Genomic analysis of European *Drosophila melanogaster* populations reveals longitudinal structure, continent-wide selection, and previously unknown DNA viruses

Martin Kapun[1,2,3,4,†,*], Maite G. Barrón[1,5,*], Fabian Staubach[1,6,§], Darren J. Obbard[1,7], R. Axel W. Wiberg[1,8,9], Jorge Vieira[1,10,11], Clément Goubert[1,12,13], Omar Rota-Stabelli[1,14], Maaria Kankare[1,15,§], María Bogaerts-Márquez[1,5], Annabelle Haudry[1,12], Lena Waidele[1,6], Iryna Kozeretska[1,16,17,], Elena G. Pasyukova[1,18], Volker Loeschcke[1,19], Marta Pascual[1,20], Cristina P. Vieira[1,10,11], Svitlana Serga[1,16], Catherine Montchamp-Moreau[1,21], Jessica Abbott[1,22], Patricia Gibert[1,12], Damiano Porcelli[1,23], Nico Posnien[1,24], Alejandro Sánchez-Gracia[1,20], Sonja Grath[1,25], Élio Sucena[1,26,27], Alan O. Bergland[1,28,§], Maria Pilar Garcia Guerreiro[1,29], Banu Sebnem Onder[1,30], Eliza Argyridou[1,25], Lain Guio[1,5], Mads Fristrup Schou[1,19,22], Bart Deplancke[1,31], Cristina Vieira[1,12], Michael G. Ritchie[1,8], Bas J. Zwaan[1,32], Eran Tauber[1,33], Dorcas J. Orengo[1,20], Eva Puerma[1,20], Montserrat Aguadé[1,20], Paul S. Schmidt[1,34,§], John Parsch[1,25], Andrea J. Betancourt[1,35], Thomas Flatt[1,2,3,†,*,§], Josefa González[1,5,†,*,§]

[1] The European *Drosophila* Population Genomics Consortium (*DrosEU*). [2] Department of Ecology and Evolution, University of Lausanne, CH-1015 Lausanne, Switzerland. [3] Department of Biology, University of Fribourg, CH-1700 Fribourg, Switzerland. [4] Current affiliations: Department of Evolutionary Biology and Environmental Sciences, University of Zürich, CH-8057 Zürich, Switzerland; Division of Cell and Developmental Biology, Medical University of Vienna, AT-1090 Vienna, Austria. [5] Institute of Evolutionary Biology, CSIC- Universitat Pompeu Fabra, Barcelona, Spain. [6] Department of Evolutionary Biology and Ecology, University of Freiburg, 79104 Freiburg, German. [7] Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, United Kingdom. [8] Centre for Biological Diversity, School of Biology, University of St. Andrews, St Andrews, United Kingdom. [9] Department of Environmental Sciences, Zoological Institute, University of Basel, Basel, CH-4051, Switzerland.[10] Instituto de Biologia Molecular e Celular (IBMC) University of Porto, Porto, Portugal. [11] Instituto de Investigação e Inovação em Saúde (I3S), University of Porto, Porto, Portugal. [12] Laboratoire de Biométrie et Biologie Evolutive, UMR CNRS 5558, University Lyon 1, Lyon, France. [13] Department of Molecular Biology and Genetics, 107 Biotechnology Building, Cornell

University, Ithaca, New York 14853, USA. [14] Research and Innovation Centre, Fondazione Edmund Mach, San Michele all' Adige, Italy. [15] Department of Biological and Environmental Science, University of Jyväskylä, Jyväskylä, Finland. [16] General and Medical Genetics Department, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine. [17] State Institution National Antarctic Center of Ministry of Education and Science of Ukraine, 16 Taras Shevchenko Blvd., 01601, Kyiv, Ukraine. [18] Laboratory of Genome Variation, Institute of Molecular Genetics of RAS, Moscow, Russia. [19] Department of Bioscience - Genetics, Ecology and Evolution, Aarhus University, Aarhus C, Denmark. [20] Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Spain. [21] Laboratoire Evolution, Génomes, Comportement et Ecologie (EGCE) UMR 9191 CNRS - UMR247 IRD - Université Paris Sud - Université Paris Saclay. 91198 Gif sur Yvette Cedex, France. [22] Department of Biology, Section for Evolutionary Ecology, Lund, Sweden. [23] Department of Animal and Plant Sciences, Sheffield, United Kingdom. [24] Universität Göttingen, Johann-Friedrich-Blumenbach-Institut für Zoologie und Anthropologie, Göttingen, Germany. [25] Division of Evolutionary Biology, Faculty of Biology, Ludwig-Maximilians-Universität München, Planegg, Germany. [26] Instituto Gulbenkian de Ciência, Oeiras, Portugal. [27] Departamento de Biologia Animal, Faculdade de Ciências da Universidade de Lisboa, Lisboa, Portugal. [28] Department of Biology, University of Virginia, Charlottesville, VA, USA. [29] Departament de Genètica i Microbiologia, Universitat Autònoma de Barcelona, Barcelona, Spain. [30] Department of Biology, Faculty of Science, Hacettepe University, Ankara, Turkey. [31] Laboratory of Systems Biology and Genetics, EPFL-SV-IBI-UPDEPLA, CH-1015 Lausanne, Switzerland. [32] Laboratory of Genetics, Department of Plant Sciences, Wageningen University, Wageningen, Netherlands. [33] Department of Evolutionary and Environmental Biology and Institute of Evolution, University of Haifa, Haifa, Israel. [34] Department of Biology, University of Pennsylvania, Philadelphia, USA. [35] Institute of Integrative Biology, University of Liverpool, Liverpool, United Kingdom.

[†] Co-correspondence: martin.kapun@uzh.ch, thomas.flatt@unifr.ch, josefa.gonzalez@ibe.upf-csic.es

[*] These authors contributed equally to this work

[§] Members of the *Drosophila* Real Time Evolution (Dros-RTEC) Consortium

Abstract

Genetic variation is the fuel of evolution, with standing genetic variation especially important for short-term evolution and local adaptation. To date, studies of spatio-temporal patterns of genetic variation in natural populations have been challenging, as comprehensive sampling is logistically difficult, and sequencing of entire populations costly. Here, we address these issues using a collaborative approach, sequencing 48 pooled population samples from 32 locations, and perform the first continent-wide genomic analysis of genetic variation in European *Drosophila melanogaster*. Our analyses uncover longitudinal population structure, provide evidence for continent-wide selective sweeps, identify candidate genes for local climate adaptation, and document clines in chromosomal inversion and transposable element frequencies. We also characterise variation among populations in the composition of the fly microbiome, and identify five new DNA viruses in our samples.

Introduction

Understanding processes that influence genetic variation in natural populations is fundamental to understanding the process of evolution (Dobzhansky 1970; Lewontin 1974; Kreitman 1983; Kimura 1984; Hudson *et al.* 1987; McDonald & Kreitman 1991; Adrian & Comeron 2013). Until recently, technological constraints have limited studies of natural genetic variation to small regions of the genome and small numbers of individuals. With the development of population genomics, we can now analyse patterns of genetic variation for large numbers of individuals genome-wide, with samples structured across space and time. As a result, we have new insight into the evolutionary dynamics of genetic variation in natural populations (e.g., Hohenlohe *et al.* 2010; Cheng *et al.* 2012; Begun *et al.* 2007; Pool *et al.* 2012; Harpur *et al.* 2014; Zanini *et al.* 2015). But, despite this technological progress, extensive large-scale sampling and genome sequencing of populations remains prohibitively expensive and too labor-intensive for most individual research groups.

Here, we present the first comprehensive, continent-wide genomic analysis of genetic variation of European *Drosophila melanogaster*, based on 48 pool-sequencing samples from 32 populations collected in 2014 (fig. 1) by the European *Drosophila* Population Genomics Consortium (*DrosEU*;

https://droseu.net). *D. melanogaster* offers several advantages for genomic studies of evolution in space and time. It boasts a relatively small genome, a broad geographic range, a multivoltine life history which allows sampling across generations on short timescales, simple standard techniques for collecting wild samples, and a well-developed context for population genomic analysis (e.g., Powell 1997; Keller 2007; Hales *et al.* 2015). Importantly, this species is studied by an extensive international research community, with a long history of developing shared resources (Larracuente & Roberts 2015; Bilder & Irvine 2017; Haudry *et al.* 2020).

Our study complements and extends previous studies of genetic variation in *D. melanogaster*, both from its native range in sub-Saharan Africa and from its world-wide expansion as a human commensal. The expansion into Europe is thought to have occurred approximately 4,100 - 19,000 years ago and into North America and Australia in the last few centuries (e.g., Lachaise *et al.* 1988; David & Capy 1988; Li & Stephan 2006; Keller 2007; Sprengelmeyer *et al* 2018; Kapopoulou *et al.* 2018a; Arguello *et al.* 2019). The colonization of novel habitats and climate zones on multiple continents makes *D. melanogaster* especially useful for studying parallel local adaptation, with previous studies finding pervasive latitudinal clines in allele frequencies (e.g., Schmidt & Paaby 2008; Turner *et al.* 2008; Kolaczkowski *et al.* 2011; Fabian *et al.* 2012; Bergland *et al.* 2014; Machado *et al.* 2016; Kapun *et al.* 2016a), structural variants such as chromosomal inversions (reviewed in Kapun & Flatt 2019), transposable elements (TEs) (Boussy *et al.* 1998; González *et al.* 2008; 2010), and complex phenotypes (de Jong & Bochdanovits 2003; Schmidt & Paaby 2008; Schmidt *et al.* 2008; Kapun *et al.* 2016b; Behrman *et al.* 2018), especially along the North American and Australian east coasts. In addition to parallel local adaptation, these latitudinal clines are, however, also affected by admixture with flies from Africa and Europe (Caracristi & Schlötterer 2003; Yukilevich & True 2008a; b; Duchen *et al.* 2013; Kao *et al.* 2015; Bergland *et al.* 2016).

In contrast, the population genomics of *D. melanogaster* on the European continent remains largely unstudied (Božičević *et al.* 2016; Pool *et al.* 2016; Mateo *et al.* 2018). Because Eurasia was the first continent colonized by *D. melanogaster* as they migrated out of Africa, we sought to understand how this species has adapted to new habitats and climate zones in Europe, where it has

111 been established the longest (Lachaise *et al.* 1988; David & Capy 1988). We analyse our data at three
112 levels: (1) variation at single-nucleotide polymorphisms (SNPs) in nuclear and mitochondrial
113 (mtDNA) genomes (~5.5 x 10^6 SNPs in total); (2) structural variation, including TE insertions and
114 chromosomal inversion polymorphisms; and (3) variation in the microbiota associated with flies,
115 including bacteria, fungi, protists, and viruses.

116

117 **Results and Discussion**

118 As part of the *DrosEU* consortium, we collected 48 population samples of *D. melanogaster* from 32
119 geographical locations across Europe in 2014 (table 1; fig. 1). We performed pooled sequencing
120 (Pool-Seq) of all 48 samples, with an average autosomal coverage ≥50x (supplementary table S1,
121 Supplementary Material online). Of the 32 locations, 10 were sampled at least once in summer and
122 once in fall (fig. 1), allowing a preliminary analysis of seasonal change in allele frequencies on a
123 genome-wide scale.

124 A description of the basic patterns of genetic variation of these European *D. melanogaster*
125 population samples, based on SNPs, is provided in the supplement (see supplementary results,
126 supplementary table S1, Supplementary Material online). For each sample, we estimated genome-
127 wide levels of π, Watterson's θ and Tajima's D (corrected for pooling; Futschik & Schlötterer 2010;
128 Kofler *et al.* 2011). In brief, patterns of genetic variability and Tajima's D were largely consistent
129 with what has been previously observed on other continents (e.g., Fabian *et al.* 2012; Langley *et al.*
130 2012; Lack *et al.* 2015, 2016), and genetic diversity across the genome varies mainly with
131 recombination rate (Langley *et al.* 2012). We also found little spatio-temporal variation among
132 European populations in overall levels of sequence variability (table 2).

133 Below we focus on the identification of selective sweeps, previously unknown longitudinal
134 population structure across the European continent, patterns of local adaptation and clines, and
135 microbiota.

136

Several genomic regions show signatures of continent-wide selective sweeps

To identify genomic regions that have likely undergone selective sweeps in European populations of *D. melanogaster*, we used *Pool-hmm* (Boitard *et al.* 2013; see supplementary table S2A, Supplementary Material online), which identifies candidate sweep regions via distortions in the allele frequency spectrum. We ran *Pool-hmm* independently for each sample and identified several genomic regions that coincide with previously identified, well-supported sweeps in the proximity of *Hen1* (Kolaczkowski *et al.* 2011), *Cyp6g1* (Daborn *et al.* 2002), *wapl* (Beisswanger *et al.* 2006), and around the chimeric gene *CR18217* (Rogers & Hartl 2012), among others (supplementary table S2B, Supplementary Material online). These regions also showed local reductions in Tajima's *D*, consistent with selection (fig. 2; fig. S1 and fig. S2; Supplementary Material online). The putative sweep regions that we identified in the European populations included 145 of the 232 genes previously identified using *Pool-hmm* in an Austrian population (Boitard *et al.* 2012; supplementary table S2C, Supplementary Material online). We also identified other regions which have not previously been described as targets of selective sweeps (supplementary table S2A, Supplementary Material online). Of the regions analysed, 64 showed signatures of selection across all European populations (supplementary table S2D, Supplementary Material online). Of these, 52 were located in the 10% of regions with the lowest values of Tajima's *D* (SuperExactTest; $p < 0.001$). These may represent continent-wide sweeps that predate the colonization of Europe (e.g., Beisswanger *et al.* 2006) or which have recently swept across the majority of European populations (supplementary table S2D).

We then asked if there was any indication of selective sweeps particular to a certain habitat. To this end, we classified the populations according to the Köppen-Geiger climate classification (Peel *et al.* 2007) and identified several putative sweeps exclusive to arid, temperate and cold regions (supplementary table S2A, Supplementary Material online). To shed light on potential phenotypes affected by the potential sweeps we performed a gene ontology (GO) analysis. For temperate climates, this analysis showed enrichment for functions such as 'response to stimulus', 'transport', and 'nervous system development'. For cold climates, it showed enrichment for 'vitamin and co-factor metabolic processes' (supplementary table S2E, Supplementary Material online). There was no

164 enrichment of any GO category for sweeps associated with arid regions.

165 Thus, we identified several new candidate selective sweeps in European populations of *D.*
166 *melanogaster*, many of which occur in the majority of European populations and which merit future
167 study, using sequencing of individual flies and functional genetic experiments.

168

169 **European populations are structured along an east-west gradient**

170 We next investigated whether patterns of genetic differentiation might be due to demographic sub-
171 structuring. Overall, pairwise differentiation as measured by F_{ST} was relatively low, particularly for
172 the autosomes (autosomal F_{ST} 0.013–0.059; *X*-chromosome F_{ST}: 0.043–0.076; Mann-Whitney U test;
173 $p < 0.001$; supplementary table S1, Supplementary Material online). The *X* chromosome is expected
174 to have higher F_{ST} than the autosomes, given its relatively smaller effective population size (Mann-
175 Whitney U test; $p < 0.001$; Hutter *et al.* 2007). One population, from Sheffield (UK), was unusually
176 differentiated from the others (average pairwise $F_{ST} = 0.027$; SE= 0.00043 vs. $F_{ST} = 0.04$; SE=
177 0.00055 for comparisons without this population and with this population only; supplementary table
178 S1, Supplementary Material online). Including this sample in the analysis could potentially lead to
179 exaggerated patterns of geographic differentiation, as it is both highly differentiated and the furthest
180 west. We therefore excluded it from the following analyses of geographic differentiation, as this
181 approach is conservative. (For details see the Supplementary Material online; including or excluding
182 this population did not qualitatively change our results and their interpretation.)

183 Despite low overall levels of among-population differentiation, we found that European
184 populations exhibit clear evidence of geographic sub-structuring. For this analysis, we focused on
185 SNPs located within short introns, with a length ≤ 60bp and which most likely reflect neutral
186 population structure (Haddrill *et al.* 2005; Singh *et al.* 2009; Parsch *et al.* 2010; Clemente & Vogl
187 2012; Lawrie *et al.* 2013). We further filtered out polymorphisms in regions of high recombination (*r*
188 > 3cM/Mb; Comeron *et al.* 2011) and restricted our analysis to SNPs at least 1 Mb away from the
189 breakpoints of common inversions, resulting in 4,034 SNPs used for demographic analysis.

We found two signatures of geographic differentiation using these putatively neutral SNPs. First, we identified a weak but significant correlation between pairwise F_{ST} and geographic distance, consistent with isolation by distance (IBD; Mantel test; $p < 0.001$; R^2=0.12, max. $F_{ST} \sim 0.045$; fig. 3A). Second, a principal components analysis (PCA) on allele frequencies showed that the three most important PC axes explain >25% of the total variance (PC1: 16.71%, PC2: 5.83%, PC3: 4.6%, eigenvalues = 159.8, 55.7, and 44, respectively; fig 3B). The first axis, PC1, was strongly correlated with longitude ($F_{1,42} = 118.08$, $p < 0.001$; table 2). Again, this pattern is consistent with IBD, as the European continent extends further in longitude than latitude. We repeated the above PCA using SNPs in four-fold degenerate sites, as these are also assumed to be relatively unaffected by selection (Akashi 1995; Halligan & Keightley 2006; supplementary fig. S3, Supplementary Material online), and found highly consistent results.

Because there was a significant spatial autocorrelation between samples (as indicated by Moran's test on residuals from linear regressions with PC1; $p < 0.001$; table 2), we repeated the analysis with an explicit spatial error model; the association between PC1 and longitude remained significant. To a lesser extent PC2 was likewise correlated with longitude ($F_{1,42} = 7.15$, $p < 0.05$), but also with altitude ($F_{1,42} = 11.77$, $p < 0.01$) and latitude ($F_{1,42} = 4.69$, $p < 0.05$; table 2). Similar to PC2, PC3 was strongly correlated with altitude ($F_{1,42} = 19.91$, $p < 0.001$; table 2). We also examined these data for signatures of genetic differentiation between samples collected at different times of the year. For the dataset as a whole, no major PC axes were correlated with season, indicating that there were no strong differences in allele frequencies shared between all our summer and fall samples ($p > 0.05$ for all analyses; table 2). For the 10 locations sampled in both summer and fall, we performed separate PC analyses for summer and fall. Summer and fall values of PC1 (adjusted R^2: 0.98; $p < 0.001$), PC2 (R^2: 0.74; $p < 0.001$) and PC3 (R^2: 0.81; $p < 0.001$) were strongly correlated across seasons. This indicates a high degree of seasonal stability in local genetic variation.

Next, we attempted to determine if populations could be statistically classified into clusters of similar populations. Using hierarchical model fitting based on the first four PC axes from the PCA mentioned above, we found two distinct clusters (fig. 3B) separated along PC1, supporting the notion

of strong longitudinal differentiation among European populations. Similarly, model-based spatial clustering also showed that populations were separated mainly by longitude (fig. 3C; using ConStruct, with K=3 spatial layers chosen based on model selection procedure via cross-validation). We also inferred levels of admixture among populations from this analysis, based on the relationship between F_{ST} and migration rate (Wright *et al.* 1951) and using recent estimates of N_e in European populations ($N_e \sim 3.1 \times 10^6$; Duchen *et al.* 2011; for pairwise migration rates see supplementary table S3, Supplementary Material online). Within the Western European cluster and between the clusters, $4N_em$ was similar ($4N_em$-WE = 43.76, $4N_em$-between = 45.97); in Eastern Europe, estimates of $4N_em$ indicate significantly higher levels of admixture, despite the larger geographic range covered by these samples ($4N_em$ = 74.17; Mann Whitney U-Test; $p < 0.001$). This result suggests that the longitudinal differentiation in Europe might be partly driven by high levels of genetic exchange in Eastern Europe, perhaps due to migration and recolonization after harsh winters in that region. However, these estimates of gene flow must be interpreted with caution, as unknown demographic events can confound estimates of migration rates from F_{ST} (Whitlock & MacCauley 1999).

In addition to restricted gene flow between geographic areas, local adaptation may explain population sub-structure, even at neutral sites, if nearby and closely related populations are responding to similar selective pressures. We investigated whether any of 19 climatic variables, obtained from the WorldClim database (Hijmans *et al.* 2005), were associated with the genetic structure in our samples. These climatic variables represent interpolated averages across 30 years of observation at the geographic coordinates corresponding to our sampling locations. Since many of these variables are highly intercorrelated, we analysed their joint effects on genetic variation, by using PCA to summarize the information they capture. The first three climatic PC axes capture more than 77% of the variance in the 19 climatic variables (supplementary table S4, Supplementary Material online). PC1 explained 36% of the variance and was strongly correlated ($r > 0.75$ or $r < -0.75$) with climatic variables differentiating 'hot and dry' from 'cold and wet' climates (e.g., maximum temperature of the warmest month, $r = 0.84$; mean temperature of warmest quarter, $r = 0.86$; annual mean temperature, $r = 0.85$; precipitation during the warmest quarter, $r = -0.87$). Conversely, PC2

(27.3% of variance explained) distinguished climates with low and high differences between seasons (*e.g.*, isothermality, $r = 0.83$; temperature seasonality, $r = 0.88$; temperature annual range, $r = -0.78$; precipitation in coldest quarter, $r = 0.79$). Both PC1 and PC2 were strongly correlated with latitude (linear regression with PC1: $R^2 = 0.48$, $p < 0.001$; PC2: $R^2 = 0.58$, $p < 0.001$) and PC2 was also weakly correlated with latitude ($R^2 = 0.11$; linear regression, $p < 0.05$) and altitude ($R^2 = 0.12$; linear regression, $p < 0.01$).

We next asked whether any of these climate PCs explained any of the genetic structure uncovered above. Pairwise linear regressions of the first three PC axes based on allele frequencies of intronic SNPs against the first three climatic PCs revealed that only one significant correlation after Bonferroni correction: between climatic PC2 ('seasonality') vs. genetic PC1 (longitude; adjusted $\alpha = 0.017$; $R^2 = 0.49$, $P<0.001$). This suggests that longitudinal differentiation along the European continent might be partly driven by the transition from oceanic to continental climate, possibly leading to local adaptation to gradual changes in temperature seasonality and the severity of winter conditions.

Interestingly, the central European division into an eastern and a western clade of *D. melanogaster* closely resembles known hybrid zones of organisms which form closely related pairs of sister taxa. These biogeographic patterns have been associated with long-term reductions of gene flow between eastern and western population during the last glacial maximum, followed by postglacial recolonization of the continent from southern refugia (Hewitt 1999). However, in contrast to many of these taxa, which often exhibit pronounced pre- and postzygotic isolation (Szymura & Barton 1986; Haas & Brodin 2005; Macholán *et al.* 2008, Knief *et al.* 2019), we found low genome-wide differentiation among eastern and western populations (average max. $F_{ST} \sim 0.045$), perhaps indicating that the longitudinal division of European *D. melanogaster* is not the result of postglacial secondary contact.

Climatic predictors identify genomic signatures of local climate adaptation

To further explore climatic patterns, and to identify signatures of local adaptation caused by climatic

differences among populations independent of neutral demographic effects, we tested for associations of SNP alleles with climatic PC1 and PC2 using BayeScEnv (de Villemereuil & Gaggiotti 2015). The total number of SNPs tested and the number of "top SNPs" (q-value < 0.05) are given in supplementary table S5A (Supplementary Material online). A large proportion of the top SNPs were intergenic (PC1: 33.5%; PC2: 32.2%) or intronic variants (PC1: 50.1%; PC2: 50.5%). Manhattan plots of q-values for all SNPs are shown in fig. 4. These figures show some distinct "peaks" of highly differentiated SNPs along with some broader regions of moderately differentiated SNPs (fig. 4). For example, the circadian rhythm gene *timeout* and the ecdysone signalling genes *Eip74EF* and *Eip75B* all lie near peaks associated with climatic PC1 ('hot/dry' vs. 'cold/wet'; fig. 4, top panels). We note that the corresponding genes have been identified in previous studies of clinal (latitudinal) differentiation in North American *D. melanogaster* (Fabian *et al.* 2012; Machado *et al.* 2016). In fact, we found a significant overlap between genes associated with PC1 and PC2 from our study and candidate gene sets from these previous studies of latitudinal clines (SuperExactTest; $p < 0.001$; Fabian *et al.* 2012; Machado *et al.* 2016). Moreover, the BayeScEnv analysis and *Pool-hmm* analysis together identify four regions with both climatic associations and evidence for continent-wide selective sweeps (supplementary table S5B-C, Supplementary Material online). Finally, four other BayeScEnv candidate genes were previously identified as targets of selection in African and North American populations based on significant McDonald-Kreitman tests (Langley *et al.* 2012; see supplementary table S5B-C, Supplementary Material online).

We next asked whether any insights into the targets of local selection could be gleaned from examining the functions of genes near the BayeScEnv peaks. We examined annotated features within 2kb of significantly associated SNPs (PC1: 3,545 SNPs near 2,078 annotated features; PC2: 5,572 SNPs near 2,717 annotated features; supplementary table S5B and C, Supplementary Material online). First, we performed a GO term analysis with GOwinda (Kofler & Schlötterer 2012) to ask whether SNPs associated with climatic PCs are enriched for any gene functions. For PC1, we found no GO term enrichment. For PC2, we found enrichment for "cuticle development", and "UDP-glucosyltransferase activity". Next, we performed functional annotation clustering with DAVID

(v6.8; Huang *et al.* 2009), and identified 37 and 47 clusters with an enrichment score > 1.3 for PC1 and PC2, respectively (supplementary table S5D-E, Supplementary Material online). PC1 was enriched for categories such as "sex differentiation" and "response to nicotine", whereas PC2 was enriched for functional categories such as "response to nicotine", "integral component of membrane", and "sensory perception of chemical stimulus" (supplementary table S5D-E, Supplementary Material online).

We also asked whether the SNPs identified by BayeScEnv show consistent signatures of local adaptation. Many associated genes (1,205) were also shared between PC1 and PC2. Some genes have indeed been previously implicated in climatic and clinal adaptation, such as the circadian rhythm genes *timeless*, *timeout*, and *clock*, the sexual differentiation gene *fruitless*, and the *couch potato* locus which underlies the latitudinal cline in reproductive dormancy in North America (e.g., Tauber *et al.* 2007; Schmidt *et al.* 2008; Fabian *et al.* 2012). Notably, these also include the major insulin signaling genes *insulin-like receptor* (*InR)* and *forkhead box subgroup O* (*foxo*), which have strong genomic and experimental evidence implicating these loci in clinal, climatic adaptation along the North America east coast (Paaby *et al.* 2010; Fabian *et al.* 2012; Paaby *et al.* 2014; Durmaz *et al.* 2019). Thus, European populations share multiple potential candidate targets of selection with North American populations (cf. Fabian *et al.* 2012; Machado *et al.* 2016; also see Božičević *et al.* 2016). We next turned to examining polymorphisms other than SNPs, i.e. mitochondrial haplotypes as well as inversion and TE polymorphisms.

Mitochondrial haplotypes also exhibit longitudinal population structure

Mitochondrial haplotypes also showed evidence of longitudinal demographic structure in European population. We identified two main alternative mitochondrial haplotypes in Europe, G1 and G2, each with several sub-haplotypes (G1.1 and G1.2 and G2.1, G2.2 and G2.3). The two sub-types, G1.2 and G2.1, are separated by 41 mutations (fig. 5A). The frequencies of the alternative G1 and G2 haplotype varied among populations between 35.1% and 95.6% and between 4.4% and 64.9%, respectively (fig. 5B). Qualitatively, three types of European populations could be distinguished based on these

haplotypes: (1) central European populations, with a high frequency (> 60%) of G1 haplotypes, (2) Eastern European populations in summer, with a low frequency (< 40%) of G1 haplotypes, and (3) Iberian and Eastern European populations in fall, with a frequency of G1 haplotypes between 40-60% (supplementary fig. S4, Supplementary Material online). Analyses of mitochondrial haplotypes from a North American population (Cooper *et al.* 2015) as well as from worldwide samples (Wolff *et al.* 2016) also revealed high levels of haplotype diversity.

While there was no correlation between the frequency of G1 haplotypes and latitude, G1 haplotypes and longitude were weakly but significantly correlated ($r^2 = 0.10$; $p < 0.05$). We thus divided the dataset into an eastern and a western sub-set along the 20º meridian, corresponding to the division of two major climatic zones, temperate (oceanic) versus cold (continental) (Peel *et al.* 2007). This split revealed a clear correlation ($r^2=0.5$; $p<0.001$) between longitude and the frequency of G1 haplotypes, explaining as much as 50% of the variation in the western group (supplementary fig. S4B, Supplementary Material online). Similarly, in eastern populations, longitude and the frequency of G1 haplotypes were correlated ($r^2 = 0.2$; $p<0.001$), explaining approximately 20% of the variance (supplementary fig. S4B, Supplementary Material online). Thus, these mitochondrial haplotypes appear to follow a similar east-west population structure as observed for the nuclear SNPs described above .

The frequency of polymorphic TEs varies with longitude and altitude

To examine the population genomics of structural variants, we first focused on transposable elements (TEs). Similar to previous findings, the repetitive content of the 48 samples ranged from 16% to 21% of the nuclear genome size (Quesneville *et al.* 2005; fig. 6). The vast majority of detected repeats were TEs, mostly long terminal repeat elements (LTRs; range 7.55 % - 10.15 %) and long interspersed nuclear elements (LINEs range 4.18 % - 5.52 %), along with a few DNA elements (range 1.16 % - 1.65 %) (supplementary table S6, Supplementary Material online). LTRs have been previously described as being the most abundant TEs in the *D. melanogaster* genome (Kaminker *et al.* 2002; Bergman *et al.* 2006). Correspondingly, variation in the proportion of LTRs best explained

352 variation in total TE content (LINE+LTR+DNA) (Pearson's $r = 0.87$, $p < 0.01$, *vs.* DNA $r = 0.58$, $p =$
353 0.0117, and LINE $r = 0.36$, $p < 0.01$ and supplementary fig. S5A, Supplementary Material online).

354 For each of the 1,630 TE insertion sites annotated in the *D. melanogaster* reference genome *v.6.04*,
355 we estimated the frequency at which a copy of the TE was present at that site using *T-lex2* (Fiston-
356 Lavier *et al.* 2015; see supplementary table S7, Supplementary Material online). On average, 56%
357 were fixed in all samples. The remaining polymorphic TEs mostly segregated at low frequency in all
358 samples (supplementary fig. S5B), potentially due to purifying selection (González *et al.* 2008; Petrov
359 *et al.* 2011; Kofler *et al.* 2012; Cridland *et al.* 2013; Blumenstiel *et al.* 2014). However, 246 were
360 present at intermediate frequencies (>10% and <95%) and located in regions of non-zero
361 recombination (Fiston-Lavier *et al.* 2010; Comeron *et al.* 2012; see supplementary table S7,
362 Supplementary Material online). Although some of these insertions might be segregating neutrally at
363 transposition-selection balance (Charlesworth *et al.* 1994; see supplementary fig. S5B, Supplementary
364 Material online), they are likely enriched for candidate adaptive mutations (Rech *et al.* 2019).

365 In each of the 48 samples, TE frequency and recombination rate were negatively correlated
366 genome-wide (Spearman rank sum test; $p < 0.01$), as has also been previously reported for *D.*
367 *melanogaster* (Bartolomé *et al.* 2002; Petrov *et al.* 2011; Kofler *et al.* 2012). This remains true when
368 fixed TE insertions were excluded (population frequency ≥95%) from the analysis, although it was
369 not statistically significant for some chromosomes and populations (supplementary table S8,
370 Supplementary Material online). In both cases, the correlation was stronger when broad-scale (Fiston-
371 Lavier *et al.* 2010) rather than fine-scale (Comeron *et al.* 2012) recombination rate estimates were
372 used, indicating that the former may best capture long-term population recombination patterns (see
373 supplementary materials and methods and supplementary table S8, Supplementary Material online).

374 We next tested whether variation in TE frequencies among samples was associated with spatially
375 or temporally varying factors. We focused on 111 TE insertions that segregated at intermediate
376 frequencies, were located in non-zero recombination regions, and that showed an interquartile range
377 (IQR) > 10 (see supplementary materials and methods, Supplementary Material online). Of these
378 insertions, 57 were significantly associated with a at least one variable of interest after multiple testing

379 correction (supplementary table S9A, Supplementary Material online): 13 were significantly
380 associated with longitude, 13 with altitude, five with latitude, three with season, and 23 insertions
381 with more than one of these variables (supplementary table S9A, Supplementary Material online).
382 These 57 TEs were mainly located inside genes (42 out of 57; Fisher's Exact Test, $p > 0.05$;
383 supplementary table S9A, Supplementary Material online).

384 The 57 TEs significantly associated with these environmental variables were enriched for two TE
385 families: the LTR *297* family with 11 copies, and the DNA *pogo* family with five copies (χ^2-values
386 after Yate's correction < 0.05; supplementary table S9B, Supplementary Material online).
387 Interestingly, 17 of the 57 TEs coincided with previously identified adaptive candidate TEs,
388 suggesting that our dataset might be enriched for adaptive insertions (SuperExactTest, $p < 0.001$),
389 several of which exhibit spatial frequency clines that deviate from neutral expectation
390 (SuperExactTest, $p < 0.001$, supplementary table S9A, Supplementary Material online; cf.; Rech *et al.*
391 2019). Moreover, 18 of the 57 TEs also show significant correlations with either geographical or
392 temporal variables in North American populations (SuperExactTest, $p < 0.001$, supplementary table
393 S9A, Supplementary Material online; cf.; Lerat *et al.* 2019).

394

395 **Inversions exhibit latitudinal and longitudinal clines in Europe**

396 Polymorphic chromosomal inversions, another class of structural variants besides TEs, are well-
397 known to exhibit pronounced spatial (clinal) patterns in North American, Australian and other
398 populations, possibly due to spatially varying selection (reviewed in Kapun & Flatt 2019; also see
399 Mettler *et al.* 1977; Knibb *et al.* 1981; Leumeunier & Aulard 1992; Hoffmann & Weeks 2007; Fabian
400 *et al.* 2012; Kapun *et al.* 2014; Rane *et al.* 2015; Adrion *et al.* 2015; Kapun *et al.* 2016a). However, in
401 contrast to North America and Australia, inversion clines in Europe remain poorly characterized
402 (Lemeunier & Aulard 1992; Kapun & Flatt 2019). We therefore sought to examine the presence and
403 frequency of six cosmopolitan inversions (*In(2L)t, In(2R)NS, In(3L)P, In(3R)C, In(3R)Mo,*
404 *In(3R)Payne*) in our European samples, using a panel of inversion-specific marker SNPs (Kapun *et al.*
405 2014). All 48 samples were polymorphic for one or more inversions (Figure 6). However, only

406 *In(2L)t* segregated at substantial frequencies in most populations (average frequency = 20.2%); all
407 other inversions were either absent or rare (average frequencies: *In(2R)NS* = 6.2%, *In(3L)P* = 4%,
408 *In(3R)C* = 3.1%, *In(3R)Mo* =2.2%, *In(3R)Payne* = 5.7%) (cf. Kapun *et al.* 2016; Kapun & Flatt
409 2019).

410 Despite their overall low frequencies, several inversions showed pronounced clinality. For all of
411 the analyses below, we tested for confounding effects of spatial autocorrelation, by asking if there was
412 significant residual spatio-temporal autocorrelation among samples; all of these test were negative,
413 except for *In(3R)C* (Moran's $I \approx 0$, $p > 0.05$ for all tests; table 3). We observed significant latitudinal
414 clines for *In(3L)P, In(3R)C* and *In(3R)Payne* (generalized linear regression, Inversion frequency ~
415 Continent * Latitude; $p < 0.001$ for all; see table 3). Clines for *In(3L)P* and *In(3R)Payne* were
416 qualitatively similar for both continents (with frequencies decreasing with latitude, $p < 0.05$ for both),
417 although all inversions differed in their frequency at the same latitude between North America and
418 Europe ($p < 0.001$ for Continent; supplementary table S10, Supplementary Material online).

419 Latitudinal inversion clines previously observed along the North American and Australian east
420 coasts (supplementary fig. S6 and supplementary table S10, Supplementary Material online; Kapun *et*
421 *al.* 2016a) have been attributed to spatially varying selection, especially in the case of *In(3R)Payne*
422 (Durmaz *et al.* 2018; Anderson *et al.* 2005; Umina *et al.* 2005; Kennington *et al.* 2006; Rako *et al.*
423 2006; Kapun *et al*. 2016a,b; Kapun & Flatt 2019). Similar to patterns in North America (Kapun *et al.*
424 2016a), we observed that clinality of the three inversion polymorphisms was markedly stronger than
425 for putatively neutral SNPs in short introns (see supplementary table S11, Supplementary Material
426 online), suggesting that these polymorphisms may be non-neutral. Together, these finding suggest that
427 latitudinal inversion clines in Europe are shaped by spatially varying selection.

428 We also detected longitudinal clines for *In(2L)t* and *In(2R)NS*, with both polymorphisms
429 decreasing in frequency from east to west (Inversion frequency ~ Latitude + Longitude + Altitude +
430 Season; $p < 0.01$; table 3; also cf. Kapun & Flatt 2019). Longitudinal clines for these two inversions
431 have also been found in North America (Kapun & Flatt 2019). One of these inversions,*In(2L)t*, also
432 changed in frequency with altitude (table 3). The longitudinal and altitudinal inversion clines did,

433 however, not deviate from neutral expectation (supplementary table S11, Supplementary Material
434 online).

435

436 **European *Drosophila* microbiomes contain Entomophthora, trypanosomatids and previously**
437 **unknown DNA viruses**

438 The microbiota can affect life history traits, immunity, hormonal physiology, and metabolic
439 homeostasis of their fly hosts (e.g., Trinder *et al.* 2017; Martino *et al.* 2017) and might thus reveal
440 interesting patterns of local adaptation. We therefore examined the bacterial, fungal, protist, and viral
441 microbiota sequence content of our samples. To do this, we characterised the taxonomic origin of the
442 non-*Drosophila* reads in our dataset using MGRAST, which identifies and counts short protein motifs
443 ('features') within reads (Meyer *et al.* 2008). We examined 262 million reads in total. Of these, most
444 were assigned to *Wolbachia* (mean 53.7%; fig. 7; supplementary table S1), a well-known
445 endosymbiont of *Drosophila* (Werren *et al.* 2008). The abundance of *Wolbachia* protein features
446 relative to other microbial protein features (relative abundance) varied strongly between samples,
447 ranging from 8.8% in a sample from Ukraine to almost 100% in samples from Spain, Portugal,
448 Turkey and Russia (supplementary table S12, Supplementary Material online). Similarly, *Wolbachia*
449 loads varied 100-fold between samples, as estimated from the ratio of *Wolbachia* protein features to
450 *Drosophila* protein features (supplementary table S12, Supplementary Material online). In contrast to
451 a previous study (Kriesner *et al*. 2016), there was no evidence for clinality of *Wolbachia* loads (p =
452 0.13, longitude; p = 0.41, latitude; Kendall's rank correlation). However, these authors measured
453 infection frequencies while we measured *Wolbachia* loads in pooled samples. Because the frequency
454 of infection does not necessarily correlate with microbial loads measured in pooled samples, we might
455 not have been able to detect such a signal in our data.

456 Acetic acid bacteria of the genera *Gluconobacter*, *Gluconacetobacter*, and *Acetobacter* were the
457 second largest group, with an average relative abundance of 34.4% among microbial protein features.
458 Furthermore, we found evidence for the presence of several genera of Enterobacteria (*Serratia,*
459 *Yersinia, Klebsiella, Pantoea, Escherichia, Enterobacter, Salmonella*, and *Pectobacterium*). *Serratia*

460 occurs only at low frequencies or is absent from most of our samples, but reaches a very high relative
461 abundance among microbial protein features in the Nicosia (Cyprus) summer collection (54.5%). This
462 high relative abundance was accompanied by an 80x increase in *Serratia* bacterial load.

463 We also detected several eukaryotic microorganisms, although they were less abundant than the
464 bacteria. We found trypanosomatids, previously reported to be associated with *Drosophila* in other
465 studies (Wilfert *et al.* 2011; Chandler & James 2013; Hamilton *et al.* 2015), in 16 of our samples, on
466 average representing 15% of all microbial protein features identified in these samples.

467 Fungal protein features make up <3% of all but three samples (from Finland, Austria and Turkey;
468 supplementary table S12, Supplementary Material online). This is somewhat surprising because
469 yeasts are commonly found on rotting fruit, the main food substrate of *D. melanogaster*, and co-occur
470 with flies (Barata *et al.* 2012; Chandler *et al.* 2012). This result suggests that, although yeasts can
471 attract flies and play a role in food choice (Becher *et al.* 2012; Buser *et al.* 2014), they might not be
472 highly prevalent in or on *D. melanogaster* bodies. One reason might be that they are actively digested
473 and thus not part of the microbiome. We also found the fungal pathogen *Entomophthora muscae* in 14
474 samples, making up 0.18% of the reads (Elya *et al.* 2018).

475 Our data also allowed us to identify DNA viruses. Only one DNA virus has been previously
476 described for *D. melanogaster* (Kallithea virus; Webster *et al.* 2015; Palmer *et al.* 2018) and only two
477 additional ones from other Drosophilid species (*Drosophila innubila* Nudivirus [Unckless 2011],
478 Invertebrate Iridovirus 31 in *D. obscura* and *D. immigrans* [Webster *et al.* 2016]). In our data set,
479 approximately two million reads came from *Kallithea* nudivirus (Webster *et al.* 2015), allowing us to
480 assemble the first complete *Kallithea* genome (>300-fold coverage in the Ukrainian sample
481 UA_Kha_14_46; Genbank accession KX130344).

482 We also found reads from five additional DNA viruses that were previously unknown
483 (supplementary table S13, Supplementary Material online). First, around 1,000 reads come from a
484 novel nudivirus closely related to both *Kallithea* virus and to *Drosophila innubila* nudivirus (Unckless
485 2011) in sample DK_Kar_14_41 from Karensminde, Denmark supplementary table S13,
486 Supplementary Material online). As the reads from this virus were insufficient to assemble the

487 genome, we identified a publicly available dataset (SRR3939042: 27 male *D. melanogaster* from
488 Esparto, California; Machado *et al.* 2016) with sufficient reads to complete the genome (provisionally
489 named "*Esparto* Virus"; KY608910). Second, we also identified two novel Densoviruses
490 (*Parvoviridae*). The first is a relative of *Culex pipiens* densovirus, provisionally named "*Viltain*
491 virus", found at 94-fold coverage in sample FR_Vil_14_07 (Viltain; KX648535). The second is
492 "*Linvill Road* virus", a relative of *Dendrolimus punctatus* densovirus, represented by only 300 reads
493 here, but with high coverage in dataset SRR2396966 from a North American sample of *D. simulans*,
494 permitting assembly (KX648536; Machado *et al.* 2016). Third, we detected a novel member of the
495 *Bidnaviridae* family, "*Vesanto* virus", a bidensovirus related to *Bombyx mori* densovirus 3 with
496 approximately 900-fold coverage in sample FI_Ves_14_38 (Vesanto; KX648533 and KX648534).
497 Finally, in one sample (UA_Yal_14_16), we detected a substantial number of reads from an
498 Entomopox-like virus, which we were unable to fully assemble (supplementary table S13,
499 Supplementary Material online).

500 Using a detection threshold of >0.1% of the *Drosophila* genome copy number, the most commonly
501 detected viruses were *Kallithea* virus (30/48 of the pools) and *Vesanto* virus (25/48), followed by
502 *Linvill Road* virus (7/48) and *Viltain* virus (5/48), with *Esparto* virus and the entomopox-like virus
503 being the rarest (2/48 and 1/48, respectively). Because *Wolbachia* can protect *Drosophila* from
504 viruses (Teixeira et al., 2008), we hypothesized that *Wolbachia* loads might correlate negatively with
505 viral loads, but found no evidence of such a correlation ($p = 0.83$ Kallithea virus; $p = 0.76$ Esparto
506 virus; $p = 0.52$ Viltain virus; $p = 0.96$ Vesanto 1 virus; $p = 0.93$ Vesanto 2 virus; $p = 0.5$ Linvill Road
507 virus; Kendall's rank correlation). Perhaps this is because the *Kallithea* virus, the most prevalent
508 virus in our data set, is not expected to be affected by *Wolbachia* (Palmer et al.,2018). Similarly, Shi
509 et al. (2018) found no link between *Wolbachia* and the prevalence or abundance of RNA viruses in
510 data from individual flies.

511 The variation in bacterial microbiomes across space and time reported here is analysed in more
512 detail in Wang *et al.* (2020); this study suggests that some of this variation is structured
513 geographically (cf. Walters *et al.* 2020). Thus, microbiome composition may contribute to phenotypic

differences and local adaptation among populations, (Haselkorn *et al.* 2009; Richardson *et al.* 2012; Staubach *et al.* 2013; Kriesner *et al.* 2016; Wang and Staubach 2018).

Conclusions

Here, we have comprehensively sampled and sequenced European populations of *D. melanogaster* for the first time (fig. 1). We find that European *D. melanogaster* populations are longitudinally differentiated for putatively neutral SNPs, mitochondrial haplotypes as well as for inversion and TE insertion polymorphisms. Potentially adaptive polymorphisms also show this pattern, possibly driven by the transition from oceanic to continental climate along the longitudinal axis of Europe. We note that this longitudinal differentiation qualitatively resembles the one observed for human populations in Europe (e.g., Cavalli-Sforza 1966; Xiao et al. 2004; Francalacci & Sanna 2008; Novembre *et al.* 2008). Given that *D. melanogaster* is a human commensal (Keller 2007, Arguello *et al.* 2019), it is thus tempting to speculate that the demographic history of European populations might have been influenced by past human migration. Outside Europe, east-west structure has been previously found in sub-Saharan Africa populations of *D. melanogaster*, with the split between eastern and western African populations having occurred ~70 kya (Michalakis & Veuille 1996; Aulard *et al.* 2002; Kapopoulou *et al.* 2018b), a period that coincides with a wave of human migration from eastern into western Africa (Nielsen *et al.* 2017). However, in contrast to the pronounced pattern observed in Europe, African east-west structure is relatively weak, explaining only ~2.7% of variation, and is primarily due to an inversion whose frequency varies longitudinally. In contrast, our demographic analyses are based on SNPs located in >1 Mb distance from the breakpoints of the most common inversions, making it unlikely that the longitudinal pattern we observe is driven by inversions.

Our extensive sampling was feasible only due to synergistic collaboration among many research groups. Our efforts in Europe are paralleled in North America by the *Dros-RTEC* consortium (Machado *et al.* 2019), with whom we are collaborating to compare population genomic data across continents. Together, we have sampled both continents annually since 2014; we aim to continue to sample and sequence European and North American *Drosophila* populations with increasing spatio-

temporal resolution in future years. With these efforts, we hope to provide a rich community resource for biologists interested in molecular population genetics and adaptation genomics.

Materials and methods

A detailed description of the materials and methods is provided in the supplementary materials and methods (see Supplementary Material online); here we give a brief overview of the dataset and the basic methods used. The 2014 *DrosEU* dataset represents the most comprehensive spatio-temporal sampling of European *D. melanogaster* populations to date (fig.1; supplementary table S1, Supplementary Material online). It comprises 48 samples of *D. melanogaster* collected from 32 geographical locations across Europe at different time points in 2014 through a joint effort of 18 research groups. Collections were mostly performed with baited traps using a standardized protocol (see supplementary materials and methods, Supplementary Material online). From each collection, we pooled 33–40 wild-caught males. We used males as they are more easily distinguishable morphologically from similar species than females. Despite our precautions, we identified a low level of *D. simulans* contamination in our sequences; we computationally filtered these sequences from the data prior to further analysis (see Supplementary Material online). To sequence these samples, we extracted DNA and barcoded each sample, and sequenced the ~40 flies per sample as a pool (Pool-Seq; Schlötterer *et al.* 2014), as paired-end fragments on a *Illumina NextSeq 500* sequencer at the Genomics Core Facility of Pompeu Fabra University. Samples were multiplexed in 5 batches of 10 samples, except for one batch of 8 samples (supplementary table S1, Supplementary Material online). Each multiplexed batch was sequenced on 4 lanes at ~50x raw coverage per sample. The read length was 151 bp, with a median insert size of 348 bp (range 209-454 bp). Our genomic dataset is available under NCBI Bioproject accession PRJNA388788. Sequences were processed and mapped to the *D. melanogaster* reference genome (v.6.12) and reference sequences from common commensals and pathogens. Our bioinformatic pipeline is available at https://github.com/capoony/DrosEU_pipeline. To call SNPs, we developed custom software (*PoolSNP*; see supplementary material and methods; https://github.com/capoony/PoolSNP), using stringent heuristic parameters. In addition, we obtained

568 genome sequences from African flies from the *Drosophila* Genome Nexus (DGN;
569 http://www.johnpool.net/genomes.html; see supplementary table S14 for SRA accession numbers).
570 We used data from 14 individuals from Rwanda and 40 from Siavonga (Zambia). We mapped these
571 data to the *D. melanogaster* reference genome using the same pipeline as for our own data above, and
572 built consensus sequences for each haploid sample by only considering alleles with > 0.9 allele
573 frequencies. We converted consensus sequences to *VCF* and used *VCFtools* (Danecek et al. 2011) for
574 downstream analyses. Additional steps in the mapping and variant calling pipeline and further
575 downstream analyses of the data are detailed in in the supplementary materials and methods
576 (Supplementary Materials online).

577

578 **Supplementary Materials**

579 Supplementary materials and methods, supplementary results and supplementary figs. S1–S13 and
580 supplementary tables S1–S18 are available at Molecular Biology and Evolution online
581 (http://www.mbe.oxfordjournals.org/).

582

583 **Acknowledgments**

584 We thank two anonymous reviewers and the editors for their helpful comments on a previous version
585 of our manuscript. We are grateful to the members of the *DrosEU* and Dros-RTEC consortia and to
586 Dmitri Petrov (Stanford University) for support and discussion. *DrosEU* is funded by a Special Topic
587 Networks (STN) grant from the European Society for Evolutionary Biology (ESEB). Computational
588 analyses were partially executed at the Vital-IT bioinformatics facility of the University of Lausanne
589 (Switzerland), the computing facilities of the CC LBBE/PRABI in Lyon (France), the bwUniCluster
590 of the state of Baden-Württemberg (bwHPC), and the University of St Andrews Bioinformatics Unit
591 which is funded by a Wellcome Trust ISSF award (grant 105621/Z/14/Z). We are grateful to Oscar
592 Gaggiotti for advice on BayeScEnv analyses.

593 **Funding**

Funder	Grant reference number	Author
University of Freiburg Research Innovation Fund 2014. Deutsche Forschungsgemeinschaft (DFG)	STA1154/4-1Project 408908608	Fabian Staubach
Academy of Finland	#268241	Maaria Kankare
Academy of Finland	#272927	Maaria Kankare
Russian Foundation of Basic Research	#15-54-46009 CT_a	Elena G. Pasyukova
Danish Natural Science Research Council	4002-00113	Volker Loeschcke
Ministerio de Economia y Competitividad	CTM2017-88080 (AEI/FEDER, UE)	Marta Pascual
CNRS	UMR 9191	Catherine Montchamp-Moreau
Vetenskapsrådet	2011-05679	Jessica Abbott
Vetenskapsrådet	2015-04680	Jessica Abbott
Emmy Noether Programme of the Deutsche Forschungsgemeinschaft,(DFG)	PO 1648/3-1	Nico Posnien
National Institute of Health (NIH)	R35GM119686	Alan O. Bergland
Ministerio de Economia y Competitividad	CGL2013-42432-P	Maria Pilar Garcia Guerreiro
Scientific and Technological Research Council of Turkey (TUBITAK)	#214Z238	Banu Sebnem Onder
ANR Exhyb	14-CE19-0016	Cristina Vieira
Network of Excellence LifeSpan	FP6 036894	Bas J. Zwaan
IDEAL	FP7/2007-2011/259679	Bas J. Zwaan
Israel Science Foundation	1737/17	Eran Tauber
National Institute of Health (NIH)	R01GM100366	Paul S. Schmidt
Deutsche Forschungsgemeinschaft (DFG)	PA 903/8-1	John Parsch
Austrian Science Fund (FWF)	P32275	Martin Kapun
Austrian Science Fund (FWF)	P27048	Andrea J. Betancourt
Biotechnology and Biological Sciences Research Council (BBSRC)	BB/P00685X/1	Andrea J. Betancourt
Swiss National Science Foundation (SNSF)	PP00P3_133641	Thomas Flatt
Swiss National Science Foundation (SNSF)	PP00P3_165836	Thomas Flatt
Swiss National Science Foundation (SNSF)	31003A_182262	Thomas Flatt
European Commission	H2020-ERC-2014CoG-647900	Josefa González

Secretaria d'Universitats i Recerca. Dept Economia i Coneixement. Generalitat de Catalunya	GRC 2017 SGR 880	Josefa González
Ministerio de Economia y Competitividad	FEDER BFU2014-57779-P	Josefa González

References

Adrian AB, Comeron JM (2013) The *Drosophila* early ovarian transcriptome provides insight to the molecular causes of recombination rate variation across genomes. *BMC Genomics*, **14**, 1-12.

Adrion JR, Hahn MW, Cooper BS (2015) Revisiting classic clines in *Drosophila melanogaster* in the age of genomics. *Trends in Genetics*, **31**, 434–444.

Akashi H (1995) Inferring weak selection from patterns of polymorphism and divergence at "silent" sites in *Drosophila* DNA. *Genetics*, **139**, 1067-1076.

Anderson AR, Hoffmann AA, McKechnie SW, Umina PA, Weeks AR (2005) The latitudinal cline in the *In(3R)Payne* inversion polymorphism has shifted in the last 20 years in Australian *Drosophila melanogaster* populations. *Molecular Ecology,* **14**, 851–858.

Arguello JR, Laurent S, Clark AG. (2019) Demographic History of the Human Commensal *Drosophila melanogaster*. *Genome Biology and Evolution* **11**:844–854.

Aulard S, David JR, Lemeunier F (2002) Chromosomal inversion polymorphism in Afrotropical populations of *Drosophila melanogaster*. *Genetic Research*, **79**, 49–63.

Barata A, Santos SC, Malfeito-Ferreira M, Loureiro V (2012) New insights into the ecological interaction between grape berry microorganisms and *Drosophila* flies during the development of sour rot. *Microbial Ecology*, **64**, 416–430.

Bartolomé C, Maside X, Charlesworth B (2002) On the Abundance and Distribution of Transposable Elements in the Genome of *Drosophila melanogaster*. *Molecular Biology and Evolution*, **19**, 926–937.

Becher PG, Flick G, Rozpędowska E *et al.* (2012) Yeast, not fruit volatiles mediate *Drosophila melanogaster* attraction, oviposition and development. *Functional Ecology*, **26**, 822–828.

Begun DJ, Holloway AK, Stevens K *et al.* (2007) Population Genomics: Whole-Genome Analysis of Polymorphism and Divergence in *Drosophila simulans*. *PLoS Biology*, **5**, e310.

620 Behrman EL, Howick VM, Kapun M *et al.* (2018) Rapid seasonal evolution in innate immunity of
621 wild *Drosophila melanogaster*. *Proceedings of the Royal Society of London B*, **285**, 20172599.
622 Beisswanger S, Stephan W, De Lorenzo D (2006) Evidence for a Selective Sweep in the *wapl* Region
623 of *Drosophila melanogaster*. *Genetics*, **172**, 265–274.
624 Bergland AO, Behrman EL, O'Brien KR, Schmidt PS, Petrov DA (2014) Genomic Evidence of Rapid
625 and STable Adaptive Oscillations over Seasonal Time Scales in *Drosophila*. *PLoS Genetics*, **10**,
626 e1004775.
627 Bergland AO, Tobler R, González J, Schmidt P, Petrov D (2016) Secondary contact and local
628 adaptation contribute to genome-wide patterns of clinal variation in *Drosophila* melanogaster.
629 *Molecular Ecology*, **25**, 1157–1174.
630 Bergman, C. M., Quesneville, H., Anxolabehere, D. & Ashburner, M. (2006) Recurrent insertion and
631 duplication generate networks of transposable element sequences in the Drosophila melanogaster
632 genome. *Genome Biology*, **7**, R112.
633 Blumenstiel JP, Chen X, He M, Bergman CM (2014) An Age-of-Allele Test of Neutrality for
634 Transposable Element Insertions. *Genetics*, **196**, 523–538.
635 Boitard S, Schlötterer C, Nolte V, Pandey RV, Futschik A (2012) Detecting Selective Sweeps from
636 Pooled Next-Generation Sequencing Samples. *Molecular Biology and Evolution*, **29**, 2177–2186.
637 Boitard S, Kofler R, Françoise P, Robelin D, Schlötterer C, Futschik A (2013) Pool-hmm: a Python
638 program for estimating the allele frequency spectrum and detecting selective sweeps from next
639 generation sequencing of pooled samples. *Mol Ecol Resour,* **13**, 337–340.
640 Boussy IA, Itoh M, Rand D, Woodruff RC (1998) Origin and decay of the P element-associated
641 latitudinal cline in Australian *Drosophila melanogaster*. *Genetica*, **104**, 45–57.
642 Božičević V, Hutter S, Stephan W, Wollstein A (2016) Population genetic evidence for cold
643 adaptation in European *Drosophila melanogaster* populations. *Molecular Ecology*, **25**, 1175–
644 1191.
645 Buser CC, Newcomb RD, Gaskett AC, Goddard MR (2014) Niche construction initiates the evolution
646 of mutualistic interactions. *Ecology Letters*, **17**, 1257–1264.

647 Caracristi G, Schlötterer C (2003) Genetic Differentiation Between American and European
648 *Drosophila melanogaster* Populations Could Be Attributed to Admixture of African Alleles.
649 *Molecular Biology and Evolution*, **20**, 792–799.
650 Cavalli-Sforza LL (1966) Population Structure and Human Evolution. *Proceedings of the Royal*
651 *Society of London B*, **164**, 362–379.
652 Chandler JA, James PM (2013) Discovery of trypanosomatid parasites in globally distributed
653 *Drosophila* species. *PLoS ONE*, **8**, e61937.
654 Chandler JA, Eisen JA, Kopp A (2012) Yeast communities of diverse *Drosophila* species:
655 comparison of two symbiont groups in the same hosts. *Applied and Environmental Microbiology*,
656 **78**, 7327–7336.
657 Charlesworth B, Sniegowski P, Stephan W (1994) The evolutionary dynamics of repetitive DNA in
658 eukaryotes. *Nature*, **371**, 215–220.
659 Cheng C, White BJ, Kamdem C *et al.* (2012) Ecological genomics of *Anopheles gambiae* along a
660 latitudinal cline: a population-resequencing approach. *Genetics*, **190**, 1417–1432.
661 Clemente F, Vogl C (2012) Unconstrained evolution in short introns? – An analysis of genome-wide
662 polymorphism and divergence data from *Drosophila*. *Journal of Evolutionary Biology*, **25**, 1975–
663 1990.
664 Cooper BS, Burrus CR, Ji C, Hahn MW, Montooth KL (2015) Similar Efficacies of Selection Shape
665 Mitochondrial and Nuclear Genes in Both *Drosophila melanogaster* and *Homo sapiens*. *G3*, **5**,
666 2165–2176.
667 Cridland JM, Macdonald SJ, Long AD, Thornton KR (2013) Abundance and distribution of
668 transposable elements in two *Drosophila* QTL mapping resources. *Molecular Biology and*
669 *Evolution*, **30**, 2311–2327.
670 Daborn PJ, Yen JL, Bogwitz MR *et al.* (2002) A single p450 allele associated with insecticide
671 resistance in *Drosophila*. *Science*, **297**, 2253–2256.
672 David JR, Capy P (1988) Genetic variation of *Drosophila melanogaster* natural populations. *Trends*
673 *in Genetics*, **4**, 106–111.

de Jong G, Bochdanovits Z (2003) Latitudinal clines in *Drosophila melanogaster*: body size, allozyme frequencies, inversion frequencies, and the insulin-signalling pathway. *Journal of Genetics*, **82**, 207–223.

Dobzhansky T (1970) *Genetics of the Evolutionary Process*. Columbia University Press.

Duchen P, Zivkovic D, Hutter S, Stephan W, Laurent S (2013) Demographic inference reveals African and European admixture in the North American *Drosophila melanogaster* population. *Genetics*, **193**, 291–301.

Durmaz E, Benson C, Kapun M, Schmidt P, Flatt T (2018) An Inversion Supergene in *Drosophila* Underpins Latitudinal Clines in Survival Traits. *Journal of Evolutionary Biology,* **31**, 1354-1364..

Durmaz E, Rajpurohit S, Betancourt N, Fabian DK, Kapun M, Schmidt P, Flatt T (2019) A clinal polymorphism in the insulin signaling transcription factor *foxo* contributes to life-history adaptation in *Drosophila*. *Evolution,* **73**, 1774-1792.

Elya C, Lok TC, Spencer QE, McCausland H, Martinez CC, Eisen MB (2018) Robust manipulation of the behavior of *Drosophila melanogaster* by a fungal pathogen in the laboratory, *eLife, 7,* e34414

Fabian DK, Kapun M, Nolte V *et al.* (2012) Genome-wide patterns of latitudinal differentiation among populations of *Drosophila melanogaster* from North America. *Molecular Ecology*, **21**, 4748–4769.

Fiston-Lavier A-S, Barrón MG, Petrov DA, González J (2015) T-lex2: genotyping, frequency estimation and re-annotation of transposable elements using single or pooled next-generation sequencing data. *Nucleic Acids Research*, **43**, e22–e22.

Fiston-Lavier A-S, Singh ND, Lipatov M, Petrov DA (2010) *Drosophila melanogaster* recombination rate calculator. *Gene*, **463**, 18–20.

Francalacci P, Sanna D (2008) History and geography of human Y-chromosome in Europe: a SNP perspective. *Journal of Anthropological Sciences*, **86**, 59–89.

Futschik A & Schlötterer C (2010) The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics*, **186**, 207–218.

González J, Karasov TL, Messer PW, Petrov DA (2010) Genome-Wide Patterns of Adaptation to Temperate Environments Associated with Transposable Elements in *Drosophila*. *PLoS Genetics*, **6**, e1000905.

González J, Lenkov K, Lipatov M, Macpherson JM, Petrov DA (2008) High Rate of Recent Transposable Element–Induced Adaptation in *Drosophila melanogaster*. *PLoS Biology*, **6**, e251.

Haas F, Brodin A. (2005). The Crow *Corvus corone* hybrid zone in southern Denmark and northern Germany. *Ibis* **147**:649–656.

Haddrill PR, Charlesworth B, Halligan DL, Andolfatto P (2005) Patterns of intron sequence evolution in *Drosophila* are dependent upon length and GC content. *Genome Biology*, **6**, R67.

Hales KG, Korey CA, Larracuente AM, Roberts DM (2015) Genetics on the Fly: A Primer on the *Drosophila* Model System. *Genetics*, **201**, 815–842.

Halligan DL, Keightley PD 2006 Ubiquitous selective constraints in the *Drosophila* genome revealed by a genome-wide interspecies comparison. *Genome Research*, **16**, 875-884.

Harpur BA, Kent CF, Molodtsova D *et al.* (2014) Population genomics of the honey bee reveals strong signatures of positive selection on worker traits. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 2614–2619.

Haselkorn TS, Markow TA, Moran NA (2009) Multiple introductions of the *Spiroplasma* bacterial endosymbiont into *Drosophila*. *Molecular Ecology*, **18**, 1294–1305.

Haudry A, Laurent S, Kapun M. 2020. Population Genomics on the Fly: Recent Advances in Drosophila. In: Dutheil JY, editor. Statistical Population Genomics. Vol. 2090. New York, NY: Springer US. p. 357–396.

Hewitt GM. (1999). Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* **68**:87–112.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol*. 25:1965–1978.

Hohenlohe PA, Bassham S, Etter PD *et al.* (2010) Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. *PLoS Genetics*, **6**, e1000862.

728 Hoffmann AA, Weeks AR (2007) Climatic selection on genes and traits after a 100 year-old invasion:
729 a critical look at the temperate-tropical clines in *Drosophila melanogaster* from eastern Australia.
730 *Genetica*, **129**, 133–147.
731 Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists
732 using DAVID bioinformatics resources. *Nature Protocols,* **4**, 44–57.
733 Hudson RR, Kreitman M, Aguadé M (1987) A test of neutral molecular evolution based on nucleotide
734 data. *Genetics*, **116**, 153–159.
735 Hutter S, Li H, Beisswanger S, De Lorenzo D, Stephan W (2007) Distinctly Different Sex Ratios in
736 African and European Populations of *Drosophila melanogaster* Inferred From Chromosomewide
737 Single Nucleotide Polymorphism Data. *Genetics*, **177**, 469–480.
738 Kaminker, J.S., Bergman, C.M., Kronmiller, B. *et al.* (2002) The transposable elements of
739 the *Drosophila melanogaster* euchromatin: a genomics perspective. *Genome Biol.*,
740 **3**, research0084.
741 Kao JY, Zubair A, Salomon MP, Nuzhdin SV, Campo D (2015) Population genomic analysis
742 uncovers African and European admixture in *Drosophila melanogaster* populations from the
743 south-eastern United States and Caribbean Islands. *Molecular Ecology*, **24**, 1499–1509.
744 Kapopoulou A, Kapun M, Pavlidis P, *et al*. (2018a) Early split between African and European
745 populations of *Drosophila melanogaster*. Preprint at *bioRxiv*, doi: https://doi.org/10.1101/340422
746 Kapopoulou A, Pfeifer S, Jensen J, Laurent S (2018b). The demographic history of African
747 *Drosophila melanogaster*. Preprint at *bioRxiv*, doi:10.1101/340406
748 Kapun M, Flatt T (2019) The adaptive significance of chromosomal inversion polymorphisms in
749 *Drosophila melanogaster*. *Molecular Ecology*, **28**, 1263-1282
750 Kapun M, Fabian DK, Goudet J, Flatt T (2016a) Genomic Evidence for Adaptive Inversion Clines in
751 *Drosophila melanogaster*. *Molecular Biology and Evolution*, **33**, 1317–1336.
752 Kapun M, Schmidt C, Durmaz E, Schmidt PS, Flatt T (2016b) Parallel effects of the inversion
753 *In(3R)Payne* on body size across the North American and Australian clines in *Drosophila*
754 *melanogaster*. *Journal of Evolutionary Biology*, **29**, 1059–1072.

Kapun M, van Schalkwyk H, McAllister B, Flatt T, Schlötterer C (2014) Inference of chromosomal inversion dynamics from Pool-Seq data in natural and laboratory populations of *Drosophila melanogaster*. *Molecular Ecology*, **23**, 1813–1827.

Keller A (2007) *Drosophila melanogaster*'s history as a human commensal. *Current Biology*, **17**, R77–R81.

Kennington JW, Partridge L, Hoffmann AA (2006) Patterns of Diversity and Linkage Disequilibrium Within the Cosmopolitan Inversion *In(3R)Payne* in *Drosophila melanogaster* Are Indicative of Coadaptation. *Genetics*, **172**, 1655 – 1663.

Kimura M (1984) *The Neutral Theory of Molecular Evolution*. Cambridge University Press.

Knibb WR, Oakeshott JG, Gibson JB (1981) Chromosome Inversion Polymorphisms in *Drosophila melanogaster*. I. Latitudinal Clines and Associations between Inversions in Australasian Populations. *Genetics*, **98**, 833–847.

Knief U, Bossu CM, Saino N, Hansson B, Poelstra J, Vijay N, Weissensteiner M, Wolf JBW. (2019). Epistatic mutations under divergent selection govern phenotypic variation in the crow hybrid zone. *Nat Ecol Evol*, **3**, 570–576.

Kofler R, Schlötterer C (2012) GOwinda: Unbiased analysis of gene set enrichment for genome-wide association studies. *Bioinformatics*, **28**, 2084-2085.

Kofler R, Betancourt AJ, Schlötterer C (2012) Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in *Drosophila melanogaster.* *PLoS Genetics*, **8**, e1002487.

Kofler R, Orozco-terWengel P, De Maio N *et al.* (2011) PoPoolation: A Toolbox for Population Genetic Analysis of Next Generation Sequencing Data from Pooled Individuals. *PLoS ONE*, **6**, e15925.

Kolaczkowski B, Kern AD, Holloway AK, Begun DJ (2011) Genomic Differentiation Between Temperate and Tropical Australian Populations of *Drosophila melanogaster*. *Genetics*, **187**, 245–260.

Kreitman M (1983) Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. *Nature*, **304**, 412–417.

783 Kriesner P, Conner WR, Weeks AR, Turelli M, Hoffmann AA (2016) Persistence of a *Wolbachia*
784 infection frequency cline in *Drosophila melanogaster* and the possible role of reproductive
785 dormancy. *Evolution*, **70**, 979–997.
786 Lachaise D, Cariou M-L, David JR *et al.* (1988) Historical Biogeography of the *Drosophila*
787 *melanogaster* Species Subgroup. In Hecht MK, Wallace B, Prance GT (Eds.) *Evolutionary*
788 *Biology* (pp. 159–225) Boston: Springer.
789 Lack JB, Cardeno CM, Crepeau MW *et al.* (2015) The *Drosophila* genome nexus: a population
790 genomic resource of 623 *Drosophila melanogaster* genomes, including 197 from a single
791 ancestral range population. *Genetics*, **199**, 1229–1241.
792 Lack JB, Lange JD, Tang AD, Corbett-Detig RB, Pool JE (2016) A Thousand Fly Genomes: An
793 Expanded *Drosophila* Genome Nexus. *Molecular Biology and Evolution*, **33**, 3308–3313.
794 Langley CH, Stevens K, Cardeno C *et al.* (2012) Genomic variation in natural populations of
795 *Drosophila melanogaster*. *Genetics*, **192**, 533–598.
796 Larracuente AM, Roberts DM (2015) Genetics on the Fly: A Primer on the *Drosophila* Model
797 System. *Genetics* **201**, 815–842.
798 Lawrie DS, Messer PW, Hershberg R, Petrov DA (2013) Strong Purifying Selection at Synonymous
799 Sites in *D. melanogaster*. *PLoS Genetics,* **9**, e1003527.
800 Lerat E, Goubert C, Guirao-Rico S, Merenciano M, Dufour A-B, Vieira C, González J (2019)
801 Population-specific dynamics and selection patterns of transposable element insertions in
802 European natural populations. *Molecular Ecology*, **28**,1506–1522.
803 Lemeunier F, Aulard S (1992). Inversion polymorphism in *Drosophila melanogaster.* In: Krimbas
804 CB, & Powell JR (Eds.), *Drosophila Inversion Polymorphism* (pp. 339–405), New York: CRC
805 Press.
806 Lewontin RC (1974) *The Genetic Basis of Evolutionary Change*. Columbia University Press.
807 Li H, Ruan J, Durbin R (2008) Mapping short DNA sequencing reads and calling variants using
808 mapping quality scores. *Genome Research*, **18**, 1851–1858.
809 Li H, Stephan W (2006) Inferring the Demographic History and Rate of Adaptive Substitution in
810 *Drosophila*. PLoS Genetics **2**, 10.

811 Machado HE, Bergland AO, O'Brien KR *et al.* (2016) Comparative population genomics of
812 latitudinal variation in *Drosophila simulans* and *Drosophila melanogaster*. *Molecular Ecology*,
813 **25**, 723–740.
814 Machado H, Bergland AO, Taylor R *et al.* (2019) Broad geographic sampling reveals predictable,
815 pervasive, and strong seasonal adaptation in *Drosophila*. Preprint at *bioRxiv*, doi:
816 https://doi.org/10.1101/337543.
817 Macholán M, Baird SJ, Munclinger P, Dufková P, Bímová B, Piálek J. (2008). Genetic conflict
818 outweighs heterogametic incompatibility in the mouse hybrid zone? *BMC Evolutionary Biology*
819 **8**:271.
820 Martino ME, Ma D, Leulier F (2017) Microbial influence on *Drosophila* biology. *Current Opinion in*
821 *Microbiology*, **38**, 165–170.
822 Mateo L, Rech GE, González J (2018) Genome-wide patterns of local adaptation in *Drosophila*
823 *melanogaster*: adding intra European variability to the map. Preprint at *bioRxiv*, doi:
824 https://doi.org/10.1101/269332
825 McDonald JH, Kreitman M (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila*.
826 *Nature*, **351**, 652–654.
827 Mettler LE, Voelker RA, Mukai T (1977) Inversion Clines in Populations of *Drosophila*
828 *melanogaster*. *Genetics*, **87**, 169–176.
829 Meyer F, Paarmann D, D'Souza M *et al.* (2008) The metagenomics RAST server - a public resource
830 for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*, **9**,
831 386.
832 Michalakis Y, Veuille M (1996) Length variation of CAG/CAA trinucleotide repeats in natural
833 populations of *Drosophila melanogaster* and its relation to the recombination rate. *Genetics*, **143**,
834 1713–1725.
835 Nielsen R, Akey JM, Jakobsson M *et al.* (2017) Tracing the peopling of the world through genomics.
836 *Nature*, **541**, 302-310.
837 Novembre J, Johnson T, Bryc K, *et al.* (2008) Genes mirror geography within Europe. *Nature*, **456**,
838 98-101.

839 Paaby AB, Bergland AO, Behrman EL, Schmidt PS (2014) A highly pleiotropic amino acid
840 polymorphism in the *Drosophila* insulin receptor contributes to life-history adaptation. *Evolution*,
841 **68**, 3395-3409.
842 Paaby AB, Blacket MJ, Hoffmann AA, Schmidt PS (2010) Identification of a candidate adaptive
843 polymorphism for *Drosophila* life history by parallel independent clines on two continents.
844 *Molecular Ecology*, **19**, 760-774.
845 Palmer WH, Medd NC, Beard PM, Obbard DJ (2018) Isolation of a natural DNA virus of Drosophila
846 melanogaster, and characterisation of host resistance and immune responses. *PLOS Pathogens*,
847 **14**, e1007050
848 Parsch J, Novozhilov S, Saminadin-Peter SS, Wong KM, Andolfatto P (2010) On the utility of short
849 intron sequences as a reference for the detection of positive and negative selection in *Drosophila*.
850 *Molecular Biology and Evolution*, **27**, 1226–1234.
851 Peel MC, Finlayson BL, McMahon TA (2007) Updated world map of the Köppen-Geiger climate
852 classification. *Hydrology and Earth System Sciences*, **11**, 1633–1644.
853 Petrov DA, Fiston-Lavier AS, Lipatov M, Lenkov K, González J (2011) Population Genomics of
854 Transposable Elements in *Drosophila melanogaster*. *Molecular Biology and Evolution*, **28**,
855 1633–1644.
856 Pool JE, Braun DT, Lack JB (2016) Parallel Evolution of Cold Tolerance Within *Drosophila*
857 *melanogaster*. *Molecular Biology and Evolution*, **34,** 349–360.
858 Pool JE, Corbett-Detig RB, Sugino RP *et al.* (2012) Population Genomics of Sub-Saharan *Drosophila*
859 *melanogaster*: African Diversity and Non-African Admixture. *PLoS Genetics*, **8**, e1003080.
860 Powell JR (1997) *Progress and Prospects in Evolutionary Biology: The Drosophila Model*. Oxford
861 University Press.
862 Quesneville H, Bergman CM, Andrieu O, Autard D, Nouaud D, et al. (2005) Combined evidence
863 annotation of transposable elements in genome sequences. *PLoS Comp Biol* **1(2)**: e22.
864 Rako L, Anderson AR, Sgrò CM, Stocker AJ, Hoffmann AA (2006) The association between
865 inversion *In(3R)Payne* and clinally varying traits in *Drosophila melanogaster. Genetica*, **128**,
866 373–384.

867 Rane RV, Rako L, Kapun M, LEE SF (2015) Genomic evidence for role of inversion *3RP* of
868 *Drosophila melanogaster* in facilitating climate change adaptation. *Molecular Ecology*, **24**,
869 2423–2432.
870 Rech GE, Bogaerts-Márquez M, Barrón MG, Merenciano M, Villanueva-Cañas JL, Horváth V,
871 Fiston-Lavier A-S, Luyten I, Venkataram S, Quesneville H, Petrov DA, González J (2019) Stress
872 response, behavior, and development are shaped by transposable element-induced mutations in
873 *Drosophila*. *PLOS Genetics,* **15**, e1007900.
874 Richardson MF, Weinert LA, Welch JJ *et al.* (2012) Population Genomics of the *Wolbachia*
875 Endosymbiont in *Drosophila melanogaster*. *PLoS Genetics*, **8**, e1003129.
876 Rogers RL, Hartl DL (2012) Chimeric genes as a source of rapid evolution in *Drosophila*
877 *melanogaster*. *Molecular Biology and Evolution*, **29**, 517–529.
878 Schlötterer C, Tobler R, Kofler R, Nolte V (2014) Sequencing pools of individuals - mining genome-
879 wide polymorphism data without big funding. *Nature Reviews Genetics*, **15**, 749–763.
880 Schmidt PS, Paaby AB (2008) Reproductive Diapause and Life-History Clines in North American
881 Populations of *Drosophila melanogaster*. *Evolution*, **62**, 1204–1215.
882 Schmidt PS, Zhu CT, Das J *et al.* (2008) An amino acid polymorphism in the *couch potato* gene
883 forms the basis for climatic adaptation in *Drosophila melanogaster*. *Proceedings of the National*
884 *Academy of Sciences of the United States of America*, **105**, 16207–16211.
885 Singh ND, Arndt PF, Clark AG, Aquadro CF (2009) Strong evidence for lineage and sequence
886 specificity of substitution rates and patterns in *Drosophila*. *Molecular Biology and Evolution*, **26**,
887 1591–1605.
888 Sprengelmeyer QD, Mansourian S, Lange JD, Matute DR, Cooper BS, Jirle EV, Stensmyr MC, Pool
889 JE. (2020). Recurrent Collection of Drosophila melanogaster from Wild African Environments
890 and Genomic Insights into Species History. *Mol Biol Evol* **37**:627–638.
891 Staubach F, Baines JF, Künzel S, Bik EM, Petrov DA (2013) Host species and environmental effects
892 on bacterial communities associated with *Drosophila* in the laboratory and in the natural
893 environment. *PLoS ONE*, **8**, e70749.

894 Szymura JM, Barton NH. (1986). Genetic analysis of a hybrid zone between the fire-bellied toads,
895 *Bombina bombina* and *B. variegata*, near Cracow in Southern Poland. *Evolution* **40**:1141–1159.
896 Tauber E, Zordan, M, Sandrelli F, Pegoraro M, Osterwalder N, Breda C, Daga A, Selmin A, Monger
897 K, Benna C, Rosata E, Kyriacou CP, Costa R (2007) Natural selection favors a newly derived
898 *timeless* allele in *Drosophila melanogaster*. *Science,* **316,**1895-1899.
899 Trinder M, Daisley BA, Dube JS, Reid G (2017) *Drosophila melanogaster* as a High-Throughput
900 Model for Host-Microbiota Interactions. *Frontiers in Microbiology*, **8**, 751.
901 Turner TL, Levine MT, Eckert ML, Begun DJ (2008) Genomic analysis of adaptive differentiation in
902 *Drosophila melanogaster*. *Genetics*, **179**, 455–473.
903 Umina PA, Weeks AR, Kearney MR, McKechnie SW, Hoffmann AA (2005) A rapid shift in a classic
904 clinal pattern in *Drosophila* reflecting climate change. *Science*, **308**, 691–693.
905 Unckless RL (2011) A DNA virus of *Drosophila*. *PLoS ONE*, **6**, e26564.
906 de Villemereuil P, Gaggiotti OE (2015) A new FST-based method to uncover local adaptation using
907 environmental variables. *Methods in Ecology and Evolution*. **6**: 1248 – 1258.
908 Walters AM, Matthews MK, Hughes R, Malcolm Jaanna, Rudman S, Newell PD, Douglas AE,
909 Schmidt PS, Chaston JM (2018) The microbiota influences the Drosophila melanogaster life
910 history strategy. bioRxiv. 471540
911 Wang Y, Kapun M, Waidele L, Kuenzel S, Bergland AO, Staubach F. (2020). Common structuring
912 principles of the *Drosophila melanogaster* microbiome on a continental scale and between host
913 and substrate. Environmental Microbiology Reports **12**:220–228.
914 Wang Y, Staubach F (2018); Individual variation of natural *D.melanogaster*-associated bacterial
915 communities, *FEMS Microbiology Letters*, **365**, fny017
916 Webster CL, Longdon B, Lewis SH, Obbard DJ (2016) Twenty-Five New Viruses Associated with
917 the Drosophilidae (Diptera). *Evolutionary Bioinformatics Online*, **12**, 13–25.
918 Webster CL, Waldron FM, Robertson S *et al.* (2015) The Discovery, Distribution, and Evolution of
919 Viruses Associated with *Drosophila melanogaster*. *PLoS Biology*, **13**, e1002210.
920 Werren JH, Baldo L, Clark ME (2008) Wolbachia: master manipulators of invertebrate biology.
921 *Nature Reviews Microbiology*, **6**, 741–751.

922 Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: $F_{ST} \neq 1/(4Nm+1)$.
923 Heredity, **82**,117–125.
924 Wilfert L, Longdon B, Ferreira AGA, Bayer F, Jiggins FM (2011) Trypanosomatids are common and
925 diverse parasites of *Drosophila*. *Parasitology*, **138**, 858–865.
926 Wolff JN, Camus MF, Clancy DJ, Dowling DK (2016) Complete mitochondrial genome sequences of
927 thirteen globally sourced strains of fruit fly (*Drosophila melanogaster*) form a powerful model
928 for mitochondrial research. *Mitochondrial DNA Part A,* **27**, 4672–4674.
929 Wright S (1951) The genetical structure of populations. *Ann Eugen* **15**, 323–354.
930 Xiao F-X, Yotova V, Zietkiewicz E *et al.* (2004) Human *X*-chromosomal lineages in Europe reveal
931 Middle Eastern and Asiatic contacts. *European Journal of Human Genetics*, **12**, 301–311.
932 Yukilevich R, True JR (2008a) Incipient sexual isolation among cosmopolitan *Drosophila*
933 *melanogaster* populations. *Evolution*, **62**, 2112–2121.
934 Yukilevich R, True JR (2008b) African morphology, behavior and phermones underlie incipient
935 sexual isolation between us and Caribbean *Drosophila melanogaster*. *Evolution*, **62**, 2807–2828.
936 Zanini F, Brodin J, Thebo L *et al.* (2015) Population genomics of intrapatient HIV-1 evolution. *eLife*,
937 **4**, e11282.

938 **Tables**

939 **Table 1. Sample information for all populations in the *DrosEU* dataset.** Origin, collection date, season and sample size (number of chromosomes: *n*) of

940 the 48 samples in the *DrosEU* 2014 data set. Additional information can be found in supplementary table S1 (Supplementary Material online).

ID	Country	Location	Coll. Date	Numbe r ID	Lat (°)	Lon (°)	Alt (m)	Season	n	Coll. name
AT_Mau_14_01	Austria	Mauternbach	2014-07-20	1	48.38	15.56	572	S	80	Andrea J. Betancourt
AT_Mau_14_02	Austria	Mauternbach	2014-10-19	2	48.38	15.56	572	F	80	Andrea J. Betancourt
TR_Yes_14_03	Turkey	Yesiloz	2014-08-31	3	40.23	32.26	680	S	80	Banu Sebnem Onder
TR_Yes_14_04	Turkey	Yesiloz	2014-10-23	4	40.23	32.26	680	F	80	Banu Sebnem Onder
FR_Vil_14_05	France	Viltain	2014-08-18	5	48.75	2.16	153	S	80	Catherine Montchamp-Moreau
FR_Vil_14_07	France	Viltain	2014-10-27	7	48.75	2.16	153	F	80	Catherine Montchamp-Moreau
FR_Got_14_08	France	Gotheron	2014-07-08	8	44.98	4.93	181	S	80	Cristina Vieira
UK_She_14_09	United Kingdom	Sheffield	2014-08-25	9	53.39	-1.52	100	S	80	Damiano Porcelli
UK_Sou_14_10	United	South Queensferry	2014-07-14	10	55.97	-3.35	19	S	80	Darren Obbard

	Kingdom									
CY_Nic_14_11	Cyprus	Nicosia	2014-08-10	11	35.07	33.32	263	S	80	Eliza Argyridou
	United									
UK_Mar_14_12	Kingdom	Market Harborough	2014-10-20	12	52.48	-0.92	80	F	80	Eran Tauber
	United									
UK_Lut_14_13	Kingdom	Lutterworth	2014-10-20	13	52.43	-1.10	126	F	80	Eran Tauber
DE_Bro_14_14	Germany	Broggingen	2014-06-26	14	48.22	7.82	173	S	80	Fabian Staubach
DE_Bro_14_15	Germany	Broggingen	2014-10-15	15	48.22	7.82	173	F	80	Fabian Staubach
UA_Yal_14_16	Ukraine	Yalta	2014-06-20	16	44.50	34.17	72	S	80	Iryna Kozeretska
UA_Yal_14_18	Ukraine	Yalta	2014-08-27	18	44.50	34.17	72	S	80	Iryna Kozeretska
UA_Ode_14_19	Ukraine	Odesa	2014-07-03	19	46.44	30.77	54	S	80	Iryna Kozeretska
UA_Ode_14_20	Ukraine	Odesa	2014-07-22	20	46.44	30.77	54	S	80	Iryna Kozeretska
UA_Ode_14_21	Ukraine	Odesa	2014-08-29	21	46.44	30.77	54	S	80	Iryna Kozeretska
UA_Ode_14_22	Ukraine	Odesa	2014-10-10	22	46.44	30.77	54	F	80	Iryna Kozeretska
UA_Kyi_14_23	Ukraine	Kyiv	2014-08-09	23	50.34	30.49	179	S	80	Iryna Kozeretska
UA_Kyi_14_24	Ukraine	Kyiv	2014-09-08	24	50.34	30.49	179	F	80	Iryna Kozeretska
UA_Var_14_25	Ukraine	Varva	2014-08-18	25	50.48	32.71	125	S	80	Oleksandra Protsenko

UA_Pyr_14_26	Ukraine	Pyriatyn	2014-08-20	26	50.25	32.52	114	S	80	Oleksandra Protsenko
UA_Dro_14_27	Ukraine	Drogobych	2014-08-24	27	49.33	23.50	275	S	80	Iryna Kozeretska
UA_Cho_14_28	Ukraine	Chornobyl	2014-09-13	28	51.37	30.14	121	F	80	Iryna Kozeretska
UA_Cho_14_29	Ukraine	Chornobyl Yaniv	2014-09-13	29	51.39	30.07	121	F	80	Iryna Kozeretska
SE_Lun_14_30	Sweden	Lund	2014-07-31	30	55.69	13.20	51	S	80	Jessica Abbott
DE_Mun_14_31	Germany	Munich	2014-06-19	31	48.18	11.61	520	S	80	John Parsch
DE_Mun_14_32	Germany	Munich	2014-09-03	32	48.18	11.61	520	F	80	John Parsch
PT_Rec_14_33	Portugal	Recarei	2014-09-26	33	41.15	-8.41	175	F	80	Jorge Vieira
ES_Gim_14_34	Spain	Gimenells (Lleida)	2014-10-20	34	41.62	0.62	173	F	80	Lain Guio
ES_Gim_14_35	Spain	Gimenells (Lleida)	2014-08-13	35	41.62	0.62	173	S	80	Lain Guio
FI_Aka_14_36	Finland	Akaa	2014-07-25	36	61.10	23.52	88	S	80	Maaria Kankare
FI_Aka_14_37	Finland	Akaa	2014-08-27	37	61.10	23.52	88	S	80	Maaria Kankare
FI_Ves_14_38	Finland	Vesanto	2014-07-26	38	62.55	26.24	121	S	66	Maaria Kankare
DK_Kar_14_39	Denmark	Karensminde	2014-09-01	39	55.95	10.21	15	F	80	Mads Fristrup Schou
DK_Kar_14_41	Denmark	Karensminde	2014-11-25	41	55.95	10.21	15	F	80	Mads Fristrup Schou
CH_Cha_14_42	Switzerland	Chalet à Gobet	2014-07-24	42	46.57	6.70	872	S	80	Martin Kapun
CH_Cha_14_43	Switzerland	Chalet à Gobet	2014-10-05	43	46.57	6.70	872	F	80	Martin Kapun

AT_See_14_44	Austria	Seeboden	2014-08-17	44	46.81	13.51	591	S	80	Martin Kapun
UA_Kha_14_45	Ukraine	Kharkiv	2014-07-26	45	49.82	36.05	141	S	80	Svitlana Serga
UA_Kha_14_46	Ukraine	Kharkiv	2014-09-14	46	49.82	36.05	141	F	80	Svitlana Serga
		Chornobyl								
UA_Cho_14_47	Ukraine	Applegarden	2014-09-13	47	51.27	30.22	121	F	80	Svitlana Serga
UA_Cho_14_48	Ukraine	Chornobyl Polisske	2014-09-13	48	51.28	29.39	121	F	70	Svitlana Serga
UA_Kyi_14_49	Ukraine	Kyiv	2014-10-11	49	50.34	30.49	179	F	80	Svitlana Serga
UA_Uma_14_50	Ukraine	Uman	2014-10-01	50	48.75	30.21	214	F	80	Svitlana Serga
RU_Val_14_51	Russia	Valday	2014-08-17	51	57.98	33.24	217	S	80	Elena Pasyukova

941

Table 2. Clinality of genetic variation and population structure. Effects of geographic variables and/or seasonality on genome-wide average levels of diversity (π, θ and Tajima's D; top rows) and on the first three axes of a PCA based on allele frequencies at neutrally evolving sites (bottom rows). The values represent F-ratios from general linear models. Bold type indicates F-ratios that are significant after Bonferroni correction (adjusted α'=0.0055). Asterisks in parentheses indicate significance when accounting for spatial autocorrelation by spatial error models. These models were only calculated when Moran's I test, as shown in the last column, was significant. *p < 0.05; **p < 0.01; ***p < 0.001.

Factor	***Latitude***	***Longitude***	***Altitude***	***Season***	***Moran's I***
$\pi_{(X)}$	4.11*	1.62	**15.23*** **	1.65	0.86
$\pi_{(Aut)}$	0.91	2.54	**27.18*** **	0.16	-0.86
$\theta_{(X)}$	2.65	1.31	**15.54*** **	2.22	0.24
$\theta_{(Aut)}$	0.48	1.44	**13.66*** **	0.37	-1.13
$D_{(X)}$	0.02	0.38	5.93*	3.26	-2.08
$D_{(Aut)}$	0.09	0.76	5.33*	0.71	-1.45
PC1	0.63	**118.08*** **(*** *)**	3.64	0.75	**4.2*** **
PC2	**4.69***	**7.15***	**11.77*** *	1.68	-0.32
PC3	0.39	0.23	**19.91*** **	0.28	1.38

949 **Table 3**. **Clinality and/or seasonality of chromosomal inversions.** The values represent *F*-ratios from generalized linear models with a binomial error
950 structure to account for frequency data. Bold type indicates deviance values that were significant after Bonferroni correction (adjusted α'=0.0071). Stars in
951 parentheses indicate significance when accounting for spatial autocorrelation by spatial error models. These models were only calculated when Moran's *I* test,
952 as shown in the last column, was significant. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$
953

Factor	***Latitude***	***Longitude***	***Altitude***	***Season***	***Moran's I***
In(2L)t	2.2	**10.09****	**43.94*****	0.89	-0.92
In(2R)NS	0.25	**14.43*****	2.88	2.43	1.25
In(3L)P	**21.78*****	2.82	0.62	3.6	-1.61
In(3R)C	**18.5***(***)**	0.75	1.42	0.04	2.79**
In(3R)Mo	0.3	0.09	0.35	0.03	-0.9
In(3R)Payne	**43.47*****	0.66	1.69	1.55	-0.89

954 **FIGURE LEGENDS**

955

956 **Fig. 1. The geographic distribution of population samples.** Locations of all samples in the 2014
957 *DrosEU* data set. The color of the circles indicates the sampling season for each location: ten of the
958 32 locations were sampled at least twice, once in summer and once in fall (see table 1 and
959 supplementary table S1, Supplementary Material online). Note that some of the 12 Ukrainian
960 locations overlap in the map.

961 **Fig. 2**. **Signals of selective sweeps in European populations.** The central panel shows the
962 distribution of Tajima's *D* in 50 kb sliding windows with 40 kb overlap, with red and green dashed
963 lines indicating Tajima's $D = 0$ and -1, respectively. The top panel shows a detail of a genomic region
964 on chromosomal arm *2R* in the vicinity of *Cyp6g1* and *Hen1* (highlighted in red), genes reportedly
965 involved in pesticide resistance. This strong sweep signal is characterized by an excess of low-
966 frequency SNP variants and overall negative Tajima's *D* in all samples. Colored solid lines depict
967 Tajima's *D* for each sample (see supplementary fig. S2 for color codes, Supplementary Material
968 online); the black dashed line shows Tajima's *D* averaged across all samples. The bottom panel shows
969 a region on *3L* previously identified as a potential target of selection, which shows a similar strong
970 sweep signature. Notably, both regions show strongly reduced genetic variation (supplementary fig.
971 S1, Supplementary Material online).

972 **Fig. 3. Genetic differentiation among European populations**. (A) Average F_{ST} among populations
973 at putatively neutral sites. The centre plot shows the distribution of F_{ST} values for all 1,128 pairwise
974 population comparisons, with the F_{ST} values for each comparison obtained from the mean across all
975 4,034 SNPs used in the analysis. Plots on the left and the right show population pairs in the lower
976 (blue) and upper (red) 5% tails of the F_{ST} distribution. (B) PCA analysis of allele frequencies at the
977 same SNPs reveals population sub-structuring in Europe. Hierarchical model fitting using the first
978 four PCs showed that the populations fell into two clusters (indicated by red and blue), with cluster
979 assignment of each population subsequently estimated by *k*-means clustering. (C) Admixture
980 proportions for each population inferred by model-based clustering with *ConStruct* are highlighted as

pie charts (left plot) or Structure plots (centre). The optimal number of 3 spatial layers (K) was inferred by cross-validation (right plot).

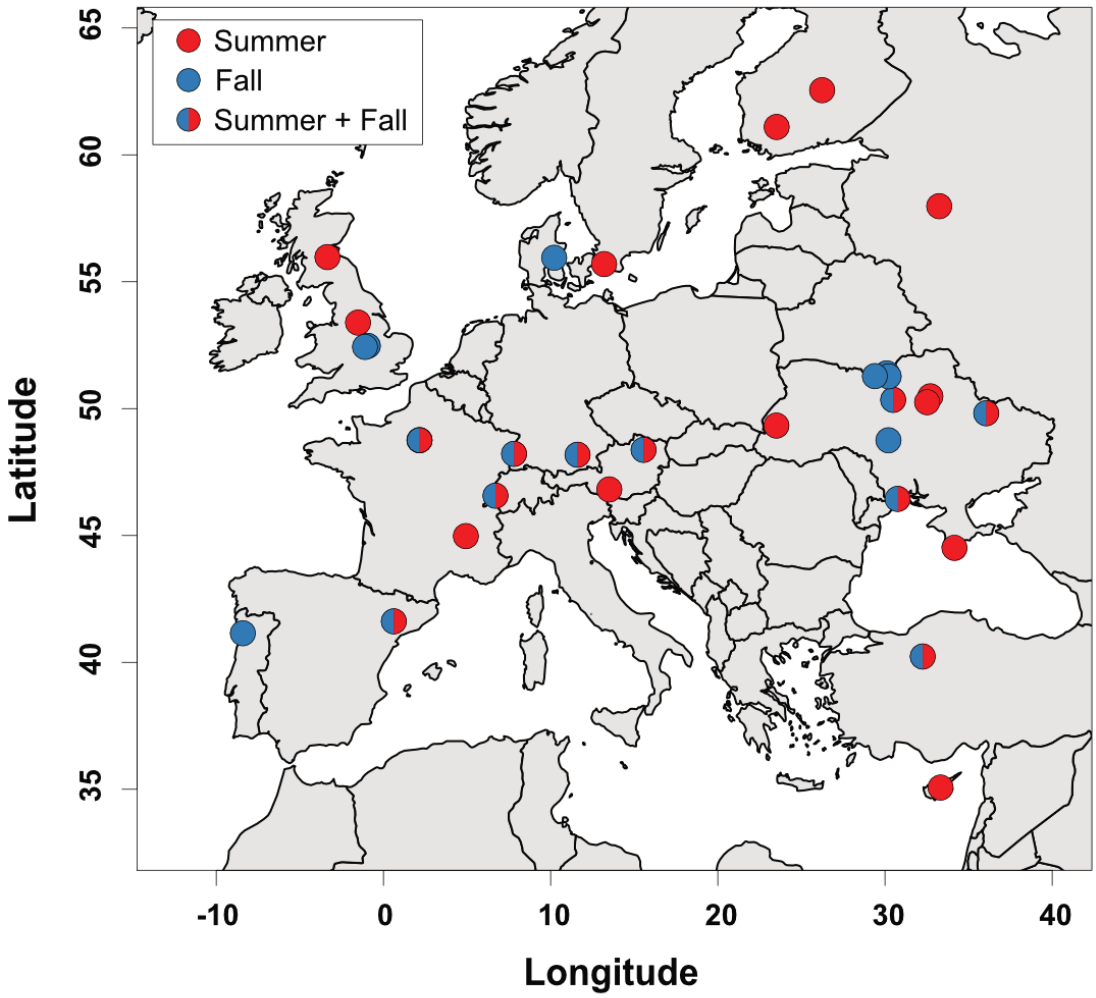
Fig. 4. Manhattan plot of SNPs with *q*-values < 0.05 in association tests with PC1 or PC2 of the bioclimatic variables. Vertical lines denote the breakpoints of common inversions. The gene names highlight some candidate genes found in our study and which have previously been identified as varying clinally by Fabian *et al.* (2012) and Machado *et al.* (2016) along the North American east coast. Note that for ease of plotting, *q*-values of 0 were set to 10% of the smallest observed *q*-value.

Fig. 5. Mitochondrial haplotypes. (A) TCS network showing the relationship of 5 common mitochondrial haplotypes; (B) estimated frequency of each mitochondrial haplotype in 48 European samples.

Fig. 6. **Geographic patterns in structural variants**. The upper panel shows stacked bar plots with the relative abundances of TEs in all 48 population samples**.** The proportion of each repeat class was estimated from sampled reads with dnaPipeTE (2 samples per run, 0.1X coverage per sample). The lower panel shows stacked bar plots depicting absolute frequencies of six cosmopolitan inversions in all 48 population samples.

Fig. 7: Microbiome. Relative abundance of *Drosophila*-associated microbes as assessed by MGRAST classified shotgun sequences. Microbes had to reach at least 3% relative abundance in one of the samples to be represented

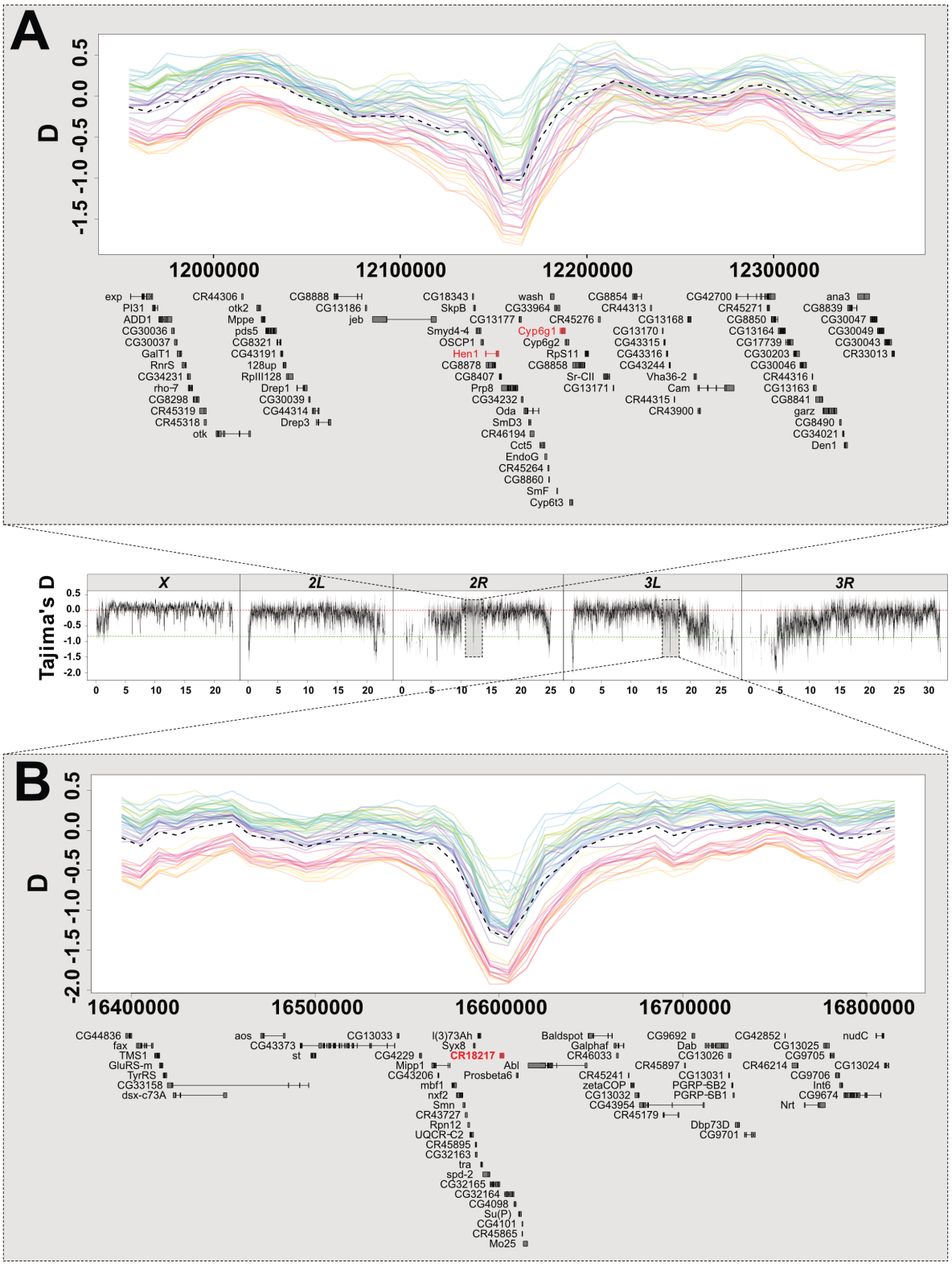
1002 **Figure 1**



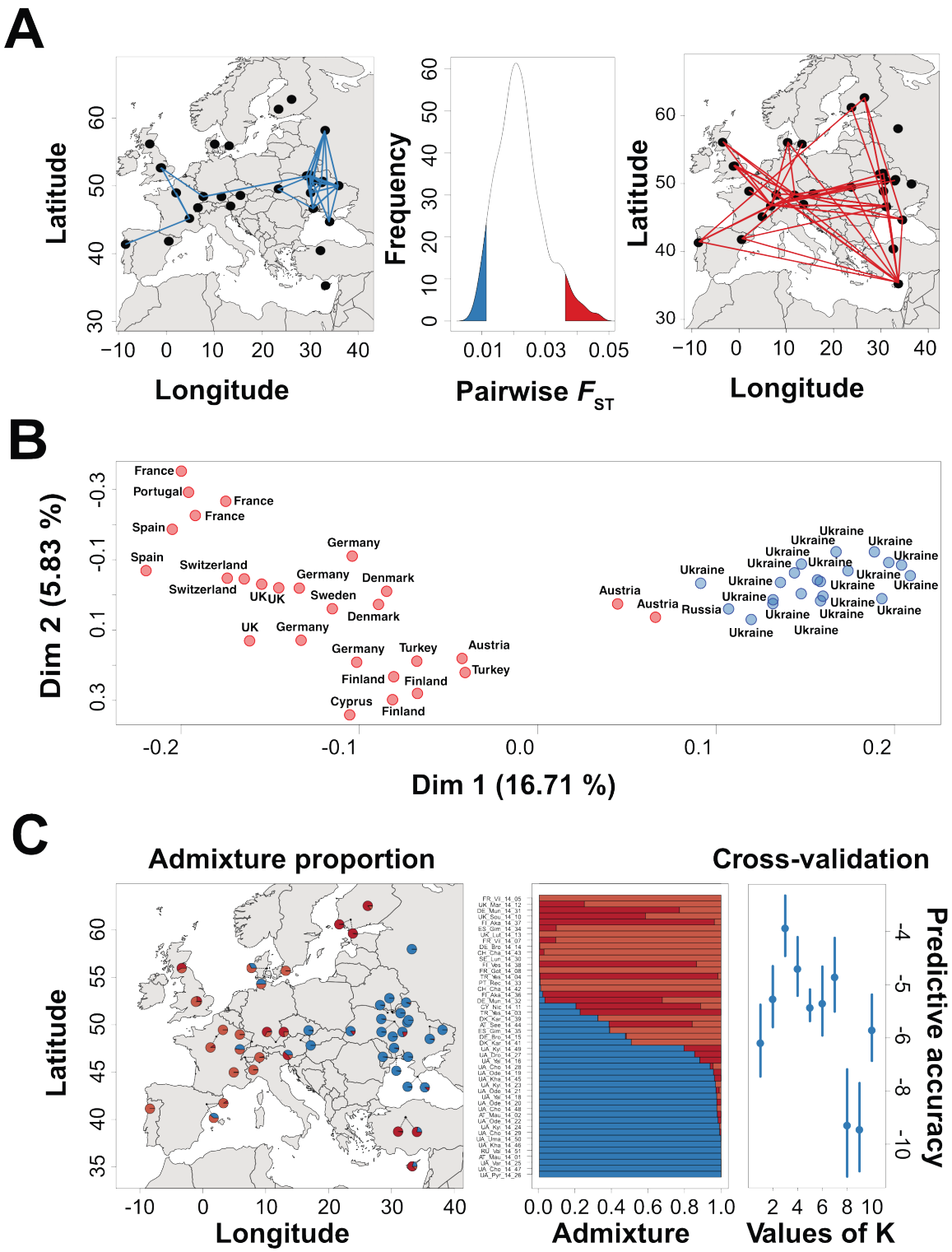
1003

1004

Figure 2



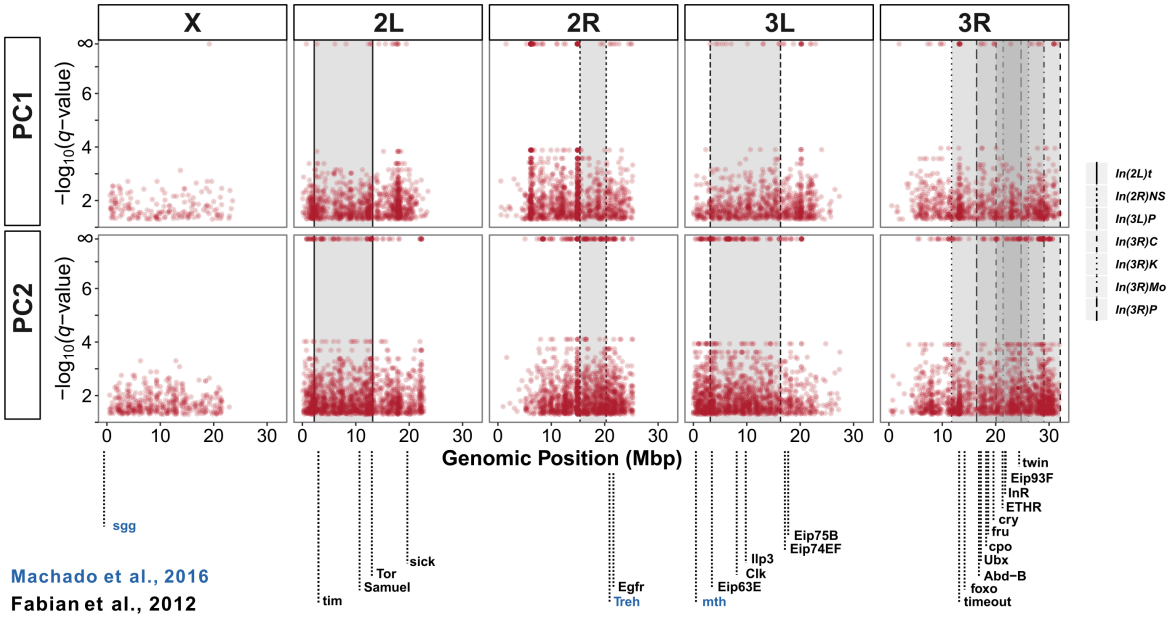
1008 **Figure 3**



1009

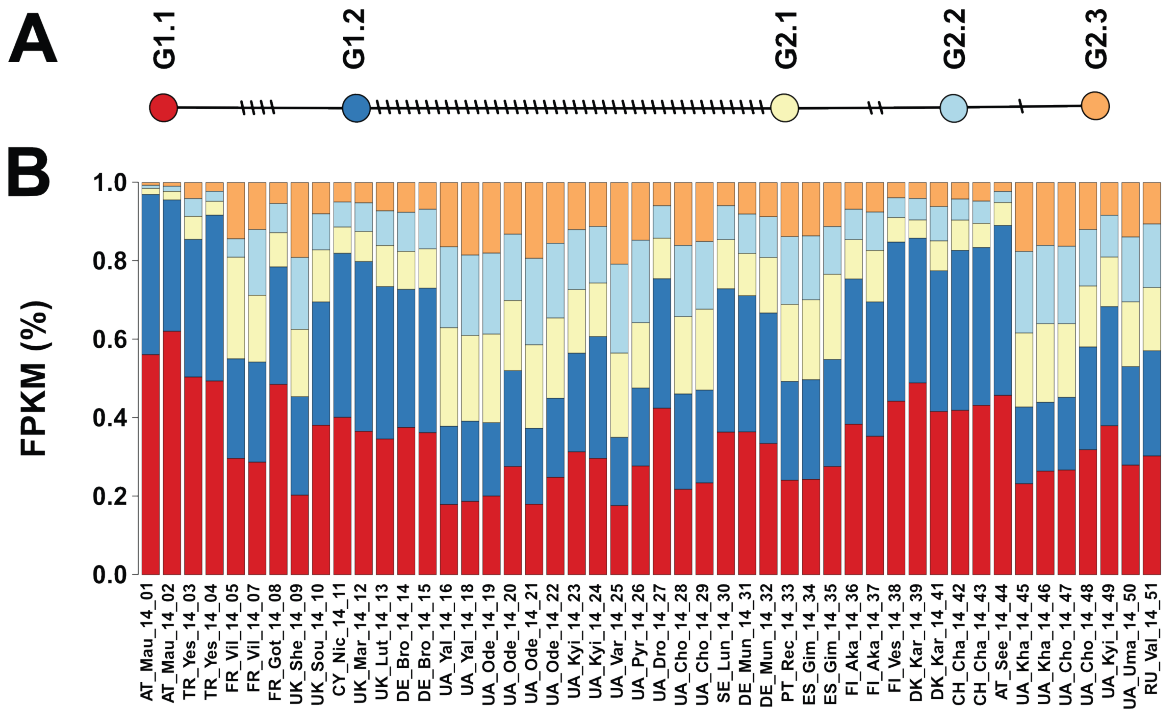
1010

1011 **Figure 4**



1012

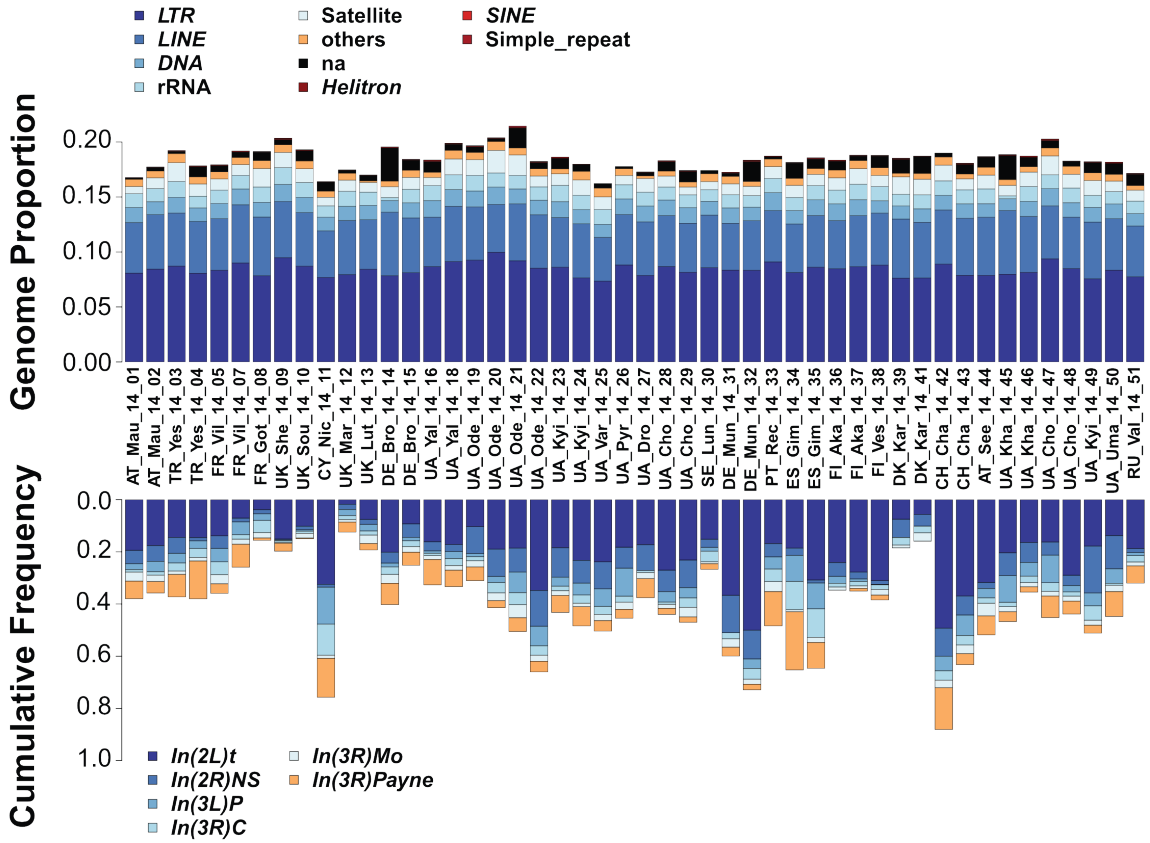
1013 **Figure 5**



1014

1015

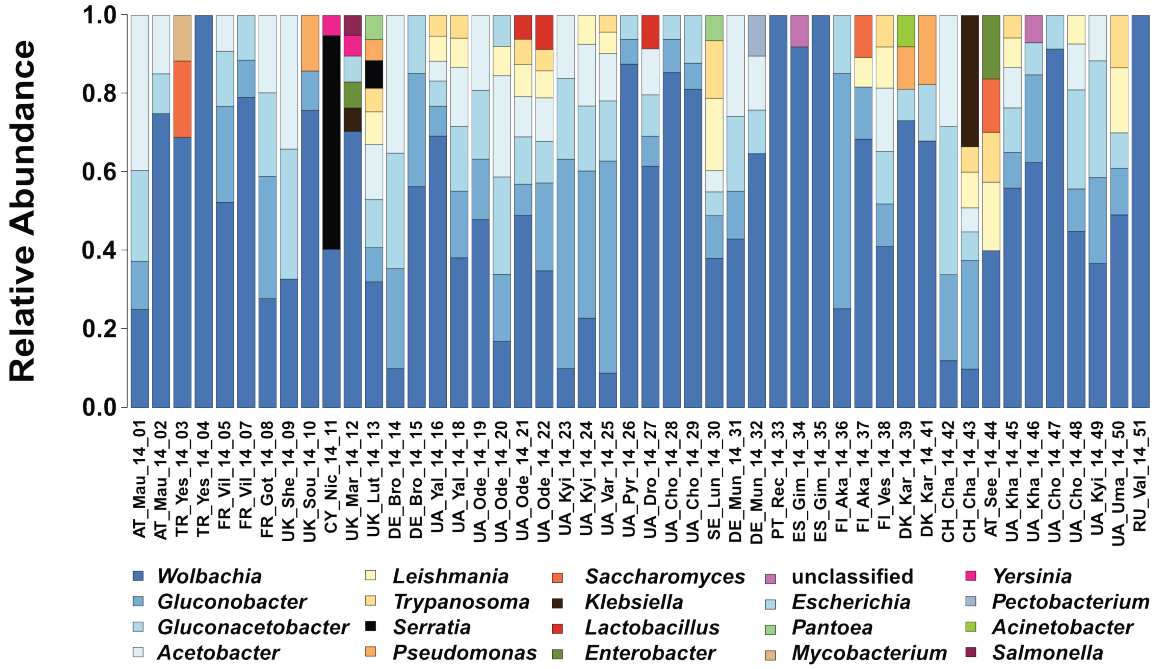
1016 **Figure 6**



1017

1018

1019 **Figure 7**



1020