Genomic signals of admixture and reinforcement between two closely related species of European sepsid flies

Athene Giesen[1], Wolf U. Blanckenhorn[1], Martin A. Schäfer[1], Kentaro K. Shimizu[1], Rie Shimizu-Inatsugi[1], Bernhard Misof[2], Oliver Niehuis[2,3], Lars Podsiadlowski[2], Heidi E. L. Lischer[1,4], Simon Aeschbacher[1], Martin Kapun[1,5]

[1]Department of Evolutionary Biology and Environmental Studies, University of Zürich, Switzerland

[2]Zoological Research Museum Alexander Koenig, Bonn, Germany

[3]Department of Evolutionary Biology and Ecology, Institute of Biology I (Zoology), Albert Ludwig University, Freiburg, Germany

[4]Interfaculty Bioinformatics Unit, University of Bern, Switzerland

[5]Division of Cell & Developmental Biology, Medical University of Vienna, Austria

Corresponding author: Martin Kapun; Winterthurerstrasse 190, CH-8057 Zürich, Switzerland; Tel.: +41 44 635 47 55; martin.kapun@ieu.uzh.ch

ACKNOWLEDGEMENTS

Re-sequencing of the genomes was performed at the Functional Genomics Center Zürich, where Jelena Kühn-Georgijevic was an important advisor. Thanks also to the Wolfpack at the University of Zürich for their support and help in maintaining fly cultures, and Patrick Rohner, Juan Pablo Busso, Nalini Puniamoorthy, Anders Kjaersgaard, Toomas Esperk and Cait Dmitriew for contributing samples. B.M., O.N. and L.P. thank the Leibniz association for granting funds to establish the Leibniz graduate school for Genomic Biodiversity Research (GBR), in which context the reference draft genomes of several fly species of the genus *Sepsis* explored in this study were sequenced and assembled. We particularly thank the GBR members Jeanne Wilbrandt, Jan Philip Øyen and Malte Petersen for their help with sample collection and for their contribution to the assembly pipeline. The University Research Priority Program 'Evolution in Action' of the University of Zürich, the Swiss National Science Foundation, and the Georges and Antoine Claraz-Donation funded the work of A.G., the Austrian Science Foundation (grant no. FWF P32275) the work of M.K.

ABSTRACT

Interspecific gene flow by hybridization may weaken species barriers and adaptive divergence, but also initiate reinforcement of reproductive isolation trough natural and sexual selection. The extent and consequences of interspecific gene flow in natural systems remain poorly understood. To assess genome-wide patterns of gene flow between the two closely related European dung fly species *Sepsis cynipsea* and *Sepsis neocynipsea* (Diptera: Sepsidae), we analyzed whole-genome resequencing data from pooled males of laboratory and field populations by means of the ABBA-BABA test. We contrasted genome-wide variation in DNA sequence differences between samples from sympatric populations of these two species in France and Switzerland with that of interspecific pairs of samples involving allopatric populations from Estonia and Italy. In the French Cevennes, we detected a relative excess of sequence identity suggesting interspecific gene flow in sympatry. In contrast, at two sites in Switzerland we observed a relative depletion of sequence identity compatible with reinforcement of species boundaries in sympatry. Our results suggest that the species boundaries between *S. cynipsea* and *S. neocynipsea* in Europe are not fixed but depend on the eco-geographic context.

KEYWORDS: ABBA-BABA test, gene flow, hybridization, introgression, reproductive isolation, speciation, sepsid flies

INTRODUCTION

According to the biological species concept, species represent groups of individuals that are reproductively isolated from other groups of individuals (Mayr 1942). Speciation entails the evolution of reproductive isolation among lineages derived from a common ancestral population and is considered completed if the diverged populations remain reproductively isolated, for example, after coming into secondary contact (Coyne & Orr, 2004; Dobzhansky, 1951; Mayr, 1963). However, many animal and plant species remain distinct entities in nature even if they hybridize and exchange genes in parapatry or sympatry (Anderson, 1949; Barton & Bengtsson, 1986; cf. DeMarais *et al*., 1992; Gante *et al*., 2016; Mallet, 2007; Nolte & Tautz, 2010; Rieseberg *et al*., 2003; Trier *et al*., 2014). Research during the past decades has indicated that hybridization not only has deleterious effects due to hybrid inferiority and negative epistasis in admixed genomes, but that it may also fuel adaptive diversification and speciation by facilitating novel combinations of alleles that become targets of divergent selection (Arnold & Meyer, 2006; Berner & Salzburger, 2015; Fontaine *et al*. 2015; Seehausen, 2004; Saetre, 2013). Thus, the genomes of recently diverged species may be viewed as mosaics of genomic regions with significant differentiation (i.e., low to no admixture) interspersed with more freely mixing regions of DNA (Nosil *et al*., 2009; Wu, 2001).

Sepsid flies (Diptera: Sepsidae) have served as model organisms in various studies on sexual selection and ecological adaptation (Blanckenhorn, 1999; Blanckenhorn *et al*., 2000; Eberhard, 1999; Kraushaar & Blanckenhorn, 2002; Parker, 1972a,b; Puniamoorthy *et al*., 2009; Pont & Meier, 2002; Rohner *et al*., 2015; Rohner, Blanckenhorn, & Puniamoorthy, 2016; Ward, 1983; Ward, Hemmi, & Rösli, 1992), and their phylogeny is well resolved (Su *et al*., 2008, 2016). Pairs of closely related species with similar ecology provide excellent opportunities to study the genomic consequences of hybridization and introgression during early stages of speciation. *Sepsis cynipsea* and *Sepsis neocynipsea* are widespread species that occur in sympatry across major parts of their natural range. While *S. cynipsea* is the most abundant sepsid species in north-central Europe laying its eggs in fresh cow dung, *S. neocynipsea* is common throughout North America, where its niche is complementary to that of *S. cynipsea* in Europe. While overall very rare in Europe, *S. neocynipsea* can be locally common at higher altitude sites, such as the Alps, where the species occurs in sympatry with *S. cynipsea* (Ozerov, 2005; Pont & Meier, 2002; Rohner, Blanckenhorn, & Puniamoorthy,

2016). Despite strong similarities in morphology and behavior (Giesen, Blanckenhorn, & Schäfer, 2017), the two species are genetically distinct. To a lesser extent this also holds for *S. neocynipsea* populations in North America vs. Europe (Baur *et al*., 2020). Previous studies showed that *S. cynipsea* and *S. neocynipsea* can hybridize under laboratory conditions despite strong pre- and post-mating isolating barriers. These barriers are mediated by assortative mating behaviors that are partly reinforced in areas where the two species occur in sympatry (Giesen *et al*., 2017, 2019). While these results imply that interspecific gene flow may occur in nature and vary with eco-geographic context, we know little about actual levels historical and ongoing gene flow between the two species in areas where they occur in sympatry or parapatry. We therefore investigated the extent of gene flow between *S. cynipsea* and *S. neocynipsea* in nature using comparative genomic methods.

In recent years, a number of novel genomic approaches have been employed in a variety of taxa to explore this classic question of interspecific gene flow in sympatry vs. allopatry from several angles (e.g. Nakazato, Warren, & Moyle, 2010; Nadeau *et al.,* 2013; Brandvain *et al.,* 2014; Bouchemousse *et al.,* 2016; Feulner & Seehausen, 2019; Kastally, Trasoletti, & Mardulyn, 2019; Moran *et al.,* 2019). We here applied ABBA-BABA tests (Green *et al*., 2010; Durand *et al*., 2011; Soraggi et al., 2018) to genome-wide allele frequency data of single nucleotide polymorphisms (SNPs) to assess the extent of genome-wide admixture between *S. cynipsea* and *S. neocynipsea* in Europe. We sequenced pools of males from laboratory and wild-caught populations of both species collected at several sites ranging from southern France to Estonia. Some of these populations occur in sympatry (e.g., in the Swiss Alps and the French Cevennes), others in allopatry (Pont & Meier, 2002). Based on hybridization opportunity alone, we expected to find higher genome-wide levels of gene flow between *S. cynipsea* and *S. neocynipsea* in geographic areas of sympatry (e.g. Nadeau *et al.,* 2013, or Martin *et al.,* 2014, for *Heliconius* butterflies). In contrast, less gene flow in sympatry than allopatry would indicate selection against gene flow and thus suggest reinforcement of reproductive barriers at sites of sympatric co-occurrence (Butlin, 1995; Noor, 1999; Coyne & Orr, 2004; e.g. Kulathinal & Singh, 2000; Massie & Makow, 2005; Giesen *et al*. 2017, 2019).

110 **MATERIALS & METHODS**

111 **Sample collection and treatment of fly cultures**

112 To study inter- and intraspecific gene flow in S. *cynipsea* and *S. neocynipsea*, we
113 compared genomic data from two different sample sources. As a first source, we used
114 laboratory populations established from single gravid females (i.e., iso-female lines) of
115 sympatric *S. cynipsea* and *S. neocynipsea* collected in summer 2013 at a site in the Swiss Alps
116 (Sörenberg) and at a site in the Swiss lowlands (Zürich; Table 1). These populations were
117 subsequently propagated in the laboratory for at least two years (>30 generations) at census
118 population sizes ranging from 10 to 100 before whole-genome resequencing. Propagation in
119 the laboratory resulted only in a moderate loss of genetic variation with respect to natural
120 populations (Supporting Information). Conditions for laboratory culturing were previously
121 described by Giesen *et al*. (2017). To obtain sufficient amounts of DNA for whole-genome
122 resequencing, we pooled 50 males per laboratory population prior to DNA extraction. As a
123 second source of samples, we additionally pool-sequenced 20 field-caught males from the
124 sampling sites in Sörenberg and Zürich as well as some additional sites in Europe (Table 1).
125 *S. cynipsea* samples from Petroia (Italy) and Pehka (Estonia) represent populations in
126 allopatry to *S. neocynipsea*, whereas samples of the two species from Zürich, the Swiss Alps
127 (Sörenberg, Hospental, Geschinen), and the Cevennes (Le Mourier) in France (Table 1)
128 represent population where the two species occur in sympatry.

129 **DNA library preparation and next generation sequencing**

130 Genomic DNA was extracted from all samples using the UltraPure
131 Phenol:Chloroform:Isoamyl alcohol (25:24:1, v/v) extraction kit (Thermo Fischer Scientific)
132 according to the manufacturer's protocol. Quantification of genomic DNA was performed
133 with a Qubit Fluorometer (Thermo Fischer Scientific; Table 1). Library preparation was
134 carried out using the TruSeq DNA PCR-Free Library Preparation (Illumina) kit according to
135 the manufacturer's protocol. Fragment-size distribution of all libraries was validated on a
136 TapeStation 2200 (Agilent Technologies, Waldbronn, Germany). Sequencing on Illumina
137 HiSeq 2500 version 4 was conducted after labeling and pooling the barcoded DNA
138 representing the four laboratory populations onto one lane to achieve a genome-wide
139 coverage of about 60x. All sequencing data are available at the short-read archive (SRA;
140 https://www.ncbi.nlm.nih.gov/sra) under the accession number PRJNA612154.

141 **Table 1:** Sampling sites and percentages of DNA reads mapping to the reference genome for
142 pool-sequenced laboratory populations (a) and field-collected flies (b). Sites where *S.*
143 *cynipeas* and *S. neocynipsea* co-occur in sympatry are printed in regular font. Sites where *S.*
144 *neocynipsea* does not occur to our knowledge are printed in italics.

Species	Sampling site	Country	Coordinates	Altitude (m)	% reads aligned to *S. thoracica*
a) Pooled laboratory males					
S. cynipsea	Sörenberg	Switzerland	46°24'8.2''N, 8°21'28.4''E	1200	48.2
	Zürich	Switzerland	47°24'0.6''N, 8°34'24.0''E	450	50.1
S. neocynipsea	Sörenberg	Switzerland	46°24'8.2''N, 8°21'28.4''E	1200	48.5
	Zürich	Switzerland	47°24'0.6''N, 8°34'24.0''E	450	50.0
b) Pooled field-collected males					
S. cynipsea	*Pehka*	*Estonia*	59°28'45.1''N, 25°44'52.3''E	5	50.6
	Petroia	*Italy*	43°15'06.4''N, 12°32'41.2''E	570	48.1
	Sörenberg	Switzerland	46°24'8.2''N, 8°21'28.4''E	1200	51.3
	Zürich	Switzerland	47°24'0.6''N, 8°34'24.0''E	450	51.3
	Cevennes	France	44° 3'30.5''N, 3°25'38.8''E	790	49.5
S. neocynipsea	Geschinen	Switzerland	46°49'23.7''N, 8°1'54.6''E	1350	50.5
	Hospental	Switzerland	46° 37'12''N, 8°34'12''E	1500	52.3
	Sörenberg	Switzerland	46°24'8.2''N, 8°21'28.4''E	1200	50.7
	Cevennes	France	44°3'30.5''N, 3°25'38.8''E	790	48.6
c) Outgroup population					
S. orthocnemis	Lenzerheide	Switzerland	46°43'45''N, 9°33'25''E	1470	46.9

145

146 **Raw read processing**

147 Qualitative validation of sequence data before and after trimming was done with FastQC
148 High Troughput Sequence QC Report version 0.11.4 (Andrews *et al*., 2011). After removal of
149 Illumina-specific adapters and trimming with *Trimmomatic* (version 0.36; Bolger, Lohse &

150 Usadel, 2014), the *S. cynipsea* and *S. neocynipsea* reads were mapped to the draft genome of
151 *S. thoracica* with *bwa mem* (version 0.7.12; Li & Durbin, 2009) using default parameters. The
152 assembly of this reference genome is part of a larger and ongoing comparative genomic study
153 of several *Sepsis* species (unpublished data*). S. thoracica* specimen used for genome
154 sequencing were originally collected from Capriasca, Ticino (CH). The draft genome, v0.1
155 was based on Oxford Nanopore long reads (app. 25x read depth), assembled with Canu
156 (Korner et al. 2017) and polished with Illumina short reads (app. 20x read depth). The
157 assembly used in this study is available under the rules of the Fort Lauderdale agreement from
158 www.cgae.de/seto_01_genome.fasta and www.cgae.de/seto_01_genome_masked.fasta.

159 PCR duplicates were removed and reads realigned around indels in each raw alignment
160 file; only reads with a PHRED-scaled mapping quality MQ of 20 and more were retained. We
161 applied the same procedure to map reads of a pool of ten inbred *S. orthocnemis* males to the
162 reference genome. We compiled the aligned reads from all population samples into a single
163 *mpileup* file using *samtools* (v. 1.3.1; Li *et al*., 2009).

164 Using the variant-caller Pool-SNP (Kapun *et al*., 2018), we identified high-confidence
165 SNPs with the following combination of SNP-calling parameters: coverage of each sample ≥
166 10x; coverage of each sample ≤ 95 % percentile of coverage distribution across contigs and
167 samples; minimum allele count of a minor allele at a SNP across all samples pooled > 20x;
168 minor allele frequency at a SNP across all samples pooled > 0.01. We only retained SNPs
169 when all samples fulfilled the above coverage threshold criteria. In addition, we removed
170 SNPs located within a 5-bp distance to indel polymorphisms that occurred in more than 20
171 copies at a given site across all pooled samples, and we ignored SNPs located within
172 repetitive regions. The resulting VCF file was converted to the SYNC file format (Kofler *et*
173 *al*., 2011), and a custom Python script was used to calculate sample-specific allele frequencies
174 for major alleles at each SNP (*sync2AF.py*; https://github.com/capoony/DrosEU_pipeline).

175

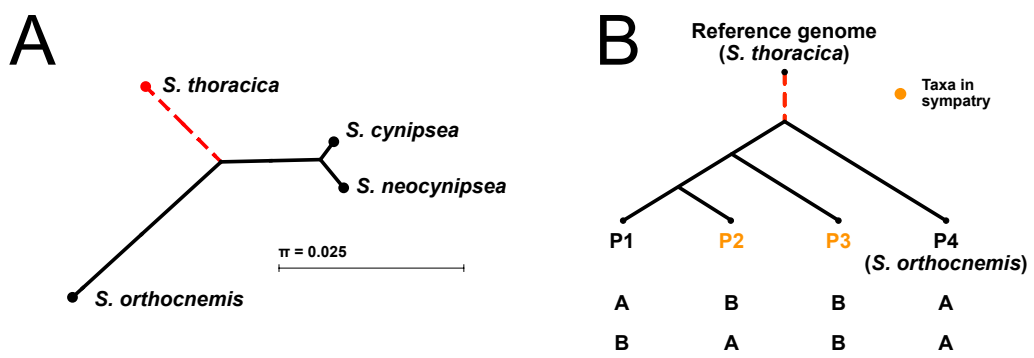
176 **Testing for introgression**

177 To test for introgression, we used the so-called ABBA-BABA test for historical gene flow
178 in the presence of incomplete lineage sorting (Green *et al*., 2010; Durand *et al*., 2011). We
179 provide a brief motivation of the approach in the following and refer to the Supporting
180 Information for details. Imagine a rooted phylogeny with four species (P1 to P4) and a
181 topology of (((P1, P2), P3), P4) as illustrated in Fig. 1B. In an alignment of one haploid
182 genome from each species, incomplete lineage sorting under neutral evolution leads to two
183 mutational configurations of the ancestral (A) and derived (B) allele, ABBA and BABA, that
184 under the infinite-sites model of mutation (Kimura, 1969) are incompatible with the species
185 topology (Fig. 1B). In the absence of gene flow between P2 and P3, ABBA and BABA
186 configurations occur with equal probability (Hudson, 1983; Tajima, 1983). In contrast,
187 historical gene flow between P2 and P3 causes an excess of ABBA configurations (Green *et*
188 *al*., 2010; Durand *et al*., 2011). This logic is captured by the *D*-statistic, a scaled difference
189 across a set of bases between counts of ABBA and BABA configurations bounded by –1 and
190 1 (Green *et al*., 2010). The ABBA-BABA test examines significant deviations of *D* from 0. A
191 significant excess of ABBA suggests evidence for gene flow between P2 and P3. A
192 significant depletion of ABBA is equivalent to an excess of ABBA if the positions of P1 and
193 P2 in the species topology are swapped, and hence either suggests gene flow between P1 and
194 P3, or a reduction in gene flow between P2 and P3 relative to gene flow between P1 and P3.

195 The original version of the ABBA-BABA test, and modifications of it, have been widely
196 applied to different types of genome-scale polymorphism data obtained with various
197 sequencing strategies, including whole genome sequencing (e.g. Green *et al*., 2010), RAD
198 sequencing (e.g., Eaton & Ree, 2013; Meier *et al*., 2017; Streicher *et al*., 2014), and exon
199 capture data (e.g., Heliconius Genome Consortium, 2012). More recently, Durand *et al*.
200 (2011) and Soraggi *et al*. (2018) extended the scope of the original test to allele frequency
201 data, and hence to unphased sequencing data (including pool-sequencing data; Schlötterer *et*
202 *al*. 2014). We implemented the extensions by Durand *et al*. (2011) and Soraggi *et al*. (2018)
203 in a Python script (https://github.com/capoony/ABBABABA-4AF), and refer to the respective
204 test statistics as D_D and D_S. Our script also computes jack-knifing *z*-scores based on a matrix
205 of allele frequencies for previously defined high-confidence SNPs (see Supporting
206 Information for details). We adopted the commonly used significance threshold of $|z| > 3$

207 (Reich *et al*., 2011; Jeong *et al*., 2016; Novikova *et al*., 2016) to identify significant
208 deviations of D_S and D_D from zero.

209 We were concerned that using the genome of a given species both as the mapping
210 reference as well as the outgroup would bias ABBA-BABA tests. We therefore mapped the
211 three ingroup and the outgroup species against the reference genome of yet another species, *S*.
212 *thoracica*. A phylogenetic analysis based on CO-II sequences by Su *et al*. (2008) indicated
213 that *S*. *thoracica* is approximately equally related to *S*. *cynipsea* and *S*. *neocynipsea* as it is to
214 the outgroup species *S*. *orthocnemis* (Fig. 1A). We therefore expected that reads of both the
215 ingroup and the outgroup species would map equally well to the reference species, which is
216 indeed what we observed (Table 1).



217

218 **Fig. 1. Putative phylogeny of sepsid species studied**. (A) Species phylogeny inferred from
219 alignments of nucleotide sequences at the Cytochrome Oxidase – Subunit II (CO-II)
220 mitochondrial locus (Genbank sequences from Su *et al*., 2016) with the Neighbor-Joining
221 method implemented in CLC Main Workbench (v. 8.1.2; https://
222 https://www.qiagenbioinformatics.com/products/clc-main-workbench/). The tree topology is
223 consistent with the phylogeny presented by Su *et al*. (2016) based on a combined
224 phylogenetic analysis of multiple nuclear and mitochondrial markers. *Sepsis orthocnemis* was
225 used as the outgroup to the focal species *S*. *cynipsea* and *S*. *neocynipsea*. Whole-genome
226 sequencing reads from all three species were aligned to the *S*. *thoracica* reference genome
227 (dashed red branches). Branch lengths are proportional to the mean number of pairwise
228 sequence differences. (B) Generic species tree assumed in all our ABBA-BABA tests for gene
229 flow among triplets of *S*. *cynipsea* and *S*. *neocynipsea* populations from various sites in
230 Europe.

231

232 We focused our analyses on the three sampling sites: Sörenberg, Zürich and Cevennes. At
233 all three sites *S. cynipsea* and *S. neocynipsea* occur in sympatry. In all ABBA-BABA tests for
234 interspecific gene flow among sympatric populations of *S. cynipsea* and *S. neocynipsea*, we
235 positioned the two focal populations as P2 and P3 ingroups in the phylogeny (in both possible
236 orders) and used various P1 ingroup populations (Fig. 1B) assumed to occur in allopatry with
237 P2 populations. We performed these permutations because Durand *et al*. (2011) showed that
238 the choice of the ingroup P1 can influence test results. All reported analyses are based on non-
239 overlapping windows of 500 consecutive SNPs, unless stated otherwise.

240 **RESULTS**

241 **Genetic diversity in laboratory and natural populations**

242 We first analyzed whole-genome resequencing data from sympatric *S. cynipsea* and *S.*
243 *neocynipsea* populations in Sörenberg and Zürich, Switzerland, two sites roughly 70 km
244 linear distance apart without obvious barriers to gene flow. At both of these sites, DNA was
245 obtained from pooled samples of males from outbred natural populations (Sörenberg *S.*
246 *cynipsea*, SoC; Sörenberg *S. neocynipsea*, SoN; Zürich *S. cynipsea*, ZuC) as well as from
247 pools of males derived from laboratory populations (Sörenberg *S. cynipsea*, ISoC; Sörenberg
248 *S. neocynipsea*, ISoN; Zürich *S. cynipsea*, IZuC; Zürich *S. neocynipsea,* IZuN). We expected
249 the genetic diversity in the laboratory populations to be strongly reduced relative to the one in
250 the respective field-caught flies. While we did observe significantly lower levels of expected
251 heterozygosity (Nei 1987) in laboratory population samples than in respective field-caught
252 flies (Mann Whitney U test, $p < 0.001$ for all three sampling sites; Fig. S1), reductions in
253 heterozygosity were much lower than expected under strong inbreeding. Specifically,
254 population genetic analyses showed that the effective population size in our laboratory
255 populations must have been on the order of 100 flies, i.e. at the upper bound of estimated
256 census sizes, to be compatible with observed reductions in heterozygosity (see "Reduction of
257 genetic diversity in laboratory populations" in the Supporting Information; Fig. S2). Our
258 laboratory populations were therefore not highly inbred and we decided to treat the
259 sequencing data from laboratory and natural populations equally in all our analyses.

260

261 **Signals of reinforcement at the Sörenberg site**

262 The majority of our tests for gene flow at the Sörenberg site revealed a genome-wide
263 deficiency of derived-allele sharing (i.e. ABBA patterns) among the local *S. cynipsea* and *S.*
264 *neocynipsea* populations (Table 2), suggesting that interspecific gene flow in sympatry
265 (among P2 and P3) is lower than interspecific gene flow among putatively allopatric
266 populations (P1 and P3). This observation was robust to our choice of the P1 ingroup
267 population, but sensitive to the version of *D*-statistic used (D_S vs. D_D), and, in the case of D_D,
268 to whether we used field or laboratory samples as P2. We first conducted tests involving two
269 *S. neocynipsea* populations from the Swiss Alps (Geschinen *S. neocynipsea*, GeN; Hospental
270 *S. neocynipsea*, HoN; Table 1) as putatively allopatric P1, and all four possible sympatric
271 interspecific pairs of field and laboratory samples from Sörenberg as P2 (*S. neocynipsea*) and
272 P3 (*S. cynipsea*). We observed a deficiency of ABBA patterns in all eight configurations (D_S
273 > 0; positive *z*-scores ranging from 0.8 to 21.2). Deviations of D_S from 0 were significant in
274 six of these eight configurations (Table 2, upper part), suggesting overall evidence for higher
275 gene flow among putatively allopatric than sympatric interspecific population pairs. While
276 GeN and HoN are both separated by high mountains (altitudes ranging from ca. 600 m asl to
277 ca. 3,200 m asl) from the Sörenberg site, they are nevertheless geographically close (ca. 67.5
278 km and ca. 46.7 km Euclidean distance for GeN and HoN, respectively), and they occur in
279 sympatry with other sepsid species, including *S. cynipsea*. We were therefore concerned that
280 our assumption of no gene flow between P1 and P3 was violated because GeN and HoN are
281 parapatric, rather than allopatric, to our focal populations in Sörenberg.

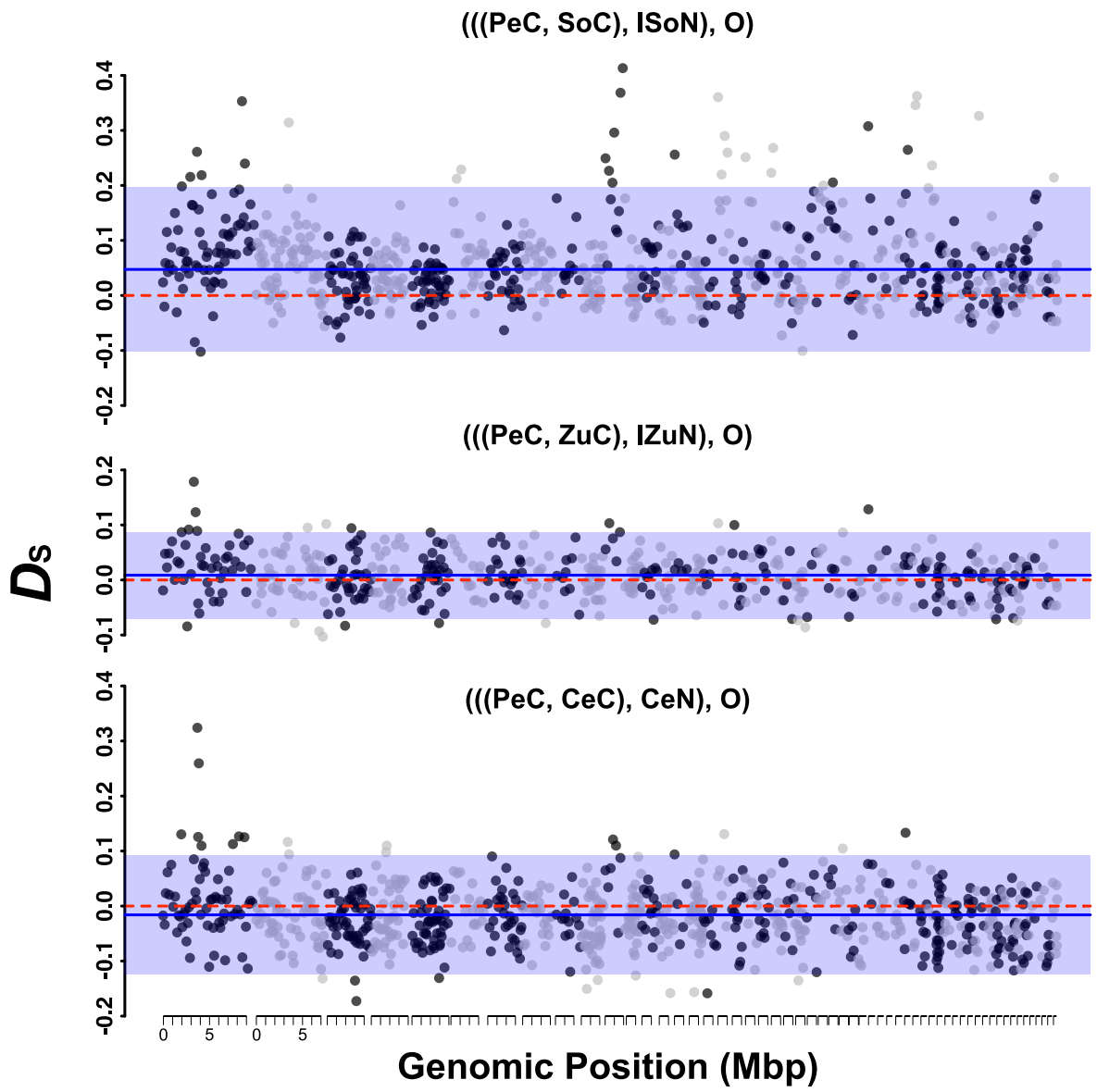
282 To address the above concern, we conducted additional ABBA-BABA tests involving P1
283 populations from geographically more remote sites in Europe. Specifically, we performed a
284 second series of tests with two alternative *S. cynipsea* populations as allopatric P1 ingroups
285 (Fig. 1B), one from Pehka, Estonia (EsC), and one from Petroia, Italy (PeC; Table 1). As
286 before, we used all possible sympatric interspecific pairs of field and laboratory samples from
287 Sörenberg as P2 (*S. cynipsea*) and P3 (*S. neocynipsea*). Contrary to our expectation based on
288 the much larger geographical distance between P1 and P3 (ca. 1,850 km and ca. 550 km for
289 EsC and PeC, respectively) than between P2 and P3 in these configurations, we still found a
290 significant deficiency of ABBA patterns in all eight tests ($D_S > 0$; *z*-scores from 4.9 to 23.4;
291 Table 2 lower part). In agreement with our previous tests involving GeN and HoN as P1, we
292 thus found consistent evidence for higher interspecific gene flow among allopatric (P1 and

293 P3) than among sympatric populations (P2 and P3). As we are much more confident that EsC
294 and PeC represent truly allopatric P1 ingroups, deficient ABBA patterns at the sympatric
295 Sörenberg site suggest a local reduction in interspecific gene flow, likely due to
296 reinforcement.

297 When repeating the ABBA-BABA tests with the *D*-statistic (D_D) suggested by Durand *et*
298 *al*. (2011), we obtained results that qualitatively agree with those obtained with the D_S statistic
299 only when using EsC and PeC as P1 ($D_D < 0$; *z*-scores from –2.3 to –9.7; significant for six of
300 eight population configurations in Table S1, lower part). With GeN and HoN as P1, in
301 contrast, we observed a deficiency of ABBA patterns ($D_D < 0$) only with the laboratory *S*.
302 *neocynipsea* samples (ISoN) as P2 (*z*-scores from –4.0 to –1.3), but not with the field-derived
303 *S. neocynipsea* samples (SoN) as P2 (*z*-scores from 1.4 to 2.6: Table S1, upper part).
304 Moreover, ABBA deficiencies with ISoN as P2 were significant only if we used field-derived
305 *S cynipsea* samples (SoC), but not if we used laboratory *S cynipsea* samples (ISoC) as P3.

306 Besides the above qualitative differences between tests based on D_S vs. D_D, we observed
307 that variances of D_D estimated by jackknifing were about one order of magnitude larger than
308 the corresponding values of D_S. The qualitatively similar results obtained with the two
309 versions of *D*-statistics makes us confident that our findings are robust. We further observed
310 that windows of elevated D_S (Fig. 2, top) were not homogeneously distributed along the
311 genome, but aggregated in clusters in several contigs. Such clustering might suggest that these
312 regions contain genomic variation involved in reinforcement caused by divergent selection
313 against gene flow between the two species.

314



315

316 **Fig. 2. Variation of D_S along the genome.** Points indicate D_S (Soraggi *et al*. 2018) for
317 genomic windows of 500 consecutive SNPs, oriented by contigs of the *S. thoracica* reference
318 genome with a length ≥ 50,000 bp (highlighted by alternating black and grey colors). Rows
319 correspond to population topologies used to test for interspecific gene flow at the three focal
320 sites in Sörenberg (SoC-ISoN; top), Zürich (ZuC-IZuN; center) and the French Cevennes
321 (CeC-CeN; bottom). For illustrative purposes, we depict mixed combinations of natural and
322 laboratory (with prefix "I") samples that yielded the clearest results when using the *S.*
323 *cynipsea* population from Petroia as P1. The horizontal dashed red and blue lines indicate the
324 neutral expectation ($D_S = 0$) and the mean genome-wide jackknife estimate, respectively. Blue
325 shading delimits mean $D_S \pm 2$ standard deviations (based on windows of 500 SNPs).

Table 2: Results of ABBA-BABA tests applied to samples originating from sympatric populations of *S. cynipsea* and *S. neocynipsea* in Sörenberg as P2 and P3 (and *vice versa*), based on D_S. The upper half of the table shows tests with two European *S. neocynipsea* populations from Geschinen (GeN) and Hospental (HoN) as P1 ingroups, and the bottom half shows the reciprocal approach with two *S. cynipsea* populations from Pehka, Estonia (EsC) and Petroia, Italy (PeC) as P1 ingroups. The first three columns indicate the phylogeny assumed in the test, with P4 always being *S. orthocnemis*; columns 4 to 6 give the genome-wide average of *D*s, its jackknifed standard deviation SD(D_S), and the corresponding *z*-score. Significant tests are shown in bold. The prefix "I" indicates pooled samples from laboratory populations.

P1	P2	P3	D_S	SD(D_S)	*z*
GeN	**ISoN**	**ISoC**	**0.018**	**2.40E−03**	**7.4**
GeN	**ISoN**	**SoC**	**0.053**	**2.51E−03**	**21.2**
GeN	SoN	ISoC	0.001	1.27E–03	0.8
GeN	**SoN**	**SoC**	**0.008**	**1.35E−03**	**5.8**
HoN	**ISoN**	**ISoC**	**0.019**	**2.34E−03**	**8.2**
HoN	**ISoN**	**SoC**	**0.051**	**2.48E−03**	**20.5**
HoN	SoN	ISoC	0.002	1.24E–03	1.9
HoN	**SoN**	**SoC**	**0.006**	**1.34E−03**	**4.2**
EsC	**ISoC**	**ISoN**	**0.013**	**2.68E−03**	**4.9**
EsC	**ISoC**	**SoN**	**0.040**	**2.62E−03**	**15.2**
EsC	**SoC**	**ISoN**	**0.039**	**1.88E−03**	**20.5**
EsC	**SoC**	**SoN**	**0.026**	**1.92E−03**	**13.4**
PeC	**ISoC**	**ISoN**	**0.023**	**2.60E−03**	**8.7**
PeC	**ISoC**	**SoN**	**0.052**	**2.77E−03**	**18.8**
PeC	**SoC**	**ISoN**	**0.047**	**2.02E−03**	**23.4**
PeC	**SoC**	**SoN**	**0.039**	**1.97E−03**	**19.8**

337 **Convergent patterns of reinforcement in Zürich**

338 Complementary to the Sörenberg site, where *S. cynipsea* and *S. neocynipsea* co–occur in
339 an alpine environment, we analyzed gene flow between sympatric *S. cynipsea* and *S.*
340 *neocynipsea* populations from a lowland site in Zürich. These analyses were based on pooled
341 laboratory samples of both species (as defined above) and one pool-sequenced field sample of
342 *S. cynipsea* (Table 1). Using the *S. neocynipsea* populations of Geschinen and Hospental as
343 P1 ingroups, we obtained weakly but significantly positive D_S values ranging from 0.016 to
344 0.055 (Table 3), indicating a deficiency of ABBA patterns. These findings are qualitatively
345 consistent with the results obtained when analyzing samples from Sörenberg in suggesting
346 reinforcement in sympatry. However, only when using the allopatric Italian Petroia (but not
347 the Estonian) *S. cynipsea* population as P1 ingroup did we also find a significant, albeit even
348 smaller deficiency of ABBA patterns (Table 3).

349 Analyzing the same data as presented in the preceding paragraph with Durand *et al.'s*
350 (2011) D_D statistic, we found only one significant result indicating a deficiency of ABBA
351 patterns ($D_D = 0.03$) with the *S. neocynipsea* population from Geschinen as P1 ingroup (Table
352 S2). However, we found a significant, yet small excess of ABBA patterns ($D_D = 0.019$) with
353 *S. cynipsea* from Pehka, Estonia, as P1. Therefore, the putative signal of reinforcement among
354 the sympatric Zürich populations is much weaker than among the sympatric populations in
355 Sörenberg. Consistent with these observations, we did not find pronounced genomic clusters
356 of elevated D_S (Fig. 2, center) as for the Sörenberg populations, which corroborates that
357 reinforcement in Zürich was weaker.

358

359 **Table 3:** Results of ABBA–BABA tests applied to samples originating from sympatric
360 populations of *S. cynipsea* and *S. neocynipsea* in Zürich as P2 and P3 (and *vice versa*) based
361 on D_S. As in Table 2, the upper half of the table summarizes tests involving two *S.*
362 *neocynipsea* populations from Geschinen (GeN) and Hospental (HoN) as P1 ingroups, and the
363 bottom half summarizes tests involving *S. cynipsea* populations from Pehka, Estonia (EsC)
364 and Petroia, Italy (PeC) as P1. Other details as in Table 2.

P1	P2	P3	D_S	SD(D_S)	z
GeN	**IZuN**	**IZuC**	**0.016**	**2.63E–03**	**6.1**
GeN	**IZuN**	**ZuC**	**0.055**	**2.79E–03**	**19.6**
HoN	**IZuN**	**IZuC**	**0.017**	**2.68E–03**	**6.5**
HoN	**IZuN**	**ZuC**	**0.055**	**2.69E–03**	**20.4**
EsC	IZuC	IZuN	0.001	3.89E–03	0.3
EsC	ZuC	IZuN	–0.002	1.54E–03	–1.4
PeC	**IZuC**	**IZuN**	**0.012**	**3.74E–03**	**3.3**
PeC	**ZuC**	**IZuN**	**0.009**	**1.66E–03**	**5.3**

365

366 **Signals of introgression in the French Cevennes populations**

367 We investigated a population pair of *S. cynipsea* and *S. neocynipsea* from the Cevennes in
368 France (790 m asl), approximately 450 km southwest of the two sampling sites in
369 Switzerland. Since no laboratory population samples were available from this site, we used
370 pool-sequenced field samples for our analyses. Interestingly, we found opposite patterns to
371 those obtained when analyzing samples from the Zürich and Sörenberg sites in Switzerland
372 (Table 4). Specifically, all tests based on D_S showed highly significant *negative* values of the
373 test statistic, suggesting an excess of ABBA patterns, and hence *more* interspecific gene flow
374 among the sympatric population pairs than among the allopatric population pairs. However,
375 none of the corresponding tests based on D_D were significant (Table S3). We nevertheless
376 consider our results robust, as we found in all four tests a corresponding trend towards
377 negative D_D, i.e. an excess of ABBA patterns consistent with results based on D_S. As for the
378 Zürich site, we did not observe genomic clustering of extreme values of D_S (Fig. 2, bottom).

379

380 **Table 4:** Results of ABBA-BABA tests applied to samples originating from sympatric
381 populations of *S. cynipsea* and *S. neocynipsea* in the French Cevennes as P2 and P3 (and *vice*
382 *versa*) based on D_S. As in Tables 2 and 3, the upper half of the table summarizes tests
383 involving two *S. neocynipsea* populations from Geschinen (GeN) and Hospental (HoN) as P1
384 ingroups, and the bottom half of the table summarizes tests involving *S. cynipsea* populations
385 from Pehka, Estonia (EsC) and Petroia, Italy (PeC) as P1. Other details as in Table 2,

P1	P2	P3	D_S	SD(D_S)	z
GeN	**CeN**	**CeC**	**–0.014**	**1.30E–03**	**–11.0**
HoN	**CeN**	**CeC**	**–0.014**	**1.36E–03**	**–10.6**
EsC	**CeC**	**CeN**	**–0.030**	**1.84E–03**	**–16.2**
PeC	**CeC**	**CeN**	**–0.016**	**1.93E–03**	**–8.3**

386

387 **DISCUSSION**

388 We applied the ABBA-BABA approach (Green *et al*., 2010; Durand *et al*., 2011; Soraggi
389 *et al*., 2018) to investigate natural patterns of admixture between a pair of closely related
390 sister species of sepsid flies for which we have detailed information on ecology, morphology,
391 life history, and behavior (Blanckenhorn, 1999; Blanckenhorn *et al*., 2000; Eberhard, 1999;
392 Ozerov, 2005; Parker, 1972a,b; Pont & Meier, 2002; Puniamoorthy *et al*., 2009; Rohner,
393 Blanckenhorn & Puniamoorthy, 2016; Ward, 1983; Ward, Hemmi & Rösli, 1992; Giesen *et*
394 *al*., 2017, 2019). Previous population genetic analyses of microsatellite markers (Baur *et al*.,
395 2020) and laboratory hybridization experiments (Giesen *et al*., 2017, 2019) suggested that *S.*
396 *cynipsea* and *S. neocynipsea* are genetically distinct but may occasionally hybridize in nature
397 when occurring in sympatry. Here we investigated this possibility by analyzing patterns of
398 DNA sequence variation among pools of laboratory as well as field-caught specimens from
399 various sampling sites across Europe where the two species occur in sympatry vs. allopatry.
400 Our study revealed qualitatively opposite, i.e. mixed patterns of interspecific gene flow in
401 sympatry depending on the sampling sites, which are discussed in detail below. Previous
402 studies of other taxa, using different genomic methods but with essentially the same logic
403 behind their approach, provided the full spectrum of potential results that we here obtained for
404 one species pair. Martin *et al*. (2014; see also Nadeau *et al*., 2013) observed *more* inter-
405 specific gene flow in sympatry than allopatry in *Heliconius* butterflies; in contrast, less gene

flow in sympatry than allopatry, suggesting reinforcement by natural or sexual selection (Butlin, 1995; Noor, 1999; Coyne & Orr, 2004), was observed for *Drosophila arizonae/mojavensis* (Massie & Makow, 2005), the monkey flower pair *Mimulus nasutus/guttatus* (Brandvain *et al.,* 2014; see also Grossenbacher & Whittall, 2011), and two sea squirt species (Tunicata: Bouchemousse *et al.,* 2016), while no geographic variation in admixture between sym- vs. allopatric populations was found for wild tomatoes (Nakazato *et al.,* 2010).

Similar to *Heliconius* (Martin *et al.,* 2014), we found evidence for an excess of interspecific gene flow in sympatry compared to allopatry in flies from the French Cevennes, based on pool-sequenced field-caught specimens (Table 4; Table S3). In contrast, at two sites in Switzerland where the two species also occur in sympatry, we observed reduced interspecific gene flow relative to interspecific pairs of parapatric and allopatric populations, similar to patterns in monkey flower, *Drosophila* and sea squirt as explained above. The reduction of interspecific gene flow was more pronounced at the Sörenberg site in the Alps (Table 2; Table S1) than at the lowland site in Zürich (Tables 3; Table S2). Consistent with a previous study showing behavioral reinforcement after enforced hybridization in the laboratory (Giesen *et al.*, 2017), the genomic data we analyzed here suggest natural selection against interspecific gene flow might also occur in nature at the two Swiss sites where both species occur in sympatry (Noor, 1999). However, as evidenced by our contrasting findings for the site in the French Cevennes, reinforcement may not be a necessary outcome where *S. cynipsea* and *S. neocynipsea* occur in sympatry. The extent of gene flow between these two dung fly sister species thus appears to covary with local environmental conditions. Well-documented differences in climate between southern France and the northern edge of the Alps might provide an ecological explanation for the contrasting patterns we observed (cf. Doebeli & Dieckmann, 2003; Rohner *et al.,* 2015), but future studies are needed to confirm this hypothesis and explore alternatives.

Laboratory hybridization experiments have shown that interspecific mating between *S. cynipsea* and *S. neocynipsea* may result in viable and fertile F_1 hybrid females and offspring in backcrosses with the parental species (Giesen *et al.*, 2017, 2019). Together with morphological and behavioral similarities among the two species, these laboratory experiments suggest that hybridization may well take place in areas of co-occurrence, such as in the French Cevennes. In other areas, including our sampling site in the Swiss Alps, the

438 subtle premating behavioral barriers and the reductions in fecundity and fertility documented
439 in the laboratory by Giesen *et al*. (2017, 2019) may combine with micro-ecological niche
440 differences that mediate spatio-temporal divergence in reproductive timing to effectively
441 prevent hybridization in nature and ultimately reinforce species boundaries. This
442 interpretation is strengthened by a recent study on Central American *Archisepsis diversiformis*
443 flies showing that mating between two disjunct populations is only evident under forced
444 laboratory conditions, while under conditions of free mate choice flies from the different
445 populations do not interbreed (Puniamoorthy 2014). Behavioral mating barriers therefore
446 seem to evolve comparatively fast (Gleason & Ritchie, 1998; Puniamoorthy *et al*., 2009;
447 Puniamoorthy, 2014), especially if species occur in sympatry and reinforcement by natural or
448 sexual selection can operate (Coyne & Orr, 2004; Seehausen, 2004; Ritchie, 2007).

449 We found a weaker signal of reinforcement between *S. cynipsea* and *S. neocynipsea* at the
450 lowland Zürich site than at the high-altitude Sörenberg site in the Swiss Alps. This difference
451 might be related to the fact that *S. neocynipsea* is rare around Zürich, in fact in European
452 lowlands in general, while it is more common at higher altitudes. If based primarily on
453 behavioral mate choice, reinforcement is expected to be stronger, and hence its evolution
454 more likely, where interspecific encounter frequencies are higher. Regardless, why *S.*
455 *neocynipsea* remains rare and restricted to highland sites in Europe, whereas it is common in
456 dry and warm lowland areas in North America, cannot be answered by our study. *Sepsis*
457 *cynipsea* is abundant around fresh dung all over Europe north of the Alps (Pont & Meier,
458 2002). It therefore might outcompete and ultimately relegate *S. neocynipsea* towards higher
459 altitudes. However, *S. cynipsea* also readily populates cold high-altitude sites (Rohner *et al*.,
460 2015), and some degree of altitudinal molecular (allozyme) differentiation can also be
461 observed (Kraushaar *et al*., 2002). It consequently remains unclear why *S. neocynipsea* can
462 better persist at high altitudes in Europe.

463 Our study on introgression among sympatric *S. cynipsea* and *S. neocynipsea* populations
464 relied on the ABBA-BABA approach, which was initially designed for the analysis of haploid
465 genotypes of one individual from each of four lineages (Green *et al*., 2010). Recent
466 extensions of this original approach to allele frequency data (Durand *et al*., 2011; Soraggi *et*
467 *al*., 2018) enabled us to apply the method to pool-sequenced laboratory and field-caught
468 specimens (see also Deitz *et al*. 2016). We have here provided a Python implementation of
469 the ABBA-BABA approaches by Durand *et al*. (2011) and Soraggi *et al*. (2018) that takes a

470 simple 2D matrix of allele frequencies as input. This implementation allows restricting the
471 ABBA-BABA analysis to high-confidence SNPs that pass a set of user-defined filters in order
472 to reduce artefacts of pool-sequencing (Schlötterer *et al*. 2014) that may produce false
473 polymorphisms due to sequencing errors (Futschik & Schlötterer 2011; Cutler & Jensen
474 2012). In addition, and in contrast to the software package ANGSD (Korneliussen *et al*.
475 2014), our implementation accommodates an outgroup (P4) taxon different from the one used
476 as a reference to call the SNPs. This distinction reduces a potential reference bias that could
477 affect sequence variation if the ingroup taxa (P1, P2, P3) strongly differed from an outgroup
478 that also served as reference (Ballouz *et al*. 2019). Our script and detailed instructions for how
479 to use it are available on GitHub (https://github.com/capoony/ABBA-BABA-4-AF; see also
480 Supporting Information).

481 In summary, our genomic study provided evidence that the extent of interspecific gene
482 flow between two closely related species of sepsid flies in sympatry relative to allopatry
483 varies geographically. Interspecific gene flow appears to be very limited at sites in the Swiss
484 Alps where the two species have perhaps locally adapted to different (micro-)ecological
485 niches, such as different breeding times or phenologies, and/or where pre- and postmating
486 reproductive barriers (Giesen *et al*., 2017) have been and are subject to stronger
487 reinforcement. It however remains unclear, if ecological and sexual selection are mediating
488 these effects. We note in closing that North American and European *S. neocynipsea*
489 populations also differ substantially based on neutral genetic markers (microsatellites), as can
490 be expected by genetic drift alone due to definitive inter-continental allopatry (Baur *et al*.,
491 2020). It is currently also unclear whether European *S. neocynipsea* are remnants of an
492 ancient Holarctic population or whether they (re-)invaded Europe from North America, where
493 the species is much more common and fills the niche that *S. cynipsea* has in Europe. Future
494 genomic research should analyze patterns of genomic differentiation across allopatric
495 populations from Europe and North America (Baur *et al*., 2020). Such analyses would
496 provide further insights into introgression patterns across the species' entire natural range, as
497 well as into the role of local (e.g. latitudinal) adaptation in shaping genome-wide patterns of
498 differentiation.

499

500 **LITERATURE CITED**

501 Anderson M.B. (1949). Sexual selection. *Princeton University Press*.

502 Andrews, S., Lindenbaum, P., Howard, B., & Ewels, P. (2011). FastQC High throughput
503 sequence QC report v. 0.10. *Babraham Bioinformatics, Cambridge, UK*.

504 Arnold M.L. & Meyer A. (2006). Natural hybridization in primates: one evolutionary
505 mechanism. *Zoology*, 109(4): 261–276.

506 Ballouz, S., Dobin, A. & Gillis, J.A. (2019). Is it time to change the reference genome? *Genome*
507 *Biology* 20: 159.

508 Barton N., & Bengtsson B.O. (1986). The barrier to genetic exchange between hybridising
509 populations. *Heredity*, 57(3): 357–376.

510 Baur J., Schäfer M.A., Blanckenhorn W.U., & Giesen A. (2020). Patterns of genetic
511 differentiation in sexual and neutral morphological traits among species and populations of dung
512 flies (Diptera: Sepsidae). *Journal of Zoological Systematics and Evolutionary Research* 58: 159–
513 173.

514 Berner D. & Salzburger W. (2015). The genomics of organismal diversification illuminated by
515 adaptive radiations. *Trends in Genetics*, 31: 491–499.

516 Blanckenhorn W.U. (1999). Different growth responses to temperature and resource limitation
517 in three fly species with similar life histories. *Evolutionary Ecology*, 13: 395–409.

518 Blanckenhorn W.U., Mühlhäuser C., Morf C., *et al*. (2000). Female choice, female reluctance to
519 mate and sexual selection on body size in the dung fly *Sepsis cynipsea*. *Ethology*, 106: 577–593.

520 Bolger A.M., Lohse M., & Usadel B. (2014). Trimmomatic: a flexible trimmer for Illumina
521 sequence data. *Bioinformatics*: btu170.

522 Bouchemousse S., Liautard-Haag C., Bierne N., & Viard F. (2016) Distinguishing
523 contemporary hybridization from past introgression with postgenomic ancestry-informative SNPs
524 in strongly differentiated *Ciona* species. *Molecular Ecology,* 25: 5527-5542.

525 Brandvain Y., Kenney A.M., Flgel L., Coop G., & Sweigart A.L. (2014). Speciation and
526 introgression between *Mimulus nasutus* and *M. guttatus*. *Plos Genetics,* 10:
527 doi:10.1371/journal.pgen.1004410

Butlin R. K. 1995. Reinforcement: an idea evolving. *Trends in Ecology & Evolution*, 10, 432-434.

Coyne J.A., & Orr H.A. (2004). Speciation. *Sinauer Associates*.

Cutler, D.J. & Jensen, J.D. (2010). To Pool, or Not to Pool? *Genetics* 186: 41–43.

Deitz, K.C., Athrey, G.A., Jawara, M., Overgaard, H.J., Matias, A. & Slotman, M.A. (2016). Genome-Wide Divergence in the West-African Malaria Vector Anopheles melas. *G3: Genes, Genomes, Genetics* 6: 2867–2879.

DeMarais B.D., Dowling T.E., Douglas M.E., *et al*. (1992). Origin of Gila seminude (Teleostei: Cyprinidae) through introgressive hybridization: implications for evolution and conservation. *Proceedings of the National Academy of Sciences* 89(7): 2747–2751.

Dobzhansky T. (1951). Genetics and the origin of species (3rd ed.). *Columbia University Press*.

Doebeli M, & Dieckmann U. (2003) Speciation along thermal gradients. Nature 421: 259–264.

Durand E.Y., Patterson N., Reich D., & Slatkin M. (2011). Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, 28(8): 2239–2252.

Eaton D.A. & Ree R.H. (2013). Inferring phylogeny and introgression using RADseq data: an example from flowering plants (Pedicularis: Orobanchaceae). *Systematic Biology*, 62(5): 689–706.

Eberhard W.G. (1999). Mating systems of sepsid flies and sexual behavior away from oviposition sites by *Sepsis neocynipsea* (Diptera: Sepsidae). *Journal of the Kansas Entomological Society,* 72(1): 129–130.

Feulner P.G.D & Seehausen O. (2018). Genomic insights into the vulnerability of sympatric whitefish species flocks. *Molecular Ecology,* 28: 615-629.

Futschik A. & Schlötterer C. (2010). The next generation of molecular markers from passively parallel sequencing of pooled DNA samples. *Genetics*, 186(1): 207–218.

Gante H.F., Matschiner M., Malmstrom M., Jakobsen K.S., Jentoft S., & Salzburger W. (2016). Genomics of speciation and introgression in Princess cichlid fishes from Lake Tanganyika. *Molecular Ecology*, 25: 6143–6161.

Giesen A, Blanckenhorn W.U., & Schäfer M.A. (2017). Behavioural mechanisms of

reproductive isolation between two hybridizing dung fly species. *Animal Behaviour*, 132: 155–166.

Giesen A., Schäfer M.A., & Blanckenhorn W.U. (2019). Geographic patterns of postzygotic isolation between two closely related widespread dung fly species (*Sepsis cynipsea* and *Sepsis neocynipsea*; Diptera: Sepsidae). *Journal of* Zoological Systematics and Evolutionary Research, 57: 80–90

Gleason, J.M. & Ritchie, M.G. 1998. Evolution of courtship song and reproductive isolation in the *Drosophila willistoni* species complex: Do sexual signals diverge the most quickly? *Evolution*, 52: 1493–1500.

Green R.E., Krause J., Briggs A.W., *et al*. (2010). A draft sequence of the Neanderthal genome. *Science*, 328(5979): 710–722.

Grossenbacher D.L. & Whittall J.B. (2011). Increased floral divergence in sympatric monkeyflowers. *Evolution,* 65: 2712-2718.

Haldane J.B.S. (1922). Sex ratio and the unisexual sterility in animal hybrids. *Journal of Genetics* 12: 101–109.

Heliconius Genome Consortium (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, 487(7405): 94–98.

Hudson R.R. (1983). Testing the Constant-Rate Neutral Allele Model with Protein Sequence Data. *Evolution* 37: 203–217.

Jeong C., Nakagome S., Di Rienzo A. (2016). Deep History of East Asian Populations Revealed Through Genetic Analysis of the Ainu. *Genetics*, 202: 261–272.

Kapun, M., Barron Aduriz, M.G., Staubach, F., *et al*. (2018). Genomic analysis of European *Drosophila* populations reveals major longitudinal structure, continent-wide selection, and unknown DNA viruses. *bioRxiv*, DOI: 10.1101/313759.

Kastally Ch., Trasoletti M., & Mardulyn P. (2019). Limited gene exchange between two sister species of leaf beetles within a hybrid zone in the Alps. *Journal of evolutionary Biology,* 32: 1406-1417.

Kimura, M. (1969). The Number of Heterozygous Nucleotide Sites Maintained in a Finite Population Due to Steady Flux of Mutations. *Genetics* 61: 893–903.

583 Kofler, R., Pandey, R.V. & Schlötterer, C. (2011). PoPoolation2: identifying differentiation
584 between populations using sequencing of pooled DNA samples (pool-seq). *Bioinformatics* 27:
585 3435–3436.

586 Koren S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H. & Phillippy, A.M. (2017).
587 Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat
588 separation. *Genome Research* 27: 722–736.

589 Kulathinal R.J. & Singh R.S. (2000). A biogeographic genetic approach for testing the role of
590 reinforcement: the case of *Drosophila pseudoobscura* and *D. persimilis*. *Evolution,* 54: 210-217.

591 Kraushaar U. & Blanckenhorn W.U. (2002). Population variation in sexual selection and its
592 effect on size allometry in two dung fly species with contrasting sexual size dimorphism.
593 *Evolution*, 56(2): 307–321.

594 Kraushaar U., Goudet J., & Blanckenhorn W.U. (2002). Geographical and altitudinal population
595 genetic structure of two dung fly species with contrasting mobility and temperature preference.
596 *Heredity*, 89: 99-106.
597
598 Korneliussen T.S., Albrechtsen A., & Nielsen R. (2014). ANGSD: analysis of next generation
599 sequencing data. *BMC bioinformatics*, 15(1): 356.

600 Li H., Handsaker B., Wysoker A., *et al*. (2009). The sequence alignment/map format and
601 SAMtools. *Bioinformatics*, 25(16): 2078–2079.

602 Li H. & Durbin R. (2009). Fast and accurate short read alignment with Burrows-Wheeler
603 transform. *Bioinformatics*, 25(14): 1754–1760.

604 Mallet J. (2007). Hybrid speciation. *Nature*, 446: 279–283.

605 Martin S.H., Dasmahapatra K.K., Nadeau N.J., *et al*. (2014). Genome-wide evidence for
606 speciation with gene flow in *Hiliconius* butterflies. *Genome Research,* 23: 1817-1828.

607 Massie K.R. & Markow T.A. (2005). Sympatry, allopatry and sexual isolation between
608 *Drosophila mojavensis* and *D. arizonae*. *Hereditas,* 142: 51-55.

609 Mayr E. (1942). Systematics and the origin of species from the viewpoint of a zoologist.
610 *Columbia University Press*.

611 Mayr E. (1963). Animal and species evolution. In: Animal species and their evolution, *Harvard*
612 *University Press*.

Meier J.I., Marques D.A., Selz O.M., *et al*. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications*, 8: 14363.

Moran, P.A., Hunt J., Mitchell Ch., Ritchie M.G., & Bailey N.W. 2019. Behavioural mechanisms of sexual isolation involving multiple modalities and their inheritance. *Journal of evolutionary Biology,* 32: 243-258.

Nadeau N.J., Martin S.H., Kozak K.M., *et al*. (2013). Genome-wide patterns of divergence in gene flow across a butterfly radiation. *Molecular Ecology,* 22: 814-826.

Nolte A. W., & Tautz D. (2010). Understanding the onset of hybrid speciation. *Trends in Genetics*, 26(2): 54–58.

Nakazato T., Warren D.L., & Moyle L.C. (2010). Ecological and geographic modes of species divergence in wild tomatoes. *American Journal of Botany,* 97: 680-693.

Nei M. (1987). *Molecular Evolutionary Genetics*. Columbia University Press.

Noor M.A.F. (1999). Reinforcement and other consequences of sympatry. *Heredity,* 83: 503-508.

Nosil P. (2009). Adaptive population divergence in cryptic color-pattern following a reduction in gene flow. *Evolution*, 63(7): 1902–1912.

Novikova Y.P. *et al*. (2016). Sequencing of the genus *Arabidopsis* identifies a complex history of non-bifurcating speciation and abundant trans-specific polymorphism. Nature Genetics 48: 1077–1082.

Ozerov A.L. (2005). World catalogue of the family Sepsidae (insect: Diptera).*Zoologicheskie Issledovania (Zoological Studies)*, 8: 1–74.

Parker G.A. (1972a). Reproductive behaviour of Sepsis cynipsea (L.) (Diptera: Sepsidae) I. a preliminary analysis of the reproductive strategy and its associated behaviour patterns. *Behaviour*, 41(1): 172–205.

Parker G.A. (1972b). Rerpductive behaviorur of Sepsis cynipsea (L.) (Diptera: Sepsidae) II. the significance of the precopulatory passive phase and emigration. *Behaviour*, 41(3): 242–250.

Pont A.C., & Meier R. (2002). *The Sepsidae (Diptera) of Europe*. Leiden, N.E.: *Fauna Entomologica Scandinavia*, 37.

Puniamoorthy N, Ismail M.R.B., Tan D.S.H., & Meier R. (2009). From kissing to belly stridulation: Comparative analysis reveals surprising diversity, rapid evolution, and much homoplasy in the mating behaviour of 27 species of sepsid flies (Diptera: Sepsidae). *Journal of Evolutionary Biology*, 22(11): 2146–2156.

Puniamoorthy N. (2014). Behavioural barriers to reproduction may evolve faster than sexual morphology among populations of a dung fly (Sepsidae). *Animal Behaviour*, 98: 139–148.

Reich, D., N. Patterson, M. Kircher, F. Delfin, M. R. Nandineni et al., 2011 Denisova admixture and the first modern human dis- persals into Southeast Asia and Oceania. *American Journal of Human Genetics,* 89: 516–528.

Rieseberg L.H., Raymond O., Rosenthal D.M., *et al*. (2003). Major ecologocial transitions in wild sunflowers facilitated by hybridization. *Science*, 301(5637): 1211–1216.

Ritchie M.G. (2007). Sexual selection and speciation. *Annual Review of Ecology, Evolution and Systematics,* 38: 79–102.

Rohner P.T., Bächli, G., Paltrinieri, L.P., Duelli, P., Obrist, M.K., Jochmann, R., & Blanckenhorn, W.U. (2015). Distribution, diversity gradients and Rapoport's elevational rule in the black scavenger flies of the Swiss Alps (Diptera: Sepsidae). *Insect Conservation and Diversity,* 8, 367–376.

Rohner P.T., Blanckenhorn W.U., & Puniamoorthy N. (2016). Sexual selection on male size drives the evolution of male-biased sexual size dimorphism via theprolongation of male development. *Evolution,* 70(6): 1189–1199.

Dafu R., Kangshan M., Lei Z., *et al*. (2016). Genomic evidence for polyphyletic origins and interlineage gene flow within complex taxa: a case study of *Picea brachytyla* in the Qinghai-Tibet Plateau. *Molecular Ecology*, 25: 2373–2386.

Saetre G.P. (2013). Hybridization is important in evolution, but is speciation? *Journal of Evolutionary Biology*, 26(2): 256–258.

Schlötterer, C., Tobler, R., Kofler, R. & Nolte, V. (2014). Sequencing pools of individuals — mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics* 15: 749–763.

Seehausen O. (2004). Hybridization and adaptive radiation. *Trends in Ecology and Evolution*,

670 19(4): 198–207.

671 Soraggi, S., Wiuf, C. & Albrechtsen, A. (2018). Powerful Inference with the D-Statistic on
672 Low-Coverage Whole-Genome Data. *G3: Genes, Genomes, Genetics* **8**: 551–566.

673 Streicher J.W., Devitt T.J., Goldberg C.S., *et al*. (2014). Diversification and asymmetrical gene
674 flow across time and space: lineage sorting and hybridization in polytypic barking frogs. *Molecular*
675 *Ecology*, 23(13): 3273–3291.

676 Su, K.F.Y., S. N. Kutty, & R. Meier (2008). Morphology versus molecules: the
677 phylogenetic relationships of Sepsidae (Dptera: Cyclorrhapha) based on morphology and
678 DNA sequence data from ten genes. Cladistics 24:902-916.

679 Su, K.F.Y., Puniamoorthy, J., Özsu, N., A. Srivathsan, & R. Meier (2016). Evolutionary
680 analysis identifies multiple genome expansions and contractions in Sepsidae (Diptera) and suggests
681 targets for future genomic research. *Cladistics* 32: 308–316.

682 Tajima, F. (1983). Evolutionary Relationship of DNA Sequences in Finite Populations.
683 *Genetics* 105: 437–460.

684 Trier C. N., Hermansen J. S., Sætre G. P., & Bailey R. I. (2014). Evidence for mitonuclear and
685 sex-linked reproductive barriers between the hybrid Italian sparrow and its parent species. *PLoS*
686 *Genet*, 10(1): e1004075.

687 Via S. & West J. (2008). The genetic mosaic suggests a new role for hitchhiking in ecological
688 speciation. *Molecular Ecology*, 17(19): 4334–4345.

689 Ward P.I. (1983). The effects of size on the mating behaviour of the dung fly *Sepsis cynipsea*.
690 *Behavioural Ecology and Sociobiology*, 13(1): 75–80.

691 Ward P.I., Hemmi J., & Rösli T. (1992). Sexual conflict on the dung fly Sepsis cynipsea.
692 *Functional Ecology*, 6(6): 649–653.

693 Wood H.M., Grahame J.W., Humphray S., *et al*. (2008). Sequence differentiation in regions
694 identified by a genome scan for local adaptation. *Molecular Ecology*, 17(13): 3123–3135.

695 Wu C.I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology*,
696 14(6): 851–856.