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], Jeanne Wilbrandt[2], Malte Petersen[2], 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

athene.giesen@gmx.net

martin.schaefer@ieu.uzh.ch

heidi.tschanz-lischer@bioinformatics.unibe.ch

kentaro.shimizu@ieu.uzh.ch

rie.inatsugi@ieu.uzh.ch

oliver.niehuis@biologie.uni-freiburg.de

b.misof@leibniz-zfmk.de

lars@cgae.de

simon.aeschbacher@ieu.uzh.ch

wolf.blanckenhorn@ieu.uzh.ch

Word count: Abstract 180; Introduction 642; Materials & Methods 885; Results 501; Discussion 1551; 50 References 1105.

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, in which context the reference draft genomes of several fly species of the genus *Sepsis* explored in this study were sequenced and assembled. 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) 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).

MATERIALS & METHODS

Sample collection and treatment of fly cultures

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

DNA library preparation and next generation sequencing

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

149 https://www.ncbi.nlm.nih.gov/sra) under the accession number (XXX; which will be provided
150 upon acceptance of the manuscript).

151 **Table 1:** Sampling sites and percentages of DNA reads mapping to the reference genome for
152 pool-sequenced laboratory populations (a) and field-collected flies (b). Sites where *S.*
153 *cynipeas* and *S. neocynipsea* co-occur in sympatry are printed in regular font. Sites where *S.*
154 *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

155

156 **Raw read processing**

157 Qualitative validation of sequence data before and after trimming was done with FastQC
158 High Troughput Sequence QC Report version 0.11.4 (Andrews *et al*., 2011). After removal of
159 Illumina-specific adapters and trimming with *Trimmomatic* (version 0.36; Bolger, Lohse &
160 Usadel, 2014), the *S. cynipsea* and *S. neocynipsea* reads were mapped to the draft genome of
161 *S. thoracica* with *bwa mem* (version 0.7.12; Li & Durbin, 2009) using default parameters. The
162 assembly of this reference genome is part of a larger and ongoing comparative genomic study
163 of several *Sepsis* species (unpublished data*). S. thoracica* specimen used for genome
164 sequencing were originally collected from Capriasca, Ticino (CH). The draft genome, v0.1
165 was based on Oxford Nanopore long reads (app. 25x read depth), assembled with Canu
166 (Korner et al. 2017) and polished with Illumina short reads (app. 20x read depth). The
167 assembly used in this study is available under the rules of the Ford Lauderdale agreement
168 from www.cgae.de/seto_01_genome.fasta and www.cgae.de/seto_01_genome_masked.fasta.

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

174 Using the variant-caller Pool-SNP (Kapun *et al*., 2018), we identified high-confidence
175 SNPs with the following combination of SNP-calling parameters: coverage of each sample ≥
176 10x; coverage of each sample ≤ 95 % percentile of coverage distribution across contigs and
177 samples; minimum allele count of a minor allele at a SNP across all samples pooled > 20x;
178 minor allele frequency at a SNP across all samples pooled > 0.01. We only retained SNPs
179 when all samples fulfilled the above coverage threshold criteria. In addition, we removed
180 SNPs located within a 5-bp distance to indel polymorphisms that occurred in more than 20
181 copies at a given site across all pooled samples, and we ignored SNPs located within
182 repetitive regions. The resulting VCF file was converted to the SYNC file format (Kofler *et*
183 *al*., 2011), and a custom Python script was used to calculate sample-specific allele frequencies
184 for major alleles at each SNP.

185

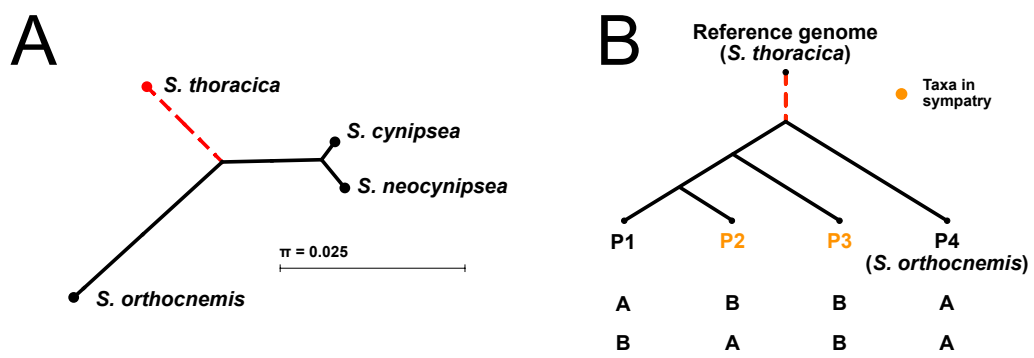
186 **Testing for introgression**

187 To test for introgression, we used the so-called ABBA-BABA test for historical gene flow
188 in the presence of incomplete lineage sorting (Green *et al*., 2010; Durand *et al*., 2011). We
189 provide a brief motivation of the approach in the following and refer to the Supporting
190 Information for details. Imagine a rooted phylogeny with four species (P1 to P4) and a
191 topology of (((P1, P2), P3), P4) as illustrated in Fig. 1B. In an alignment of one haploid
192 genome from each species, incomplete lineage sorting under neutral evolution leads to two
193 mutational configurations of the ancestral (A) and derived (B) allele, ABBA and BABA, that
194 under the infinite-sites model of mutation (Kimura, 1969) are incompatible with the species
195 topology (Fig. 1B). In the absence of gene flow between P2 and P3, ABBA and BABA
196 configurations occur with equal probability (Hudson, 1983; Tajima, 1983). In contrast,
197 historical gene flow between P2 and P3 causes an excess of ABBA configurations (Green *et*
198 *al*., 2010; Durand *et al*., 2011). This logic is captured by the *D*-statistic, a scaled difference
199 across a set of bases between counts of ABBA and BABA configurations bounded by –1 and
200 1 (Green *et al*., 2010). The ABBA-BABA test examines significant deviations of *D* from 0. A
201 significant excess of ABBA suggests evidence for gene flow between P2 and P3. A
202 significant depletion of ABBA is equivalent to an excess of ABBA if the positions of P1 and
203 P2 in the species topology are swapped, and hence either suggests gene flow between P1 and
204 P3, or a reduction in gene flow between P2 and P3 relative to gene flow between P1 and P3.

205 The original version of the ABBA-BABA test, and modifications of it, have been widely
206 applied to different types of genome-scale polymorphism data obtained with various
207 sequencing strategies, including whole genome sequencing (e.g. Green *et al*., 2010), RAD
208 sequencing (e.g., Eaton & Ree, 2013; Meier *et al*., 2017; Streicher *et al*., 2014), and exon
209 capture data (e.g., Heliconius Genome Consortium, 2012). More recently, Durand *et al*.
210 (2011) and Soraggi *et al*. (2018) extended the scope of the original test to allele frequency
211 data, and hence to unphased sequencing data (including pool-sequencing data; Schlötterer *et*
212 *al*. 2014). We implemented the extensions by Durand *et al*. (2011) and Soraggi *et al*. (2018)
213 in a Python script (https://github.com/capoony/ABBABABA-4AF), and refer to the respective
214 test statistics as D_D and D_S. Our script also computes jack-knifing *z*-scores based on a matrix
215 of allele frequencies for previously defined high-confidence SNPs (see Supporting
216 Information for details). We adopted the commonly used significance threshold of $|z| > 3$

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

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



227

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

241

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

250 **RESULTS**

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

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

270

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

272 The majority of our tests for gene flow at the Sörenberg site revealed a genome-wide
273 deficiency of derived-allele sharing (i.e. ABBA patterns) among the local *S. cynipsea* and *S.*
274 *neocynipsea* populations (Table 2), suggesting that interspecific gene flow in sympatry
275 (among P2 and P3) is lower than interspecific gene flow among putatively allopatric
276 populations (P1 and P3). This observation was robust to our choice of the P1 ingroup
277 population, but sensitive to the version of *D*-statistic used (D_S vs. D_D), and, in the case of D_D,
278 to whether we used field or laboratory samples as P2. We first conducted tests involving two
279 *S. neocynipsea* populations from the Swiss Alps (Geschinen *S. neocynipsea*, GeN; Hospental
280 *S. neocynipsea*, HoN; Table 1) as putatively allopatric P1, and all four possible sympatric
281 interspecific pairs of field and laboratory samples from Sörenberg as P2 (*S. neocynipsea*) and
282 P3 (*S. cynipsea*). We observed a deficiency of ABBA patterns in all eight configurations (D_S
283 > 0; positive *z*-scores ranging from 0.8 to 21.2). Deviations of D_S from 0 were significant in
284 six of these eight configurations (Table 2, upper part), suggesting overall evidence for higher
285 gene flow among putatively allopatric than sympatric interspecific population pairs. While
286 GeN and HoN are both separated by high mountains (altitudes ranging from ca. 600 m asl to
287 ca. 3,200 m asl) from the Sörenberg site, they are nevertheless geographically close (ca. 67.5
288 km and ca. 46.7 km Euclidean distance for GeN and HoN, respectively), and they occur in
289 sympatry with other sepsid species, including *S. cynipsea*. We were therefore concerned that
290 our assumption of no gene flow between P1 and P3 was violated because GeN and HoN are
291 parapatric, rather than allopatric, to our focal populations in Sörenberg.

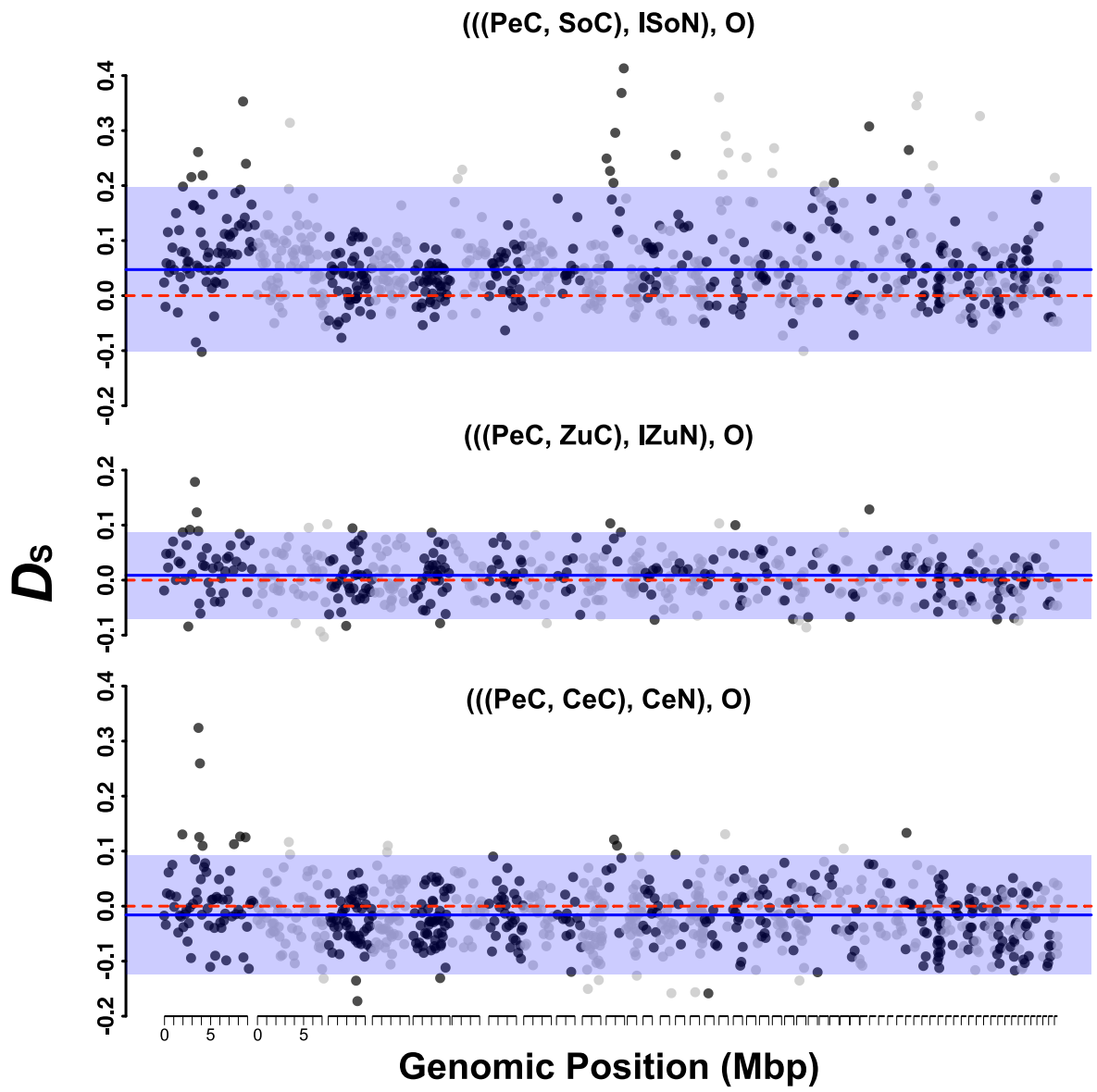
292 To address the above concern, we conducted additional ABBA-BABA tests involving P1
293 populations from geographically more remote sites in Europe. Specifically, we performed a
294 second series of tests with two alternative *S. cynipsea* populations as allopatric P1 ingroups
295 (Fig. 1B), one from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

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

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

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

324



325

326 **Fig. 2. Variation of D_S along the genome.** Points indicate D_S (Soraggi *et al*. 2018) for
327 genomic windows of 500 consecutive SNPs, oriented by contigs of the *S. thoracica* reference
328 genome with a length ≥ 50,000 bp (highlighted by alternating black and grey colors). Rows
329 correspond to population topologies used to test for interspecific gene flow at the three focal
330 sites in Sörenberg (SoC-ISoN; top), Zürich (ZuC-IZuN; center) and the French Cevennes
331 (CeC-CeN; bottom). For illustrative purposes, we depict mixed combinations of natural and
332 laboratory (with prefix "I") samples that yielded the clearest results when using the *S.*
333 *cynipsea* population from Petroia as P1. The horizontal dashed red and blue lines indicate the
334 neutral expectation ($D_S = 0$) and the mean genome-wide jackknife estimate, respectively. Blue
335 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**

347 **Convergent patterns of reinforcement in Zürich**

348 Complementary to the Sörenberg site, where *S. cynipsea* and *S. neocynipsea* co–occur in
349 an alpine environment, we analyzed gene flow between sympatric *S. cynipsea* and *S.*
350 *neocynipsea* populations from a lowland site in Zürich. These analyses were based on pooled
351 laboratory samples of both species (as defined above) and one pool-sequenced field sample of
352 *S. cynipsea* (Table 1). Using the *S. neocynipsea* populations of Geschinen and Hospental as
353 P1 ingroups, we obtained weakly but significantly positive D_S values ranging from 0.016 to
354 0.055 (Table 3), indicating a deficiency of ABBA patterns. These findings are qualitatively
355 consistent with the results obtained when analyzing samples from Sörenberg in suggesting
356 reinforcement in sympatry. However, only when using the allopatric Italian Petroia (but not
357 the Estonian) *S. cynipsea* population as P1 ingroup did we also find a significant, albeit even
358 smaller deficiency of ABBA patterns (Table 3).

359 Analyzing the same data as presented in the preceding paragraph with Durand *et al.'s*
360 (2011) D_D statistic, we found only one significant result indicating a deficiency of ABBA
361 patterns ($D_D = 0.03$) with the *S. neocynipsea* population from Geschinen as P1 ingroup (Table
362 S2). However, we found a significant, yet small excess of ABBA patterns ($D_D = 0.019$) with
363 *S. cynipsea* from Pehka, Estonia, as P1. Therefore, the putative signal of reinforcement among
364 the sympatric Zürich populations is much weaker than among the sympatric populations in
365 Sörenberg. Consistent with these observations, we did not find pronounced genomic clusters
366 of elevated D_S (Fig. 2, center) as for the Sörenberg populations, which corroborates that
367 reinforcement in Zürich was weaker.

368

369 **Table 3:** Results of ABBA–BABA tests applied to samples originating from sympatric
370 populations of *S. cynipsea* and *S. neocynipsea* in Zürich as P2 and P3 (and *vice versa*) based
371 on D_S. As in Table 2, the upper half of the table summarizes tests involving two *S.*
372 *neocynipsea* populations from Geschinen (GeN) and Hospental (HoN) as P1 ingroups, and the
373 bottom half summarizes tests involving *S. cynipsea* populations from Pehka, Estonia (EsC)
374 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**

375

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

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

389

390 **Table 4:** Results of ABBA-BABA tests applied to samples originating from sympatric
391 populations of *S. cynipsea* and *S. neocynipsea* in the French Cevennes as P2 and P3 (and *vice*
392 *versa*) based on D_S. As in Tables 2 and 3, the upper half of the table summarizes tests
393 involving two *S. neocynipsea* populations from Geschinen (GeN) and Hospental (HoN) as P1
394 ingroups, and the bottom half of the table summarizes tests involving *S. cynipsea* populations
395 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**

396

397 **DISCUSSION**

398 We applied the ABBA-BABA approach (Green *et al*., 2010; Durand *et al*., 2011; Soraggi
399 *et al*., 2018) to investigate natural patterns of admixture between a pair of closely related
400 sister species of sepsid flies for which we have detailed information on ecology, morphology,
401 life history, and behavior (Blanckenhorn, 1999; Blanckenhorn *et al*., 2000; Eberhard, 1999;
402 Ozerov, 2005; Parker, 1972a,b; Pont & Meier, 2002; Puniamoorthy *et al*., 2009; Rohner,
403 Blanckenhorn & Puniamoorthy, 2016; Ward, 1983; Ward, Hemmi & Rösli, 1992; Giesen *et*
404 *al*., 2017, 2019). Previous population genetic analyses of microsatellite markers (Baur *et al*.,
405 2020) and laboratory hybridization experiments (Giesen *et al*., 2017, 2019) suggested that *S.*
406 *cynipsea* and *S. neocynipsea* are genetically distinct but may occasionally hybridize in nature
407 when occurring in sympatry. Here we investigated this possibility by analyzing patterns of
408 DNA sequence variation among pools of laboratory as well as field-caught specimens from
409 various sampling sites across Europe where the two species occur in sympatry vs. allopatry.
410 Our study revealed qualitatively opposite, i.e. mixed patterns of interspecific gene flow in
411 sympatry depending on the sampling sites, which are discussed in detail below. Previous
412 studies of other taxa, using different genomic methods but with essentially the same logic
413 behind their approach, provided the full spectrum of potential results that we here obtained for
414 one species pair. Martin *et al*. (2014; see also Nadeau *et al*., 2013) observed *more* inter-
415 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

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

459 We found a weaker signal of reinforcement between *S. cynipsea* and *S. neocynipsea* at the
460 lowland Zürich site than at the high-altitude Sörenberg site in the Swiss Alps. This difference
461 might be related to the fact that *S. neocynipsea* is rare around Zürich, in fact in European
462 lowlands in general, while it is more common at higher altitudes. If based primarily on
463 behavioral mate choice, reinforcement is expected to be stronger, and hence its evolution
464 more likely, where interspecific encounter frequencies are higher. Regardless, why *S.*
465 *neocynipsea* remains rare and restricted to highland sites in Europe, whereas it is common in
466 dry and warm lowland areas in North America, cannot be answered by our study. *Sepsis*
467 *cynipsea* is abundant around fresh dung all over Europe north of the Alps (Pont & Meier,
468 2002). It therefore might outcompete and ultimately relegate *S. neocynipsea* towards higher
469 altitudes. However, *S. cynipsea* also readily populates cold high-altitude sites (Rohner *et al*.,
470 2015), and some degree of altitudinal molecular (allozyme) differentiation can also be
471 observed (Kraushaar *et al*., 2002). It consequently remains unclear why *S. neocynipsea* can
472 better persist at high altitudes in Europe.

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

480 simple 2D matrix of allele frequencies as input. This implementation allows restricting the
481 ABBA-BABA analysis to high-confidence SNPs that pass a set of user-defined filters in order
482 to reduce artefacts of pool-sequencing (Schlötterer *et al*. 2014) that may produce false
483 polymorphisms due to sequencing errors (Futschik & Schlötterer 2011; Cutler & Jensen
484 2012). In addition, and in contrast to the software package ANGSD (Korneliussen *et al*.
485 2014), our implementation accommodates an outgroup (P4) taxon different from the one used
486 as a reference to call the SNPs. This distinction reduces a potential reference bias that could
487 affect sequence variation if the ingroup taxa (P1, P2, P3) strongly differed from an outgroup
488 that also served as reference (Ballouz *et al*. 2019). Our script and detailed instructions for how
489 to use it are available on GitHub (https://github.com/capoony/ABBA-BABA-4-AF; see also
490 Supporting Information).

491 In summary, our genomic study provided evidence that the extent of interspecific gene
492 flow between two closely related species of sepsid flies in sympatry relative to allopatry
493 varies geographically. Interspecific gene flow appears to be very limited at sites in the Swiss
494 Alps where the two species have perhaps locally adapted to different (micro-)ecological
495 niches, such as different breeding times or phenologies, and/or where pre- and postmating
496 reproductive barriers (Giesen *et al*., 2017) have been and are subject to stronger
497 reinforcement. It however remains unclear, if ecological and sexual selection are mediating
498 these effects. We note in closing that North American and European *S. neocynipsea*
499 populations also differ substantially based on neutral genetic markers (microsatellites), as can
500 be expected by genetic drift alone due to definitive inter-continental allopatry (Baur *et al*.,
501 2020). It is currently also unclear whether European *S. neocynipsea* are remnants of an
502 ancient Holarctic population or whether they (re-)invaded Europe from North America, where
503 the species is much more common and fills the niche that *S. cynipsea* has in Europe. Future
504 genomic research should analyze patterns of genomic differentiation across allopatric
505 populations from Europe and North America (Baur *et al*., 2020). Such analyses would
506 provide further insights into introgression patterns across the species' entire natural range, as
507 well as into the role of local (e.g. latitudinal) adaptation in shaping genome-wide patterns of
508 differentiation.

509

510 **LITERATURE CITED**

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

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

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

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

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

520 Baur J., Schäfer M.A., Blanckenhorn W.U., & Giesen A. (2020). Patterns of genetic
521 differentiation in sexual and neutral morphological traits among species and populations of dung
522 flies (Diptera: Sepsidae). *Journal of Zoological Systematics and Evolutionary Research* 58: 159–
523 173.

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

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

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

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

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

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

538 Butlin R. K. 1995. Reinforcement: an idea evolving. *Trends in Ecology & Evolution*, 10, 432-
539 434.

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

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

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

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

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

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

550 Durand E.Y., Patterson N., Reich D., & Slatkin M. (2011). Testing for ancient admixture

551 between closely related populations. *Molecular Biology and Evolution*, 28(8): 2239–2252.

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

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

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

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

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

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

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

566 Giesen A., Schäfer M.A., & Blanckenhorn W.U. (2019). Geographic patterns of postzygotic
567 isolation between two closely related widespread dung fly species (*Sepsis cynipsea* and *Sepsis*
568 *neocynipsea*; Diptera: Sepsidae). *Journal of* Zoological Systematics and Evolutionary Research,
569 57: 80–90

570 Gleason, J.M. & Ritchie, M.G. 1998. Evolution of courtship song and reproductive isolation in
571 the *Drosophila willistoni* species complex: Do sexual signals diverge the most quickly? *Evolution*,
572 52: 1493–1500.

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

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

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

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

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

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

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

588 Kastally Ch., Trasoletti M., & Mardulyn P. (2019). Limited gene exchange between two sister
589 species of leaf beetles within a hybrid zone in the Alps. *Journal of evolutionary Biology,* 32: 1406-
590 1417.

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

593 Kofler, R., Pandey, R.V. & Schlötterer, C. (2011). PoPoolation2: identifying differentiation
594 between populations using sequencing of pooled DNA samples (pool-seq). *Bioinformatics* 27:
595 3435–3436.

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

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

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

604 Kraushaar U., Goudet J., & Blanckenhorn W.U. (2002). Geographical and altitudinal population
605 genetic structure of two dung fly species with contrasting mobility and temperature preference.
606 *Heredity*, 89: 99-106.
607
608 Korneliussen T.S., Albrechtsen A., & Nielsen R. (2014). ANGSD: analysis of next generation
609 sequencing data. *BMC bioinformatics*, 15(1): 356.

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

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

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

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

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

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

621 Mayr E. (1963). Animal and species evolution. In: Animal species and their evolution, *Harvard*
622 *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*,

680 19(4): 198–207.

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

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

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

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

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

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

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

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

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

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

705 Wu C.I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology*,
706 14(6): 851–856.