# Genomes from 25 historical *Drosophila melanogaster* specimens illuminate adaptive and demographic changes across more than 200 years of evolution

Short title (70 char): Museum Fly Genomes Reveal Recent Adaptive and Demographic History

Max Shpak[1], Hamid R. Ghanavi[2], Jeremy D. Lange[1^], John E. Pool[1@], Marcus C. Stensmyr[2,3@]

[1]Laboratory of Genetics, University of Wisconsin–Madison, USA

[2]Department of Biology, Lund University, Sweden

[3]Max Planck Center on Next Generation Insect Chemical Ecology

[^]Present address: Genus PLC, Waunakee, WI, USA

[@]Corresponding authors: jpool@wisc.edu, marcus.stensmyr@biol.lu.se

**ABSTRACT**

The ability to perform genomic sequencing on long-dead organisms is opening new frontiers in evolutionary research. These opportunities are especially profound in the case of museum collections, from which countless documented specimens may now be suitable for genomic analysis. Here, we report 25 newly sequenced genomes from museum specimens of the model organism *Drosophila melanogaster*, including the oldest extant specimens of this species. By comparing historical samples ranging from the early 1800s to 1933 against modern day genomes, we document evolution across thousands of generations, including time periods that encompass the species' initial occupation of northern Europe and an era of rapidly increasing human activity. At the genome-wide level, we find that historical flies from the same time and place show much greater evidence for relatedness than flies from modern collections, and some show evidence of inbreeding as well, potentially reflecting either much smaller local population sizes in the past or else the specific circumstances of the collections. We also find that the Lund, Sweden population underwent local genetic differentiation during the early 1800s to 1933 interval (potentially due to accelerated drift) but then became more similar to other European populations thereafter (potentially due to increased migration). Within each time period, our temporal sampling allows us to document compelling candidates for recent natural selection. In some cases, we gain insights regarding previously implicated selection candidates, such as *ChKov1*, for which our inferred timing of selection favors the hypothesis of antiviral resistance over insecticide resistance. Other candidates are novel, such as the circadian-related gene *Ahcy*, which yields a selection signal that rivals that of the DDT resistance gene *Cyp6g1*. These insights deepen our understanding of recent evolution in a model system, and highlight the potential of future museomic studies.

## INTRODUCTION

Museum collections around the world contain billions of specimens collected over the last centuries. These collections serve as a window back in time – to an era before industrialization and modern agricultural practices – and could as such provide insights into issues ranging from insecticide resistance to climate change. The analysis of so-called historical DNA (hDNA), *i.e.* extracted genetic material from the last centuries, from museum samples accordingly holds much promise (Card et al. 2021; Raxworthy & Smith 2021). For example, whole genome analysis of historical museum samples has been used to e.g. document temporal genomic erosion in the endangered Grauer's and mountain gorillas (van der Valk et al. 2019), and to decipher the evolutionary response of rapidly changing environmental conditions in montane chipmunks (Bi et al. 2019).

A handful of prior studies have targeted whole genome sequences from museum insect specimens. These have included proof-of-concept sequencing studies (Staats et al. 2013; Korlević et al. 2021), and investigations of the taxonomic status of museum specimens (Mikheyev et al. 2017; Grewe et al. 2021), with one study including butterfly specimens up to 150 years old (Cong et al. 2021). A few insect museomic studies have sequenced multiple individuals to examine potential targets of recent positive selection, such as focusing on the responses of honey bees to the introduction of a parasitic mite (Mikheyev et al. 2015; Parejo et al. 2020) and the emergence of the Colorado potato beetle as a crop pest (Cohen et al. 2022). In short, museomics has enabled previously unreachable research topics to be examined and brought new relevance to the myriad of specimens collected and carefully curated over the centuries in natural history collections worldwide. However, there remains much untapped potential in this area, especially in terms of sequencing larger numbers of specimens of greater antiquity, in order to extend the temporal horizon and the range of questions that museomic studies can address.

*D. melanogaster* has long been a primary model system for population genetics, including the identification of targets of selective sweeps. Recently, the availability of data from multiple time points has begun to improve the prospects for distinguishing natural selection

65 from neutral evolution in this system. Genome sequencing of wild-caught and laboratory-
66 maintained *Drosophila melanogaster* individuals has enabled the study of genomic evolution
67 across seasonal (*e.g.* Bergland et al. 2014; Machado et al. 2021) and decadal (Lange et al.
68 2022) time-scales, making the species an important model for temporal genomic study.
69 However, no previous study has investigated genome-scale variation from historical *D.*
70 *melanogaster* specimens, nor from any species of such a minute animal. We here set out to
71 sequence genomes from historical *D. melanogaster* specimens using minimally destructive
72 techniques. Our ultimate goal was to identify changes in genome-wide diversity and specific
73 genetic loci showing signs of having differentiated over time. Although not an agricultural
74 pest, nor a disease vector – and accordingly not an intentional target of control methods or a
75 casualty of altered agricultural practices – as a strict human commensal outside of its native
76 African range, the drastic changes brought upon the planet since the dawn of the
77 Anthropocene (here defined as the onset of the industrial revolution ca. 1780) should have left
78 its marks in the genome of the fly. The roughly two hundred year time-frame of our analysis
79 here should also encompass the earliest stages of this species' adaptation to a novel high
80 latitude environment (Keller 2007). Hence, museomic study of European *D. melanogaster*
81 offers the potential to reveal which genes contributed to adaptive evolution during this pivotal
82 time period. Furthermore, this interval also offers unique opportunities to investigate the
83 demographic unknowns of early versus recent northern European *D. melanogaster*
84 populations (such as the their local size and connectivity). Finally, for no other species do we
85 have as detailed understanding of the genetic organization, molecular function, and the
86 physiological and behavioral manifestations of most genes, making this fly an ideal subject for
87 temporal genomic investigations.

88 The insect collections held in Lund and Stockholm are among the oldest in the world
89 and contain many important samples, including the specimens used to erect the genus
90 *Drosophila* (Fallén, 1814-1826). The Swedish collections span the whole Anthropocene, from
91 the first steam engines to widespread industrialization and urbanization, to altered agricultural
92 practices and the introduction of mass-produced chemical insecticides, to globalization and
93 climate change (**Figure 1A**). The oldest material in the Swedish museums are four specimens

from the collection of Carl Fredrik Fallén (1764-1830; **Figure 1B**) – professor in natural history at Lund university and the authority behind genus *Drosophila*. These flies are kept in the natural history museum in Stockholm but were collected in Lund. No date is provided, nor is the collector specified, but the specimens were probably collected before 1810 (see below). From about the same time stems three specimens from the collection of Johan Wilhelm Zetterstedt (1785-1874; **Figure 1C**), which are kept in Lund. Zetterstedt was a student of Fallén, and from 1822 professor in natural history at Lund University. This material comprises the lectotype (not examined here) and two paralectotypes of *Drosophila approximata*, a junior synonym of *D. melanogaster*. The two paralectotypes we have examined are labelled Lund and Smol (*i.e.* Småland) respectively, the latter indicating the province north of Scania. No collector is specified. From the collection in Copenhagen there are two specimens from Passau, Germany (**Figure 1D**) collected by Joseph Waltl (1805-1888), physician and naturalist, and from 1835 a docent and later professor in natural history at the Philosophisch-Theologische Hochschule Passau. These flies are similarly not dated but likely collected in the 1850s. Precise collection locality is unknown. Also from Copenhagen, there is one specimen from the latter half of the 19$^{th}$ century collected by Rasmus William Traugott Schlick (1839-1916) on northeast Zealand, Denmark (**Figure 1E**). Again, no date is provided, but Schlick collected extensively in this region in the 1880s. Lastly, we have a set of 15 specimens from the collection in Lund (**Figure 1F**), collector unknown, labelled "Zootis 1933" – the informal name of the former Zoological Department at Lund University (**Figure 1G**).

**Figure 1: Historical context of the museum specimens and time interval encompassed by the present study.**

(A) Timeline of the Anthropocene showing the chronology of the analyzed fly samples, notable human events, and climate trends in Sweden (Temperature data from SMHI, Sweden).

(B) One of the four *D. melanogaster* specimens from the Fallén collection at the Swedish Natural History Museum (Stockholm). Insert shows the label (in Fallén's handwriting) accompanying the specimen. Scale bar: 1 mm. Photo: C. Fägerström.

(C) One of three *D. melanogaster* specimens from the Zetterstedt collection at the Biological Museum (Lund). Insert depicts the attached label. Scale bar: 1 mm. Photo: C. Fägerström.
(D) *D. melanogaster* specimen from the Natural History Museum of Denmark (Copenhagen), collected by Joseph Waltl in Passau, Germany. Scale bar: 1 mm. Photo: C. Fägerström.
(E) *D. melanogaster* specimen from the Natural History Museum of Denmark (Copenhagen), collected by William Schlick on NE Själland, Denmark. Scale bar: 1 mm. Photo: C. Fägerström.
(F) 15 *D. melanogaster* specimens from Lund, kept in collections of the Biological Museum (Lund). Scale bar: 5 mm. Photo: C. Fägerström.
(G) The old department of Zoology building at Lund University. Photo: Lund University.
(H) Lydbecksa Gården in Lund, where Fallén had his lodgnings until 1811. Photo: Lund University.
(I) Sandgatan 5 in Lund, where Zetterstedt lived. Photo: Lund University.
(J) A representation of the 6 contemporary population samples included in the present analysis. Photo: M. Stensmyr

The precise timing, location, and collector for the old Swedish material is not easy to decipher. The Fallén samples are accompanied by a single (contemporary) note, in what looks like Fallén's handwriting, stating (translated) "*Imago and larvae in sawdust [*unintelligible*] can of raisins*". If the flies were caught by Fallén himself, a potential locality would then be his lodgings at Lidbeckska gården in central Lund (**Figure 1H**), and that the material was collected around 1807-1809, based on the known whereabouts of Fallén (who after 1811 spent considerable time at his wife's family's mansion at Esperöd, near Simrishamn). From his personal correspondence, the first mention of *Drosophila* can be found in a letter to Zetterstedt dated 14 September 1810 (Zetterstedt archive, Lund University). The samples in Zetterstedt's collection are likewise difficult to precisely date. From Zetterstedt's *Diptera Scandinaviae disposita et descripta* we can read that Zetterstedt collected *D. approximata* in Lund (possibly at his home at Sandgatan 5 in central Lund; **Figure 1I**). The specimens from Småland were attributed to Sven Ingemar Ljungh (1757-1828), civil servant and prominent naturalist, who resided at Skärsjö manor (Jönköping County, Småland – roughly 200 km from Lund). Whether

151 or not the specimens in the collection are the ones mentioned in his treatise on Scandinavian
152 Diptera we cannot know, but if so, the flies would likely have been caught within a decade or
153 so of Fallén's samples. It should be noted that Zetterstedt kept many of Fallén's specimens in
154 his collection, and moreover, that given the considerable age of these collections, labels might
155 have been lost or details obscured over the years, and lastly, although both Fallén and
156 Zetterstedt kept type specimens, detailed labelling was evidently not considered all that
157 important.

158 All in all, the Swedish and Danish museum collections contain 26 *D. melanogaster*
159 specimens suitable for analysis. The old Swedish specimens are to the best of our knowledge
160 the earliest *D. melanogaster* available in any collection, likely predating the Meigen holotypes
161 kept in Paris by two decades or more. We here provide whole-genome analysis from these
162 flies, obtaining from most of them relatively complete and high quality genomes, and compare
163 them against modern fly genomes (**Figure 1J**). This analysis may span more than 3,000 fly
164 generations (Turelli and Hoffmann 1995; Pool 2015), analogous to studying human evolution
165 across more than 75,000 years (approximately corresponding to the time period in which
166 modern humans first colonized Eurasia from Africa, e.g. Mallick et al 2016, Jouganos et al
167 2017, Yang and Fu 2018, Wohns et al 2022). Examining changes in genomic diversity across
168 two roughly century-scale time intervals, we find that the relationship between the Lund
169 population and other European populations has changed over time. We identify a variety of
170 strong candidates for the action of positive selection in each time interval, providing temporal
171 context for previously known cases of selection, while also identifying novel putative selection
172 targets. These findings illustrate the potential of museomic studies to deepen our
173 understanding of recent evolution in a rapidly changing world.

174

175 **RESULTS**

176 **Most historical flies yielded genomes of comparable quality to contemporary flies**

177 For proof-of-principle, we first attempted to extract DNA from one of the 1933 "Zootis"
178 specimens (**Figure 1F**). Briefly, the whole fly – still attached to the pin – was first transferred to

179 a lysate buffer containing Proteinase K and left at 56°C for 2 hours, after which the fly was
180 removed, washed, dehydrated, and returned to the collection (**Figure 2A**). The extraction
181 process left the fly largely unharmed, except for a slight loss in coloration and some shrinkage
182 of the abdomen (**Figure 2B**). The extracted DNA was subsequently prepared for Illumina short
183 read sequencing. As expected, the DNA was highly fragmented with an average fragment
184 length of about ~50 bp. The Illumina run yielded ~24 million reads, which we subsequently
185 mapped onto a *D. melanogaster* reference genome. The ensuing historic genome was 91%
186 complete with a 15.7X mean sequencing depth. In short, the employed method was indeed
187 able to extract enough DNA to piece together largely complete genomes, while doing minimal
188 harm to the specimen.
189
190 **Figure 2: Sequencing of historical *Drosophila* specimens yields genomes with slightly**
191 **reduced genomic coverage and some evidence of cytosine deamination.**
192 (A) Illustration of the process of extracting and sequencing DNA from pinned insect specimens.
193 Flies were submerged in a lysate buffer for two hours, then put through three washing steps
194 and dried, before being restored to the museum repository. Genomic libraries were then
195 prepared and sequenced using the Illumina NovaSeq 6000 platform. Further details appear
196 in the Methods section.
197 (B) Images of the appearance of a representative specimen (from 1933) before and after the
198 DNA extraction protocol, illustrating minor changes in coloration. Photo: C. Fägerström.
199 (C) The fraction of genomic coverage is plotted as a function of mean per-site read depth for
200 each historical fly genome, and for a representative set of contemporary inbred line genomes
201 from Lyon, France. For a given level of depth most historical fly genomes achieve levels of
202 genomic coverage only slightly lower than contemporary genomes.
203 (D) The relationship between depth and unique (singleton) variant-calling. Only the lowest
204 quality historical fly genomes show notable elevations in the rate of singleton variants called.
205 (E) The fraction of singleton variants called as adenine or thymine is elevated for historical fly
206 genomes, particularly those with greater singleton rates, potentially due to cytosine
207 deamination.

208

209 Having verified the method, we next proceeded to process the other specimens for
210 DNA extraction and sequencing. Out of the 25 specimens examined, we were able to generate
211 genomes from all but one. While the mean depth of coverage per site varies among the
212 historical fly genomes (**Figure 2C**), the median value among these genomes was
213 approximately 20X, which is comparable to typical *D. melanogaster* genomes generated from
214 contemporary source material (Lack et al. 2016). We find, however, that with similar mean
215 depth, genomic coverage tends to be slightly lower in historical versus contemporary
216 genomes (often with only 2-3% less of the genome covered; **Figure 2C**, **Table S1**).

217 The modest reduction in genomic coverage in most of the historical fly genomes may
218 relate to their consistently shorter insert sizes – typically only about 50 bp (**Table S1**), compared
219 to contemporary genomes which are typically sequenced from fragments of a few hundred bp
220 in length. The shorter DNA fragments from the old specimens are expected to limit our ability
221 to align into repetitive genomic regions, and hence such regions may not receive base calls in
222 the historical genomes, leading to somewhat reduced genomic coverage as observed above.
223 Genomes from the 1800s flies had mean insert sizes averaging 47.6 bp, while those from the
224 1933 collection averaged 50.8 bp. Although the distributions of mean insert sizes overlap
225 between time points, they nevertheless differ significantly ($P$ = 0.0188; Mann Whitney two-
226 tailed $U$ test). There was a marginal correlation between fragment length and depth of
227 coverage (correlation $r$ = 0.40, $P$ = 0.055). A stronger and significant association was found
228 between fragment length and genomic coverage ($r$ = 0.63, $P$ = 0.0001).

229 The shorter length of historical fly DNA fragments may be due to DNA degradation,
230 which may also explain the reduced depth and genomic coverage obtained from a minority of
231 the examined specimens. Among the nine 1800s samples, four genomes were characterized
232 by a low average read depth of < 10 reads per site and/or a low genomic coverage of < 80%.
233 For the 1933 samples, one sample was immediately excluded from further analyses due to
234 extremely low coverage and read depth. Most samples with low coverage also had low mean
235 read depth (**Figure 2C**; **Table S1**), and the overall correlation between these quality metrics
236 was statistically significant ($r$ = 0.404, $P$ = 0.018). We note that all seven low-quality genomes

237 (as defined above) were from female flies (out of 14 total females), whereas all 11 males yielded
238 higher coverage/depth genomes. This sex difference is statistically significant (two-tailed
239 binomial $P = 0.0057$) and may be due to female flies being larger and potentially experiencing
240 greater decay before dehydration.

241 Regarding the six historical fly genomes retained with lower sequencing depth,
242 inferences of diploid genotype are not statistically robust at sites for which an individual has
243 few reads. Therefore, we treated the lower quality genomes as though they were haploid for
244 autosomal and female X chromosomes by assigning only a single, most frequent nucleotide at
245 each site. Consequently, there were effectively 14 rather than 18 autosome alleles for the
246 1800s samples and 28 rather than 30 for the 1933 autosomes. Similarly, treating each low-
247 quality female X as haploid reduced allele counts to 11 and 21 for 1800s and 1933, respectively
248 (from what would otherwise be X chromosome allele counts of 15 and 23).

249 The DNA degradation inferred above could also lead to incorrect base identification
250 via anomalous sample-specific variants at individual sites. Therefore, we assessed whether
251 singleton variants (unique to individual genomes) were enriched in samples with low depth
252 and coverage. Indeed, we found a higher relative incidence of singleton variants unique to
253 genomes with lower genomic coverage (two-tailed binomial $P < 0.01$; Figure 2B) and to a
254 lesser extent those with low depth as well ($P = 0.05$).

255 One of the most common types of nucleotide degradation is cytosine deamination,
256 which can lead to spurious thymine enrichment, which has been documented extensively in
257 ancient DNA samples (e.g. Dabney et al 2013). We therefore assessed the extent to which A/T
258 nucleotides are over-represented among singletons from each genome, and we found a
259 higher fraction of A/T sites in singletons among the historical samples than in modern genomes
260 (**Figure 2D, E; Table S1**). Whereas the modern genomes had singleton A/T frequencies of
261 ~55%, many of the historical genomes' singletons were 60-74% AT, presumably due to
262 cytosine deamination. Although most historical genomes do not show meaningfully elevated
263 singleton rates, our findings regarding singleton AT% justify the exclusion of singleton alleles
264 for analyses that are sensitive to rare alleles and small frequency differences, such as
265 demographic inferences based on allele frequency change over time. In contrast, analyses

266 that search for window-scale outliers for elevated allele frequency differentiation between
267 samples should be little-affected by low rates of spurious singleton variants.
268
269 **The historical fly genomes show signs of relatedness and inbreeding**
270 A preliminary analysis of pairwise genetic distances among the genomes indicated that some
271 specific pairs of individuals from within the early 1800s and 1933 Lund samples had much
272 greater similarity than expected. For example, as seen in **Table S2**, the raw (Hamming)
273 distances among X chromosomes within and among modern European populations are
274 typically 0.1-0.14. In contrast, there are 28 pairs of historical genomes with pairwise distances
275 below 0.08 (as low as 0.038 in one case). These results suggested relatedness among some
276 historical flies, which should lead to substantial tracts of Identity By Descent (IBD; **Figure 3A**).
277 Because the inclusion of related individuals in estimates of allele frequency generates
278 artifactual non-independent sampling, these redundancies should be identified and
279 eliminated from downstream population genetic analyses.
280
281 **Figure 3: Significant levels of identity-by-descent (IBD) were observed within and**
282 **between many historical fly genomes, reflecting inbreeding and relatedness.**
283 (A) Illustration of the genealogical relationships that may result in IBD due to relatedness
284 between individuals (color1) or inbreeding within individuals (color2).
285 (B) For the 1800s Sweden-Lund and Germany-Passau samples, the proportions of each
286 genome detected as IBD due to apparent inbreeding (given along the diagonal) were low or
287 zero. Above the diagonal, the proportions of each pairwise genomic comparison detected
288 as IBD due to apparent relatedness varied within the Lund sample from zero to over 75% (the
289 latter indicating a close familial relationship), while the two Passau genomes showed low-
290 level relatedness IBD.
291 (C) The same quantities are given for the Lund 1933 population sample, revealing some
292 genomes with high levels of inbreeding, and a broad spectrum of relatedness among
293 genomes.

294 Since the probability of IBD detection is different for diploid and haploid sequences, males
295 (with a single X chromsome) are given in italics. Lower quality genomes treated as haploids
296 are shown in bold. IBD regions were masked in subsequent analyses.
297
298 We identified IBD regions by partitioning each chromosome arm into windows of
299 approximately 100 kb length, and for each pair of individuals from the same population
300 sample, testing whether pairwise distance in a given window was closer to simplified
301 expectations for unrelated sequences (*i.e.* nucleotide diversity) or else IBD (*i.e.* zero for two
302 haploid sequences, or intermediate values for IBD involving diploid sequences; see Methods).
303 In line with the chromosome-wide distances cited above, we inferred many IBD windows
304 spanning various lengths, up to whole chromosome arms (**Table S2; Table S3**), while **Figure**
305 **3B and C** summarize the proportion of autosomal windows with IBD between every pair of
306 genomes in the 1800s and 1933 samples. These indicated up to sibling-level relatedness
307 among some flies in both the 1800s and 1933 Swedish samples, and a much lower level of
308 relatedness between the two specimens from 1840s Germany. The 1933 specimens, which
309 are all mounted together, were likely collected at the same time, at a guess from leftover fruits
310 on the third-floor lunch room. The six Swedish flies from the early 1800s also all show
311 relatedness, with some pairs even at sibling-level IBD. The most plausible explanation would
312 be that in spite of museum records tentatively linking them with three different scientists, these
313 specimens were all from the same Lund collection event (see Discussion).
314 In addition to IBD between genomes, we found within-genome IBD due to inbreeding
315 – several samples from 1933 had long regions of chromosomes with depleted heterozygosity
316 (**Table S2; Table S4**). Two 1933 Lund samples had especially extensive regions of near-zero
317 heterozygosity (covering 45% and 36% of their genomes; **Figure 3C**), consistent with mating
318 between close relatives, and several other historical genomes had homozygosity across 20-
319 40% of specific chromosome arms. Most of the genomes, however, showed no evidence of
320 inbreeding (**Figure 3B, C**),
321 The pairwise and within-genome IBD indicates that many of the individuals collected
322 for the historical were closely related and that (at least among the 1933 Lund flies) inbreeding

323 was prevalent. These observations contrast with the low levels of relatedness observed from
324 contemporary population samples, even when multiple fly lines founded from the same trap
325 site were analyzed (Lack et al. 2016). The elevated IBD of the historical fly genomes could have
326 resulted from sampling methodology and/or from a lower population density of flies in earlier
327 eras (see Discussion).

328 Following IBD and inbreeding masking, we were left with 9-11 autosomal alleles for the
329 combined 1800s samples on average rather than 14, and 10-14 rather than 28 for the 1933
330 samples. The average sample size of X chromosome alleles was then 8 for the 1800s set and
331 11 for 1933, rather than 12 and 20. Based on the mosaic pattern of window masking due to
332 intergenomic and intragenomic IBD across individuals on each chromosome arm, post-
333 filtering sample size varied around these averages, and downstream analyses included sample
334 size thresholds for site inclusion (see Methods).

335

336 **Inversion frequencies appear to increase with time**

337 We then sought to examine how within-population genetic diversity and between-population
338 genetic differentiation have changed across more than two centuries. We first examined
339 frequencies of known polymorphic inversions (specifically *(2L)t, (2R)Ns, (3L)P, (3R)K*, and
340 *(3R)Mo)*, since inversions are known to significantly impact genomic diversity in this species
341 (*e.g.* Corbett-Detig and Hartl 2012; Kapun et al. 2016). We used a previously reported list of
342 inversion-associated SNPs (Kapun et al. 2016), a majority of which appeared to be inversion-
343 associated in our analysis as well (see Methods and **Table S5**). Our 24 historical flies comprise
344 a total of 192 sampled autosomal arms (as well as 37 X chromosomes), yet we inferred that
345 only a single autosomal arm from the 1933 sample carried an inversion (being heterozygous
346 for *(2L)t*), whereas no inversions were detected in any 1800s genome (**Table S5**). Inversion
347 frequencies were however, also somewhat low in the modern Lund data, with only *(2L)t* (at
348 13%) giving a non-zero frequency estimate. The inferred frequency changes in *(2L)t* were
349 statistically significant when modern Lund was compared to all historical genomes (Fisher exact
350 test $P$=0.048), though not when Lund was compared to either historical sample separately
351 ($P$=0.226 vs. 1800s and $P$=0.203 vs. 1933). These results hint at increasing inversion

352 frequencies through time, but further temporal sampling will be needed to evaluate the
353 previous suggestion of a recent African origin for most inversions now present in European *D.*
354 *melanogaster* (Corbett-Detig and Hartl 2012; Pool et al. 2012).
355
356 **Genomic diversity suggests transiently elevated genetic differentiation**
357 We then focused our SNP-based analysis of genetic diversity and differentiation on the X
358 chromosome, in order to minimize the influence of inversions, which are very rare on this
359 chromosome in European populations (*e.g.* Kapun et al. 2016). Patterns of nucleotide diversity
360 ($\pi$) in Lund appeared to show a decline with time (**Figure 4A**), which could reflect the action
361 of genetic drift. However, we note that damage-induced errors may inflate diversity estimates,
362 especially for the 1800s sample (as potentially indicated by its slightly elevated $D_{xy}$ values). In
363 addition, the modern Lund sample represents a published pool-seq data set, analyzed via a
364 distinct pipeline including the masking of rare variants, which may deflate its diversity estimate.
365
366 **Figure 4: Genetic differences among historical and modern fly populations suggest**
367 **transient differentiation in the Lund population.**
368 (A) Genetic differentiation between population samples is summarized by $F_{ST}$ (below
369 diagonal, red heat map), and by the rate of pairwise sequence differences $D_{xy}$ (above
370 diagonal, blue heat map). Along the diagonal, the rate of pairwise sequence differences
371 within populations (nucleotide diversity) is given. Data reflects the X chromosome only (to
372 avoid effects of inversions), with down-sampling to four alleles per population at each site.
373 The Lund population appears to show decreasing diversity with time, subject to caveats
374 regarding historical genome quality and the differentially-processed Lund 2010s pool-seq
375 data. When Lund allele frequencies are compared to other European populations (via $F_{ST}$),
376 differentiation increases between the 19$^{th}$ and 20$^{th}$ century Lund samples, then decreases
377 again between the 20$^{th}$ and 21$^{st}$ century samples.
378 (B) Prinicple Components Analysis (PCA) results are shown for the same individually-
379 sequenced population samples included in (A), in terms of loadings from the first two principle
380 components. Here again, the Lund 1933 population shows elevated differentiation, whereas

381 genomes from the three 1800s locations cluster with modern European genomes. These
382 results reflect an analysis of X chromosome variation, with each population sample
383 downsampled to five individuals, with non-inbred historical genomes lacking close relatedness
384 chosen for this analysis.
385
386 Patterns of genetic differentiation (as indicated by $F_{ST}$) revealed a curious temporal
387 trajectory (**Figure 4A**). When Lund samples are compared to modern samples from around
388 Europe, between the 1800s and 1900s sampling points, Lund becomes more differentiated
389 from other populations. In contrast, between the 1900s and 2000s sampling points, Lund
390 became more similar to other modern European populations. Concordant patterns were
391 observed from Principal Components Analysis (PCA) based on X-linked variation (**Figure 4B**;
392 **Table S6**). Here, 1800s Lund genomes were seen to cluster with modern European genomes,
393 whereas the 1933 Lund genomes form a distinct cluster. This indication of transiently increased
394 local genetic differentiation in Lund could indicate a history in which genetic drift led to
395 increased allele frequency divergence from more southerly European populations, but
396 subsequently elevated migration led to genetic homogenization between European
397 populations (see Discussion).
398 We further investigated a simple model of the demographic history of Lund by using
399 population genetic simulations to fit single-population models of genetic drift. We varied the
400 effective population size ($N_e$) for a given time interval and compared the mean difference in
401 allele frequency at non-singleton SNPs between a given pair of time points from simulated
402 versus empirical data, using a Bayesian approach to identify the best-fitting $N_e$ for each time
403 interval, for each chromosome arm separately. In contrast to longer-term estimates of $N_e$ in
404 this species, which are often on the order of one million (e.g. Sprengelmeyer et al. 2020), our
405 point estimates of Lund $N_e$ were only 2,000-2,500 diploid individuals for the 1800s-1933 time
406 interval, and 2,400-3,200 for the 1933-2010s interval (**Table S7**). These values are of the same
407 order of magnitude as an estimate of 9,500 from a northeast United States population between
408 1975 and 2015 (Lange et al. 2022). Estimates from both studies share however, a key limitation
409 in assuming that genetic drift (along with random sampling variance) is responsible for all

410 observed frequency differences. To address how well this assumption holds, we compared $\pi$
411 from coalescent simulations based on our estimated $N_e$ values to those from the empirical data,
412 using a demographic model previously estimated for a French population (Sprengelmeyer et
413 al. 2020) as a starting point for the pre-1800s history. While this history allowed us to match
414 Lund ~1809 $\pi$ reasonably well, our low estimates of $N_e$ yielded simulated values of $\pi$ much
415 lower than those observed in our empirical data. For example, **Table S7** shows that for arm
416 3L, simulated $\pi$ for 1933 and 2010s Lund was less than half of observed $\pi$. These results
417 indicated that drift-only models did not provide an accurate approximation of the Lund
418 population's demographic history, and that true $N_e$ values were probably greater than our
419 estimates. Instead, migration may have played an important role in shifting allele frequencies
420 without decreasing $\pi$, as suggested above. Estimating a reasonable spatiotemporal
421 demographic model that incorporates both local population sizes and geographic patterns of
422 genetic structure and gene flow may require more extensive sampling of population genomic
423 data across space and time.
424
425 **Genome scans to detect recent shifts in SNP and structural variant frequencies**
426 One principal goal of this study is to identify instances of elevated genetic differentiation
427 between time points that may reflect the action of recent positive selection. Our SNP-based
428 genome scan focused on Population Branch Excess (PBE; Yassin et al. 2016), a statistic that
429 quantifies elevated allele frequency differentiation in a focal population compared to two other
430 reference/outgroup populations. PBE builds upon the $F_{ST}$-based Population Branch Statistic
431 (PBS; Yi et al. 2010) but is more tailored to detecting selection that is specific to the focal
432 population. To search for selection between the early 1800s and 1933, we defined the 1800s
433 samples as the focal population and included the Lund 1933 and Lund 2010s samples as
434 outgroups. To focus on the 1933 to 2010s interval, Lund 1933 was the focal population and
435 Lund 2010s and France were the outgroups. PBE was evaluated in diversity-scaled windows
436 averaging 4.6 kb in length. In the absence of a suitable demographic model to provide a null
437 hypothesis for the extent of genetic differentiation, we focused our analysis on the top 1% of
438 window PBE values. However, we emphasize that any outlier-based genome scan may entail

439 both false positives and false negatives. Our SNP-based PBE scans revealed notable outlier
440 peaks reflecting elevated allele frequency change for each time period (**Figure 5A, B**). Overall,
441 we identified 190 outlier regions for the 1800s-1933 time interval (referred to below as scan A)
442 and 173 for the 1933-2010s interval (scan B), with some of these regions incorporating multiple
443 outlier windows (**Table S8**). Since not all outliers may represent true positive targets of recent
444 natural selection, we focus here on loci that showed the most extreme frequency changes
445 across each century, referring to outlier regions based on their ranked maximal window PBE
446 values.
447
448 **Figure 5: Allele frequency changes across centuries reveal candidates for positive**
449 **selection.**
450 (A) Allele frequency change for ~5 kb genomic windows, focusing on the 1800s-1933 time
451 interval (scan A), as quantified by Population Branch Excess (PBE). Genes within top outlier
452 regions that are discussed in the text are indicated.
453 (B) PBE is shown for the same genomic windows, focused on the 1933-2010s time interval
454 (scan B).
455 (C) Window-level PBE is shown for a roughly 50 kb region focusing on the top outlier region
456 from scan A, with the position of the known selection target *ChKov1* indicated.
457 (D) Window-level PBE is shown for a roughly 100 kb region focusing on the top outlier region
458 from scan B, with the position of the known selection target *Cyp6g1* indicated.
459
460 We complemented this SNP-based approach with a parallel statistical methodology to
461 search for copy number variants (CNVs). We treated each population sample's mean read
462 depth within a given window (scaled relative to its chromosome arm average) as a quantitative
463 trait, and calculated $Q_{ST}$ (a quantitative trait analog of $F_{ST}$; Lande 1992; Spitze 1993) based on
464 variation in read depth within and between two population samples. We then used the
465 pairwise $Q_{ST}$ values among a trio of populations to calculate an analog of PBE (here referred to
466 as QPBE), which will be elevated when the focal population has a scaled mean window
467 sequencing depth that is much different in comparison to the two outgroup populations,

468 potentially indicating a CNV at contrasting frequency in the focal population. It is possible that
469 locus-specific heterogeneity in DNA degradation might also influence depth and thus QPBE
470 values involving historical samples. Therefore, we highlight only the top result from each time
471 interval (with further top-1% outliers given in **Table S9**), in part because both of these top
472 outliers involve loci with previously-detected structural variants. For the 1800s to 1933 period,
473 our top depth QPBE window contained only the gene *Tequila*, a serine protease involved in
474 insulin/glucose regulation, lifespan, and memory (Colomb et al. 2009; Huang et al. 2015), for
475 which polymorphic transposon insertions have created new introns (Chakraborty et al. 2018).
476 For the 1933 to 2010s interval, the top window contained only the gene *Tenascin accessory*,
477 involved in synapse organization (Hong et al. 2012; Mosca et al. 2012), for which polymorphic
478 transposons are associated with expression differences (Rech et al. 2022) and balancing
479 selection has been suggested (Croze et al. 2017). Otherwise, we refer to QPBE outliers only
480 where they intersect with the SNP-based PBE outliers discussed below.
481
482 **Top SNP-based outliers inform classic examples of positive selection**
483 In both time intervals, the top window PBE value was also associated with the widest outlier
484 region, which is consistent with relatively strong positive selection. For the 1800s-1933
485 interval, the top outlier (labelled A1) was a 31 kb region that included the gene *CHKov1* (**Figure**
486 **5C**), which is thought to encode an ecdysteroid kinase (Scanlan *et al.* 2020). This gene is known
487 to be associated with a protein-truncating transposon insertion under recent positive selection
488 in non-African *D. melanogaster* populations, which was found to correlate with resistance to an
489 organophosphate pesticide (Aminetzach *et al.* 2005). Transposon insertion and gene
490 duplication at the *CHKov1/CHKov2* locus has, however, also been found to correlate with
491 resistance to infection by the *Drosophila melanogaster* sigmavirus (DmelSV) (Magwire *et al.*
492 2011). Given the lack of widespread insecticide usage during this period, the timing of
493 selection indicated by our analysis rather favor the antiviral role of this locus as a potential
494 selective advantage. Measurements of the substitution rate in DmelSV indicate *D.*
495 *melanogaster* acquired this rhabdovirus around 200 years ago (Carpenter *et al.* 2007), which
496 predicts that this time interval should have included key stages in the evolution of resistance

497 to DmelSV. Consistent with this hypothesis, we also found the gene *refractory to Sigma P*
498 (*ref(2)p*) within outlier region A20. Ref2p activates the Toll pathway and has been shown to
499 confer resistance to DmelSV infections (Bangham *et al.* 2007).

500 *CHKov1* and *ref(2)p* are of particular interest because mutational variants associated
501 with viral resistance have been previously identified (Wayne et al. 1996; Magwire et al. 2011).
502 In *CHKov1*, several SNPs have alleles in linkage disequilibrium with a transposable element
503 insertion in the gene (Aminetzach et al. 2005) associated with viral resistance (Magwire et al.
504 2011). Based on seven such SNPs scored in our data, we inferred that the resistant haplotype
505 increased from 16.7% in the 1800s sample to 100% in the 1933 and 2015 Lund samples. The
506 occurrence of three resistance-associated haplotypes (carried in heterozygous form by two
507 early 1800s Lund flies, as well as the later Zealand sample) could indicate that selection on this
508 allele began prior to the collection of our 1800s flies. The fixation of the resistant haplotype in
509 the 1933 and 2010s Lund samples suggests a complete (or nearly complete) sweep, mirroring
510 findings from some but not all recently-collected *D. melanogaster* population samples
511 (Aminetzach et al. 2005).

512 The outlier PBE score for *ref(2)p* for the 1800s time interval appeared not to be driven
513 by the short, complex deletion identified by Wayne et al. (1996). Based on inspecting reads,
514 this variant appeared to exist in two early 1800s Lund flies and two 1933 Lund flies, implying
515 modest frequency change between temporal samples. The highest SNP-level PBE score at this
516 gene was at a non-synonymous variant over 1 kb downstream from the complex deletion
517 (2L:19544138, a threonine/serine polymorphism).

518 The top outlier for the 1933-2010s interval (denoted region B1) spanned 75 kb and
519 included a known target of insecticide resistance evolution, *Cytochrome P450 6g1* (*Cyp6g1*,
520 shown in **Figure 5D**). This gene is associated with resistance to dichlorodiphenyl-
521 trichloroethane (DDT) and other insecticides in *D. melanogaster*. As with *ChKov1*, there is prior
522 evidence for positive selection associated with transposon insertions and gene duplication
523 (Daborn et al. 2002; Chung 2007; Schmidt et al. 2010; Battlay et al. 2016). Here, the novel and
524 widespread deployment of DDT, introduced in 1944, provides a clear hypothesis for a selective
525 pressure driving the observed frequency changes at SNPs linked to the locus in question.

526

527 **A subset of top outliers show narrow peaks of genetic differentiation**

528 Both of the above outlier regions were relatively broad, and some of the other top outliers also
529 showed less-broad plateaus of elevated genetic differentiation between temporal samples. In
530 contrast, several other top outliers showed more narrowly localized signals of elevated genetic
531 differentiation. At least in spatial analyses of local adaptation, broader intervals of genetic
532 differentiation are more likely to be associated with selection favoring a single initial haplotype
533 (resulting in a hard sweep), whereas narrow signals of frequency change may result from
534 selection favoring multiple initial haplotypes, resulting in a "soft sweep" (da Silva Ribeiro et al.
535 2022). Narrow differentiation signals are of particular interest based on their potential to
536 indicate not just a specific gene but often a small set of candidate variants which may include
537 the target of selection. **Figure 6A-D** depicts four narrow SNP-level patterns of allele frequency
538 change focused on the 1800s-1933 time interval, whereas **Figure 7A-D** depicts four such
539 examples for the 1933-2010s interval (all of which were among the top 11 regions for window
540 PBE for their time interval). Focusing on the earlier period, a few SNPs upstream of *hector* (*hec*;
541 in region A3), a calcitonin-like neuropeptide receptor involved in male courtship behavior (Li
542 et al. 2011), showed elevated PBE values (**Figure 6A**). In region A8, a cluster of SNPs within a
543 5′ UTR intron showed the highest PBE values at *rutabaga* (*rut*; **Figure 6B**), a calcium-responsive
544 adenyl cyclase that influences learning, memory, lifespan, circadian rhythm, ethanol tolerance,
545 and response to heat and oxidative stress (Levin et al. 1992; Tong et al. 2007; Donlea et al.
546 2009; Qin and Dubnau 2010; Scheunemann et al. 2012). We also detected a narrow set of
547 SNPs within region A7 at the lysozyme genes *LysC* and *LysD* (**Figure 6C**), a locus for which a
548 gene duplication and inversion polymorphisms have been reported (Chakraborty et al. 2018),
549 and which was found to be strongly differentiated between Swedish and Italian *D.*
550 *melanogaster* populations (Mateo et al. 2018). In contrast, no function is known for *CG17032*
551 (in region A5), which also featured a cluster of high PBE SNPs (**Figure 6D**), the most extreme
552 of which was a synonymous variant (at R5 position 3L:15953092). Among other notable 1800s-
553 1933 outliers, the A2 region encompassed *sturkopf*, a lipid droplet-associated protein that
554 regulates growth and stress response (Werthebach et al. 2019). A polymorphic duplication of

555 this gene has been associated with low genetic diversity (Cardoso-Moreira et al. 2016) and this
556 window was detected by our depth qPBE scan as well (**Table S9**). Region A4 included *cabut*
557 (*cbt*), a transcription factor that regulates metabolic and circadian responses to nutrition
558 (Bartok et al. 2015).
559
560 **Figure 6: Narrow SNP-level intervals of differentiation indicate potential targets of**
561 **natural selection from the 1800s-focused scan.**
562 PBE for each SNP is plotted across selected top outlier regions that showed narrow peaks of
563 differentiation. These peaks occurred in or near:
564 (A) *hector* (hec) in region A3,
565 (B) *rutabaga* (rut) in region A8,
566 (C) *Lysozyme C* (*LysC*) and *Lysozyme D* (*LysD*) in region A7, and
567 (D) *CG17032* in region A5.
568
569 For the 1933-2010s interval, a window focused on *Adenosylhomocysteinase* (*Ahcy*; in
570 region B2) yielded a PBE value nearly as extreme as that observed at *Cyp6g1*, but with a far
571 narrower genomic scale (**Figure 7A**). Curiously, the top SNP PBE value occurs within a protein-
572 coding exon, but at a synonymous variant. *Ahcy* regulates methionine metabolism, histone
573 methylation, and lifespan (Parkhitko et al. 2016). It is preferentially expressed in circadian-
574 regulating pacemaker neurons (Rivas et al. 2021), and its role in circadian regulation has been
575 confirmed in mice (Greco et al. 2020). Despite the strength of this signal, *Ahcy* does not appear
576 to have been detected as a top selection candidate from previously published scans of
577 contemporary genomic variation. The SNP with the highest site PBE in *Ahcy* is associated with
578 a synonymous variant (at R5 position 3L:15953092). We note that within the next-highest
579 region (B3, not plotted), *Lysine demethylase 4B* (*Kdm4B*) also impacts histone methylation and
580 circadian rhythm (Shalaby et al. 2019). For the narrow signal at region B8, the highest PBE
581 SNPs were observed at the upstream end of *retinal degeneration B* (*rdgB*; **Figure 7B**), involved
582 in phototransduction, olfaction, and phospholipid transport (Woodard et al. 1992; Milligan et
583 al. 1999; Gardner et al. 2012). Notably, the narrow PBE signal at region B11 (**Figure 7C**)

584 centered on a gene that was recently detected as a top outlier in the 1975-2015 US temporal
585 population genomic study cited above (Lange et al. 2022); *Hexosaminidase 2* (*Hexo2*) encodes
586 a sperm-associated protein that may play a role in fertilization (Cattaneo et al. 2006; Intra et al.
587 2017). Whereas another narrowly-focused top outlier (region B5) centered on the non-coding
588 RNA gene *CR43835* (**Figure 7D**), for which no function is known.
589
590 **Figure 7: Narrow SNP-level intervals of differentiation indicate potential targets of**
591 **natural selection from the 1900s-focused scan.**
592 PBE for each SNP is plotted across selected top outlier regions that showed narrow peaks of
593 differentiation. These peaks occurred in or near:
594 (A) *Adenosylhomocysteinase* (*Ahcy*, or *Ahcy13*) in region B2,
595 (B) *retinal degeneration B* (*rdgB*) in region B8,
596 (C) *Hexosaminidase 2* (*Hexo2*) in region B11, and
597 (D) *ncRNA-CR43835* in region B5.
598
599 As hinted by the exclusively X-linked examples shown in **Figure 7**, for the 1933-2010s
600 scan specifically, there was an abundance of X-linked loci among the top outliers: although the
601 X chromosome constitutes less than one fifth of the analyzed genome, 12 of the top 15 outlier
602 regions were X-linked (**Table S8**). Explanations of this enrichment could include greater X
603 chromosome effects of sampling variance (although this effect should have been stronger for
604 the 1800s scan), genetic drift, or positive selection (potentially facilitated by X chromosome
605 hemizygosity; Charlesworth et al. 2018). However, neither scan was meaningfully enriched for
606 X-linked loci if the full list of outliers was considered (with the X contributing 21.7% of outliers
607 for scan A and 18.6% for scan B). The excess of top outliers specifically for scan B could indicate
608 that more of the X chromosome's selection in this interval was from hard sweeps (as suggested
609 by Harris and Garud 2023) and therefore more readily detectable from window genetic
610 differentiation (da Silva Ribeiro et al. 2022). However, the narrower peaks of differentiation
611 shown in **Figure 7** could indicate a role of X-linked soft sweeps as well.
612

613 **Further insights into previously implicated selection targets**

614 Three other top outliers for the 1933-2010s interval included genes previously implicated in
615 recent positive selection in European *D. melanogaster*. *Prosap* (in region B6) is thought to
616 encode a structural component of the postsynaptic density (Leibl and Featherstone 2008) and
617 affects circadian rhythm (Harbison et al. 2019). This gene was previously found to display
618 signals of parallel adaptation across cold-adapted *D. melanogaster* populations (Pool et al.
619 2017), and it harbors a polymorphic transposon insertion associated with signals of selection
620 (Rech et al. 2022). Region B7 centered on *phantom* (*phm*), which encodes a cytochrome P450
621 protein involved in ecdysteroid biosynthesis (Warren et al. 2004). Variation at this gene
622 revealed a selective sweep signal in a European *D. melanogaster* population (Orengo and
623 Aguadé 2007), and the protein displays rapid evolution between species (Good et al. 2014).
624 Further down the list, the region B21 contained *polyhomeotic-proximal* (*ph-p*), a
625 developmental and chromatin-modifying gene (*e.g.* Wang et al. 2006; Enderle et al. 2011),
626 which also displayed evidence of a selective sweep in a European population (Beisswanger
627 and Stephan 2008), associated with altered thermosensitivity of gene expression (Voigt et al.
628 2015). For these outliers, our findings offer insights regarding the timing of positive selection.
629

630 **Gene Ontology analysis reveals shifts in membranes and metabolism**

631 We also applied a Gene Ontology (GO) enrichment analysis, using the permutation approach
632 initially described in Langley et al. (2012), separately to the outliers detected from each
633 temporal scan. From the 1800s-1933 scan, some of the most enriched categories related to
634 cell shape and morphogenesis, including actin-mediated contraction and membrane
635 invagination (**Table S10**). Enrichment was also observed for membrane (and vesicle)
636 organization, compatible with the existence of population differences in membrane lipid
637 composition (Cooper et al. 2014) and this time interval coinciding with the species' earliest
638 known occupation of northern Europe (Keller 2007). Curiously, "response to DDT" was also
639 enriched, due to three cytochrome P450 outlier genes other than *Cyp6g1*, potentially
640 reflecting pre-insecticide roles of insecticide-related genes (as with the *ChKov* locus described
641 above). For the 1933-2010s interval, insecticide-related GO categories were not enriched

642 among outliers. Instead, metabolic processes dominated the top results (including amide,
643 fatty acid, lipid, and peptide metabolism), and categories related to cell division were enriched
644 as well (**Table S10**). As with membrane composition, metabolism is also known to vary
645 significantly between *D. melanogaster* populations from contrasting thermal environments
646 (*e.g.* Cogni et al. 2015; Brown et al. 2018).
647

648 **DISCUSSION**

649 Here, we report sequenced genomes from more than two dozen historical specimens of
650 *Drosophila melanogaster*. Comparing these historical genomes to modern population
651 genomic data, we have quantified allele frequency changes across roughly three thousand
652 generations of evolution, and these analyses have yielded a range of insights regarding recent
653 evolution in this model species.
654

655 **Illuminating Known and Novel Candidates for Recent Adaptive Evolution**

656 Above, we describe genomic scans for outliers reflecting elevated SNP frequency changes
657 across two separate time intervals (early 1800s to 1933, and 1933 to 2010s). While we report
658 outliers reflecting the upper 1% of genomic windows for our focal statistic, we focus above on
659 the most extreme of these outlier signals. This focus is motivated by uncertainty regarding the
660 actual number of loci under selection in each time interval, which we do not attempt to estimate
661 due to limitations in the data and in our ability to identify a realistic neutral model. Among the
662 top outliers discussed above, none appeared within the full list of outliers for the alternate time
663 period. This lack of outlier overlap could indicate that selective pressures differed between
664 eras, or that adaptive events tended to complete within a given time period rather than
665 spanning between them.

666 The strongest signals of positive selection revealed by our study added to our
667 understanding of known selection targets, particularly with regard to the timing of their
668 evolution. The top outlier from the 1933 to 2010s interval, *Cyp6g1*, was known to be the target
669 of a particularly strong selective sweep related to DDT resistance, and thus provides a strong
670 proof of concept for our ability to detect loci under recent selection. However, our inferred

671 timing of selection for our top 1800s to 1933 outlier, *CHKov1*, served to favor one adaptive
672 hypothesis (resistance to Sigma virus) over another (insecticide resistance). A second outlier,
673 *ref(2)p*, reinforced the potential importance of Sigma virus as a selective pressure during our
674 earlier time interval. As cited above, a significant proportion of our other top outliers also
675 centered on previously reported selection targets. In addition to the above three outliers,
676 these included *LysC/LysD* and *sturkopf* for the 1800s to 1933 time period, and *Hexo2*, *ph-p*,
677 *phm*, and *Prosap* for the 1933 to 2010s interval. The number of our top outliers identified by
678 previous studies is a testament to the depth of previous population genomic studies in *D.*
679 *melanogaster*, particularly for European populations.

680 And yet, the majority of top outlier regions contained no previously known selection
681 candidate. A prime example is *Ahcy* from the 1930s to 2010s scan, which showed a striking
682 level of window frequency change comparable to *Cyp6g1*, in spite of the more narrowly
683 focused peak at *Ahcy*. The role of *Ahcy* in circadian regulation connects it with other top
684 outliers, especially the functionally related *Kdm4B* (the third highest outlier in this scan after
685 *Cyp6g1* and *Ahcy*). Both *Ahcy* and *Kdm4B* alter histone methylation, an epigenetic mechanism
686 known to regulate fly circadian rhythm (*e.g.* Shalaby et al. 2018). In addition, *Prosap* from this
687 same 1933 to 2010s time interval, and *cbt* and *rut* from the 1800s to 1933 scan, also represent
688 top outliers with circadian functions. The evolution of circadian behavior in high latitude *D.*
689 *melanogaster* populations is well known (*e.g.* Kyriacou et al. 2008; Svetec et al. 2015). Changes
690 in day length help regulate reproductive diapause (Saunders et al. 1989), a key adaptation in
691 *D. melanogaster* populations from strongly seasonal climates (Schmidt et al. 2005; Erickson et
692 al. 2020). The northerly location of Sweden entails pronounced seasonal variation in day
693 length, which could heighten local selective pressures on daylight-responsive behaviors.
694 While each of the above five genes have been functionally linked to circadian regulation, to
695 our knowledge none have been linked to the evolution of circadian behavior (although *Prosap*
696 does have documented signatures of local adaptation; Pool et al. 2017). Hence, these genes
697 represent strong candidates for future investigations into the molecular basis of behavioral
698 evolution.

699 In theory, alleles promoting adaptation to the local environment should have been
700 adaptive in the earlier time period, and yet three of these five circadian genes were only found
701 as outliers in the later interval. It is possible that each local population initially possessed only
702 a subset of adaptive variants, and others arrived later via migration. Alternatively, in the 1800s
703 to 1933 interval, the frequencies of causative variants at some circadian genes may have still
704 been early in their rise, poising them to rise more quickly in the 1933 to 2010s interval. Another
705 possibility is contingency – some circadian variants may have only become beneficial after
706 other adaptive changes had occurred due to epistatic interactions among circadian genes.
707

708 **Temporal analyses reveal the complexity of demographic history**

709 Analyses involving our museum fly genomes also yielded novel insights regarding the
710 demographic history of a local *D. melanogaster* population from Lund, Sweden. We found that
711 between the early 1800s and 1933, Lund genomic diversity became more differentiated from
712 other modern European populations with respect to allele frequencies, and yet this pattern
713 reversed between the 1933 and 2010s, with Lund regaining similarity to other European
714 populations. This unexpected pattern of transient local differentiation suggests that distinct
715 evolutionary forces had predominant influences on genomic diversity between these time
716 points. As elaborated below, a primary hypothesis for this shifting pattern is that important
717 changes in population dynamics occurred between these eras.

718 The increased genetic differentiation of the early 1800s to 1933 period could reflect an
719 important influence of genetic drift, which could have occurred for multiple reasons. This
720 interval appears to reflect the earliest days of *D. melanogaster* occupying northern Europe
721 (Keller 2007), perhaps as shifting human activities only just permitted the species to survive in
722 this region, and hence we might predict that initial local population sizes were low. The climate
723 during this period was also colder than at present (**Figure 1A**), which may have also limited
724 local population sizes. In agreement with this hypothesis, Zetterstedt commented in his
725 treatise on Scandinavian Diptera with regard to *D. melanogaster*, "*habitat in Svecia meridionali*
726 *& media rarius*", *i.e.* in Southern and Middle Sweden rare. Furthermore, the species is
727 conspicuously absent from Carl Linnaeus' *Fauna Svecica* (Linnaeus 1746; 1761), nor mentioned

728 in Johann Christian Fabricius' *Genera insectorum* (Fabricius 1776). While these factors indicate
729 that local population size was likely low in the early 1800s, we cannot exclude an important
730 influence of natural selection in the observed increase in genome-wide population
731 differentiation, as European fly populations adapted to novel environments.

732 The hypothesized low historical population sizes of *D. melanogaster* in Lund could also
733 contribute to the inferred prevalence of relatedness among individuals within each historical
734 population sample, whereas flies from modern population samples generally show much less
735 relatedness, even when multiple individuals or lines obtained from the same fly trap are
736 analyzed (*e.g.* Lack et al. 2016). Furthermore, a notable fraction of flies from the 1933 sample
737 show evidence of inbreeding, indicating that their parents were related to some degree. Such
738 observations should be more likely in a relatively small and isolated local population. However,
739 the precise circumstances of these collections are not entirely known, and details of sampling
740 methodology (*e.g.* if multiple flies eclosing from the same piece of fruit were collected, or if
741 flies were raised in captivity for observation) may also contribute to the relatedness and
742 inbreeding we infer. Such uncertainties represent an inherent limitation in studying specimens
743 that are accompanied by finite documentation.

744 In contrast, the genomic homogenization between Lund and other European regions
745 that occurred between 1933 and the 2010s could reflect increased migration through time, in
746 conjunction with reduced genetic drift in fly populations that may have grown with time. Both
747 of those demographic changes might be predicted based on concurrent changes in human
748 activity during the 20$^{th}$ century. This time period featured a five-fold increase in the human
749 population of Lund and the surrounding region, along with a warming climate (**Figure 1A**),
750 both of which may have been conducive to population growth in this human commensal insect
751 (and hence reduced drift), even as new challenges such as insecticides emerged. In parallel,
752 increased human transportation, particularly the expanded shipping of fruit and other
753 commodities, would be expected to facilitate increased long-distance dispersal of *D.*
754 *melanogaster*. While these hypothesized demographic changes could be sufficient to explain
755 the genetic homogenization of the 1933 to 2010s time period, here again, we cannot exclude

756 a meaningful role for natural selection (such as selective sweeps spreading between regions)
757 in contributing to reduced differentiation.

758 The shifting population history potentially implied by our results raises questions about
759 the typical consideration of demographic processes in population genetic studies. Usually,
760 data is only available from one or more modern populations, and by necessity, fairly simple
761 neutral models of population history are fit to the data. However, we find that when genomes
762 from multiple temporal samples are available, we see hints of important historical changes in
763 population size and gene flow that conventional demographic inference would be unlikely to
764 capture. How well simple demographic models can reflect key aspects of complex
765 demographic models, and to what degree they can recapitulate the effects of complex
766 demographies on genomic diversity, is largely unknown. Regardless, it is clear that
767 ancient/historical genomes have strong potential to inform our understanding of the histories
768 of natural populations, just as they have for our own species (*e.g.* Yang and Fu 2018; Wohns et
769 al. 2022) and for other taxa (*e.g.* van der Valk et al. 2019; Cohen et al. 2022).

770

771 **Implications of sequencing historical *D. melanogaster* specimens**

772 This project contributes to the rapidly growing field of historical genomics and the study of
773 temporal population genomics in natural populations. To our knowledge, our samples include
774 the oldest pinned insects from which genome sequences have been reported (extending the
775 temporal horizon of insect museomics from the late 1800s back to the early 1800s). Overall,
776 96% of our specimens yielded analyzable genomes, and 73% yielded genomes of relatively
777 high quality. These results are encouraging for future museomic studies of small invertebrates.

778 To our knowledge, *D. melanogaster* is the smallest animal from which historical
779 genomes have been obtained. Surprisingly, that small size may have even been an advantage.
780 Male *D. melanogaster* are smaller than females (roughly 1 mg vs. 1.8 mg wet weight; Hillesheim
781 and Stearns 1991), and yet all 11 males in our study yielded high quality genomes, whereas
782 only half of the 14 females gave comparable outcomes. It is possible that less voluminous
783 insects such as male *D. melanogaster* experience a faster rate of desiccation relative to decay,
784 which should improve the quality and quantity of available DNA.

785 The genomes of these museum specimens have added to our knowledge of not only
786 recent evolution in *D. melanogaster*, but also the specimens themselves. Beyond confirming
787 the sex of each fly, we inferred from relatedness that the Swedish early 1800s flies, which had
788 been linked to three different scientists and two different localities, were actually from the same
789 time and place. These results may reflect widespread sharing of specimens among
790 contemporary scientists, and/or imperfect record keeping through time. Whether these flies
791 were collected by Zetterstedt (who received his doctoral degree from Lund University in 1808)
792 or by his mentor Fallén (who was primarily active before 1810) is difficult to say. We note that
793 all of these Swedish early 1800s flies – including the one labelled Småland – share the same
794 type of pin (similar to the those used by Fallén). Intriguingly, the lectotype specimen (also
795 labelled Småland) – which we were not granted permission to sequence – is, however, mounted
796 with a distinctly different pin. The lectotype may accordingly be one of Ljungh's flies, as
797 described in Zetterstedt's *Diptera Scandinaviae*, unlike the paralectotype that evidently was
798 caught in Lund, despite the label stating otherwise. The exact collection date of the sequenced
799 flies is unknown, although it may have occurred before 1810 (see Introduction). Given the
800 apparent acquisition of one of these flies by Ljungh, his death in 1828 puts a latest possible
801 bound on the collection date. These circumstances imply that the flies examined here are the
802 oldest surviving specimens of *D. melanogaster*, predating those described by Meigen (1830).
803 The expanse of existing knowledge regarding the biology and genome of *D.*
804 *melanogaster* has greatly enhanced the insights we have been able to draw from our museum
805 fly genomes. In turn, our results have implications for the extensive *Drosophila* research
806 community. We have identified various loci that represent likely targets of adaptive evolution
807 within specific recent time intervals, and in some cases, these genes have been found to impact
808 traits relevant to known selective pressures in the recent history of *D. melanogaster* (*e.g.*
809 circadian regulation). Hence, our study provides a wealth of candidates for functional evolution
810 that can be investigated in subsequent laboratory studies, through leverage of the molecular,
811 genetic, and transgenic tools available in this model system. Although museum specimens for
812 this (or any) species are finite, further historical genomic studies in *D. melanogaster* have the
813 potential to greatly inform the spatial and temporal dynamics of adaptive evolution and

814 population history, broadening the range of insights that can be obtained from genomic
815 variation in this population genetics model species.
816
817 **METHODS:**
818 **Samples and DNA extraction**
819 Historical specimens of *Drosophila melanogaster* were sampled from the entomological
820 collections of Biological Museum (Lund, Sweden: MZLU), Swedish Museum of Natural History
821 (Stockholm, Sweden: NRM), and National History Museum Denmark (Copenhagen, Denmark:
822 NHMD) (**Table S1**). DNA was extracted from the sampled specimens using the QIAamp DNA
823 Micro kit (Qiagen). Given the fragility of these old samples, special attention was given to
824 preserve their integrity. The samples were placed in a microtube without removing the pin. In
825 cases where the pin was too long to fit in the microtube, or in the case the specimen was not
826 mounted near the end, the pin was shortened. The lysis buffer was carefully added to avoid
827 damaging the specimens due to surface tension of the buffer prior to the addition of the
828 Proteinase K. The microtubes were incubated at 56°C for 2 hours without shaking. The buffer
829 containing the DNA was transferred to a new microtube, and 1 mL of MilliQ water was added
830 to the tube containing the fly to wash away the buffer. After 30 min the water was transferred
831 to a new microtube and stored at -20°C for future DNA extraction if needed. Then the fly was
832 washed in successive ethanol solutions of increasing concentrations (1 mL, 50%, 70%, 80%,
833 99%), again paying extra attention to not harm the specimen. And finally, 500 µl of acetone
834 was added to the specimen, which was then placed to dry in a suspended position. The rest
835 of the DNA extraction followed the protocol, except for the final elution which was performed
836 in two steps. In the first step, 50 µl of elution buffer was added to the columns, which were left
837 to incubate at room temperature for 15 min prior to centrifugation. For the second step, 50 µl
838 of elution buffer was added and the columns were incubated at 70°C for 15 min prior to
839 centrifugation.
840
841 **Library Prep and Sequencing**

842 The quality of the extracted DNA was visually investigated on an agarose gel (1.5%). cThe
843 library prep followed mainly the protocol of Twort *et al*. (2021). Briefly, the fragmented DNA
844 was blunt-end repaired with T4 Polynucleotide Kinase (New England Biolabs) and then purified
845 with the MinElute Purification Kit (Qiagen). Next followed adaptor ligation, reaction
846 purification, and adapter fill in. The resulting products were subsequently indexed with unique
847 dual indexes. Indexing PCR was performed in ten independent reactions to reduce
848 amplification bias. Each PCR reaction consisted of 12-18 cycles depending on the
849 concentration. Indexing PCR reactions were pooled prior to purification with Sera-Mag
850 SpeedBeads™ carboxylate-modified hydrophilic (Sigma-Aldrich). An initial bead
851 concentration of 0.5X was used to remove long fragments that were likely contaminants of
852 fresh DNA. Libraries were selected using a 1.8X bead concentration to size select the intended
853 library range of ~300 bp. The resulting libraries were quantified and quality checked using the
854 Quanti-iT™ PicoGreen™ dsDNA assay and Bioanalyzer 2100 (Agilent technologies). The final
855 indexed libraries were pooled together prior to sequencing on an Illumina Novaseq 6000
856 platform at the Swedish National Genomics Institute (NGI) in Stockholm (2x150 bp, S4 flow
857 cell). After observing that sequenced fragments only averaged about 50 bp in length,
858 subsequent sequencing of most specimens with 2x50 bp reads was performed by the
859 University of Wisconsin Biotechnology Center using an Illumina NovaSeq 6000 (**Table S1**).
860

861 **Genome Alignment and Variant Calling**

862 The procedures for alignments and variant calling used to process the museum genomes
863 largely followed the *Drosophila* Genome Nexus (DGN) pipeline as described in Lack et al.
864 (2015). Using this pipeline with the release 5.57 *D. melanogaster* reference genome allowed
865 newly-assembled museum genomes to be compared against published data from modern
866 population samples of individual genomes. Modern genomes from France and Egypt were
867 reported in the subsequent DGN publication (Lack et al. 2016). A sample from the Netherlands
868 was also aligned in that study, and originally sequenced by Grenier et al. (2015). Population
869 samples from Sweden (Stockholm) and Italy were sequenced by Mateo et al. (2018) and
870 assembled using the DGN pipeline for this study. The (distinct) processing of published

871 pooled sequencing (pool-seq) data from Lund, Sweden (Kapun et al. 2021) is described in the
872 next section.
873 Following read trimming using fastp (v0.21.0, Chen et al. 2018), we followed the DGN
874 pipeline to map, filter, and call genotypes from the sequenced museum genomes. While the
875 pipeline is described in detail in the previously-cited papers, we note its key feature of two
876 rounds of mapping prior to genotyping. The initial mapping of reads to the reference genome
877 is performed with bwa -aln v.0.5.9 (Li and Durbin 2010) and stampy (Lunter and Goodson. This
878 step is followed by indel and variant-calling using GATK (McKenna et al 2010, Depristo et al
879 2011). The SNPs and indels were mapped onto the (release 5) reference genome, to create a
880 more robust target reference for a second round of alignment of the reads with bwa -aln and
881 stampy. It is only after this second realignment and a shift of coordinates around indels to
882 match the reference genome that genotyping for further downstream analyses was performed.
883 Diploid genotyping was initially performed at all sites, except for the X chromosomes
884 from male flies. The sex of the flies in the museum samples was not identified prior to
885 sequencing. It was identified by calculating the ratios of average read depth in the X
886 chromosome versus the autosomes, which is expected to be approximately 1:1 in females and
887 1:2 in males. All perl scripts referenced in this section and elsewhere were executed on perl
888 v5.34.0, all python scripts were executed on python 3.8.6 (except for the stampy aligner, which
889 was run on python 2.7.18-3 to ensure compatibility), and the R scripts were run using R v.4.2.2.
890 For those museum genomes with a total coverage of < 80% and average per-site read
891 depth of < 10, we concluded that diploid genotype calls could not be confidently made. These
892 were instead treated as "pseudo-haploid" with only a single nucleotide called per site;
893 UnifiedGenotyper (GATK, McKenna et al 2010) was run in haploid mode for all chromosomes
894 for these low-quality genomes. When more than a single base was identified at a (pseudo-
895 haploid) site, we retained the most frequent allele (or selected one at random in cases of a tie).
896 For diploid genotypes, we required at least 25% of reads to represent a minor allele, otherwise
897 the site was classified as homozygous and assigned the more frequent allele.
898
899 **Processing and Analysis of Pool-seq Data**

900 Published pool-seq data (Kapun et al. 2021) was obtained from modern three Lund population
901 samples: one from summer 2014 (SRR 5647735) and two from summer 2015 (SRR8439151,
902 8439156). Each was derived from 40 wild-caught males. For each of these samples, alignment
903 and post-mapping processing was performed using the same procedures as for individual
904 genomes in the first round of the DGN pipeline, and the three pool-seq sample alignments
905 were then merged to a single bam file. Because of the aggregate, multi-individual nature of
906 these sequence data, frequency estimation was performed using PoolSNP (Kapun et al. 2020).
907 We imposed an additional constraint of requiring there to be at least 5 instances of a minor
908 allele for a site to be identified as polymorphic when filtering the VCFs, in order to minimize
909 mislabeling read base-calling errors as SNPs. Additionally (and obviously), because individual
910 genomes cannot be identified from pool-seq data and thus could not be used to create a
911 revised reference genome, only a single round of alignments and genotyping was performed
912 on the pool-seq data.

913 Pool-seq data introduces several statistical sampling artifacts that are not an issue with
914 individual sequencing. Specifically, the number of reads contributed by each individual varies,
915 so that some genomes are over-represented and others are under-represented. To deal with
916 this, we calculate an effective allele (lineage) count (Ferretti et al. 2013), i.e. the number of $j$
917 unique lineages at a site given $n_r$ reads and $n_c$ the number of chromosomes in the sample pool
918 ($n_c$ = 240 for autosomes, $n_c$ = 120 for X chromosomes because all of the 120 flies sampled in
919 the pool-seq study were male). The effective count $j$ follows a distribution based on Stirling
920 numbers of the second kind $S(n_r,j)$, *i.e.* the number of ways in which $n_r$ reads can be partitioned
921 into $j$ lineages. The probability of $j$ independent lineages in the sample is then:

922 $$P(j|,n_c,n_r) = S(n_r,j)\frac{n_c!}{(n_c-j)!\ n_c^{n_r}}$$

923 The effective number of distinct lineages (*i.e.* de facto allele count per site) was estimated from
924 the expectation of this distribution. We found a maximum pool size of 183 for the autosomes
925 (median = 145) and of 95 for ChrX (median = 70). Allele frequencies at each site were
926 multiplied by the pool size to generate the appropriate down-sampled counts for population
927 genetic analysis.

928

929 **Assessment of Genomic Data Quality**

930 The total number of reads and the average read length was assessed from each bam file using
931 CollectInsertSizeMetrics from the Picard suite of tools. Based on the depth and coverage
932 criteria given above, the low-quality genomes that were treated as pseudo-haploid in the
933 analysis pipeline were 18SL4, 18SL5, 18SL6, 18SL12 (from 1800s), 19SL3, and 19SL7 (from
934 1933). One sample from 1933, 19SL2, was of such low quality that it was excluded from any
935 further processing or analysis.

936 To determine whether low quality sequences contributed a significant number of false
937 positive variants in each population sample, we tallied the number of "singletons" - instances
938 where a given genome had an allele otherwise not observed among at least five total alleles.
939 The fraction of singleton alleles that were A or T bases was calculated to determine if cytosine
940 deamination accounted for many of the singletons, and whether this phenomenon was
941 correlated with sample quality. Singleton minor alleles were excluded from certain
942 downstream analyses, such as demographic inference, while a total allele count of at least four
943 was required for all analyses unless otherwise indicated. We also calculated the Pearson
944 correlations between read depth and coverage (as well as between read depth and the fraction
945 of singletons contributed by each genome, and between the fraction of singletons and the A/T
946 content) to quantify the association between these quality metrics.

947

948 **Identification of Identity By Descent**

949 Identity by descent (IBD) among certain genomes in both the 1800s and 1933 Lund samples
950 was initially suspected based on highly variable pairwise differences among genomes. To
951 identify specific regions of individual genomes with IBD, we began by partitioning
952 chromosomes into "windows" defined by having 25,000 SNPs in an ancestral range Zambia
953 population sample (Lack et al. 2015), averaging approximately 100 kb in length.

954 We then sought to account for genome-specific differences in genetic distances, such as
955 due to sample quality. For each chromosome window of the *i*th genome, a background
956 genetic distance $d_i$ is the median Hamming distance between the *i*th genome and unrelated

957 modern Lyon samples. This genome-specific individual distance estimate was used to obtain
958 a distance correction factor $m_i$ that accounts for its tendency to yield higher or lower distances
959 than others from its population sample. The baseline population-specific distances $D$ for 1800s
960 and 1933 Lund were estimated as the median pairwise distance among individuals with no
961 obvious IBD based on the Hamming distance matrix. Pairwise distances were compared to this
962 baseline, scaled with respect to the Lyon outgroup distance, as follows.

963 For two genomes $i,j$ collected in the same location and time point, the chromosome
964 window Hamming distance is $d(i,j)$, while $m_i$, $m_j$ respectively are the scaling factors specific to
965 sequences $i,j$ – calculated from $d_i, d_j$ divided by the median pairwise distance between all of the
966 Lund sequences (of a given time point) and the Lyon sequences.

967 A. For the rescaled distance $D'(i,j) = d(i,j)m_i m_j$, a chromosome window is considered to be
968 IBD between two diploid sequences (autosomes or female X chromosomes) if $D'(i,j) <$
969 $(7/8)D$. The reasoning is that if one allele from each individual is IBD, then ¼ of pairwise
970 allelic comparisons will yield zero distance due to this IBD, and hence the overall
971 expected distance is ¾ of that from a non-IBD pair. For haploid to haploid (male X
972 chromosome comparison) the same rule with a factor of ½ is applied, while for haploid
973 to diploid (male to female X chromosome comparison) a factor of ¾ is used.

974 B. When pairwise IBD regions are identified, this window is masked in one genome of the
975 pair. The masking is prioritized so that low-quality genomes are preferentially masked
976 when compared to a higher quality genome in order to minimize data loss.

977 C. To avoid false negatives due to IBD regions covering partial windows, when a window
978 was masked, the proximate half of the "upstream" and the "downstream" window was
979 also masked. While this practice may mask some non-IBD regions, it errs on the
980 conservative side for avoiding oversampling IBD alleles (since IBD covering more than
981 half of a flanking window would be likely to satisfy the above masking criteria and cause
982 the full window to be masked).

983 In addition to IBD among genomes, there was also evidence for depleted
984 heterozygosity as a consequence of inbreeding in a subset of genomes from both the 1800s
985 and (especially) the 1933 Lund populations, *i.e.* intragenomic IBD. In non-inbred genomes,

986 the heterozygosity should approximately equal nucleotide diveristy ($\pi$). Therefore, to identify
987 enrichment of homozygosity in inbred genomes, we compared the observed heterozygosity
988 per window to the median pairwise distance among non-IBD pairs from the same population
989 in that window. If the within-window heterozygosity of a sample was less than $\pi/2$, the window
990 was flagged as IBD and a randomly selected allele of the two in the window was masked.
991 We included samples believed to originate from other sites and time points (18DZ5
992 from Zealand, Denmark and 18SL6, which was labelled as being from Småland, Sweden) in the
993 IBD analyses, because the latter showed low divergence when compared to some of the 1800s
994 Lund samples. There was unambiguous evidence of close relatedness between 18SL6 and
995 18SL10, suggesting that the former was in fact from Lund and thus could be grouped with this
996 population in subsequent analyses. In contrast, 18DZ5 showed roughly 1% of its genome with
997 apparent IBD with one Lund genome, and lesser levels with two others. As these could be
998 false positives due to conservative filtering criteria, we continued to treat this sample as being
999 from Zealand rather than Lund in demographic analyses based on location, while at the same
1000 time masking that small region of the genome (18SL12) that showed low, IBD-like divergence
1001 with respect to 18DZ5.
1002
1003 **Analysis of Genetic Differentiation**
1004 Temporal and spatial divergence among different populations (defined by time and location)
1005 was assessed with analyses of population genetic distances. Our analysis focused on X
1006 chromosome sequences (to the exclusion of low recombination regions defined by Pool et al.
1007 2017) because of the rarity of sex chromosome inversions in *Drosophila*, and only considered
1008 polymorphic sites without missing data in any of the analyzed samples.
1009 To avoid distortions due to sample sizes, the number of alleles taken from each
1010 population (*i.e.* the 1933 Lund and the modern European and Egyptian samples) was down-
1011 sampled to match the 4 alleles from the IBD-masked 1800s Lund allele counts. This entailed
1012 selecting 4 random alleles at each site for the 1933 Lund population, the 2010s Lund pool-seq
1013 population, and the other modern populations, sampling independently at each analyzed site.

1014 This down-sampled data was used to calculate nucleotide diversity within each population, as
1015 well as between-population genetic distance ($D_{xy}$) and $F_{ST}$ (Reynolds et al. 1983).
1016 Genetic differentiation among the populations was also characterized using principal
1017 components analysis (PCA). Down-sampling of the Lund 1933 and modern samples was
1018 necessary to avoid over-weighting larger samples. However, because individual haplotypes
1019 were projected into the PCA space, the pool-seq Lund sample had to be excluded. To increase
1020 the number of sites with no missing data, we sampled non-masked X chromosomes that had
1021 no detectable pairwise IBD from the Hamming distance matrices from the 1800s and 1933
1022 populations. Five such unrelated X chromosomes were identified from the 1800s samples:
1023 18SL11, 18SL13, 18DZ5, 18GP9, and 18GP25, which were matched by the 5 unrelated 1933
1024 Lund samples 19SL8, 19SL15, 19SL16, 19SL20, and 19SL23. Five X chromosomes were
1025 sampled from each of the modern European and Egyptian outgroup populations as well, for a
1026 total of 35 chromosomes (5 from 7 populations). For each of the 25,094 polymorphic sites with
1027 no missing data across all 35 samples in the analyzed regions of the X, the frequency of the
1028 minor allele at each site was calculated. An R-mode PCA was then used to compute the 35
1029 eigenvectors of the correlation matrix.
1030
1031 **Inversion Polymorphism**
1032 To assay for inversions, we used the list of inversion-linked SNP alleles identified by
1033 Kapun (2016, 2020) for the following common autosomal inversions: *In(2L)t*, *In(2R)Ns*, *In(3L)P*,
1034 *In(3R)K*, *In(3R)Mo*, and *In(3R)P*. The frequency of each inversion was estimated from the median
1035 frequency of the inversion-associated alleles among their respective loci. For the Lund pool-
1036 seq data, the requirement of > 4 minor alleles per site to call minor alleles was relaxed to
1037 identify low-frequency inversions, *i.e.* even singleton minor alleles were included.
1038 We used the median rather than the mean allele frequency because some alleles may
1039 be absent due to missing data. Additionally, not all of the alleles identified in Kapun (2016) as
1040 inversion-associated appeared to be fully linked to inversions in our data set. There were
1041 several instances where a small fraction of inversion-linked alleles are present in the museum
1042 samples or modern Lund (*e.g.* for *In(3R)P*, 4 of the 20 inversion-linked alleles were present at

1043 low frequency in the modern Lund pool-seq data), while the absence of inversion-linked alleles
1044 at the majority of sites suggested that the inversion itself is absent. In such cases, the use of
1045 mean allele frequency as a proxy for inversions would result in a non-zero estimated inversion
1046 frequency that is much lower than the reciprocal of the number of chromosomes sampled in
1047 the population. The estimated frequencies for each inversion in modern Lund versus the
1048 historical populations were compared using Fisher exact tests on karyotype counts, using R.
1049

1050 **Demographic Estimation**

1051 Using allele frequency data from 1800s, 1933, and 2010s Lund, we estimated local effective
1052 population size ($N_e$) for the 1809-1933 and 1933-2015 intervals by implementing a simple drift-
1053 only neutral model of evolution (thus assuming no effect of migration, selection, or other
1054 processes on allele frequency changes). To estimate $N_e$ for these two time intervals, we
1055 compared the mean changes in allele frequencies across observed segregating sites to
1056 changes in allele frequencies simulated under a Fisher-Wright (FW) model of random genetic
1057 drift in discrete, non-overlapping generations. Specifically, we implemented a Bayesian
1058 approach to optimize $N_e$ that minimizes the difference between observed $\overline{\Delta p}$, the mean
1059 absolute value difference in allele frequency across sites between 1809 and 1933 (and
1060 between 1933 and 2015) and the $\overline{\Delta p}$ simulated under a FW model of genetic drift.

1061 We initialized the simulations by calculating the allele frequencies at polymorphic loci
1062 at each time point in the Lund populations. Regions of low recombination, as defined by Pool
1063 et al. (2017) were excluded, as were singleton variants and non-Lund 1800s genomes. The
1064 sample frequencies $p_i$ in the 1809 population (or 1933 for the second time interval) at the $i$th
1065 locus were used to approximate the initial allele frequencies $f_i$ in the population using Bayesian
1066 estimation. The prior of $f_i$ was based on a mutation-drift equilibrium with a site mutation rate
1067 of 5.21e-10 per generation (Huang et al. 2016) and a candidate haploid effective population
1068 size $N^*$. The posterior probabilities of $f_i$ were calculated prior and the observed frequencies
1069 at each site. The initial allele frequencies in each simulation replicate were generated by
1070 drawing $n$ alleles equal to the number in each sample. Using a uniform prior instead resulted
1071 in very similar distributions of $f_i$.

1072 Population allele frequencies were initialized to $f_i(0)$ at simulation start time,
1073 representing 1809 and 1933 for the first and second time intervals, respectively. The
1074 frequencies in generation t+1 were generated by drawing a binomial sample ~ $bin(f_i(t), N^*)$ at
1075 each site, where $f_i = x_i/N$ for $x_i$ count of what was the minor allele at $t = 0$. The model assumes
1076 a sufficiently large gene pool for sampling with replacement and treating each locus as
1077 statistically independent (*i.e.* under linkage equilibrium). This sampling is repeated for 1860
1078 (non-overlapping) generations for 1809-1933 and 1230 generations for 1933-2015, *i.e.* for 15
1079 generations per year (Turelli and Hoffmann 1995; Pool 2015). In the final (census) generation,
1080 an autosomal allele count $n = 14$ ($n = 9$ for ChrX) was drawn to represent the 1933 Lund sample,
1081 giving terminal sample allele frequencies $f_i^*(T)$ based on ~ $bin(f_i(T), n)$.
1082 Much the same method was used for estimating the best fit $N_e$ for the 1933-2015 time
1083 interval, with an important exception. Because the 2010s Lund samples contributed pool-seq
1084 data, the additional sample variance that is introduced by sampling of both individuals and
1085 reads must be incorporated into the simulation. In the terminal (census) generation, the
1086 simulations implemented a two-step sampling model consistent with the nature of pool-seq
1087 data. Namely, we sampled 240 alleles (corresponding to the 120 flies) in the census generation
1088 ~ $bin(f_i(T), 240)$. This created a pool of 240 alleles where the frequency of the (initial) minor
1089 allele frequency was $f_i'(T)$, with an expectation equal to $f_i(T)$. To simulate the final sample set,
1090 we drew $n$ alleles from this pool-seq subset of individuals (where $n$ is the effective number of
1091 alleles per site sampled in the pool-seq data, with median values of 145 for autosomes and 70
1092 for ChrX), *i.e.* ~ $bin(f_i'(T), n)$. This final draw gave terminal sample allele frequencies $f_i^*(T)$.
1093 The absolute differences between final and initial sample allele frequencies in the
1094 simulations $\Delta p_i = |f_i^*(T) - p_i(0)|$ were averaged across polymorphic sites, giving $\overline{\Delta p}$. We
1095 optimized our estimate of $N_e$ by running the FW simulations over 1000 replicates over a range
1096 of $N^*$={1000…20000} in increments of 1000 and selecting the value minimizing the difference
1097 between simulated and observed $\overline{\Delta p}$ . This estimate was further refined by considering $N^*$ in
1098 five increments of 100 on either side of the coarse-grained optimum.
1099 We evaluated the accuracy our estimates of effective population size ($N_1$, $N_2$) for the
1100 time intervals 1809-1933 and 1933-2015 by comparing $\pi$ observed in the 1800s, 1933, and

1101 2010s Lund samples to that of simulated populations of sizes $N_1$, $N_2$ at these time points. The
1102 simulations were initiated by sampling $N_1$ alleles from a source European population based on
1103 the demographic history of *D. melanogaster* inferred in Sprengelmeyer et al. (2020). The size
1104 $N_1$ Lund sample then experienced genetic drift (with no gene flow or natural selection) for 1860
1105 generations (scaled in units of $4N_e$) until 1933, at which point the population size was set to $N_2$
1106 and the simulations continued for a further 1230 generations. This neutral demographic model
1107 was implemented using the program ms (Hudson 1990), with the mutation rate cited above,
1108 along with a crossing-over rate of 1.09e-8, a gene conversion rate of 6.25e-8, and mean gene
1109 conversion tract length of 518 bp Comeron et al. (2012). We focused ms simulations on
1110 chromosome arm 3L because of the favorable sample sizes and relative rarity of inversion
1111 polymorphisms on this arm, and the availability of a parameterized demographic model for its
1112 evolutionary history in Europe.
1113
1114 **Population Branch Excess Scan for Positive Selection**
1115 To identify regions of the genome that may contain sites that may have experienced directional
1116 selection in the time between 1800s, 1900s, and modern samples, we used Population Branch
1117 Excess (PBE; Yassin et al. 2016), a statistic that aims to identify loci with particularly strong allele
1118 frequency differentiation in a focal population, in light of data from two outgroup populations.
1119 PBE is based on the Population Branch Statistic (PBS; Yi et al. 2010), which essentially estimates
1120 the length of the focal population's branch (in terms of allele frequency differentiation). PBE is
1121 determined by a comparison of the focal population's observed PBS value to that expected
1122 based on the lengths of the non-focal populations' branches at the same locus, along with the
1123 average lengths of the focal and non-focal branches across loci (Yassin et al. 2016). Compared
1124 to PBS, PBE is more focused on population-specific differentiation, and will not return a large
1125 value at loci where all population branches are long (*e.g.* due to background selection or
1126 positive selection in all populations).
1127 These statistics are based on log-transformed window $F_{ST}$ values $d = -\log(1-F_{ST})$
1128 between each pair of populations. Here as in other studies, window $F_{ST}$ was calculated by
1129 summing the numerator terms and summing the denominator terms of the Reynolds et al. 1983

1130 $F_{ST}$ estimator, and obtaining their ratio. Chromosome windows were defined to contain 250
1131 Zambia non-singleton SNPs, corresponding to approximately 4.6 kb windows on average. We
1132 performed a separate scan for each time interval. To focus on the 1800s, we made Lund 1800s
1133 the focal population, and used Lund 1933 and Lund 2010s as the non-focal populations. To
1134 focus on the 1900s, we made Lund 1933 the focal population, and used Lund 2010s and Lyon
1135 (2010) as the outgroups. Although the Lund 2010s sample comes from (distinctly-processed)
1136 pool-seq data, the presence of one other individually-sequenced population sample among
1137 the non-focal populations should prevent any artefactual differences between individual and
1138 pool-seq data from registering as differentiation specific to the focal population.
1139 Given that a viable demographic model was not available to provide a distribution of
1140 PBE values under a null hypothesis, outlier PBE values were defined based on the upper 1%
1141 quantile of the window distribution (which contained 240 windows genome-wide). Because
1142 strong selection can create linkage disequilibrium that extends across multiple windows, we
1143 aggregated adjacent or near-adjacent PBE-significant windows (separated by no more than 2
1144 non-outlier windows). Although SNP-level genetic differentiation can sometimes allow the
1145 detection of soft sweeps that do not register as outlier in window-level scans, our sample sizes
1146 of museum genomes were not sufficient to enable this approach (da Silva Ribeiro et al. 2022).
1147 Instead, we used site PBE as a heuristic for identifying sites and genes of interest within
1148 window-based PBE outlier regions.
1149
1150 **Gene Ontology Enrichment**
1151 We performed a gene ontology (GO) enrichment analysis on genes located PBE outlier regions
1152 from the 1800s- and 1900s-focused genome scans, in order to establish whether certain
1153 functional categories (*e.g.* key metabolic or developmental pathways) were disproportionately
1154 over-represented among genes that may differ between the historical and modern
1155 populations. The GO analysis methods followed the approach of Pool et al. (2012), which
1156 involves randomly permuting the genomic locations of outlier regions, to account for variation
1157 in gene length and the genomic clustering of functionally related genes. Raw p-values were
1158 computed from 10,000 random permutations, based on the proportion of reprlicates with an

1159 equal or greater number of outlier regions associated with a given GO category as observed
1160 in the empirical data.
1161
1162 **Copy Number Variation**
1163 To identify potential copy number variation (CNV) with frequency differences among
1164 populations, we adopted an approach analogous to the one used above for SNP-based branch
1165 statistics. We searched for CNV differentiation among samples by comparing the mean read
1166 depth per genomic window, on the assumption that duplication (or deletion) present at
1167 elevated frequency in a given population sample will result in higher (or lower) depth of
1168 coverage, all else being equal.
1169 Because these analyses compare read depths in museum samples with short fragments
1170 to read depths in modern samples with much longer fragments, we used trimmomatic (Bolger
1171 2014) to reduce the read length from the modern Lyon and Stockholm samples to 50 bp. We
1172 then reran the relevant portions of the DGN assembly pipeline, while treating the truncated
1173 reads as unpaired in the mapping. We used the same genomic windows as in the SNP PBE
1174 analysis, and for each population, we calculated the mean read depth for each window across
1175 all sites in that interval. Because genomes vary in their mean depth of coverage, we corrected
1176 for such differences by dividing the mean read depth of each window by the mean depth over
1177 the entire chromosome arm. These rescaled window read depths were treated as a
1178 quantitative trait. The quantitative trait analog to the fixation index $Q_{ST}$ (Lande 1992, Whitlock
1179 1999), was computed for read depths among populations based on variances in read depth,
1180 as $Q_{ST} = \sigma_{GB}^2/(\sigma_{GB}^2 + 2\sigma_{GW}^2)$, where $\sigma_{GW}^2$ is the within-population genetic trait variance and $\sigma_{GB}^2$ is
1181 the among-sample genetic trait variance.
1182 By further analogy with allele frequency-based branch statistics, we calculated the
1183 quantitative trait equivalents of PBS and PBE using $d = -\log(1-Q_{ST})$ as the $Q_{ST}$-based distance
1184 metric. To distinguish these trait-based branch statistics from $F_{ST}$-based measures, we refer to
1185 them as QPBS and QPBE. These branch statistics were calculated for the 1800s population
1186 using 1933 Lund and modern Lyon as outgroups. For 1933 as the focal population, Lyon and
1187 Stockholm were used as the modern outgroups (since this approach is not applicable to pool-

seq data such as the Lund 2010s sample). As with SNP-based PBE, we identified outlier windows in the upper 1% quantiles, and we aggregated these into regions by merging outlier windows separated by no more than one non-outlier window.

**Data Availability Statement:**

All new sequence data used in the present study can be found in NIH SRA Bioproject PRJNA945389. All scripts used in the present study can be found at: https://github.com/mshpak76/Museum_Drosophila_Genomes

**Acknowledgments:**

We wish to thank kind assistance with obtaining samples from Biologiska Museet (Lund, Sweden) and Naturhistoriska Museet (Stockholm, Sweden), and in particular Niklas Whalberg and Rune Bygebjerg. We also thank other members of our labs for providing valuable suggestions as well as relevant data and code from previous projects.

Funding was provided from Vetenskapsrådet (2019-03536) and the Max Planck Center on Next Generation Insect Chemical Ecology to MCS, and from NIH NIGMS award 144 AAH7544 to JEP.

**Author contributions:**

HRG and MCS collected data. MS, JDL, HRG, and JEP analyzed data. MS, JEP, and MCS wrote the manuscript.

**Declaration of interests:**

All authors declare no financial interests.

1213 **REFERENCES:**

1214

1215 Aminentzach YT, Macpherson JM, Petrov DA. 2005. Pesticide resistance via transposition-
1216 mediated adaptive gene truncation in *Drosophila*. Science 209: 764-767.

1217

1218 Bangham J, Obbard DJ, Kim KW, Haddrill PR, Jiggins FM. 2007. The age and evolution of an
1219 antiviral resistance mutation in *Drosophila melanogaster*. Proceedings of the Royal Society B:
1220 Biological Sciences 274:2027-2034.

1221

1222 Bartok O, Teesalu M, Ashwall-Fluss R, Pandey V, Hanan M, Rovenko BM, Poukkula M, Havula E,
1223 Moussaieff A, Vodala S et al. 2015. The transcription factor Cabut coordinates energy
1224 metabolism and the circadian clock in response to sugar sensing. EMBO Journal 34: 1538-
1225 1553.

1226

1227 Battlay P, Schmidt JM, Fournier-Level A, Robin C. 2016. Genomic and transcriptomic
1228 association identify a new insecticide resistance phenotype for the selective sweep at the
1229 *Cyp6g1* locus of *Drosophila melanogaster*. G3 6:2573-2481.

1230

1231 Beisswanger S, Stephan W. 2008. Evidence that strong positive selection drives
1232 neofunctionalization in the tandomly duplicated polyhomeotic genes in *Drosophila*.
1233 Proceedings of the National Academy of Sciences USA 105:5447-5452.

1234

1235 Bergland AO, Behrman EL, O'Brien KR, Schmidt PS, Petrov DA. 2014. Genomic evidence of
1236 rapid and stable adaptive oscillations over seasonal time scales in *Drosophila*. PLoS Genetics
1237 10:e1004775.

1238

1239 Bi KE, Linderoth T, Singhal S, Vanderpool D, Patton JL, Nielsen R, Moritz C, Good JM. 2019.
1240 Temporal genomic contrasts reveal rapid evolutionary responses in an alpine mammal during
1241 recent climate change. PLoS Genetics 15:e1008119.

1242

1243 Brown EB, Torres J, Bennick RA, Rozzo V, Kerbs A, DiAngelo JR, Keen AC. 2018. Variation in
1244 sleep and metabolic function is associated with latitude and average temperature in
1245 *Drosophila melanogaster*. Ecology and Evolution 8:4084-4097.

1246

1247 Card DC, Shapiro B, Giribet G, Moritz C, Edwards SV. 2021. Museum genomics. Annual Review
1248 of Genetics 55:633-659.

1249

1250 Cardosa-Moreira M, Arguello JR, S. Gottipati S, Harshman LG, Grenier JK, Clark AG. 2016.
1251 Evidence for the fixation of gene duplications by positive selection in *Drosophila*. Genome
1252 Research 26:787-798.

1253

1254 Carpenter JA, Obbard DJ, Maside X, Jiggins FM. 2007. The recent spread of a vertically
1255 transmitted virus through populations of *Drosophila melanogaster*. Molecular Ecology
1256 16:3947-3954.

1257

1258 Cattaneo F, Pasini ME, Intra J, Matsumoto M, Briani F, Hoshi M, Perotti ME. 2006. Identification
1259 and expression analysis of *Drosophila melanogaster* genes encoding beta-hexosaminidases of
1260 the sperm plasma membrane. Glycobiology 16:786-800.

1261

1262 Chakraborty M, VanKuren NW, Zhao R, Zhang X, Kalsow S, Emerson JJ. 2018. Hidden genetic
1263 variation shapes the structure of functional elements in *Drosophila*. Nature Genetics 50:20-25.

1264

1265 Charlesworth B, Campos JL, Jackson BC 2018. Faster-X evolution: theory and evidence from
1266 *Drosophila*. Molecular Ecology 27:3753-3771.

1267

1268 Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor.
1269 Bioinformatics 34:i884-i890.

1270

1271 Chung H, Bogwitz MR, McCart C, Andrianopoulos A, Ffrench-Constant RH, Batterham P,
1272 Daborn PJ. 2007. Cis-regulatory elements in the accord retrotransposon result in tissue-
1273 specific expression of the *Drosophila melanogaster* insecticide resistance gene *Cyp6g1*.
1274 Genetics 175:1071-1077.
1275
1276 Cogni R, Kuczynski K, Lavington E, Koury S, Behrman EL, O'Brien KR, Schmidt PS, Eanes WF.
1277 2015. Variation in *Drosophila melanogaster* central metabolic genes appears driven by natural
1278 selection both within and between populations. Proceedings of the Royal Society of London B
1279 282:20142688.
1280
1281 Cohen ZP, François O, Schoville SD. 2022. Museum genomics of an agricultural super-pest, the
1282 Colorado Potato Beetle, Leptinotarsa decemlineata (Chrysomelidae), provides evidence of
1283 adaptation from standing variation. Integrative and Comparative Biology 62:1827-1837.
1284
1285 Colomb J, Kaiser L, Chabaud M-A, Preat T 2009. Parametric and genetic analysis of *Drosophila*
1286 appetitive long-term memory and sugar motivation. Genes Brain and Behavior 8:407-415.
1287
1288 Comeron JM, Ratnappan R, Bailin S. 2012. The many landscapes of recombination in
1289 *Drosophila melanogaster*. PloS Genetics 8(10):e1002905.
1290
1291 Cong Q, Shen J, Zhang J, Li W, Kinch LN, Calhoun JV, Warren AD, Grishin NV. 2021. Genomics
1292 reveals the origins of historical specimens. Molecular Biology and Evolution 38:2166-2176.
1293
1294 Cooper BS, Hammad LA, Montooth KL. 2014. Thermal adaptation of cellular membranes in
1295 natural populations of *Drosophila melanogaster*. Functional Ecology 28:886-894.
1296
1297 Corbett-Detig RB and Hartl DL 2012. Population genomics of inversion polymorphisms in
1298 *Drosophila melanogaster*. PLoS Genetis 8:e1003056.
1299

1300 Croze M, Wollstein A, Bozicevic V, Zivkovic D, Stephan W, Hutter S. 2017. A genome-wide scan
1301 for genes under balancing selection in *Drosophila melanogaster*. BMC Evolutionary Biology
1302 17:15
1303
1304 Dabney J, Meyer M, Paabo S. 2013. Ancient DNA damage. Cold Spring Harbor Perspectives
1305 in Biology 5:a012567.
1306
1307 Daborn PJ, Yen JL, Bogwitz MR, Le Goff G, Feil E, Jeffers S, Tijet N, Perry T, Heckel D, Batterham
1308 P et al. 2002. A single p450 allele associated with insecticide resistance in *Drosophila*. Science
1309 297:2253-2256.
1310
1311 da Silva Ribeiro T, Galvan JA, Pool JE. 2022. Maximum SNP $F_{ST}$ outperforms full-window
1312 statistics for detecting soft sweeps in local adaptation. Genome Biology and Evolution
1313 14:evac143
1314
1315 Depristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel
1316 G, Rivas MA, Hanna M et al. 2011. A framework for variation discovery and genotyping using
1317 next-generation DNA sequencing data. Nature Genetics 43:491-498.
1318
1319 Donlea JM, Ramanan N, Shaw PJ. 2009. Use-dependent plasticity in clock neurons regulates
1320 sleep need in *Drosophila*. Science 324:105-108.
1321
1322 Enderle D, Beisel C, Stadler MB, Gerstung M, Ahtri P, Paro R. 2011. Polycomb preferentially
1323 targets stalled promoters of coding and noncoding transcripts. Genome Research 21:216-226.
1324
1325 Erickson PA, Weller CA, Song DY, Bangerter AS, Schmidt P, Bergland AO. 2020. Unique
1326 genetic signatures of local adaptation over space and time for diapause, an ecologically
1327 relevant complex trait, in *Drosophila melanogaster*. PLoS Genetics 16:e1009110.
1328

1329 Ferretti L, Ramos-Onsins SE, Perez-Enciso M. 2013. Population genomics from pool
1330 sequencing. Molecular Ecology 22:5561-5576.
1331
1332 Fabricius JC. 1776. Genera insectorum: eorumque characteres naturales secundam numerum,
1333 figuram, situm et proportionem, omnium partium oris adiecta mantissa specierum nuper
1334 detectarum. Litteris MF Bartschii.
1335
1336 Garner K, Hunt AN, Koster G, Somerharju P, Groves E, Li M, Raghu P, Holic R, Cockcroft S. 2012.
1337 Phosphatidylinositol transfer protein, cytoplasmic 1 (PITPNC1) binds and transfers
1338 phosphatidic acid. Journal of Biological Chemistry 287: 32263-32276.
1339
1340 Good RT, Gramzow L, Battlay P, Sztal T, Batterham P, Robin C. 2014. The molecular evolution
1341 of cytochrome P450 genes within and between *Drosophila* species. Genome Biology and
1342 Evolution 6:1118-1134.
1343
1344 Greco CM, Cervantes M, Fustin JM, Ito K, Ceglia N, Samad M, Shi J, Koronowski KB, Forne I,
1345 Ranjit S, Gaucher J. 2020. S-adenosyl-l-homocysteine hydrolase links methionine metabolism
1346 to the circadian clock and chromatin remodeling. Science Advances 16:eabc5629.
1347
1348 Grewe F, Kronforst MR, Pierce NE, Moreau CS. 2021. Museum genomics reveals the Xerces
1349 blue butterfly (*Glaucopsyche xerces*) was a distinct species driven to extinction. Biology Letters
1350 17:20210123.
1351
1352 Harbison ST, Kumar S, Huang W, McCoy LJ, Smith KR, Mackay TFC .2019. Genome-wide
1353 association study of circadian behavior in *Drosophila melanogaster*. Behavioral Genetics
1354 49:60-82.
1355
1356 Harris M, Garud NR. 2023. Enrichment of hard sweeps on the X chromosome in *Drosophila*
1357 *melanogaster*. Molecular Biology and Evolution 40:msac268

1358

1359 Hillesheim E, Stearns SC. 1991. The responses of *Drosophila melanogaster* to artificial selection
1360 on body weight and its phenotypic plasticity in two larval food environments. Evolution
1361 45:1909-1923.

1362

1363 Hong W, Mosca TJ, Luo L. 2012 Teneurins instruct synaptic partner matching in an olfactory
1364 map. Nature 484:201-207.

1365

1366 Huang CW, Wang H-D, Bai H, Wu M-S, Yen J-H, Tatar M, Fu T-F, Wang P-Y. 2015. Tequila
1367 regulates insulin-like signaling and extends lifespan in *Drosophila melanogaster*. Journal of
1368 Gerontology A Biological and Medical Sciences 70:1461-1469

1369

1370 Huang W, Lyman RF, Lyman RA, Carbone MA, Harbison ST, Magwire MM, Mackay TFC et al.
1371 2016. Spontaneous mutations and the origin and maintenance of quantitative genetic
1372 variation. eLife 5:e14625

1373

1374 Intra J, Veltri C, De Caro D, Perotti ME, Pasini ME. 2017. In vitro evidence for the participation
1375 of *Drosophila melanogaster* sperm Beta-N-actylglucasaminidases in the interactions with
1376 glycans carrying terminal N-acetylglucosamine residues on the egg's envelopes. Archives of
1377 Insect Biochemistry and Physiology 96:arch21403

1378

1379 Jouganous J, Long W, Ragsdale AP, Gravel S. 2017. Inferring the joint demographic history of
1380 multiple populations: beyond the diffusion approximation. Genetics 206:1549-1567.

1381

1382 Kapun M, Fabian DK, Goudet J, Flatt T. 2016. Genomic evidence for adaptive inversion clines
1383 in *Drosophila melanogaster*. Molecular Biology and Evolution 33: 1317-1336.

1384

1385 Kapun M, Barron MG, Staubach F, Obbard DJ, Wiberg AW, Vieira J, Goubert C, Rota-Stabelli
1386 O, Kankare M, Bogaerts-Marquez M, et al. 2020. Genomic analysis of European *Drosophila*

1387 *melanogaster* population reveals longitudinal structure, continent-wide selection, and
1388 previously unknown DNA viruses. Molecular Biology and Evolution 37:2661-2678
1389
1390 Kapun M, Nunez JCB, Bogaerts-Marequz M, Murga-Moreno J, Paris M, Outten J, Coronado-
1391 Zamora M, Tern C, Rota-Stabelli O, Garcia Guerreiro MP, et al. 2021. *Drosophila* evolution over
1392 space and time (DEST): a new population genomics resource. Molecular Biology and Evolution
1393 38(12):5782-5805.
1394
1395 Keller A. 2007. *Drosophila melanogaster*'s history as a human commensal. Current Biology 17:
1396 R77-81.
1397
1398 Korlević P, McAlister E, Mayho M, Makunin A, Flicek P, Lawniczak MK. 2021. A Minimally
1399 Morphologically Destructive Approach for DNA Retrieval and Whole-Genome Shotgun
1400 Sequencing of Pinned Historic Dipteran Vector Species. Genome Biology and Evolution
1401 13:evab226.
1402
1403 Kyriacou CP, Peixoto AA, Sandrelli F, Costa R, Tauber E. 2008. Clines in clock genes: fine-tuning
1404 circadian rhythms to the environment. Trends in Genetics 24:124-132.
1405
1406 Lack JB, Cardeno CM, Crepeau MW, Taylor W, Corbett-Degit RB, Stevens KA, Langley CH, Pool
1407 JE. 2015. The Drosophila Genome Nexus: a population genomic resource of 623 *Drosophila*
1408 *melanogaster* genomes, including 197 from a single ancestral range population. Genetics
1409 199:1229-1241
1410
1411 Lack JB, Lange JD, Tang AD, Corbett-Detig RB, Pool JE. 2016. A thousand fly genomes: an
1412 expanded *Drosophila* Genome Nexus. Molecular Biology and Evolution 33:3308-3313.
1413
1414 Lande R 1992. Neutral theory of quantative genetic variance in an island model with local
1415 extinction and colonization. Evolution 46:381-389.

1416

1417 Lange JD, Hastide H, Lack JB, Pool JE. 2022. A population genomic assessment of three
1418 decades of evolution in a *Drosophila* population. Molecular Biology and Evolution 39:msab368
1419

1420 Langley CH, Stevens S, Cardeno C, Lee YCG, Schrider DR, Pool JE, Langley SA, Suarez C,
1421 Corbett-Detig RB, Kolaczkowski B et al. 2012. Genomic variation in natural populations of
1422 *Drosophila melanogaster*. Genetics 192: 533-598.
1423

1424 Levin LR, Han PL, Huang PM, Feinstein PG, Davis RL Reed RR. 1992. The *Drosophila* learning
1425 and memory gene *rutabaga* encodes a Ca2+/Calmodulin-responsive adenylyl cyclase. Cell
1426 68:479-489.
1427

1428 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler tranform.
1429 Bioinformatics 25:1754-1760.
1430

1431 Li Y, Hoxha V, Lama C, Dien BH, Vo CN, Dauwalder B. 2011. The hector G-protein coupled
1432 receptor is required in a subset of fruitless neurons for male courtship behavior. PLoS One
1433 6:e28269
1434

1435 Liebl FL, Featherstone DE. 2008. Identification and investigation of *Drosophila* postsynaptic
1436 density homologs. Bioinformatics and Biology Insights, 2:375-388
1437

1438 Linnaeus C. 1746. Fauna Svecica sistens animalia Sveciae regni: qvadrupedia, Aves, Amphibia,
1439 Pisces, Insecta, Vermes, distributa per classes and ordines, genera and species. Cum
1440 differentiis specierum, synonymis auctorum, nominibus incolarum, locis habitationum,
1441 descriptionibus Insectorum. Laurentius Salvius, Stockholm.
1442

1443 Linnaeus C. 1761. Fauna Svecica sistens animalia Sveciae regni: qvadrupedia, Aves, Amphibia,
1444 Pisces, Insecta, Vermes, distributa per classes and ordines, genera and species. Cum

1445 differentiis specierum, synonymis auctorum, nominibus incolarum, locis habitationum,
1446 descriptionibus Insectorum (2nd edition). Laurentius Salvius, Stockholm.
1447
1448 Lunter G, Goodson M. 2011. Stampy: a statistical algorithm for sensitive and fast mapping of
1449 Illumina sequence reads. Genome Research 21: 936-939.
1450
1451 Machado HE, Bergland AO, Taylor R, Tilk S, Behrman E, Dyer K, Fabian DK, Flatt T, González J,
1452 Karasov TL, Kim B, et al. 2021. Broad geographic sampling reveals the shared basis and
1453 environmental correlates of seasonal adaptation in *Drosophila*. eLife 10:e67577.
1454
1455 Magwire MM, Bayer F, Webster CL, Can C, Jiggins FM. 2011. Successive increases in the
1456 resistance of *Drosophila* to viral infection through a transposon insertion followed by
1457 duplication. PLoS Genetics 7:e1002337.
1458
1459 Mallick S, Li H, Lipson M, Gymrek M, Racimo F, Zhao M, Chennagiri N, Nordenfelt S, Tandon
1460 A, Skoglund P et al. 2016. The Simons Genome Diversity Project: 300 genomes from 142
1461 diverse populations. Nature 538:201-206.
1462
1463 Mateo L, Rech GE, Gonzalez J. 2018. Genome-wide patterns of local adaptation in Western
1464 European *Drosophila melanogaster* natural populations. Scientific Reports 8:16143
1465
1466 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler
1467 D, Gabriel S, Daly M, DePristo M. 2010. The genome analysis toolkit: a mapreduce framework
1468 for analyzing next-generation DNA sequencing data. Genome Research 20: 1297-303.
1469
1470 Meigen JW. 1830. Systematische Beschreibung der Bekannten Europaischen Zweiflugeligen
1471 Insekten, Vol. 6. Theil Schulze, Vienna.
1472

1473 Mikheyev AS, Tin MM, Arora J, Seeley TD. 2015. Museum samples reveal rapid evolution by
1474 wild honey bees exposed to a novel parasite. Nature Communications. 6:7991.
1475
1476 Mikheyev AS, Zwick A, Magrath MJ, Grau ML, Qiu L, Su YN, Yeates D. 2017. Museum genomics
1477 confirms that the Lord Howe Island stick insect survived extinction. Current Biology 27:3157-
1478 3161.
1479
1480 Milligan SC, Alb JG, Elaine RB, Bankaitis VA, Hyde DR. 1997. The phosphatidylinositol transfer
1481 protein domain of *Drosophila* retinal degeneration B protein is essential for photoreceptor cell
1482 survival and recovery from light stimulation. Journal of Cell Biology 139: 351-363.
1483
1484 Mosca TJ, Hong W, Dani VS, Favaloro V, Luo L. 2012. Trans-synaptic teneurin signalling in
1485 neuromuscular synapse organization and target choice. Nature 484:237-241.
1486
1487 Orengo DJ, Aguade M. 2007. Genome scans of variation and adaptive change: extended
1488 analysis of a candidate locus close to the *phantom* gene region in *Drosophila melanogaster*.
1489 Molecular Biology and Evolution 24:1122-1129.
1490
1491 Parejo M, Wragg D, Henriques D, Charriere JD, Estonba A. 2020. Digging into the genomic
1492 past of Swiss honey bees by whole-genome sequencing museum specimens. Genome Biology
1493 and Evolution. 12:2535-2551.
1494
1495 Parkhito AA, Binari R, Zhang N, Asara JM, Demotis F, Perrimon N. 2016. Tissue-specific down-
1496 regulation of S-adenosyl-homocysteine via suppression of dAhcyL1/dAhcyL2 extends health
1497 span and life span in *Drosophila*. Genes and Development 30:1409-1422.
1498
1499 Pool JE, Corbett-Detig RB, Sugino RP, Stevens KA, Cardeno CM, Crepeau MW, Duchen P,
1500 Emerson JJ, Saelao P, Begun DJ, Langley CH. 2012. Population genomics of sub-Saharan

1501 *Drosophila melanogaster*: African diversity and non-African admixture. PLoS Genetics
1502 8:e1003080.
1503
1504 Pool JE. 2015. The mosaic ancestry of the *Drosophila* Genetic Reference Panel and the *D.*
1505 *melanogaster* reference genome reveals a network of epistatic fitness interactions. Molecular
1506 Biology and Evolution 32:3236-3251.
1507
1508 Pool JE, Braun DT, Lack JB. 2017. Parallel evolution of cold tolerance within *Drosophila*
1509 *melanogaster*. Molecular Biology and Evolution 34:349-360.
1510
1511 Raxworthy CJ, Smith BT 2021. Mining museums for historical DNA: advances and challenges
1512 in museomics. Trends in Ecology and Evolution 36:1049-1060.
1513
1514 Rech GE, Radio S, Guirao-Rico S, Aguilera L, Horvath V, Green L, Lindstadt H, Jamilloux V,
1515 Quesneville H, Gonzales J, et al. 2022. Population-scale short read sequencing uncovers
1516 transposable elements associated with gene expression variation and adaptive signatures in
1517 *Drosophila*. Nature Communications 13:1948.
1518
1519 Reynolds J, Weir BS, CC Cockerham CC. 1983. Estimation of the coancestry coefficient: basis
1520 for a short-term genetic distance. Genetics 105: 767-779.
1521
1522 Rivas GBS, Zhou J, Merlin C, Hardin PE. 2021. Clockwork orange promotes CLOCK-CYCLE
1523 activation via the putative *Drosophila* ortholog of clock interacting protein circadian. Current
1524 Biology 31:4207-4219.
1525
1526 Saunders DS, Henrich VC, Gilbert LI. 1989. Induction of diapause in *Drosophila melangoaster*:
1527 photoperiodic regulation and the impact of arrhythmic clock mutations on time measurement.
1528 Proceedings of the National Academy of Sciences (USA) 86:3748-3752.
1529

1530 Scanlan JL, Glendhill-Smith RS, P Battley P, Robin C. 2020. Genomic and transcriptomic
1531 analyses in Drosophila suggest that the ecdysteroid kinase-like (EcKL) gene family encodes the
1532 'detoxification-by-phosphorylation' enzymes of insects. Insect Biochemistry and Molecular
1533 Biology 123: 103429
1534
1535 Scheunemann L, Jost E, Richlitzki A, Day JP, Sebastian S, Thum AS, Efetova M, Davies S-A,
1536 Schwarzel M. 2012. Consolidated and labile door memory are separately encoded in the
1537 *Drosophila* brain. Journal of Neuroscience 32:17163-17171
1538
1539 Schmidt JM, Good RT, Appleton B, Sherrad J, Raymart GC, Bogwitz MR, Martin J, Daborn PJ,
1540 Goddard ME, Batterham, Robin C. 2010. Copy number variation and transposable elements
1541 feature in recent ongoing adaptation at the *Cyp6g1* locus. PLoS Genetics 6:e1000998
1542
1543 Schmidt PS, Matzkin L, Ippolito M, Eanes WF. 2005. Geographic variation in diapause
1544 incidence, life-history traits, and climatic adaptation in *Drosophila melanogaster*. Evolution
1545 59:1721-1732.
1546
1547 Shalaby NA, Pinzon JH, Narayanan AS, Jin EG, Ritz MP, Dove RJ, Wolfenberg H, Rodan AR,
1548 Buszczak M, Routhenfluh A. 2018. JmjC domain proteins modulate circadian behaviors and
1549 sleep in *Drosophila*. Scientific Reports 8:815
1550
1551 Spitze K 1993. Population structure in *Daphnia obtusa*: quantitative genetic and allozyme
1552 variation. Genetics 135:367-374.
1553
1554 Sprengelmeyer QD, Mansourian S, Lange JD, Matute DR, Cooper BS, Jirle EV, Stensmyr MC,
1555 Pool JE. 2020. Recurrent collection of *Drosophila melanogaster* from wild African environments
1556 and genomic insights into species history. Molecular Biology and Evolution 37:2775.
1557

1558 Staats M, Erkens RH, van de Vossenberg B, Wieringa JJ, Kraaijeveld K, Stielow B, Geml J,
1559 Richardson JE, Bakker FT. 2013. Genomic treasure troves: complete genome sequencing of
1560 herbarium and insect museum specimens. PLOS ONE 8:e69189.
1561
1562 Svetec N, Zhao L, Saelo P, Chiu JC Begun DJ. 2015. Evidence that natural selection maintains
1563 genetic variation for sleep in *Drosophila melanogaster*. BMC Evolutionary Biology 15:316
1564
1565 Tong JJ, Schriner SE, McCleary D, Day BJ, Wallace DC 2007. Life extension through
1566 neurofibromin mitochondrial regulation and antioxidant therapy for neurofibromatosis-1 in
1567 *Drosophila melanogaster*. Nature Genetics 39:476-485.
1568
1569 Turelli M, Hoffmann AA. Cytoplasmic incompatibility in *Drosophila simulans*: dynamics and
1570 parameter estimates from natural populations. Genetics 140:1319-1338.
1571
1572 Van Der Valk T, Díez-del-Molino D, Marques-Bonet T, Guschanski K, Dalén L. 2019. Historical
1573 genomes reveal the genomic consequences of recent population decline in eastern gorillas.
1574 Current Biology 29:165-170.
1575
1576 Voigt S, Laurent S, M Litovchenko, W. Stephan W. 2015. Positive selection at the polyhomeotic
1577 locus led to decreased thermosensitivity of gene expression in temperate Drosophila
1578 melanogaster. Genetics 200:591-599.
1579
1580 Wang J, Lee CH, Lin S, Lee T. 2006. Steroid hormone-dependent transformation of
1581 polyhomeotic mutant neurons in the Drosophila brain. Development 133:1231-1240.
1582
1583 Warren JT, Petryk A, Marques G, Parvy J-P, Shinoda T, Itoyama K, Kobayashi J, Jarcho M, Li Y,
1584 O'Connor MB, Dauphin-Villemart C, Gilbert LI. 2004. Phantom encodes the 25-hydroxylase of
1585 *Drosophila melanogaster* and *Bombyx mori*: a P450 enzyme critical in ecdysone biosynthesis.
1586 Insect Biochemistry and Molecular Biology 34:991-1010.

1587

1588 Werthebach M, Stewart FA, Gahlen A, Mettler-Altmann T, Akhtar I, Maas-Enriquez K, Droste A,
1589 Eichmann TO, Poschmann G, Stuhler K, Beller M. 2019. Control of *Drosophila* growth and
1590 survival by the lipid droplet-associated protein CG9186/Sturkopf. Cell Reports 26:3726-3740.
1591

1592 Whitlock MC 1999. Neutral additive genetic variation in a metapopulation. Genetical Research
1593 74:215-221.
1594

1595 Wohns AW, Wong Y, Jeffery B, Akbari A, Mallick S, Pinhasi R, Patterson N, Reich D, Kelleher J,
1596 McVean G. 2022. A unified genealogy of modern and ancient genomes. Science
1597 375:eabi8264.
1598

1599 Woodard C, E Alcorta E, J Carlson 1992. The *rdgB* gene in *Drosophila*: a link between vision
1600 and olfaction. Journal of Neurogenetics 8: 17-31
1601

1602 Yang MA, Fu Q. 2018. Insights into modern human prehistory using ancient genomes. Trends
1603 in Genetics 34:184-96.
1604

1605 Yassin A, Debat V, Bastide H, Pool JE. 2016. Recurrent specialization on a toxic fruit in an island
1606 *Drosophila* population. Proceedings of the National Academy of Sciences (USA) 113:4771-
1607 4776.
1608

1609 Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, Xu X, Jian H, Vinckenbosch N,
1610 Korneliussen TS, et al. 2010. Sequencing fifty human exomes reveals adaptation to high
1611 altitude. Science 329: 75-79.
1612

1613 Zetterstedt JW. 1842-1860. Diptera Scandinaviae disposita et descripta. Lundae ex Officina
1614 Lundbergiana. Vol 4-14: sumtibus rebus

1615 **Supplementary Information:**

1616

1617 **Table S1. Sample Information and Genomic Sequencing Statistics.** Available collection
1618 information is given for each museum specimen. All early 1800s Sweden flies (including the
1619 one labelled Småland) were ultimately inferred to have come from the same collection event.
1620 Sequencing statistics including read count and lengths, depth of coverage, genomic coverage,
1621 singleton rate, and proportion of singletons with A or T are also given. Sample 19SL2 was
1622 excluded from most analyses because of its low sequencing depth and genomic coverage.
1623 The sex of each specimen was inferred from the relative levels of sequencing depth on the X
1624 chromosome and the autosomes. Genomic characteristics of a representative set of genomes
1625 from modern material (from Lyon, France - Lack et al. 2016 MBE) are provided, as depicted in
1626 Figure 1.

1627

1628 **Table S2: Evidence of identity-by-descent from relatedness and inbreeding and genomic**
1629 **masking proportions.**

1630 **Tab A:** Mean pairwise genetic distance between X chromosomes from museum and modern
1631 genomes (excluding sites with missing data from any individual). Unusually low values are
1632 highlighted. Mean distance against a panel of modern genomes is indicated at the bottom.

1633 **Tab B:** Proportion of each genome masked due to pairwise IBD with other genomes, given as
1634 a fraction of the total genome and each chromosome arm.

1635 **Tab C:** Proportion of each genome masked due to inbreeding IBD, as indicated by runs of
1636 minimal heterozygosity, for each full genome and each chromosome arm.

1637

1638 **Table S3. Regions of each museum fly genome that were masked due to relatedness**
1639 **IBD.** Release 5 positions are given in separate tabs for each chromosome arm.

1640

1641 **Table S4. Regions of each museum fly genome that were masked due to inbreeding IBD.**
1642 Release 5 positions are given in separate tabs for each chromosome arm.

1643

1644 **Table S5. Results of Principle Components Analysis of Genetic Structure.** Loadings for
1645 each genome are given for all 35 principle components. The X chromomsome was analyzed
1646 to avoid the influence of autosomal inversions. Up to 5 apparently unrelated individuals were
1647 included per population sample.
1648
1649 **Table S6. The estimated frequency of each of the most common inversions in museum**
1650 **and modern populations.** Inversion frequency estimates and results from each SNP reported
1651 to be associated with each inversion (by Kapun et al. 2016) are given in separate tabs.
1652 Estimation of inversion frequency was based on the median of SNP frequencies, in order to
1653 avoid spurious non-zero estimates.
1654
1655 **Table S7. Effective population size estimates from drift-only models do not predict**
1656 **observed nucleotide diversity.**
1657 **Top:** Mean frequency differences and haploid effective population size estimates from drift-
1658 only models for each Lund time interval.
1659 **Bottom:** Comparison of nucleotide diversity from simulations (using estimated Ne) versus
1660 observed data, for arm 3L.
1661
1662 **Table S8. Outlier regions from read depth QPBE as a potential indicator of structural**
1663 **variants at differing frequencies between samples.**
1664 **Tab A:** For the 1800s time interval (1800s as focal population, Lund 1933 and Lyon as
1665 outgroups), QPBE outlier regions based on the upper 1% quantile and overlapping genes are
1666 given.
1667 **Tab B:** For the 1900s time interval (Lund 1933 as focal population, Lyon and Stockholm as
1668 outgroups), QPBE outlier regions based on the upper 1% quantile and overlapping genes are
1669 given.
1670
1671 **Table S9. Outlier regions from PBE reflecting windows with SNP frequencies at strongly**
1672 **differing frequencies between samples.**

1673 **Tab A:** For the 1800s time interval (1800s as focal population, Lund 1933 and Lund 2010s),
1674 PBE outlier regions based on the upper 1% quantile and overlapping genes are given.
1675 **Tab B:** For the 1900s time interval (Lund 1933 as focal population, Lund 2010s and Lyon as
1676 outgroups), PBE outlier regions based on the upper 1% quantile and overlapping genes are
1677 given.
1678
1679 **Table S10. Gene Ontology (GO) Enrichment based on PBE Outlier Regions.**
1680 **Tab A:** Gene ontology enrichment for the 1800s-focused PBE outliers.
1681 **Tab B:** Gene ontology enrichment for the 1900s-focused PBE outliers.
1682 For each GO category, the total number of genes in the analyzed regions is given, along with
1683 the number of outlier regions associated with these genes, and the identities of these outlier-
1684 overlapping genes. A permutation P value is also given for each GO category, which does not
1685 attempt to correct for the number of GO categories tested.

**Figure 1.**

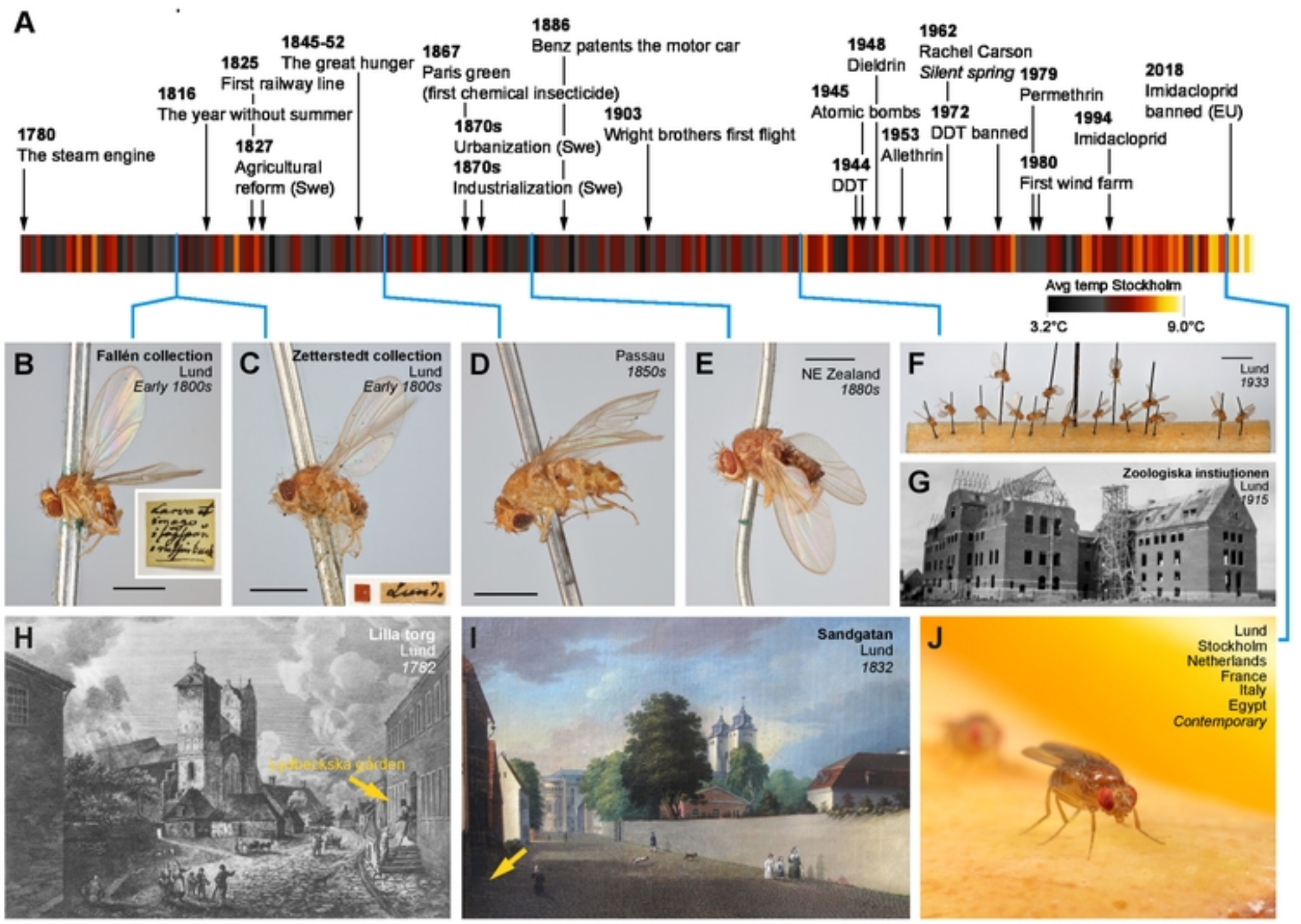



Figure 1

**Figure 2.**

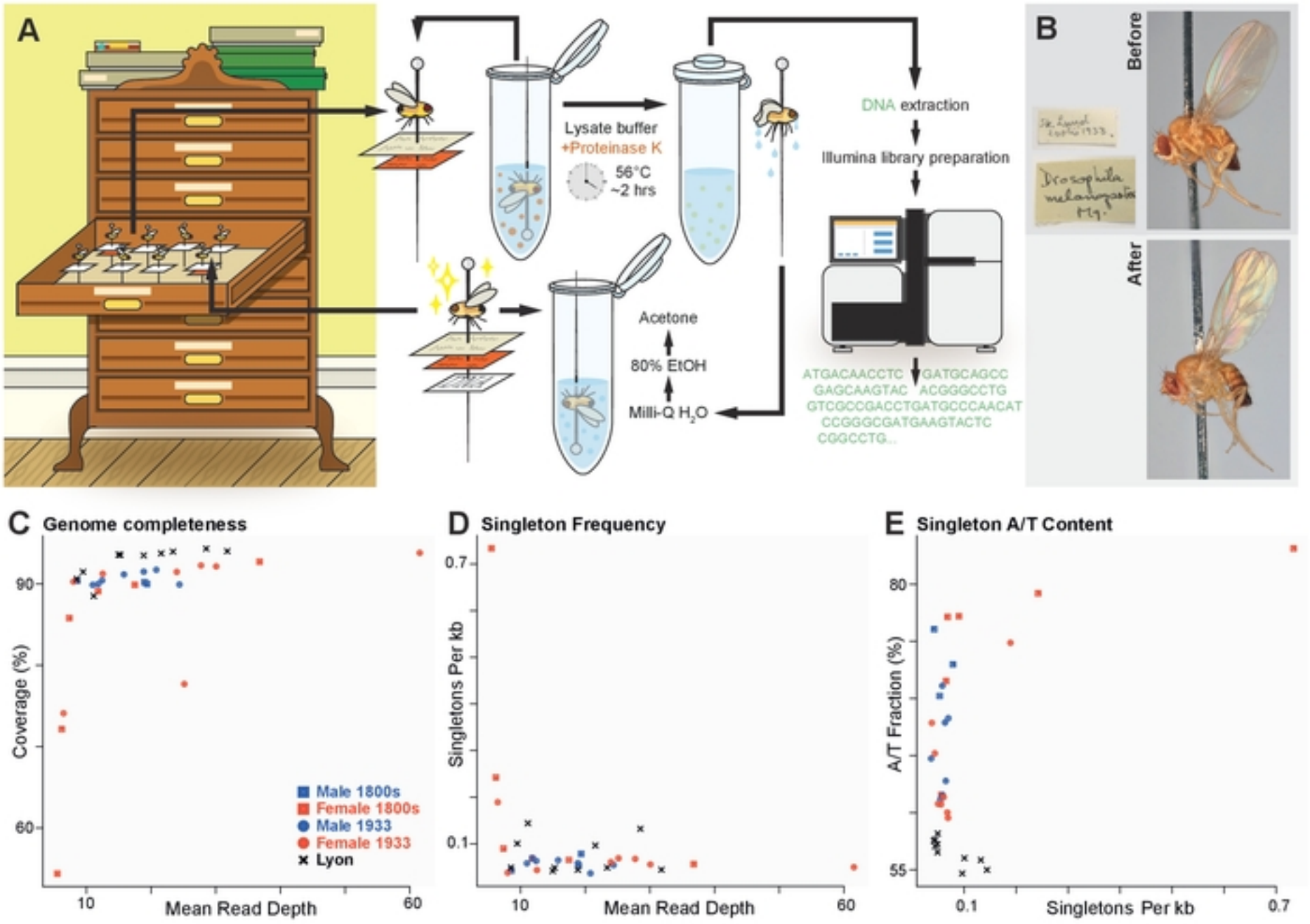
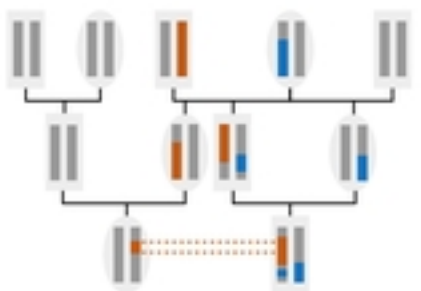



Figure 2

## Figure 3.

**A**

**Identity-By-Decent mapping**



Relatedness IBD
Inbreeding IBD

**Inbreeding and relatedness - 1800s genomes**

| | 18DZ5 | 18SL4 | 18SL6 | 18SL12 | *18SL10* | *18SL11* | 18SL13 | 18GP9 | 18GP25 |
|---|---|---|---|---|---|---|---|---|---|
| **18DZ5** | 0.000 | 0.000 | 0.003 | 0.014 | 0.007 | 0.000 | 0.000 | 0.000 | 0.000 |
| **18SL4** | | 0.000 | 0.145 | 0.140 | 0.783 | 0.000 | 0.371 | 0.000 | 0.000 |
| **18SL6** | | | 0.000 | 0.151 | 0.029 | 0.191 | 0.134 | 0.000 | 0.000 |
| **18SL12** | | | | 0.000 | 0.029 | 0.077 | 0.042 | 0.000 | 0.000 |
| *18SL10* | | | | | 0.019 | 0.040 | 0.337 | 0.000 | 0.000 |
| *18SL11* | | | | | | 0.000 | 0.319 | 0.000 | 0.000 |
| 18SL13 | | | | | | | 0.006 | 0.000 | 0.000 |
| 18GP9 | | | | | | | | 0.000 | 0.060 |
| 18GP25 | | | | | | | | | 0.000 |

Proportion of IBD

**Lower quality genome**
Diploid X
*Haploid X*

**B**

**Inbreeding and relatedness - 1933 genomes**

| | 19SL3 | 19SL7 | *19SL1* | *19SL8* | *19SL14* | *19SL15* | *19SL16* | *19SL17* | 19SL18 | *19SL19* | 19SL20 | *19SL21* | 19SL22 | 19SL23 | 19SL24 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **19SL3** | 0.000 | 0.068 | 0.037 | 0.015 | 0.006 | 0.185 | 0.011 | 0.000 | 0.079 | 0.042 | 0.181 | 0.007 | 0.050 | 0.115 | 0.194 |
| **19SL7** | | 0.000 | 0.261 | 0.027 | 0.429 | 0.187 | 0.183 | 0.029 | 0.082 | 0.084 | 0.056 | 0.308 | 0.153 | 0.278 | 0.007 |
| *19SL1* | | | 0.091 | 0.087 | 0.272 | 0.123 | 0.277 | 0.076 | 0.375 | 0.011 | 0.082 | 0.389 | 0.083 | 0.149 | 0.003 |
| *19SL8* | | | | 0.006 | 0.059 | 0.052 | 0.080 | 0.007 | 0.000 | 0.040 | 0.003 | 0.043 | 0.010 | 0.003 | 0.000 |
| *19SL14* | | | | | 0.206 | 0.081 | 0.200 | 0.163 | 0.054 | 0.042 | 0.030 | 0.380 | 0.216 | 0.281 | 0.016 |
| *19SL15* | | | | | | 0.079 | 0.284 | 0.176 | 0.019 | 0.190 | 0.024 | 0.060 | 0.169 | 0.098 | 0.081 |
| *19SL16* | | | | | | | 0.030 | 0.091 | 0.226 | 0.006 | 0.042 | 0.205 | 0.243 | 0.248 | 0.015 |
| *19SL17* | | | | | | | | 0.084 | 0.000 | 0.055 | 0.090 | 0.140 | 0.039 | 0.118 | 0.051 |
| 19SL18 | | | | | | | | | 0.355 | 0.022 | 0.124 | 0.358 | 0.102 | 0.083 | 0.269 |
| *19SL19* | | | | | | | | | | 0.454 | 0.054 | 0.139 | 0.064 | 0.056 | 0.010 |
| 19SL20 | | | | | | | | | | | 0.000 | 0.321 | 0.159 | 0.044 | 0.325 |
| *19SL21* | | | | | | | | | | | | 0.000 | 0.180 | 0.344 | 0.216 |
| 19SL22 | | | | | | | | | | | | | 0.037 | 0.072 | 0.057 |
| 19SL23 | | | | | | | | | | | | | | | 0.015 |
| 19SL24 | | | | | | | | | | | | | | | 0.000 |

Figure 3

# Figure 4.

**A**

**Genetic distance**

| | *Lund 1800s* | *Lund 1933* | *Lund 2010s* | *Lyon* | *Stockholm* | *Netherlands* | *Italy* | *Egypt* |
|---|---|---|---|---|---|---|---|---|
| *Lund 1800s* | ***0.00319*** | 0.00410 | 0.00387 | 0.00400 | 0.00425 | 0.00402 | 0.00410 | 0.00425 |
| *Lund 1933* | 0.0585 | **0.00260** | 0.00353 | 0.00373 | 0.00392 | 0.00373 | 0.00374 | 0.00391 |
| *Lund 2010s* | 0.0455 | 0.0676 | **0.00234** | 0.00346 | 0.00420 | 0.00347 | 0.00402 | 0.00356 |
| *Lyon* | 0.0301 | 0.0651 | 0.0414 | **0.00263** | 0.00384 | 0.00365 | 0.00369 | 0.00381 |
| *Stockholm* | 0.0438 | 0.0632 | 0.0420 | 0.0399 | **0.00291** | 0.00385 | 0.00386 | 0.00391 |
| *Netherlands* | 0.0474 | 0.0792 | 0.0588 | 0.0527 | 0.0522 | **0.00255** | 0.00370 | 0.00377 |
| *Italy* | 0.0435 | 0.0581 | 0.0402 | 0.0400 | 0.0336 | 0.0568 | **0.00269** | 0.00378 |
| *Egypt* | 0.0886 | 0.0794 | 0.0701 | 0.0807 | 0.0549 | 0.0842 | 0.0652 | **0.00262** |

$F_{ST}$

Rate of pairwise sequence differenes ($D_{xy}$)

Genetic differentiation summarized by $F_{ST}$

**Diagonal:** Rate of pairwise sequence differences within populations (π)

**B**

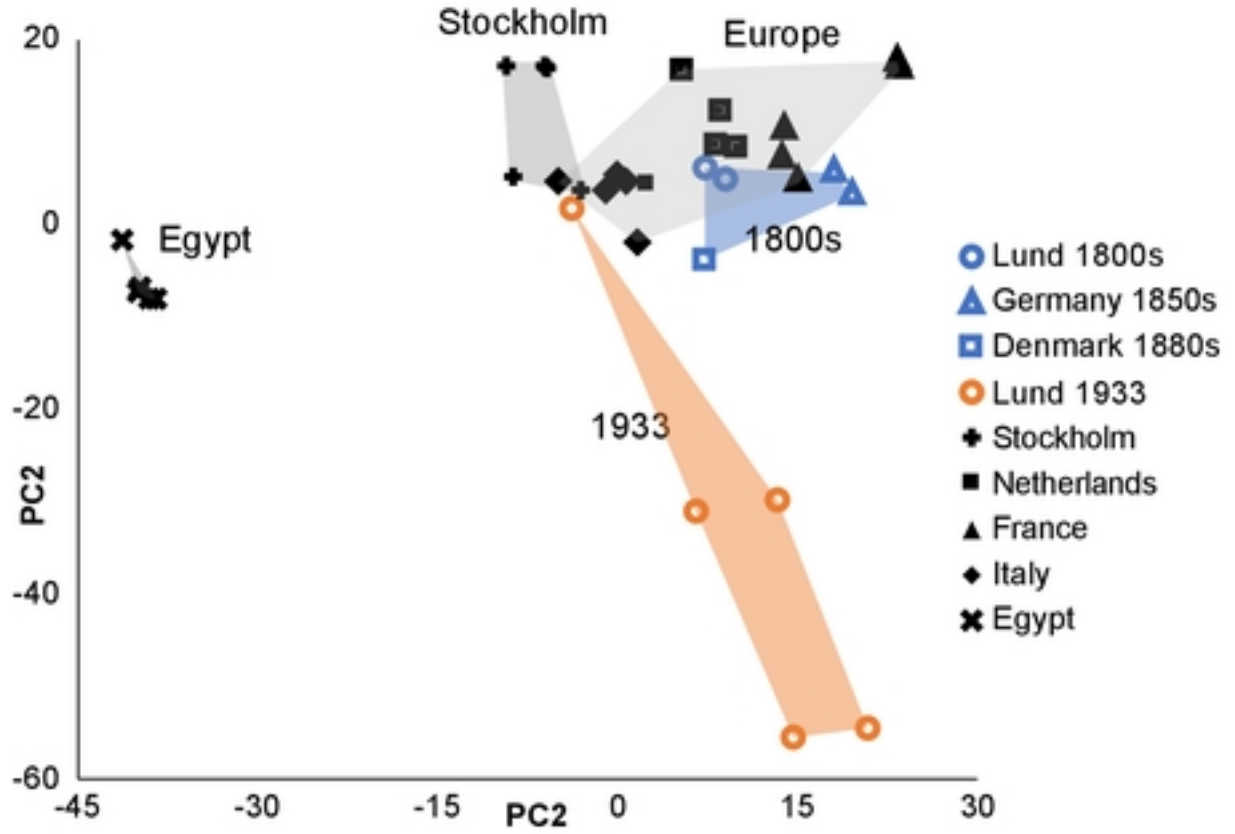



Figure 4

**Figure 5.**

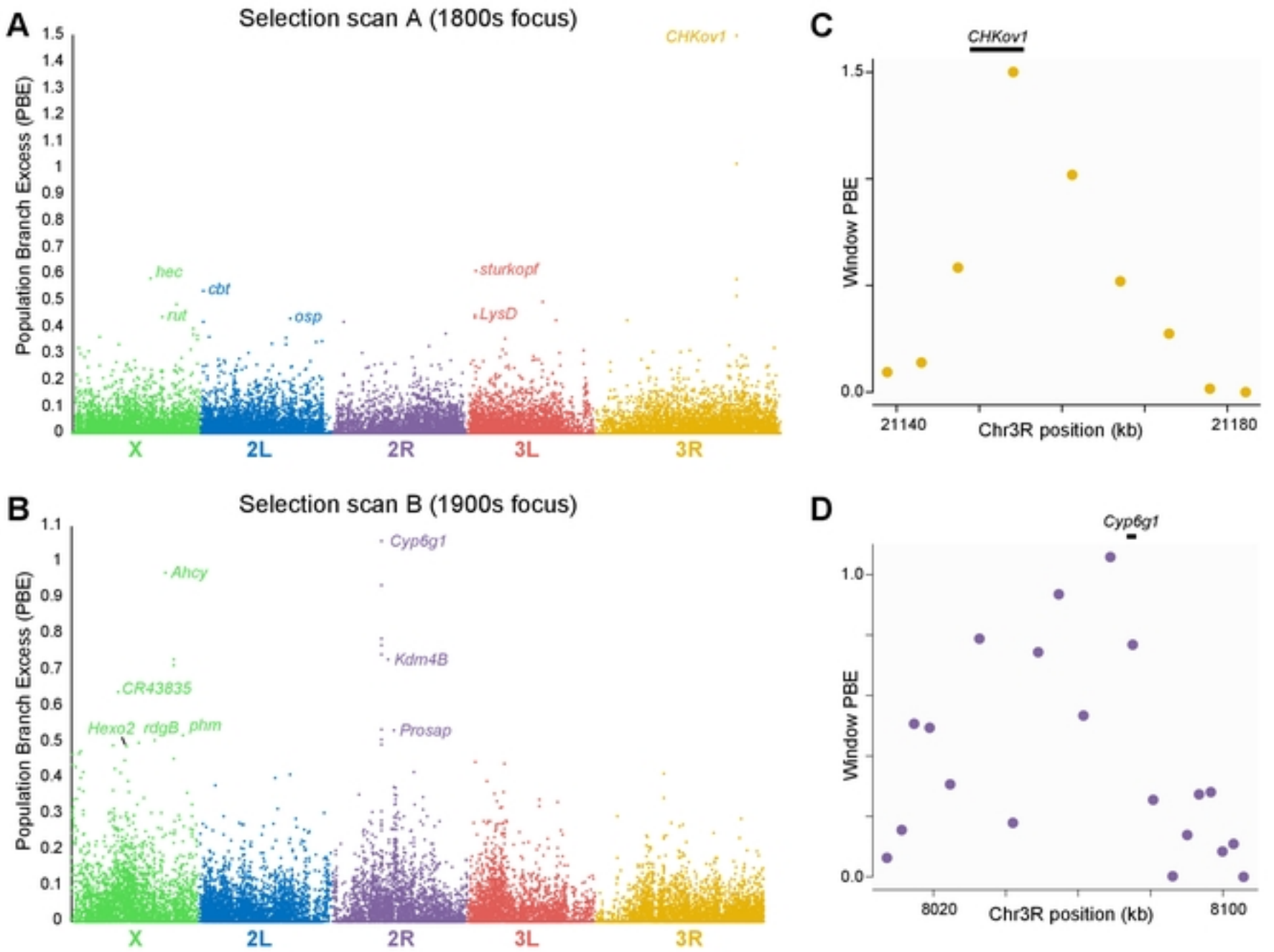



Figure 5

## Figure 6.

### 1800s-focused scan

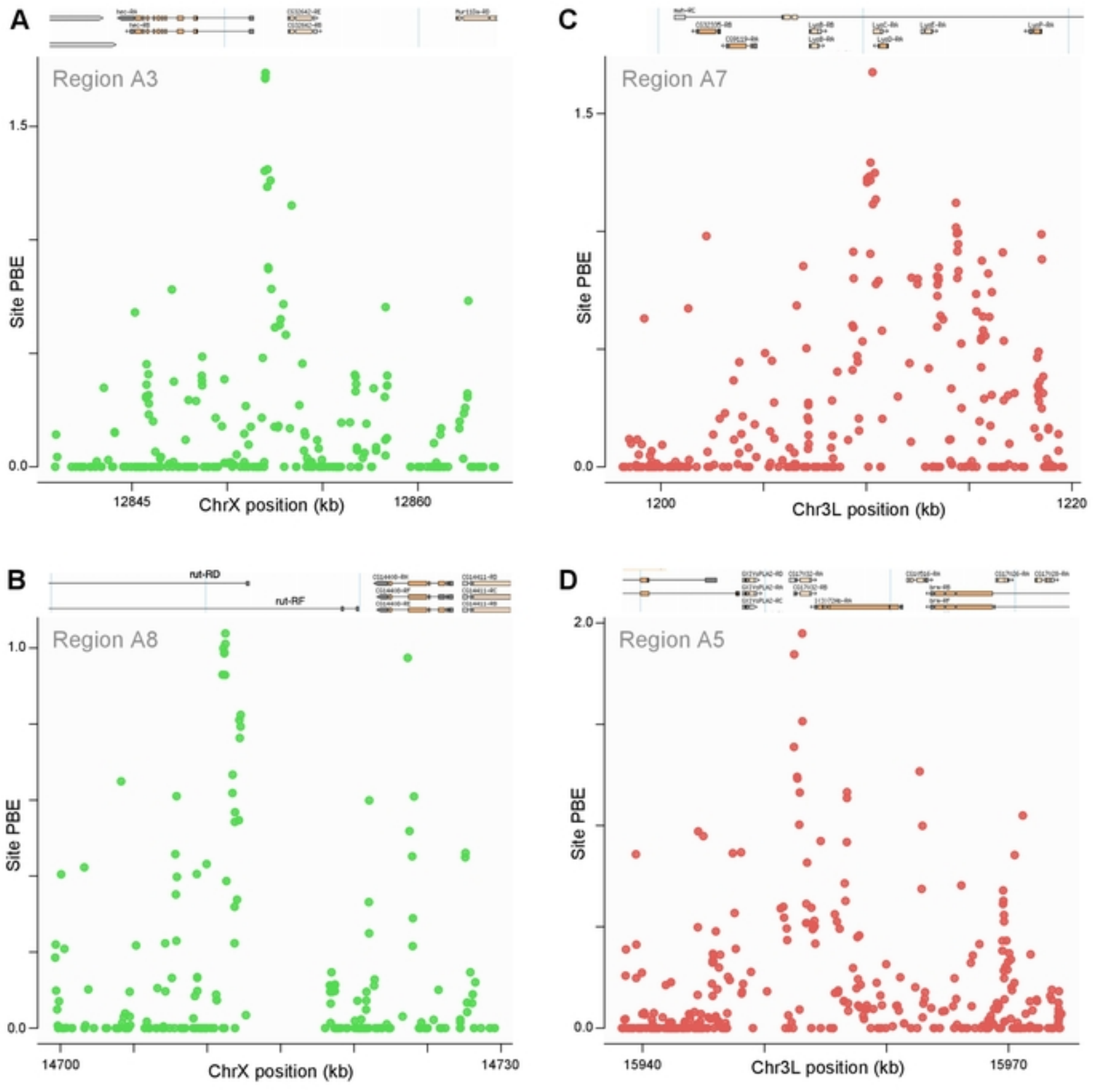



Figure 6

**Figure 7.**

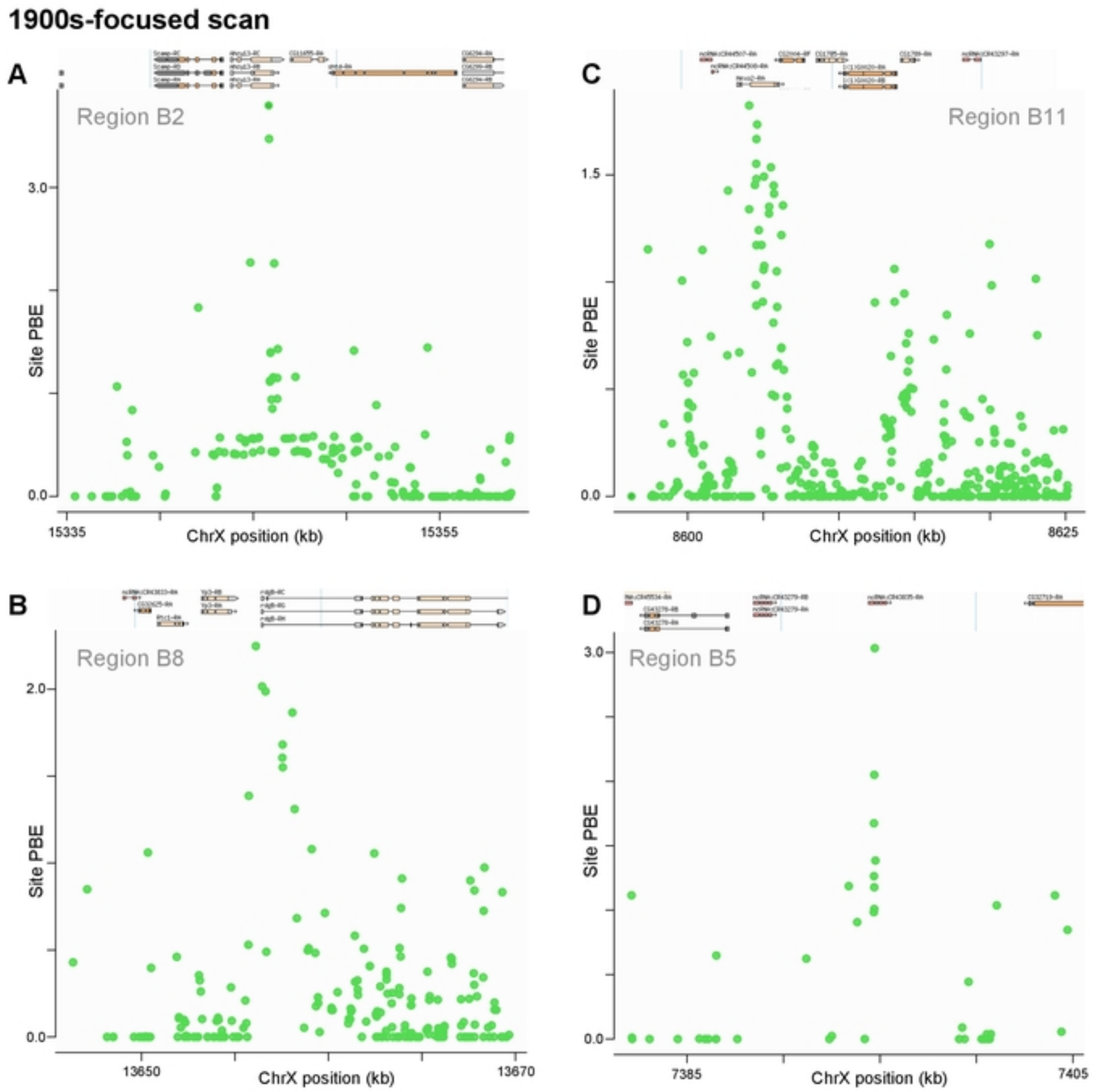



Figure 7