61	6,58	35,48	10,14	16,93	11,44
H3_23	11479794	58,25	11,87	81,21	2,43	72,22	2,78
H3_24	1595036	9,07	6,86	32,42	12,91	15,59	11,93
H3_25	1226276	5,24	3,49	22,84	10,62	9,04	5,92
H3_3	3618722	10,05	6,81	35,66	9,95	17,48	11,63
H3_4	5376922	9,64	6,13	38,00	9,20	17,48	10,72
H3_5	2067214	9,42	6,98	30,28	11,14	15,42	11,46
H3_6	2009604	8,06	5,55	31,51	11,04	14,09	9,68
H3_7	4064222	11,07	6,54	39,40	9,87	19,38	11,24
H4_1	3548298	8,65	5,84	31,27	9,11	15,11	10,18
H4_10	552482	9,89	6,40	37,69	8,66	18,16	11,31
H4_11	5733808	9,18	5,92	35,78	8,64	16,47	10,22
H4_12	7167150	9,15	5,93	33,76	7,76	15,97	9,95
H4_13	4776204	8,47	5,64	33,99	9,16	15,20	9,90
H4_14	5244092	9,18	5,61	34,91	8,89	15,73	9,17
H4_15	5328548	9,09	5,99	35,46	8,77	16,34	10,34
H4_16	4745940	9,05	5,93	35,16	9,37	16,06	10,16
H4_17	4627788	8,97	6,04	36,06	9,67	16,31	10,61
H4_18	4769868	9,05	6,09	35,02	9,37	15,92	10,30
H4_19	5945332	9,29	5,92	35,99	8,76	16,57	10,06
H4_2	3105138	7,60	5,19	30,45	9,56	13,72	9,40
H4_20	2468690	8,19	5,72	28,86	9,65	14,10	9,89
H4_23	11653212	55,07	8,35	75,91	2,95	66,00	3,43
H4_24	1338108	7,50	5,57	27,12	11,46	12,86	9,64
H4_25	4395294	7,95	5,03	32,93	8,55	14,67	9,00
H4_3	4622760	9,43	6,36	35,79	9,19	16,85	11,05

Supplementary Table 2 (continued)

Sample	# reads	bwaaln %mapping	bwaaln %mapping after cleaning	bwamem %mapping	bwamem %mapping after cleaning	bowtie2 %mapping	bowtie2 %mapping after cleaning
H4_4	5517244	10.59	6.68	40.23	9.15	19.29	11.67
H4_5	4391746	9.53	6.34	35.42	9.17	16.84	10.89
H4_6	4046232	9.46	6.89	31.87	9.24	16.03	11.57
H4_7	6356752	9.74	6.44	33.82	7.92	16.76	10.77
H5_1	3298886	9.02	5.71	32.83	9.72	15.39	9.74
H5_10	5081914	12.06	7.78	41.29	9.75	21.08	13.14
H5_11	4703488	12.02	7.37	41.11	9.77	20.85	12.44
H5_12	3043574	9.18	5.46	33.08	10.13	15.39	9.16
H5_13	6538038	11.77	7.39	42.46	9.00	21.14	12.60
H5_14	6684662	10.71	6.63	38.64	8.58	18.84	11.05
H5_15	5420298	12.40	7.92	43.06	9.61	22.08	13.45
H5_16	5952828	12.70	7.74	45.44	9.62	23.00	13.18
H5_17	4179616	9.50	6.47	34.82	9.64	16.84	11.37
H5_18	1411660	7.65	5.55	28.90	11.92	13.37	9.84
H5_19	2480920	10.23	7.20	34.04	11.35	17.26	12.06
H5_2	1653414	8.16	5.57	30.43	11.67	14.11	9.80
H5_20	1116284	9.13	6.59	31.82	13.82	15.44	11.46
H5_23	18271766	54.33	9.60	77.04	2.84	66.52	3.75
H5_24	3591724	10.00	7.02	35.40	10.25	17.38	12.11
H5_25	7765956	8.80	5.16	35.81	7.22	16.21	8.97
H5_3	3291736	8.97	5.74	34.59	9.89	15.99	10.22
H5_4	4695248	9.53	5.93	36.53	9.10	17.00	10.38
H5_5	5711234	9.98	5.98	37.39	8.31	17.90	10.32
H5_6	7341442	11.29	7.01	42.06	8.63	20.34	11.79
H5_7	4597664	11.47	7.03	40.98	10.25	19.88	11.88
H6_1	2422322	10.01	6.45	32.99	10.15	16.45	10.48
H6_10	3678322	9.16	5.96	33.76	9.10	16.09	10.35
H6_11	4835974	10.24	6.31	35.17	8.12	17.58	10.43
H6_12	2923382	9.44	5.57	31.94	8.84	15.76	9.04
H6_13	2946738	8.12	5.57	28.11	8.74	13.40	9.15
H6_14	5013760	11.21	7.10	38.81	9.16	19.50	11.82
H6_15	2432634	9.53	5.75	34.61	9.82	16.63	10.01
H6_16	2305636	12.09	8.47	41.51	12.80	21.18	14.62
H6_17	4257362	12.12	7.87	40.84	10.02	20.94	13.13
H6_18	3861956	11.04	7.52	38.80	10.12	19.62	12.98
H6_19	2455440	8.93	6.03	32.87	10.72	15.40	10.32
H6_2	4958876	9.18	5.73	34.47	8.23	16.20	9.85
H6_20	3754426	10.64	6.32	40.35	10.28	19.27	11.46
H6_23	1588874	48.68	10.79	73.42	3.31	61.76	4.47
H6_24	2432588	9.99	6.97	35.11	11.11	17.14	11.90
H6_25	2458392	8.36	5.69	34.94	11.67	15.10	10.09
H6_3	5215798	10.37	6.52	37.77	8.51	18.32	11.01
H6_4	3308266	9.07	6.08	34.12	9.68	15.82	10.43
H6_5	4806846	9.38	5.52	37.18	8.54	17.20	9.78
H6_6	3440968	8.97	5.76	34.58	9.48	15.81	9.99
H6_7	2268384	9.61	5.91	33.37	10.21	16.29	9.93
H7_1	2317186	10.65	6.75	41.47	12.09	19.32	12.09
H7_10	3712668	9.43	6.22	39.79	10.40	17.55	11.26
H7_11	3212386	10.28	6.44	41.08	10.80	18.61	11.27
H7_12	4285912	10.76	6.41	42.10	9.84	19.30	11.12
H7_13	3319194	10.19	6.63	42.87	11.07	19.04	12.02
H7_14	3522578	10.33	6.37	40.40	10.33	18.52	11.02
H7_15	2914808	10.26	6.57	42.19	11.12	19.22	12.03
H7_16	5124778	9.31	5.84	40.18	9.04	17.58	10.49
H7_17	3146916	10.21	6.75	40.06	10.57	18.80	12.14
H7_18	3667748	11.20	6.94	44.59	10.38	21.15	12.51
H7_19	3516984	10.39	6.57	42.01	10.59	19.18	11.67
H7_2	2566484	9.54	6.12	39.02	11.40	17.51	11.22
H7_20	2620428	9.95	6.36	41.49	11.62	18.87	11.98
H7_23	15278012	55.33	11.05	79.97	2.67	69.15	3.10
H7_24	690894	9.02	6.99	35.96	17.41	16.34	13.06
H7_25	541886	11.19	8.31	39.56	19.37	19.38	14.68
H7_3	4162812	9.47	5.83	41.71	9.88	17.77	10.46
H7_4	4348734	11.16	6.75	46.25	10.18	20.87	11.94
H7_5	2620972	11.21	7.37	41.79	11.43	20.06	12.94
H7_6	2578934	10.59	6.70	43.65	11.78	19.83	12.26
H7_7	2863004	11.03	6.98	42.85	11.52	20.05	12.35
H8_1	713192	9.77	6.01	29.69	12.94	15.40	9.17
H8_10	3824422	10.27	6.66	37.38	9.66	18.09	11.45
H8_11	3818800	10.49	6.70	35.21	9.15	17.97	11.14
H8_12	3643828	10.28	6.25	34.60	9.09	17.27	10.20
H8_13	2970156	9.97	6.63	34.53	9.98	17.08	11.23
H8_14	4685686	11.32	6.89	40.08	9.35	19.76	11.43
H8_15	6608828	9.99	6.09	38.64	8.21	17.81	10.20
H8_16	1559348	12.40	8.68	44.84	15.52	22.42	15.73
H8_17	5784810	11.21	6.57	42.41	8.74	20.45	11.32
H8_18	6324854	11.29	6.75	42.77	8.55	20.72	11.52
H8_19	2900834	10.59	6.94	37.23	10.73	18.36	11.76
H8_2	3886454	9.49	5.80	35.83	9.11	16.79	10.06
H8_20	3762006	10.83	6.85	38.28	9.70	18.79	11.69
H8_23	16152434	52.48	10.99	76.55	2.82	65.87	3.73
H8_24	1491276	8.84	6.04	31.36	11.40	15.12	10.36
H8_25	1957834	9.00	6.33	32.99	11.69	15.82	11.07
H8_3	4273384	10.95	6.90	39.25	9.37	19.32	11.69
H8_4	4453366	10.97	6.75	39.09	9.19	19.09	11.32
H8_5	3598844	10.73	7.01	37.04	9.76	18.50	11.71
H8_6	3726878	9.68	6.09	37.37	9.56	17.72	10.87
H8_7	3620186	10.52	6.45	37.46	9.74	18.21	10.91
HF1_1	468986	17.15	8.21	44.20	17.05	24.68	11.06
HF1_10	5898566	10.07	3.98	42.58	7.58	17.76	6.70
HF1_11	68276	39.02	18.75	68.09	53.89	19.91	16.91
HF1_12	8475212	9.61	4.16	39.02	6.47	16.75	6.64
HF1_13	5836816	9.35	4.47	39.70	7.86	16.10	7.15
HF1_14	3273376	10.56	5.74	38.23	9.91	17.31	8.76
HF1_15	6774896	9.51	4.27	40.27	7.36	16.92	7.18
HF1_16	3104514	11.53	6.99	42.69	11.54	19.95	11.81
HF1_17	9943940	10.06	4.07	42.24	6.37	17.94	6.64
HF1_18	48134	13.09	9.13	40.08	25.42	20.25	13.70
HF1_2	4103496	8.98	4.81	36.56	8.70	15.47	7.97
HF1_3	3295688	8.27	3.40	36.81	8.42	14.50	5.73
HF1_4	14383772	10.50	4.26	44.34	5.94	18.76	6.86
HF1_5	1089206	9.71	5.95	37.51	14.65	16.32	9.93
HF1_6	9825776	9.87	4.94	39.93	6.49	17.39	8.07
HF1_7	3998344	11.34	6.37	42.36	10.11	19.20	10.21
HF2_1	364834	14.88	8.40	42.86	21.67	21.34	11.24
HF2_10	1407532	9.55	5.79	41.33	15.05	16.50	9.77
HF2_11	638522	11.67	6.83	40.73	18.16	18.23	10.04
HF2_12	2496028	9.90	5.18	38.90	11.37	16.26	8.16
HF2_13	1195648	9.51	6.13	40.15	15.95	15.98	10.07
HF2_14	3146516	10.14	5.54	40.95	11.08	17.06	8.77
HF2_15	1714208	9.11	5.22	39.69	13.44	15.54	8.62
HF2_16	3154778	10.66	6.35	43.33	11.67	18.35	10.47
HF2_17	2372966	9.56	5.50	40.97	12.79	16.34	9.04
HF2_18	2735676	10.26	5.88	42.68	12.34	17.53	9.56
HF2_2	1376192	10.05	5.90	38.88	14.33	16.22	9.14
HF2_3	5349778	10.18	5.81	43.08	9.67	17.88	9.49
HF2_4	2454848	10.39	6.26	43.13	13.34	17.68	10.12
HF2_5	2491360	9.69	5.67	41.10	12.63	16.36	9.09
HF2_6	1754070	8.74	5.69	38.67	14.25	15.35	9.77
HF2_7	538050	13.29	7.67	43.21	19.34	20.21	11.12
mean		11.18	6.94	37.08	12.11	18.29	11.02
sd		6.73	1.98	8.08	4.82	7.92	2.80

Supplementary Table 3

Erebia embla

	Adjusted R ²	Anova P value
Year	0.0096691685	0.002**
MEM1	0.0044102882	0.04*
MEM2	0.0005274334	
MEM3	0.0026585175	
MEM4	0.0019595386	0.002**
MEM5	0.0010806029	0.026*
MEM6	0.0007008614	
MEM7	-0.0005027576	
MEM8	0.0003807166	
all	0.01522548	0.001***

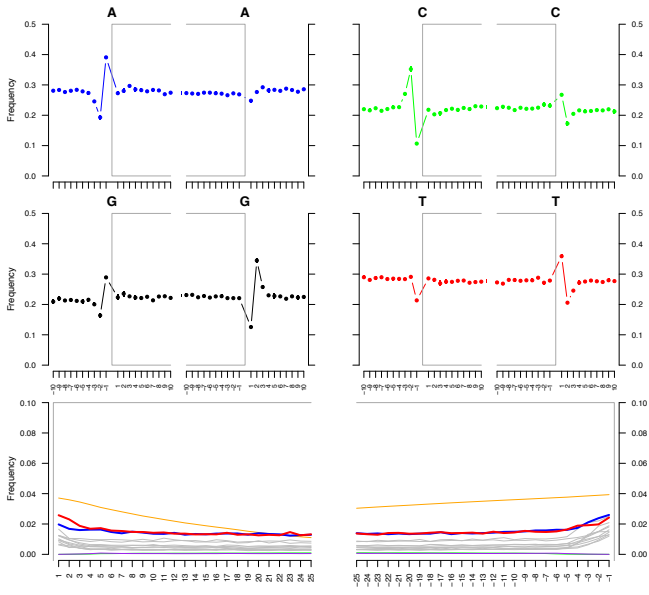
Lycaena helle

	Adjusted R ²	Anova P value
Year	6.020e ⁻⁰³	0.002**
MEM1	1.895e ⁻⁰³	
MEM2	1.387e ⁻⁰³	0.022*
MEM3	1.550e ⁻⁰³	
MEM4	2.422e ⁻⁰⁵	
MEM5	1.998e ⁻⁰⁴	
MEM6	1.365e ⁻⁰³	0.044*
MEM7	3.792e ⁻⁰⁴	
MEM8	-9.290e ⁻⁰⁴	
all	0.009318473	0.001***

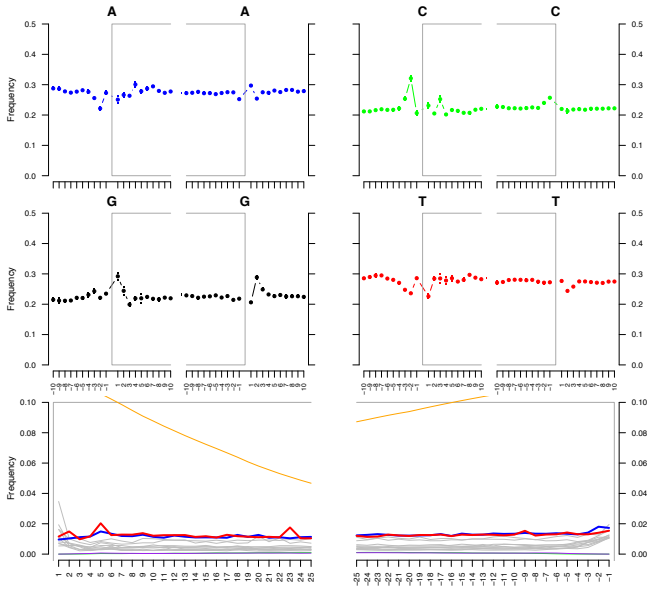
Supplementary Figure 1

Erebia embla

Historical samples

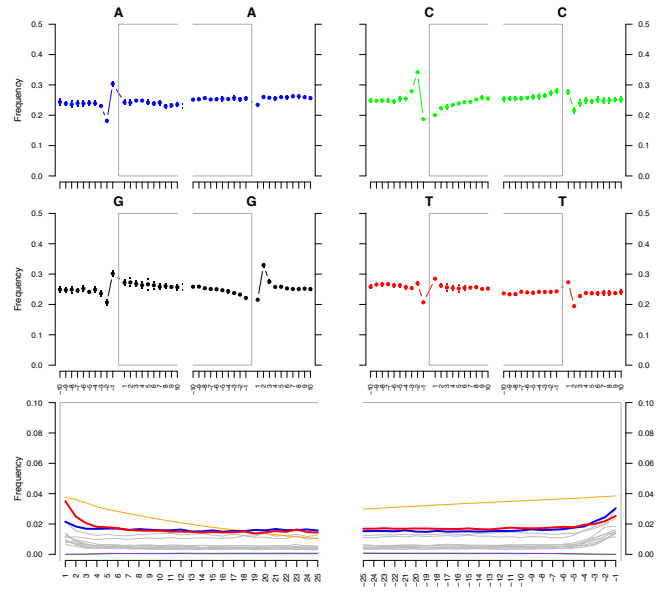


Fresh samples

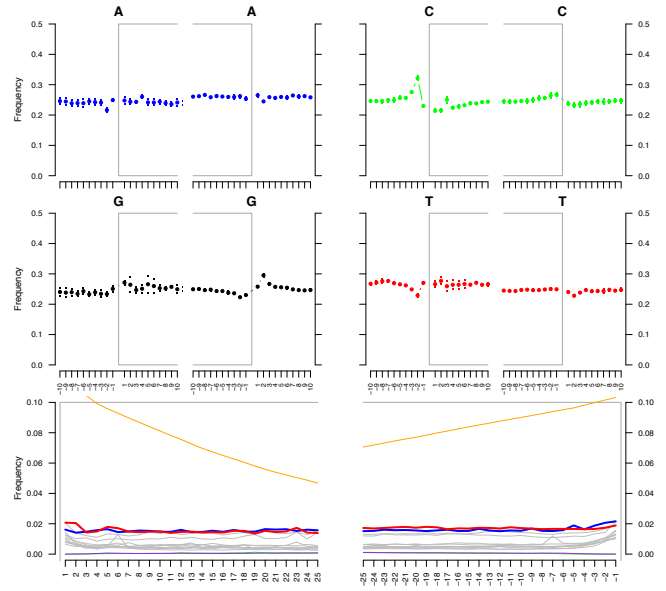


Lycaena helle

Historical samples



Fresh samples



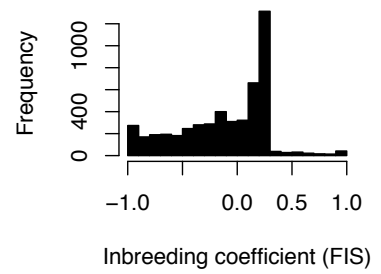
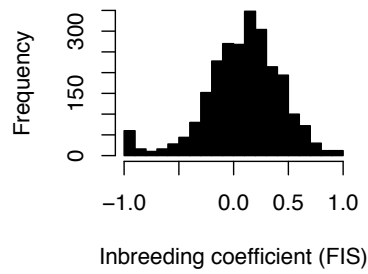
Erebia embla **Supplementary Figure 2**

freebayes

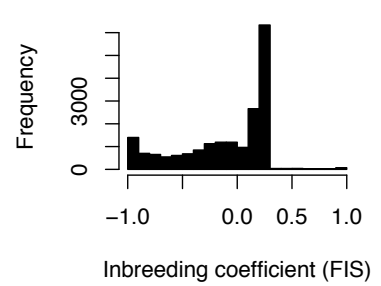
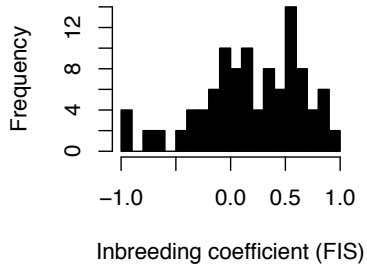
varscan

SNP calling tool
Mapping tool

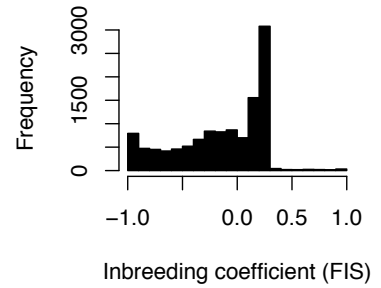
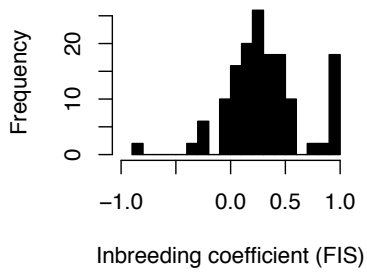
bwa aln



bwa mem



bowtie2



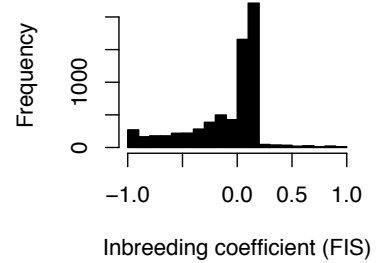
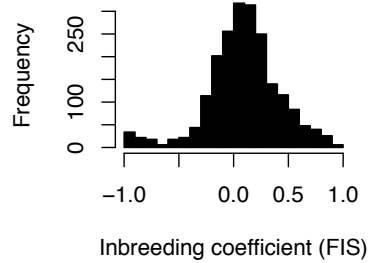
Lycaena helle

freebayes

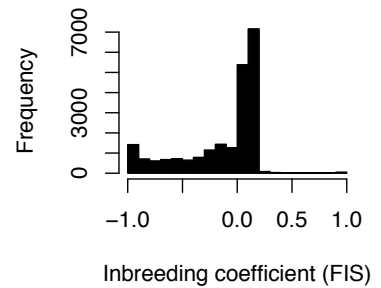
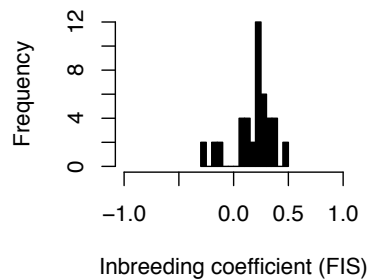
varscan

SNP calling tool
Mapping tool

bwa aln



bwa mem



bowtie2

